**Integrated Data Analysis – Probabilistic Data Fusion in Fusion**

*Andreas Dinklage*
*Max-Planck-Institut für Plasmaphysik, Greifswald*

In fusion research, data analysis typically needs to combine different information at different levels of uncertainty: experimental data, physics quantities, constraints and technical parameters enter the analysis process which must result in a picture compliant with physics considerations. Beyond the combination, weighting different sources of information is a prerequisite to derive errors and to assess the quality of measurements to be sufficient to verify or to falsify physical models.

As for any experimental science, errors are mandatorily required for any proper analysis of experimental data. An additional specific issue in fusion research is the combination of spatially and temporally heterogeneous measurements. Those result in global figures (e.g. stored energy), line integrated data (e.g. from interferometer signals), local measurements (e.g. Thomson scattering or charge-exchange recombination spectroscopy) or combinations of which. A second specific feature of fusion diagnostics results from the *remote-sensing* character of measurements – many measurements in fusion do not directly result in physics quantities but require a plasma physics model for the interpretation of signals (e.g. radiation transport in electron cyclotron emission). In magnetic confinement fusion, spatially heterogeneous measurements are mapped on common coordinates. A usual approach for the plasma core is to use magnetic equilibria which potentially need to be inferred from the data which are to be mapped. Moreover, the derivation of transport relevant gradients also requires the correct inclusion of proper coordinates.

In order to treat typical problems in the analysis of fusion data, the Bayesian interpretation of probability theory is a mathematical framework which allows one to treat consistently the aforementioned issues. Bayesian approaches consider probability as a description of uncertainties. Conceptually simple, probabilistic data analysis relies on *Bayes theorem* which reads - applied on the interpretation of data $d$ to retrieve parameters $p$ (incl. the quantities to be measured with background considerations (summarized by $I$)):

$$P(p|d,I) \propto P(d|p,I) \times P(p|I)$$

to result in the conditional (so-called *posterior*) probability $P(p|d,I)$. This outcome of the analysis is gives the most likely measured value $p$ given the measurement $d$. The uncertainty encoded in $P(p|d,I)$ is reflected by the shape of the posterior and reduced errors can be retrieved by marginalized distributions. A specific strength of Bayesian analysis results from the fact that any quantity entering the measurement can be coherently considered to be uncertain. The uncertainty of measurements is encoded in the likelihood $P(d|p,I)$ which involves a data descriptive forward model $D = f(p)$ (synthetic diagnostics). The Integrated Data Analysis approach[1] aims to quantify the prior probabilities $P(p|I)$ with physics information.

The lecture introduces concepts and applications of Integrated Data Analysis employing Bayesian analysis methods. Beginning with the analysis of single diagnostics, the lecture describes the integrated approach and introduces the concept of *meta-diagnostics* as a superset of single diagnostics units. The lecture also gives a first guidance to numerical techniques typically applied in probabilistic analysis approaches.

[1] R. Fischer, A. Dinklage, E. Pasch, Plasma Phys. Control. Fusion **45**, 1095 (2003)