Integrated Approaches in Fusion Data Analysis

A. Dinklage*, R. Fischer†, H. Dreier*, J. Svensson* and Yu. Turkin*

*Max-Planck-Institut für Plasmaphysik, EURATOM-Association,
Wendelsteinstr. 1, 17491 Greifswald, Germany,
†Boltzmannstr. 2, 85748 Garching, Germany

Abstract. The concept of integrated data analysis in nuclear fusion requires the linkage of data and physical information. Summarizing the key steps for the analysis of transport in the core plasma, benefits of probabilistic modelling of single diagnostics are discussed. Concepts for full diagnostics models consisting of several diagnostics modules and linkage through mapping procedures are given in figures of Bayesian graphical models. Coupling to theory codes is demonstrated by the error estimation of neoclassical error analysis allowing a quantitative physical model validation. As an inverted use of the integrated data analysis approach, goals for the design of diagnostics and *sets* of diagnostics (meta-diagnostics) are outlined.

INTEGRATED DATA ANALYSIS IN PLASMA PHYSICS

Introduction

Data analysis in magnetic fusion research, a field in plasma physics, is inference for a large, complex system. Size is given by several dozens up to even hundreds of different measurements, each of them are afflicted by a large number of uncertain nuisance parameters. Complexity arises in fusion data analysis in a sense that a large number of interdependent measurements are to be combined. Synergistic effects then result, e.g., in enhancing significance even in quantities not directly measured through correlations (see next Section). The final goal of fusion research, however, can be summarized within one general statement: it is the demonstration of a sustainable, safe source of energy with nearly unlimited resources.

What is the relationship of this abstract goal to data analysis? For the demonstration of a fusion power plant, a number of demanding physical issues are to be resolved[1] which are also of general physical interest because the understanding of hot plasmas needs to combine several disciplines in modern physics. The demonstration of fusion relevant parameters by validated experimental investigations requires the combination of many heterogeneous information sources and physical models.

---

[1] An example is the understanding of transport mechanisms which lead to enhanced particle and energy degradation ultimately affecting the estimated size of a burning plasma. A second example is the investigation of plasma behavior on different time scales, where the largest time scales relevant for quasi-stationary operation (magnetic configuration relaxation time $\tau_{L/R} \approx 20\,\mathrm{s}$ and thermal equilibration time scale $\tau_{\mathrm{eq}} \approx$ several min, figures are estimated for W7-X) become accessible for fusion relevant parameters in large devices only.

Selected topics from data analysis in magnetic fusion research are summarized aiming at the preparation of data analysis modules for Wendelstein 7-X (W7-X) [1], a stellarator device presently under construction in Greifswald, Germany. The selection of subjects is motivated from one of the key issues in hot plasma physics, i.e., the interpretation of energy and particle transport from experimental investigations of the plasma core. This analysis is usually done sequentially in several steps and, although, error propagation is possible in principle, it is frequently difficult to be performed due to complicated interdependencies. These interdependencies also allow to increase the significance of the outcome and can serve for consistency checks and validation of physical models. The framework of Bayesian probability theory meets with the requirements of data linkage and consistency checks because probability theory allows one to incorporate additional information straightforwardly. This leads to enhanced significance of parameters to be inferred. In contrast to the usually employed Gaussian statistics (with independence assumptions), the Bayesian framework also allows for a treatment of so-called systematic errors. Throughout this paper several examples will be discussed which demonstrate access to formerly unaccessible parameters. In addition, uncertainties of derived, not directly measured quantities can be determined. In this paper focus is laid on coupling of modelling codes to data analysis (which will be motivated in the next section).

Particular interest in the integrated approach arises for burning plasma experiments due to limited diagnostics access and for the optimization of *sets* of diagnostics forming *meta-diagnostics* as it will be desirable under the conditions of such experiments [2].

An integrated approach in a sense that data are directly used for physical model selection is far beyond the present status of our research. But a demonstration of central steps showing the realisability of this concept can already be given. Nonetheless, the required number of data analysis models is large [3] (about 20 units projected as the start-up set for W7-X). Moreover, the benefits of the probabilistic approach for the design of specific diagnostics units can be used for diagnostics performance optimization under some constraints [4]. Probability theory offers several advantages which turn out to be beneficial for data analysis on different stages as well as for the design of diagnostics. The major benefit for fusion research is a concise combination of uncertain data — ultimately leading to the use of interdependencies.

