EFDA–JET–CP(08)01/01

E. Ronchi, S. Conroy, E.A. Sundén, G. Ericsson, A. Hjalmarsson, C. Hellesen,
M.G. Johnson, M. Weiszflog and JET EFDA contributors

A Neural Networks Framework for Real-Time Unfolding of Neutron Spectroscopic Data at JET

A Neural Networks Framework for Real-Time Unfolding of Neutron Spectroscopic Data at JET

E. Ronchi, S. Conroy, E.A. Sundén, G. Ericsson, A. Hjalmarsson, C. Hellesen, M.G. Johnson, M. Weiszflog and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Association EURATOM-VR, Uppsala University, SE-75120 Uppsala, Sweden
[2]EURATOM-UKAEA Fusion Association, Culham Science Centre, OX14 3DB, Abingdon, OXON, UK
** See annex of M.L. Watkins et al, "Overview of JET Results ", (Proc. 21 st IAEA Fusion Energy Conference, Chengdu, China (2006)).*

Preprint of Paper to be submitted for publication in Proceedings of the HTPD High Temperature Plasma Diagnostic 2008, Albuquerque, New Mexico. (11th May 2008 - 15th May 2008)

ABSTRACT.
A determination of fast ion population parameters such as intensity and kinetic temperature is important for fusion reactors. This becomes more challenging with finer time resolution of the measurements, since the limited data in each time slice causes increasing statistical variations in the data. This paper describes a framework using Bayesian-regularized Neural Networks (NN) designed for such a task. The method is applied to the TOFOR 2.5MeV fusion neutron spectrometer at JET. NN training data is generated by random sampling of variables in neutron spectroscopy models. Ranges and probability distributions of the parameters are chosen to match the experimental data. Results have shown good performance both on synthetic and experimental data. The latter was assessed by statistical considerations and by examining the robustness and time consistency of the results. The regularization of the training algorithm allowed for higher time resolutions than simple forward methods. The fast execution time makes this approach suitable for tEal-time analysis with a time resolution limit in the μs time scale.

1. INTRODUCTION

Measurement of thermal and supra thermal ion population parameters such as kinetic temperature and intensity is of interest for nuclear fusion reactors. At the Joint European Torus (JET) these can be estimated from neutron energy spectra with the 2.5MeV Time-Of-Flight Optimized For Rate (TOFOR) fusion neutron spectrometer [1]. Since experimental data comes folded with the response function of the instrument, an inverse process is necessary to extract the required information. Neural Networks (NN) [2] are here used for this purpose, with real-time capabilities. Operation at high time resolution (e.g. ≤100-200ms with TOFOR) requires the NN to cope with low statistics in each time slice. In this case, proper regularization is required to prevent overfitting on the random Poisson fluctuations in each data bin.

2. GENERATION OF SYNTHETIC DATA & THE NEURAL NETWORKS FRAMEWORK

A parametric Neutron Emission Spectroscopy (NES) model is used to generate synthetic data to train the NN. This includes components generated by the various heating schemes used, i.e. Ohmic, Ion Cyclotron Resonance Heating (ICRH) and Neutral Beam Injection (NBI). In addition, a back-scatter (SCATT) component accounts for neutrons that reach TOFOR after scattering one or more times in the reactor chamber. Random time-of-flight coincidences are also accounted for in the model depending on the count rate. The full model features $\nu = 6$ degrees of freedom and the guidelines are explained in further detail in [3,4,5]. The first component describes the neutron emission from thermalized deuterons in the plasma. This is here referred to as Bulk and depends on two parameters, namely a relative intensity (0-100%) A_{BULK} and an ion temperature T_{BULK}. The ICRH component describes the neutron emission generated by this type of heating and it depends on a intensity and a ion tail temperature, namely A_{ICRH} and T_{ICRH}. The NBI component includes contributions of all 16 Positive Ion Neutral beam Injectors (PINI) at JET. From diagnostic data on each PINI power and energy the relative contribution of each injector can be estimated. This permits to fix the overall NBI neutron energy spectral shape leaving out a single free parameter, i.e. the intensity A_{NBI}. Finally, the SCATT component also depends on a single parameter (A_{SCATT}). Generation of synthetic energy spectra (y) is performed

