PREPARED FOR THE U.S. DEPARTMENT OF ENERGY,
UNDER CONTRACT DE-AC02-76CH03073

PPPL-3880
UC-70

PPPL-3880

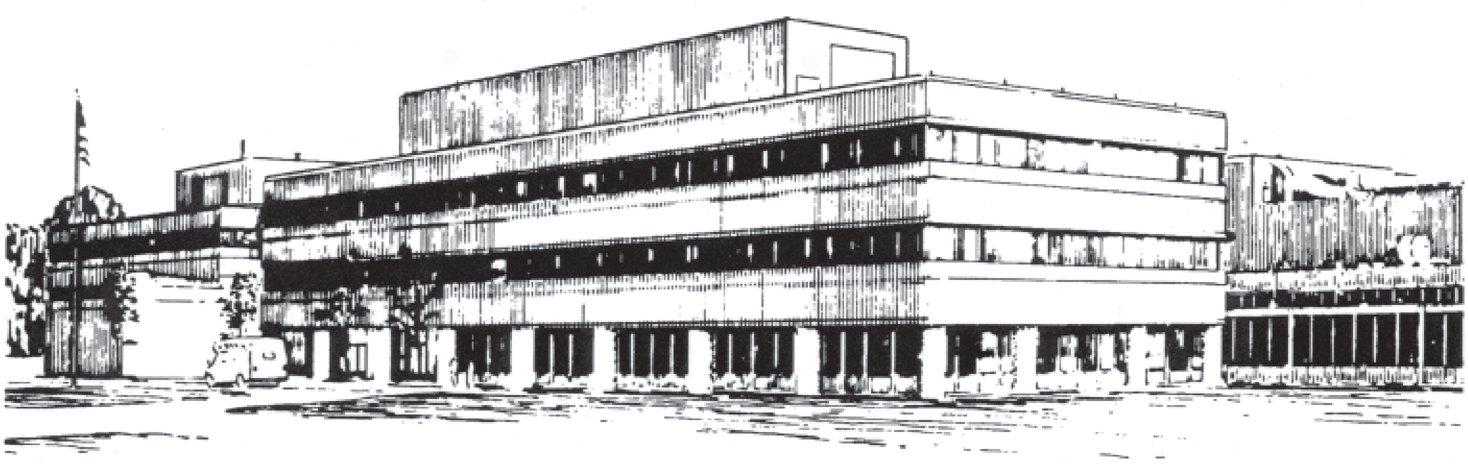
DbAccess: Interactive Statistics and Graphics for Plasma Physics Databases

by

W. Davis and D. Mastrovito

October 2003

PPPL
PRINCETON PLASMA PHYSICS LABORATORY



PRINCETON PLASMA PHYSICS LABORATORY
PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY

DbAccess: Interactive Statistics and Graphics for Plasma Physics Databases

W. Davis*, D. Mastrovito
Princeton Plasma Physics Laboratory, P.O. Box 451, Princeton, NJ 08543-0451

Abstract
DbAccess is an X-windows application, written in IDL®, meeting many specialized statistical and graphical needs of NSTX Plasma Physicists, such as regression statistics and the analysis of variance. Flexible "views" and "joins," which include options for complex SQL expressions, facilitate mixing data from different database tables. General Atomics Plot Objects add extensive graphical and interactive capabilities. An example is included for plasma confinement-time scaling analysis using a multiple linear regression least-squares power fit.

Keywords: Database; Data visualization; Plasma Physics; Regression; NSTX

1. Background

The National Spherical Torus Experiment (NSTX) began fusion experiments at the Princeton Plasma Physics Laboratory (PPPL) in February, 1999 [1]. During an experimental cycle, or "shot," a plasma is produced and over 200 megabytes of data are acquired from instruments in dozens of subsystems, hosted by Unix, VMS and Windows computers. A typical run day has around 40 shots. NSTX runs for between 12 and 20 weeks a year. Over 1 terabyte of raw and analyzed NSTX data currently resides on disk.

Most raw and analyzed data from NSTX are stored in MDSplus (Stillerman [2], Davis, [3]), a data acquisition and storage system used at several fusion facilities world-wide. Generalized tools allow straightforward access to data from individual shots. Batch jobs run between shots, or off-line, to calculate and write summary information, consisting of tens or hundreds of values per shot, to an Microsoft® SQL (MS-SQL) Server 2000 database.

2. Reasons for Using a Database

Plasma physics databases are used to explore the statistical relationships between various measured and/or calculated quantities. Examples include stability limit studies and energy confinement scaling (Christiansen [4]). Through regression analysis of database values, "scaling laws" are used to identify important aspects of fusion plasmas and to help design new fusion devices. "Global" parameters (characterizing an entire plasma, such as "maximum plasma current") are useful for confirming theoretical explanations of important phenomena. "Local" properties, such as gradients within a plasma, can be compared directly to theoretical predictions [Hoang [5]).

