29th EPS Conference on Plasma Phys. and Contr. Fusion Montreux, 17-21 June 2002 ECA Vol. ***26B****, P-5.103 (2002)*

Steps towards an integrated data analysis: Basic concepts and Bayesian analysis of Thomson scattering data

A. Dinklage[1], R. Fischer[2], G. Kühner[3], H. Maaßberg[1], E. Pasch[3], J. Svensson[3]

[1]*Max-Planck-Institut für Plasmaphysik, TI Greifswald, EURATOM Association, Wendelsteinstr.1, 17491 Greifswald, Germany*
[2]*Centre for Interdisciplinary Plasma Science, Boltzmannstr. 2, 85748 Garching, Germany*
[3]*Max-Planck-Institut für Plasmaphysik, EURATOM Association, Boltzmannstr. 2, 85748 Garching, Germany*

Abstract
Aspects of the integration of different data sources are discussed with respect to improvement of reliability and significance of experimental results. Thomson scattering data have been analysed by means of Bayesian probability theory. This much improved statistical model for the evaluation of the data can be used for assessment and improvement of the diagnostics. A Bayesian graphical model has been developed for the linkage of different diagnostics, with the goal of arriving at a fully *integrated data analysis*.

Background
Consistency checks and validation of experimental data from different diagnostic sources is a persistent problem in magnetically confined plasma research. If compared to efforts in diagnostic hardware development, data analysis turns out to be frequently a sideline activity, in particular if the analysis has to combine information from different sources. Therefore, the goal of our investigation is to explore possibilities for integration of available physics information in order to enhance the significance and the reliability of experimental data from fusion devices. The notion of integrated data analysis has to be understood not only as a sophisticated combination of available data by statistical means, but also as the usage of expert knowledge of all parties involved in the data validation process and to integrate theoretical considerations and modelling. In order to provide platforms for communication and to minimize efforts in data administration, computer science becomes an essential part of the integration concept. Many of the positive aspects emerging from a well organized data management have been proven on large scale devices like JET [1].

A great problem for integration of the analysis of fusion data comes from the extreme heterogeneity of the information sources. E.g., one has to deal with many different diagnostics on different spatial and time scales as wells as theoretical considerations, like equilibrium calculations or transport modelling. Our investigations start with a rigorous error analysis of data (here Thomson scattering (TS) data for the electron density n_e and the electron temperature T_e), which will later be combined with other measurements (e.g. spatial profiles of n_e and T_e on flux coordinates from TS, ECE and lithium beam measurements), to give a full joint evaluation of diagnostic data. To achieve this goal we have formulated a framework which will allow us to extract the most reliable profile information from different sources, including a consistent mapping on magnetic surfaces. Moreover, the statistical models developed for the error analysis are used for diagnostic improvement and diagnostic design. Outstanding tools for tackling these issues are given by Bayesian probability theory.

Evaluation of Thomson scattering data by means of Bayesian probability theory

For a systematic statistical modelling of Thomson scattering, Bayesian probability theory (BPT, [2,3]) was employed. The main advantages of BPT result from the possibility to derive the probability density function (pdf) of the quantities of interest, to treat nuisance parameters in a concise way and to include expert knowledge (priors). The analysis starts from Bayes theorem, which in our case links the *marginal posterior* pdf $P(T_e,n_e|d, \sigma, I)$ to find an electron density n_e and temperature T_e, given the data d, its uncertainties σ and some additional information I, to the *likelihood* $P(d|,T_e,n_e,\sigma,I)$ and the so-called *prior* $P(T_e,n_e,|I)$:

$$P(T_e,n_e\,|\,d,\sigma,I) = P(d\,|\,T_e,n_e,\sigma,I)P(T_e,n_e\,|\,I)/P(d\,|\,I)$$

For our purposes, the evidence $P(d|I)$ serves for normalization only. The major objective of our investigation, which is to determine reliable uncertainty measures of the experimental results allowing one to combine different diagnostics sources, is reflected by the marginal posterior probability density functions of the quantities of interest. It must be emphasized that much care and effort has been taken for a detailed investigation of the uncertainties of all parameters that enter the statistical model.

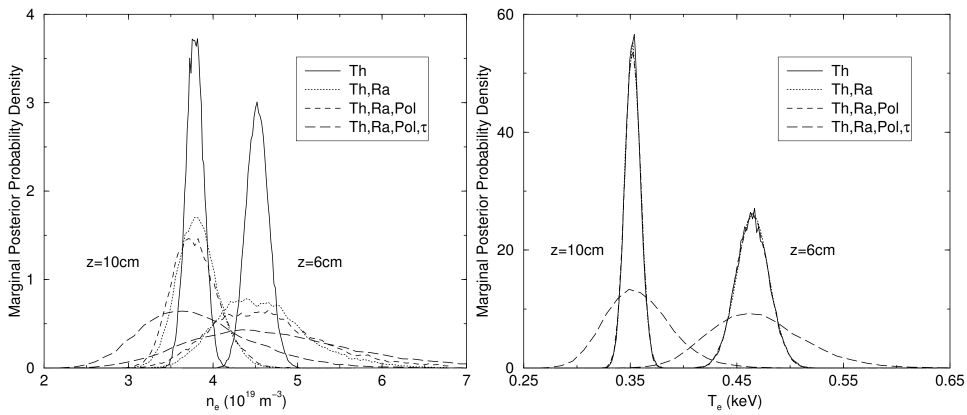

Fig. 1: Marginal probability densities for n_e (left panel) and T_e (right panel) at two spatial positions for a Wendelstein 7-AS plasma # 47894. Different lines depict uncertainties included in the analysis. The dotted and the dashed lines are hidden for the T_e analysis because inclusion of both the Raman scattering errors (Ra) and the polarization factors (Pol) are negligible if compared to the errors in the Thomson data (Th) and spectral sensitivities of the polychromators (τ).