by randomly sampling each of the *v* free variables from given probability distributions functions and boundaries. These are chosen from plasma physics and empirical principles. The synthetic time-of-flight data (m) is calculated by folding the corresponding energy spectrum with the TOFOR response function (R) and by normalizing the area to a given number of counts (N). The addition of random coincidences and the application of Poisson errors complete the process and yield a vector of integer Poisson-perturbed data (m_0) analogous to the experimental data. N has been sampled uniformly across the TOFOR typical operating range, which is typically 200-10,000 counts/spectrum. The NN training is performed using m_0 as input and the free parameters as output. This generates a network of 151 inputs (time bins between 45 and 98 ns flight time) and *v* outputs. Two hidden layers of 16 and 10 neurons have been used. The network architecture chosen is feed-forward with tan-sigmoidal hidden layers and linear outputs. The training algorithm used is a Bayesian-regularized version of the Levenberg-Marquardt method [6,7] and was applied to 24,000 events. Other 24,000 events were used for validation and 100,000 for testing. Four networks have been trained on Ohmic, ICRH, NBI and ICRH+NBI heating schemes including the back-scatter component in all cases. In these cases *v* equals to 3, 5, 4 and 6 respectively. Application time for each NN is typically ≤5μs.

3. ANALYSIS

Performance assessment for numerical methods must keep in consideration the numerical degeneracy of the problem. This is when two significantly different sets of parameters generate time-of-flight spectra that are statistically indistinguishable at a given level of counts and thus cannot be resolved. This implies multiple solutions and analysis based on simple error evaluation of each i-th parameter, e.g. $\in_i = Output_i^{NN} - Correct_i$, becomes inadequate. For this reason a different metric has been adopted to evaluate the statistical plausibility of the NN output. This is here referred to as Σ metric and scales each $\in_i$ with the confidence interval of the i-th parameter near the solution (σ_i) , as in Eq. (1). The confidence intervals have been estimated using Cash statistic (*Cstat*) [8] by projecting the $\Delta Cstat = +1$ ellipsoid onto the parameter axes and a typical 68.3% level has been chosen. This is analogous to the projection of the $\Delta\chi^2 = +1$ ellipsoid for the case of normal errors and higher statistics. Since σ_i is generally not symmetric, the positive-going confidence interval σ_i^+ i is used when $\in_i$ is positive, and the negative-going interval σ_i^- otherwise, as in Eq. (1).

$$\Sigma_i = \frac{Output_i{}^{NN} - Solution_i}{\sigma_i} = \frac{\in_i}{\sigma_i} \quad i = 1,\ldots,v$$

$$\sigma_i = \begin{cases} 68.3\%\ \sigma_i^+ & \in_i \geq 0 \\ 68.3\%\ \sigma_i^- & \in_i < 0 \end{cases} \qquad (1)$$

Σ values between -1 and $+1$ indicate that the NN output is within the given interval of confidence and should be considered statistically plausible even if $\in_i$ is large, as in degenerate cases. Figure 1 shows results of the Σ metric calculated on synthetic data for the most complex model (NBI+ICRH) over the whole range of counts and event rates. Only 20,000 randomly selected spectra have been processed with the metric for computational time reasons.