All of the raw data and most of the analyzed data for NSTX is accessible from disk through an MDSplus data server. Instead of keeping data in separate databases, we could simply read it from MDSplus, and recalculate the desired parameters before each analysis. One reason not to do this is the significant speed increases when retrieving processed summary information from our databases. For example, extracting the peak plasma current for 1000 shots directly from MDSplus takes about 4 minutes, but takes only a few seconds from a database. The space

* Corresponding author. E-mail: bdavis@pppl.gov, Phone: (609)-243-2546, Fax: (609)-243-3086

currently used by our databases is small compared to that of the raw data, so additional disk space is not an issue.

Another advantage of storing values in databases is that an expert can process the raw data and make decisions about inclusion and data validity. For example, for confinement-time scalings, only shots that reached a "steady state" are desired.

3. Reasons for Using DbAccess

Previous experiments at PPPL made extensive use of databases using the locally-developed LOCUS[6] and MINGL[7] programs. These powerful tools were written in FORTRAN, use a command line interface, and only produce Tektronix graphics, so their continued use and maintenance is unattractive. After data is entered into a database, commercial packages, such as EXCEL® or JMP® [8], can be used for sophisticated analysis, including the multiple linear regression shown later as an example. However, these packages can be time consuming to learn and awkward for routine, multi-step use. They also do not incorporate data-loading features.

DbAccess is an X-windows application, written in IDL® [9], which provides a point-and-click interface for many specialized statistical and graphical needs of NSTX Physicists, such as regression statistics and the analysis of variance. Multiple linear least squares fits used for confinement-time scaling and other analyses, involve just a few clicks in DbAccess. Constraints on data selection and graphical representations are straightforward. Flexible "views" and "joins" are possible, including the optional use of complex SQL expressions. IDL scripts may be generated automatically for loading databases from batch jobs. The integration of a plotting package from General Atomics (GA Plot Objects[10]) adds extensive interactive graphics capabilities (and documentation). DbAccess can use ODBC drivers (currently requiring IDL's Dataminer option), or routines included in the MDSplus distribution, to connect to a remote or local database.

4. Data Stored in NSTX Databases

There is no centralized control over the content of the physics database for NSTX, which makes it easy for users create tables and access data. A table named "contents" describes the other tables in the database and many of the table "columns" which may be of general interest. Some of the tables in the NSTX database are routinely accessed by many people, such as those used for the Logbook application; these include descriptive and evaluative comments about each plasma shot. Some researchers create their own database tables, and some specify algorithms for programmers to implement. The Survey table, for example, lists 20 general plasma parameters at 3 times-of-interest: 1) near the beginning of "ramp-up", 2) during the "flat-top," and, 3) at the time of the maximum plasma current (excluding times of disruptions or spikes). The EFIT table contains about 70 characteristic values from the common equilibrium fitting code, Efit [11], at the times of maximum plasma current, stored energy and beta (the ratio of plasma energy to magnetic field energy). The Haccess table contains about 80 columns relevant to H-mode studies [12]. About 20 other tables contain more specialized data. Databases can be loaded interactively or from files using DbAccess, but more typically, batch jobs containing IDL scripts write to a database table after each shot. All data is readable by everyone on our network.

5. Multiple Linear Regression Analysis Example

The following example examines the effect of various discharge parameters on NSTX energy confinement. After starting DbAccess at PPPL and double-clicking on the tau2002 database in the left-hand column, pulling down on the “Data Access” menu produces the widget in Fig. 1.

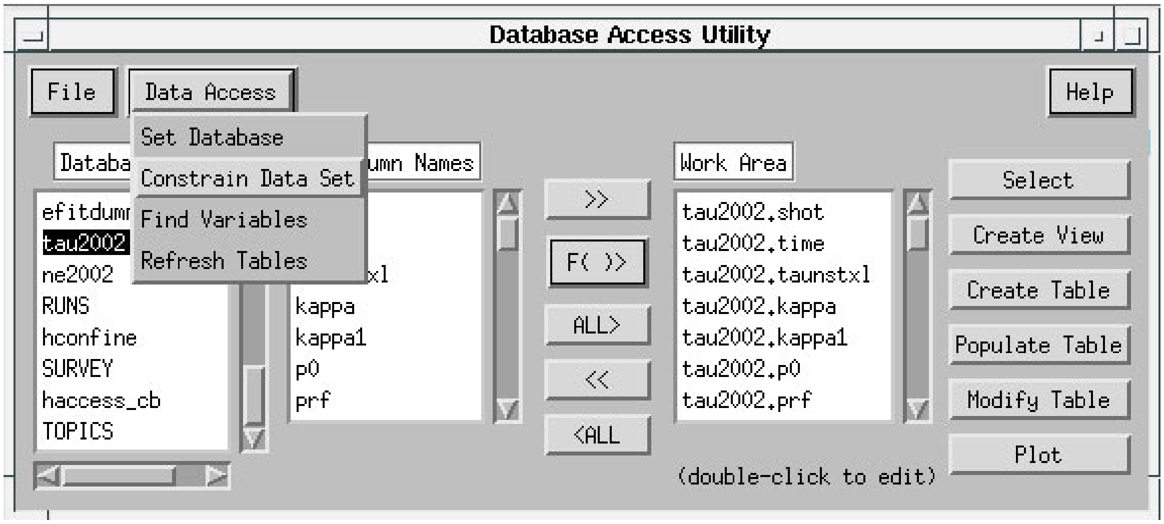


Fig. 1. DbAccess Startup Widget

Constraints on the data to be included may be selected from the widget shown in Fig. 2. In this case, L-mode plasmas with no RF power input are specified.