After a discussion of the general background, examples in fusion data analysis are discussed along with considerations for the design of meta-diagnostics. Since the scope of this conference is devoted to probabilistic methods for analysis, particular plasma physical remarks are left to footnotes.

Topics in fusion data validation

An overview of typical steps to be performed in transport analysis is summarized in the flow chart Fig. 1.

What makes data analysis in magnetic fusion particularly demanding? Besides the aforementioned size and the complexity of data analysis also the proper inclusion of

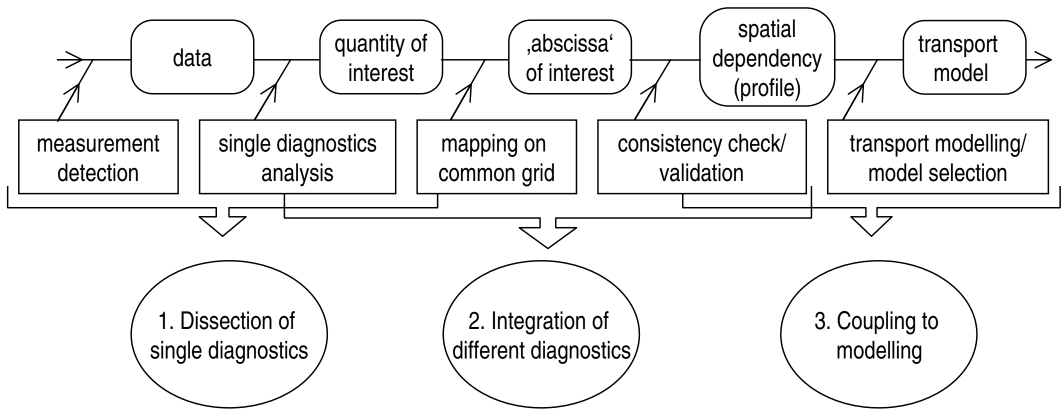

FIGURE 1: Flowchart for interpretative transport analysis for the core plasma in magnetic fusion devices. The rounded boxes represent (processed) data, the boxes indicate processing actions. Please note that edges and nodes of this diagram may represent multiples of data and processing steps. For the purpose of clarity, feeding back edges indicating iterations are omitted. The ellipses correspond to proof-of-principle steps discussed in this paper.

physical considerations leads to iterations including several modelling results[2]. But even the outcome of individual diagnostics frequently requires detailed physical modelling[3].

As an essential prerequisite for many steps shown in Fig. 1, the coupling of modelling to data analysis is required on different stages of data analysis and the following list summarizes different motivations for this step.

- **Diagnostic methods** themselves may require a detailed physical modelling (see footnote 3).
- The **linkage of different measurements** from the core of the plasma is given by coordinate transformations onto a common grid given by the surfaces of constant magnetic flux, which are a prerequisite for confinement. These surfaces are affected by the plasma through currents flowing in the plasma itself and require the determination of the magneto-hydrodynamics equilibrium [4].

---

[2] As an example for interdependent analysis steps, the validation of profile data, i.e. spatially resolved data resembling those profiles and typically approximated by parametric models for further analysis (typically temperature and density), can be considered. In order to link plasma core profile data from different lines of sight, the outcome is to be mapped on a common grid given by the magnetic surfaces. These surfaces depend on the temperature and the density in turn. Moreover, the shape of the surfaces is significantly affected by internal plasma currents, which needs to be modelled with profiles as input data.

[3] An example is the analysis of data from electron cyclotron emission which is usually approximated by the assumption of local black body radiation. However, in low density discharges, radiation transport of that radiation needs to be considered; the assumption of local emission leads to non-trivial systematic uncertainties. A different example is the effect of non-thermal or non-isotropic electron distribution functions deviating from the usually assumed Maxwellian distribution functions.

[4] Which is the solution of the force balance equation $\nabla p = \vec{j} \times \vec{B}$ (where p, $\vec{j}$, $\vec{B}$ are the plasma pressure, the current density and the magnetic field, respectively). A detailed solution may become very complicated for the three dimensional case as required for W7-X.