When performance on experimental data is to be assessed, the metric can no longer be calculated

as the exact solution is unknown. In these conditions the NN output can be used to generate the corresponding energy spectrum which can be folded with the response function ($m_{NN} = R \cdot y_{NN}$) and compared with the data. The reduced Cstat of mNN can then be used as quality-of-fit estimator of the NN solution, albeit indirectly. Figure 2 illustrates the statistical probability densities of the reduced Cstat under the three main JET plasma auxiliary heating schemes. The solid lines relate to the synthetic data, while the dashed refer to actual TOFOR experimental data processed with NN. Each NN solution was also used as a starting guess to find the set of variables yielding the minimum *Cstat* (MINC) within the same parameter boundaries used for the synthetic data. This is shown with a dotted line. All recorded TOFOR data up to JET discharge number 70700 was used for the test with a dynamic time-slice criterion aimed to achieve a given number of counts ($N_C = 4000$) per time slice. Any change in heating scheme, however, forced the termination of the time slice even if N_C had not yet been reached. 100 ms was set as the minimum time slice duration for computational reasons and the value chosen for N_C is rather typical for TOFOR. Due to the relatively few ICRH-only time slices at JET and to their lower neutron yield, the ICRH curves in Figure 2 (top panel, dashed and dotted traces) are more jagged than in the case of NBI or ICRH+NBI.

Figure 3 provides further insight into the results obtained with MINC on all TOFOR time slices, regardless of the heating scheme. In particular, the statistical distribution of TBULK is normalized to the maximum value and shown as solid (NN) and dashed (MINC) lines. It can be seen that despite featuring similar Cstat distributions as in Figure 2, NN and MINC show significant differences in the shape of the statistical distribution of the individual variables. In particular, it can be seen that a large fraction of MINC samples for T_{BULK} saturates either at the minimum or maximum limit used for this parameter (1−40keV). This aspect is discussed in further detail in Section IV.

An example of time-resolved results of NN is presented in Figure 4 for JET Pulse No: 69392 using two different dynamic time slicing criteria aimed at $N_C = 1500$ counts/spectrum (above) and 10,000 counts/spectrum (below). The sampling rates in these cases is typically ≤100−200ms and ≤700−1400 ms respectively for TOFOR depending on the nuclear fusion rate. All heating amplitudes sum to 100% while the scattering contribution is intended as a fraction of the whole spectrum. This yields a total sum of the amplitudes at 100% + A_{SCATT} to highlight the integral nature of the latter. All heating-compatible NN are applied to each time slice and the one producing the lowest reduced Cstat is selected. This is referred to as the Adaptive Model Selection (AMS) and partitions the time slices into regions, each processed with a particular network. In case of two or more networks producing very similar Cstat, such as during the transitions between regions, the one producing the least model changes is selected. Pulse No: 69392 was chosen due to the large high-energy tails produced by the ICRH heating. In fact, both NBI and ICRH heating were active during the displayed time window. However, the ICRH contribution is found to be very small until 6s. Between 6-6.5s both components show significant signatures in the data but in the region between 6.5 and 10s the ICRH contribution dominates over the beams. This effect is correctly displayed by the NN results. In particular, it can be seen how AMS automatically selects the NBI, NBI+ICRH and ICRH networks in the three intervals mentioned above instead of forcing the NBI+ICRH model on all time slices. This permits to reduce the degeneracy of the model and avoids trying to resolve components that do not leave significant differences in the data. Tests

performed without AMS indicated significantly poorer Cstat performance and worse time consistency of the traces, validating these statements. Repeating the analysis with different dynamic time-slicing target counts/spectrum yielded consistent results. In fact, it can be seen that the panels associated with N_C = 10,000 in Figure 4 produce results that closely relate to a moving time average on the case N_C = 1500 suggesting that the method is robust with respect to the way data is divided into spectra.

