Integrated Data Analysis to expand measurement capability

Whitepaper submitted to DOE Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences
Primary topic: F (Data management, analysis, and assimilation)
Secondary topic: C (Whole device modeling, especially validation)
Oral presentation requested if time available

D. J. Den Hartog, M. E. Galante, L. M. Reusch, M. D. Nornberg, and the MST Team
University of Wisconsin–Madison, djdenhar@wisc.edu, April 2015

Challenge:	Data produced by large fusion experiments such as ITER will be simultaneously more valuable (a burning plasma) and more difficult to obtain (limited diagnostic set operating in a nuclear environment).
Approach:	Integrated Data Analysis (IDA) provides a framework to deal with measurement limitations, and maximizes the usefulness of the information recorded by a set of diagnostics.
Impact:	Extraction of maximum scientific information from each measurement will enable the complete and rigorous comparison of experiment to simulation characteristic of a validation effort.
Recommendation:	Support development of IDA expertise, with particular emphasis on new methods of multi-diagnostic and multi-parameter analysis. This will require personnel and project support, and computational resources similar to that required for simulation work.

Measurements enable understanding of a physical system. But experimental measurements often have shortcomings, shortcomings that become particularly apparent when the measurement result is applied to a validation effort. Measurements are always limited in scope, scale, and resolution. Uncertainties are always present, and while statistical uncertainties are often estimated, systematic uncertainties are usually much more difficult to quantify. In addition, models are often required to transform raw data into useful measurements, implying that such models must themselves be understood and validated.

Integrated Data Analysis (IDA) provides methods to deal with measurement limitations, and maximizes the usefulness of the information recorded by a set of diagnostics.[1] To achieve the most useful and most rigorous comparison of experiment to simulation, IDA should be developed as an integral part of a validation effort. The goal of IDA is to combine data from heterogeneous and complementary diagnostics, considering all dependencies within and between diagnostics, in order to obtain the most reliable results in a transparent and standardized way. IDA exploits the redundancy of complementary diagnostics to help resolve data inconsistencies and maximize the value of experimental measurements to the validation process.

In addition, IDA enables formation of "meta-diagnostics," which combine information from various instruments to produce unique measurements. For example, standard single-instrument techniques to measure Z_{eff} on the Madison Symmetric Torus (MST) all failed to provide robust and reliable measurement. However, data from many instruments contain some information about Z_{eff} (e.g., soft x-ray (SXR) tomography, visible and x-ray spectroscopy, Thomson scattering background, neutral beam attenuation, direct CHERS measurements of impurity densities, *inter alia*). IDA techniques have recently been applied to combine information from two diagnostics (CHERS and SXR) to produce a most probable estimate of Z_{eff}.[2]

As implemented on MST and elsewhere, IDA is accomplished using a Bayesian framework. This framework enables combination of information from complementary diagnostics in a rigorous way, and production of the most probable estimate of a physical quantity. IDA proceeds with a series of steps:

- Identify uncertainties and quantify with probability distribution functions (PDF)
- Combine all relevant information within a probabilistic framework

  - include diagnostic models and prior knowledge
  - develop a forward model for measurement
  - marginalize out nuisance parameters such as systematic effects
- Search parameter space, which is often high-dimensional
- Final result is the marginal posterior PDF of the quantity of interest
  - mean value of PDF is estimate of parameter of interest
  - variance of PDF is a measure of the uncertainty of this parameter

Figure 1 shows an example measurement of Z_{eff} in the core region of MST, along with the one-sigma uncertainty. Quantitative model discrimination is also part of IDA, and has been applied to determine that the effect of potential unknown low-*Z* and mid-*Z* impurities on this measurement is likely minimal.

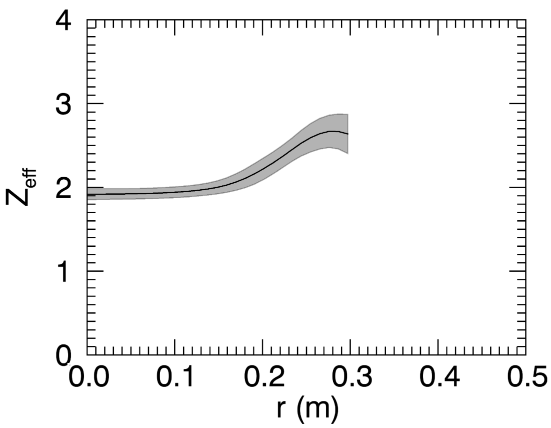

Fig. 1. Most probable radial profile of Z_{eff} in the core of MST, produced by integrating data from CHERS and SXR tomography measurements using the techniques of IDA. Lack of SXR data outside of $r/a = 0.3$ currently constrains the measurement to the core region; it is not an inherent limit of the technique.

Integrated Data Analysis requires both computational and instrumentation knowledge, often requiring an individual to straddle experiment and simulation, and is an "analyst" skillset likely to be in increasing demand as fusion experiments grow in size and complexity. In particular, the severe limitations imposed by operation of diagnostics in fusion nuclear environments means that maximum scientific value must be extracted from each measurement. Data produced by large fusion experiments is extremely valuable and difficult to obtain (e.g., ITER[3]). Techniques such as IDA may even be critical to real-time feedback ITER operation, as individual diagnostics may not always provide the measurements necessary to maintain plasma stability.

The developing techniques of IDA have already been applied to data from TJ-II,[4] ASDEX,[5] JET,[6] and other experiments. It is one of the topics of a new series of IAEA Technical Meetings.[7] U. S. fusion scientists should increase engagement with this existing community of IDA experts, and contribute to it. Continuing development of analyst expertise should be supported, particularly at universities where analysts will be trained. Where appropriate, connections should be made to other implementations of IDA techniques, such as planetary search in the space science community. Integral to this effort in the fusion context must be increased support for diagnostic development and operation, as high-quality raw data is the essential input to a fusion science IDA effort. IDA projects should emphasize development of new methods of multi-diagnostic and multi-parameter analysis; this will require substantial computational support, similar in magnitude to that required for simulation work. Computational support will take two forms: large-scale computing power, and development of algorithms to efficiently search high-dimensional probability space. Finally, IDA technical development should be supported as a key part of a larger validation effort, and well-integrated into this effort.

---

[1] R. Fischer, A. Dinklage, E. Pasch, Plasma Phys. Controlled Fusion **45**, 1095 (2003).
[2] M. E. Galante, L. M. Reusch, D. J. Den Hartog, P. Franz, J. R. Johnson, M. B. McGarry, M. D. Nornberg, and H. D. Stephens, submitted to *Nuclear Fusion*.
[3] A. J. H. Donné *et al*., Nucl. Fusion **47**, S337 (2007).
[4] B. Ph. van Milligen, T. Estrada, E. Ascasíbar, D. Tafalla, D. López-Bruna, A. López Fraguas, J. A. Jiménez, I. García-Cortés, A. Dinklage, and R. Fischer, Rev. Sci, Instrum. **82**, 073503 (2011).
[5] R. Fischer, C. J. Fuchs, B. Kurzan, W. Suttrop, E. Wolfrum, and ASDEX Upgrade Team, Fusion Sci. Technol. **58**, 675 (2010).
[6] J. Svensson, O. Ford, D.C. McDonald, A. Meakins, A. Werner, M. Brix, A. Boboc, M. Beurskens, and JET EFDA Contributors, Contrib. Plasma Phys. **51**, 152 (2011).
[7] IAEA First Technical Meeting on Fusion Data Processing, Validation and Analysis (Nice, France, 2015); http://irfm.cea.fr/TMFDPVA15/.