- **Access to barely measurable quantities** which have to be derived from analysis steps requires appropriate treatment of uncertainties[5].
- The transport analysis can be supplemented by observations acting in the sequential analysis as in- and outcome (see Section ’Coupling of modelling to data analysis’). A systematic analysis of **error propagation** (from observations to the outcome) in typically used transport **modelling** codes (e.g. particle and energy fluxes) is frequently hard to be done due to complex interdependencies.
- The capability of **quantitative model selection** of physical hypotheses, which can be treated within the Bayesian probability theory.

APPLICATIONS OF THE BAYESIAN PROBABILITY THEORY ON DATA VALIDATION

Examples of successful applications of Bayesian probability theory in fusion research are given in Refs. [5, 6, 7]. The goal of the research discussed in this paper are proof-of-principle studies; each of which benefits from a probabilistic treatment. Moreover, a possible link of those steps appears to become possible since the integrated analysis of several diagnostics is in the probabilistic framework a formal step. The work is strictly guided by the requirements resulting from physical questions to be resolved with W7-X.

Dissection of single diagnostics set-ups

The prerequisite for integrated approaches are reliable error analyses of the atomic data sources, namely diagnostics units. Consequently, a systematic statistical modelling of a Thomson scattering system[6] as an example was performed [8].

Along with the demonstration of practicability (the implementation was validated with previous analyses and the software was used for the final experimental campaign of Wendelstein 7-AS, predecessor of W7-X), the full error treatment allows to quantify correlations of density (n_e) and temperature (T_e) results. In particular, it is the correlation of the parameter uncertainties which became accessible by the consequent probabilistic modelling. This correlation represents the fact, that the data are functions of both the density and the temperature. Within the results of analyses (see, e.g., Ref. [9]), these correlations appear as probability density functions along isobars, i.e. along the product $n_e \times T_e$.

The probabilistic model is presently used for studies aiming at the design of a future Thomson scattering system on W7-X [4]. An additional benefit of the probabilistic model for the diagnostician are the quantitative assessment of all nuisance parameters which were used for diagnostics hardware improvement.

---

[5] An example is the determination of internal current densities, such as the bootstrap current density to be validated with neoclassical modelling.

[6] Thomson scattering is a technique to measure the electron density and the electron temperature by means of laser scattering on electrons in a hot plasma.

Consequent modelling of single diagnostic units is the elementary step of integrated data analysis. It is worth to be mentioned, that the major workload is required for setting up the single diagnostic models requiring detailed exploration of the error statistics of nuisance parameters.

Using interdependencies: Integrated data analysis

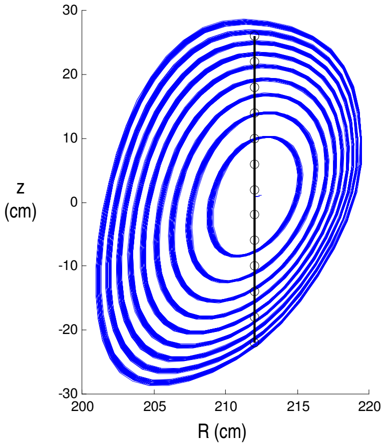

FIGURE 2: Cross section through magnetic surfaces of a configuration of W7-AS. The uncertainty of the flux surfaces is indicated by Markov-Chain Monte Carlo samples of the representing surfaces. Circles indicate the position of spatial channels of the Thomson scattering system.

Once probabilistic models of single diagnostics become available, the combination of different results is an obvious issue. Within our proof-of-principle studies this was done in two steps. First, a single spatial channel of the Thomson scattering system was combined with additional measurements referring to this position in the plasma. Second, the joint analysis of several diagnostics acting on different lines of sight or integration volumes within a given magnetic configuration was performed [10]. The latter step is crucial for further analyses and is particularly difficult because it requires methods for data consistency checks. It is important to note, that inconsistencies may be due to systematic uncertainties, but also frequently due to inappropriate physical models of the measurement. It is a frequent misunderstanding that probability theory resolves physical issues rather than being the tool of choice to resolve inconsistencies.