4. DISCUSSION

The introduction of the Σ metric allowed for evaluation of the method's performance within the limits posed by numerical degeneracy from low statistics and finite energy resolution of the instrument. The non-linear and integral type of the model forced the confidence intervals σ_i to be estimated numerically. From Figure 1 it can be seen that the majority of NN results falls within the 68.3% confidence interval around the solution, with distributions centered on zero and standard deviation of approximately one σ_i in all cases. This is an indication that NN has made efficient use of the information present in the experimental data. All distributions are rather symmetric except for AICRH and ASCATT for which the peak is shifted by a fraction of σ_i. Empirical evidence suggests that this is a consequence of the probability densities (i.e. peaked Gaussians instead of uniform) from which these have been sampled. The statistical plausibility of the NN results is also illustrated by Figure 2, with reduced *Cstat* distributions on synthetic data (solid line) centered on unity for all the networks trained. When NN is applied to TOFOR experimental data, however, the distribution appears shifted rightwards (dashed line). This is a symptom of poorer match between the NN output and the experimental data and can be due to two causes. The first lies in the training set while the second depends on the TOFOR response function as well as on the NES modelused to generate the synthetic data. Any error in these would compromise the reconstruction of the experimental data with the sets of components included in the model. Tests with different NN training sets and configurations have been carried out without significant improvement on experimental data indicating that the model could be the major cause for the performance gap. Additional evidence arises by examining the MINC curves (dotted lines) in Figure 2 and by comparing the traces in Figure 3. These show that the NN results could in fact be improved in terms of reduced Cstat, but this does not necessarily imply better physical results. For instance, Figure 3 shows that NN applied to all TOFOR time slices produce two peaks in T_{BULK}. The first peak around 2−7keV is usually what is expected from hot plasma at JET with auxiliary heating. The second peak between 20 and 35keV is typically due to the thermalization of the NBI component and other slowing-down contributions. By contrast, the distribution of MINC lies more flat and tends to accumulate at the extremities, in contrast with physical predictions. Removal of upper boundaries for each variable with MINC decreased the reduced Cstat shift even further, but caused unphysical results such as T_{BULK} reaching 100s keV and MeV regions. Attempts to use more advanced models have been made, for instance by adding a second ICRH component at lower energy, yielding 8 parameters instead of 6. Once again the shift in the reduced Cstat distribution was found to decrease, but the low-energy ICRH component became often degenerate with the NBI and bulk components at lower statistics. This caused a net loss of quality despite the higher complexity of the NES model. The failure to use Cstat as the sole performance index on experimental data forced the reliability of NN to be analyzed further using additional empirical

evidence. In particular, the consistency found in Figure 4 indicated that NN results are robust with respect to the way data is divided into time slices. Furthermore, the relatively small oscillations over time found for the case N_C = 1500 suggest that the NN output is time consistent even at high time resolutions. The temporal behavior of T_{BULK} seems to be the only exception, showing deeper oscillations. This is a symptom of higher uncertainty on this variable and it was foreseen especially at lower statistics. The bulk component is in fact dependent on the thermalization process of both NBI and ICRH components and it is usually dominated by the direct auxiliary heating con- tributions, as in between 7 and 9.5s in Figure 4. More counts are therefore required to resolve this component under these circumstances, as in Figure 4 (bottom).

CONCLUSIONS

This paper has presented a NN framework that can provide statistically accurate results within the boundaries posed by finite counts in the data and by the finite resolution of the instrument response function. This was proved on synthetic data by analyzing the NN errors with respect to the amplitude of the confidence intervals of the parameters near the solution. This permitted the assessment of NN performance independently of the numerical degeneracy of the problem. Performance on experimental data from the TOFOR fusion neutron spectrometer at JET required joint use of reduced Cash statistic, self-consistency and time-consistency criteria, with satisfac- tory results. The use of adaptive model selection permitted to limit numerical degeneracy and thus improve the time consistency and reliability of the NN results. The method was here applied to the TOFOR but can be extended to any instrument of which the response function is known and a parametric representation of the data is possible. The application time of NN is on the order of few μs. This is orders of magnitude below the typical time resolution of neutron spectrometers at JET and the method is thus suitable for real-time analysis.

ACKNOWLEDGEMENTS

This work was performed under the auspices of the European Fusion Development Agreement with financial support from the Swedish Research Council and EURATOM.