For a systematic derivation of probability density functions from diagnostic data, the evaluation of n_e and T_e profiles from Thomson scattering on Wendelstein 7-AS (spatially resolved, 0.5 J, 10 ns pulsed Nd:YAG laser based TS at 20 Hz, discrete spectral channel detection) employing BPT was performed. It was shown that BPT reproduced results from different analysis techniques which have been proven for validity in the parameter regimes chosen for comparison [4]. Moreover, due to systematic inclusion of all accessible uncertainties that affect the outcome – or, in a different terminology, by the discussion of a complete statistical model of the diagnostics – substantial improvements with respect to sensitivity could be achieved by the BPT analysis, which contributes to investigations in parameter regimes [5] the diagnostics was originally not designed for.

BPT can be employed for assessing and improvement of diagnostic capabilities [6]. In figure 1 the outcome of the BPT analysis – the marginal pdf for n_e and T_e – is shown. In this representation the influence of different nuisance parameters can be visualized. E.g. the T_e measurement is barely affected by the Raman calibration and the polarization dependence of the detectors, but spectral sensitivities τ strongly affect the uncertainties. The latter result is valid for n_e as well, where τ even shifts the maximum of the T_e pdf. These considerations may be used to quantify the effect of any improvement related to a given nuisance parameters. Moreover, figure 1 shows that the resulting pdfs may be non-Gaussian, but Gaussians are an essential prerequisite for error propagation laws. A detailed discussion of the results may be found in [4].

Joint evaluation of profile data

The Bayesian framework is especially well suited for the combination of different sources of information. The reason for this is the direct association of uncertainties not only with observations, but also with the unknown parameters of a model. Since unknown parameters are described directly by probability distributions, relationships between the unknown parameters of a model can be stated explicitly. Such relationships are easily formalised and visualized by the use of Bayesian graphical models [7], where nodes represent unknown parameters or observations, and the directed edges represent probabilistic or logical dependencies between the nodes. By applying Bayes theorem [2], the joint posterior probability distribution of the unknown nodes given the observed nodes can then be calculated. In figure 2 such a model is shown for an integrated solution of the problem of combining profile measurements of electron temperature and density from two different Thomson scattering diagnostics. The mapping on magnetic surfaces is here an integral part of the solution, since it will both depend on and influence the inferred profiles. To demonstrate the generality of the method, an extra non-profile diagnostic signal (the total plasma energy) has been added as an auxiliary information source.

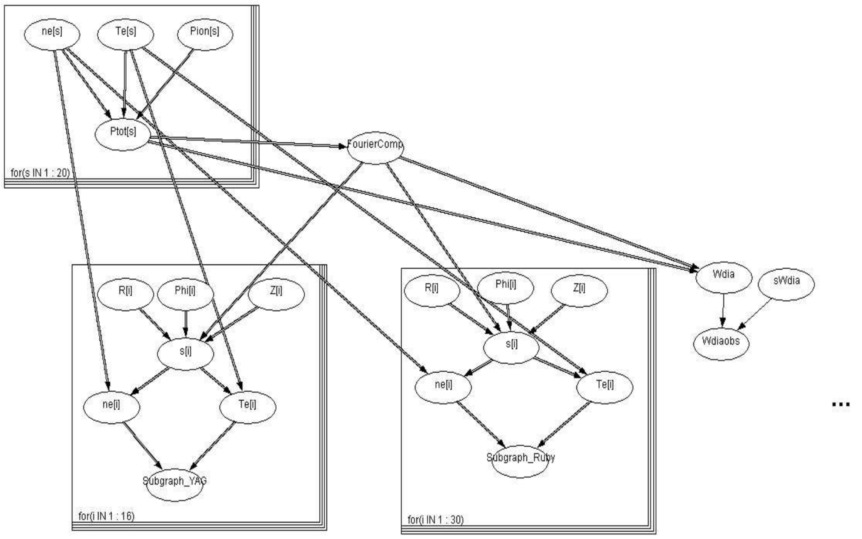

Fig. 2: Bayesian graph for joint evaluation of electron temperature and density profiles, including a mapping on magnetic surfaces. The bottom sections represent subgraphs for different diagnostics. Integration of further diagnostics is done by adding further subgraphs representing those information sources.

Conclusion

BPT provides outstanding tools for statistical modelling allowing the combination of uncertainties from different sources. Deep cooperation with diagnosticians result in error analyses which allow for a quantitative comparison of data from different diagnostics. The feasibility of this approach was shown for a Thomson scattering diagnostics. On the basis of this thorough error analysis, the uncertainties due to combination of different diagnostics sources becomes quantifiable. Again, the Bayesian framework allows for a rigorous combination of different data. Realisation of that approach is the next step towards an *integrated data analysis*.

References

[1] G. Cordey, D. McDonald, et al., *private communications*

[2] J.J. Jeffreys, *Theory of Probability*, Oxford University Press, 1939.

[3] D.S. Sivia, *Data Analysis – A Bayesian Tutorial*, Clarendon Press, 1996.

[4] R. Fischer, C. Wendland, A. Dinklage, et al., Plasma Phys. Contr. Fusion (accepted).

[5] K. Mc Cormick, et al., Phys. Rev. Lett. (2002) (at press).

[6] H. Bindslev, Rev. Sci. Instrum. 70, 1093 (1999).

[7] Steffen L. Lauritzen, *Graphical Models*, Oxford University Press, 1996.