The result of the first step – the combination of Thomson scattering data with temperature data from a different diagnostics – demonstrates the benefit of the integrated approach through inclusion of correlations. Additional temperature measurements were used as prior probability density for the Thomson scattering model [11]. As to be expected for an independent (traditional) treatment of errors in density and temperature, the inclusion of temperature information enhances the significance of the combined T_e measurement. But – as a benefit from the use of interdependencies – also the uncertainty in density is diminished due to non-linear (isobaric) $n_e - T_e$ correlation. This result cannot be obtained in a (traditional) independent treatment of the uncertainties; the Bayesian way is straightforward by assignment of the prior.

The result of the second step – the analysis of a *’full diagnostics model’* of (the mothballed predecessor device of W7-X,) W7-AS, linking Thomson scattering, interferometry and diamagnetic measurements) – gives access to further model uncertainties. Methodologically, the linkage of different models of diagnostics was performed through Bayesian graphical models, in which different plates (modules) represent diagnostics models. Technically, the models can be linked by network protocol wrapped software modules (WebServices) allowing the use of already well validated software packages [10]. In that respect, our approach has to differ from the `AutoBayes` package [12] which generates codes from graphical models directly. And due to the required workload, the inclusion of sophisticated modelling codes is hardly to be expected to be formulated in an appropriate graphical model.

Within the W7-AS model, different diagnostics are combined and the linkage of different models is given by a geometric relationship through mapping on toroidal magnetic surfaces described by the magnetic configuration. A toroidal cross section, a Poincaré plot of field lines, is shown in Fig. 2 for the toroidal position of the Thomson scattering diagnostic. Since stellarator devices (like W7-AS and W7-X) have a three dimensional magnetic topology, the cross section at different toroidal positions alters making a transformation of diagnostics results from laboratory coordinates to flux surfaces necessary. Then data can be linked under the assumption that the temperature and density are constant on flux surfaces.

Since the magnetic configuration depends on the measured quantities as well, the uncertainties of that data results in uncertainties of the mapping. The width of the uncertain flux surfaces is shown in Fig. 2 and demonstrates a possible way to access the uncertainty of flux surfaces which is of obvious interest for the determination of transport.

But this example is not the most important benefit of the proposed approach. Since any parameter of the diagnostics is accessible, the impact of any uncertainty can be determined quantitatively within the model. This outcome is a valuable tool for detecting inconsistencies. Moreover, the capability to treat systematic uncertainties allows one to quantify the degree of consistency of mismatching measurements, e.g. how much line integrals from interferometry differ from integrated spatially resolved measurements. Again, physical inconsistencies requiring reformulation of the model of the measurement cannot be resolved by the analysis of a given model, but the probabilistic results of which can be used for answering the question, how far away the interpretation of data is from reality.

Coupling modelling to data analysis

In the previous section, the inclusion of modelling results is a crucial step in plasma physics data analysis. Here, as an example for the coupling of modelling to data analysis, particle and heat transport analysis from experimentally derived profiles is investigated.

For this purpose, density and temperature observations jointly mapped on magnetic coordinates are considered. Moreover, observations of the radial electric field were available as well and it is the goal of a 'super-fit' to determine the particle and heat fluxes most consistent with *all* observations which enter the thermodynamic forces $(T(r) \rightarrow \nabla T,\, n(r) \rightarrow \nabla n,\, E_r(r))$.

The usual procedure is to represent density and temperature data as a parametric fit to the density and temperature observations, then to pass the data to a transport analysis code and to compare the outcome with additional observations; here with measurements of the radial electric field. The latter step is regarded as a validation step which has formally no consequence for the actual input data passed to the transport code. Hence, the sequential results are the fluxes for independent maximum likelihood parametric models for the density and temperature data. Those results are bold lines in Fig. 3.

The reference for comparison is the dotted line which indicates the fluxes derived from all external sources. These fluxes are derived independently from our analysis.