REFERENCES

[1]. M.G.J. et al., doi:10.1016/j.nima.2008.03.010 (2008).
[2]. C. Bishop, *Neural Networks for Pattern Recognition* (Springer, 1996).
[3]. H. Brysk, *Plasma Phys.* **15**, 611 (1973).
[4]. H.H. et al., *Rev. Sci. Instrum.* **72**, 832 (2001).
[5]. M. T. et al., Uppsala Univ. ISSN 1401-6269 (2000).
[6]. D. Marquardt, SIAM J. Appl. Math. **11**, 431 (1963).
[7]. D. McKay, *Neural Comput*. **4** n5, 720 (1992).
[8]. W. Cash, ApJ **228**, 939 (1979).

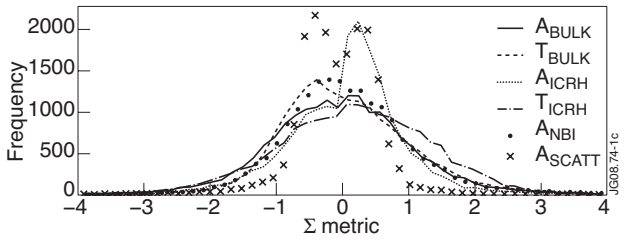

Figure 1: Σ metric for the 6 variables of the full NES model (BULK+RF+NBI+SCATT) on synthetic data.

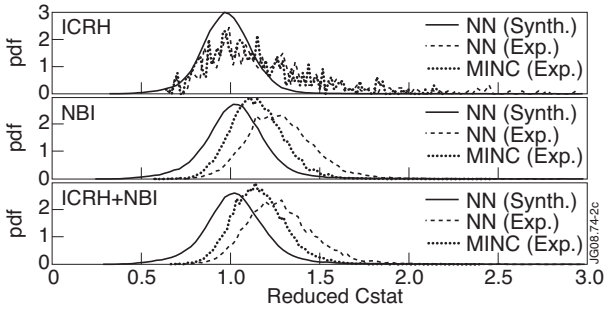

Figure 2: Reduced Cstat probability distributions for synthetic and experimental TOFOR data in the three main JET auxiliary heating schemes.

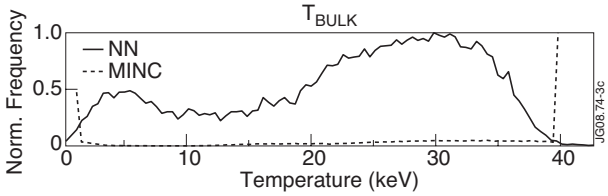

Figure 3: Comparison of T_{BULK} normalized statistical frequencies for NN and MINC on all TOFOR data up to JET Pulse No: 70700.

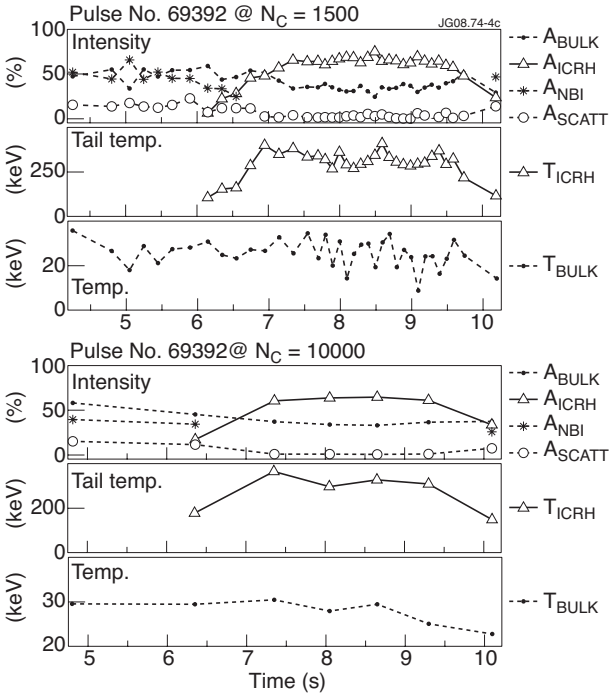

Figure 4: (Top three panels). Results from NN processing of JET Pulse No: 69392 with dynamic time-slicing at N_C =1500 counts. (bottom three panels) Same discharge processed with dynamic time-slicing at N_C = 10,000 counts.