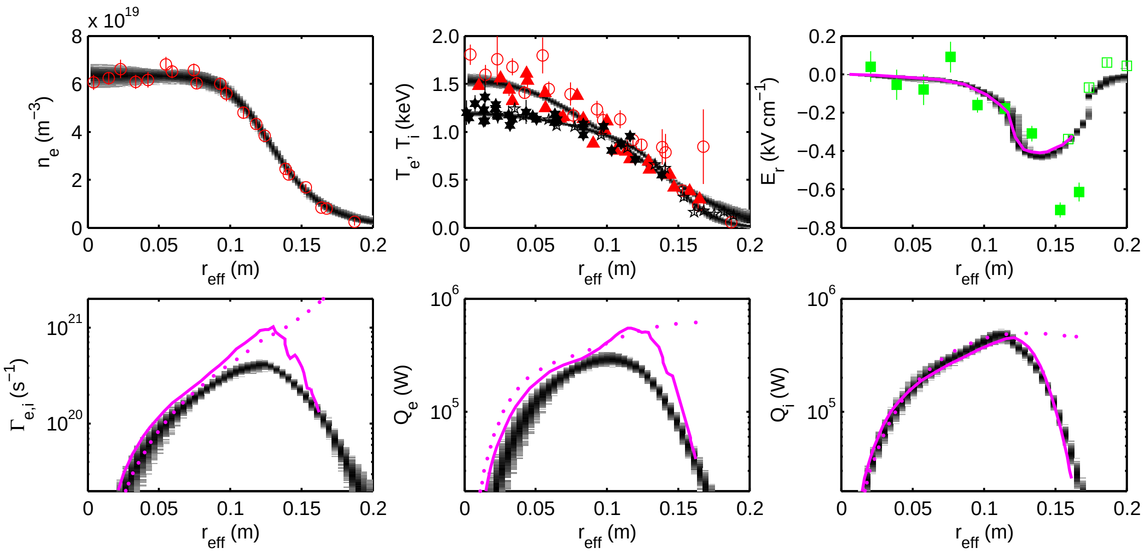

FIGURE 3: Results of a neoclassical transport analysis on W7-AS data. The shaded patches represent histograms of MCMC posterior samples of an integrated observation descriptive model. More details are to be found within the text.

The integrated approach uses all data within the fitting procedure, particularly including the radial electric field. Hence, the consistency of data is reflected directly by the outcome. Fig. 3 (the shade of the color represents the log posterior) shows a posterior probability sample employing the same parametric model as used in the sequential approach. Sampling also indicates the uncertainty of the fitted model and data consistency checks can be quantitatively performed.

E.g., some of the observations of the radial electric field at about $r_{\mathrm{eff}} \approx 0.16\,\mathrm{m}$ (cf. Fig. 3) are clearly outside the error margins. Second, the electron temperature (upper curve in the T_e, T_i plot in Fig. 3) appears to be systematically too low (but within the observation error margins) either indicating systematic overestimation of Thomson scattering data (circles), or too small error bars of electron cyclotron measurements (triangles), or systematic mismatch of temperature measurements due to their measurement principle[7], or an inappropriate transport model.

A discussion of the latter issue can be done with reference results of heat and particle fluxes (dotted lines in Fig. 3) which indicate that the underlying physical and data descriptive (fitting) model are consistent with the observations in the plasma core (up to $r_{\mathrm{eff}} < 0.10\,\mathrm{m}$ which is the plasma core) although a small systematic underestimation of fluxes appears even there. For larger radii, however, the model fails to describe the reference data. This outcome is in agreement with physical considerations because the transport model used here considers neoclassical transport only, whereas anomalous transport mechanisms known to act at the plasma edge are not included.

The physical conclusion to be drawn from the results displayed in Fig. 3 is that the

---

[7] Assuming for both analyses Maxwellian energy distributions, which are probed at different energies for the two measurements.

anomalous transport at the edge of the analyzed W7-AS discharge does barely affect the neoclassical electric field and anomalous transport processes are expected to be ambipolar. As an outlook, models for anomalous transport will be incorporated aiming at the comparison of physical models by means of Bayesian model selection techniques.

Design of meta-diagnostics

Apart from different data analysis issues, the data analysis models are also explored with respect to design capabilities. A discussion of results for the Thomson scattering module is presented at this conference [4], in this paper the analogy of the concept of integrated analysis on design is outlined.

Appropriate groups of diagnostics are said to form a meta-diagnostic and the design of which is an extension of single diagnostics design aiming at the use of interdependencies. The combination of benefits of set-ups and the compensation of single set-ups drawbacks is to be achieved by design criteria including robustness (e.g. in harsh environments) or effort. It is the goal of the meta-diagnostics design, to define utility functions (as an exemplary reference we would like to refer to [13]) oriented at physical questions to be resolved. An example differently discussed in this paper is the analysis of transport which requires accurate measurements of $\nabla T/T$, $\nabla n/n$ and E_r at those positions where these quantities have the most relevant impact. Expectations for those regions are to be derived from artificial data. The beginning of that studies, however, is the design of smallest possible sets of diagnostics in order to evaluate the information gain to be expected[8].

CONCLUSIONS

Concluding, methods for an interpretative integrated modelling were demonstrated and elementary steps, such as the probabilistic modelling of single diagnostics and the linkage of different diagnostics were discussed. The integration is based on the Bayesian probability theory. Applicability of the methods was tested on W7-AS data in order to prepare data analysis tools for Wendelstein 7-X. The Bayesian framework allows one to include physics results at different stages of complexity of data analysis.

ACKNOWLEDGMENTS

Helpful discussions with C.D. Beidler, J. Geiger, S. Gori, H.-J. Hartfuß, M. Hirsch, H. Maaßberg, R. Preuss, U. v. Toussaint and A. Werner are gratefully acknowledged.

---

[8] A possible example is the combination of the spatially resolved measurement of electron density by Thomson scattering, which may suffer from calibration, with robust but line integrated interferometry measurements.

The authors would like to express their gratitude to Prof. V. Dose - his inspiration and pioneering work for fusion applications are the prerequisite for the research communicated in this paper.

REFERENCES

1. Grieger, G., Lotz, W., Merkel, P., Nührenberg, J., Sapper, J., Strumberger, E., Wobig, H., Team, W.-X., Burhenn, R., Erckmann, V., Gasparino, U., Giannone, L., Hartfuß, H. J., Jänicke, R., Kühner, G., Ringler, H., Weller, A., Wagner, F., and Team, W.-A., *Phys. Fluids*, **B4**, 2081–2091 (1992).
2. Fischer, R., and Dinklage, A., *Rev. Sci. Instrum.*, **75**, accepted for publication (2004).
3. Hartfuss, H.-J., Brakel, R., Endler, M., Geist, T., Grigull, P., Hofmann, J., Junker, J., Kick, M., Kühner, G., Niedermeyer, H., Ringler, H., Teubel, A., Wagner, F., and Weller, A., *Rev. Sci. Instrum.*, **68**, 1244–1249 (1997).
4. Fischer, R., "Bayesian Experimental Design," this conference.
5. Preuss, R., Maraschek, M., Zohm, H., and Dose, V., "Bayesian analysis of magnetic island dynamics," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by C. Williams, AIP, Melville, NY, 2003, vol. CP659, pp. 124–131.
6. Dose, V., Neuhauser, J., Kurzan, B., Murmann, H., Salzmann, H., and Team, A. U., *Nuclear Fusion*, **41**, 1671–1685 (2001).
7. Dose, V., Preuss, R., and von der Linden, W., *Phys. Rev. Lett.*, **81**, 3407–3410 (1998).
8. Fischer, R., Wendland, C., Dinklage, A., Gori, S., Dose, V., and the W7-AS Team, *Plasma Phys. Control. Fusion*, **44**, 1501–1519 (2002).
9. Dinklage, A., Fischer, R., and Svensson, J., *Fus. Science Techn.*, **46**, September (2004).
10. Svensson, J., Dinklage, A., Geiger, J., Werner, A., and Fischer, R., *Rev. Sci. Instrum.*, **75**, accepted for publication (2004).
11. Fischer, R., Dinklage, A., and Pasch, E., *Plasma Phys. Control. Fusion*, **45**, 1095–1111 (2003).
12. Fischer, B., Hajian, A., Knuth, K., and Schumann, J., "Automatic Derivation of Statistical Data Analysis Algorithms: Planetary Nebulae and Beyond," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by G. Erickson and Y. Zhai, AIP, Melville, NY, 2004, vol. CP707, pp. 276–296.
13. T.Loredo, "Bayesian Adaptive Exploration," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by G. Erickson and Y. Zhai, AIP, Melville, NY, 2004, vol. CP707, pp. 330–346.