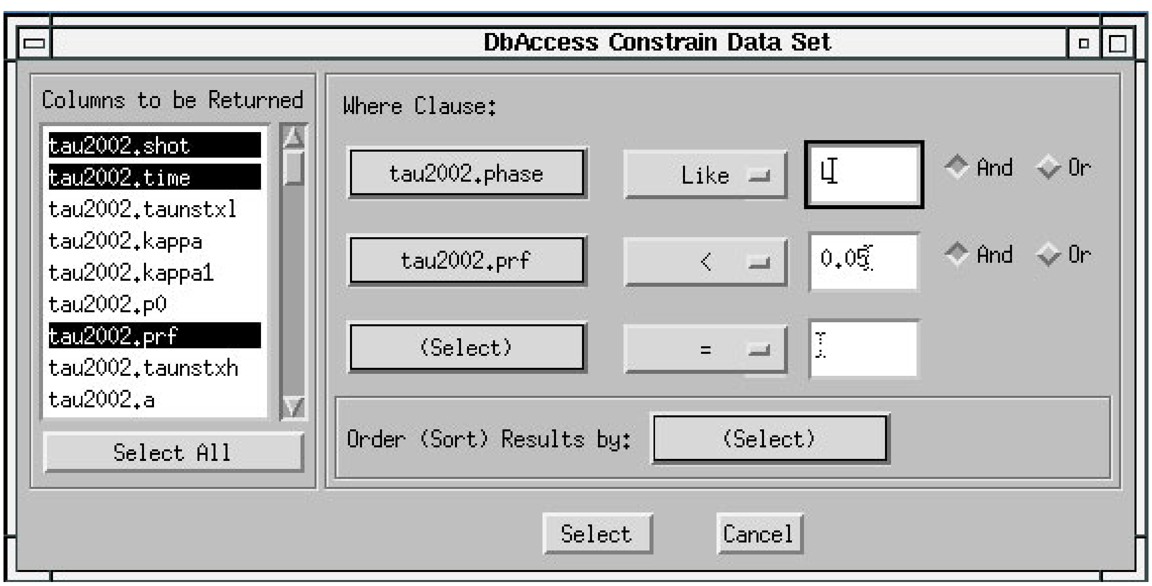


Fig. 2. DbAccess Constrain Data Widget

Clicking on Select in Fig. 2 results in a widget displaying a scrollable table of the data. From that widget, you may save the data to a file, print it, plot it, select and remove rows or click on “Analyze,” which brings up the “DbAccess Model” widget (Fig. 3).

The form of a multiple linear regression model is $y = a_0 + a_1*x_1 + a_2*x_2 + \ldots$ X_1, x_2, etc., are the “effects” chosen for the model (by selecting a parameter in the main list, and clicking on the Add button). Scaling laws in plasma physics tend to be of the form $y = a_0 * x_1^{a1} * x_2^{a2} \ldots$ This exponential equation can be linearized by taking the natural log of it before computing the multiple linear regression, and exponentiating the resulting equation afterwards. This occurs if the “Use Powers” box is checked in the DbAccess Model window.

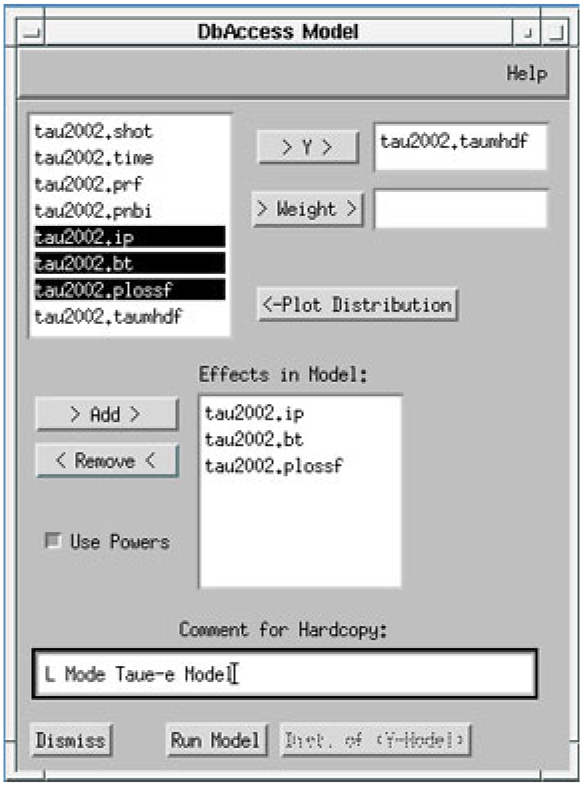
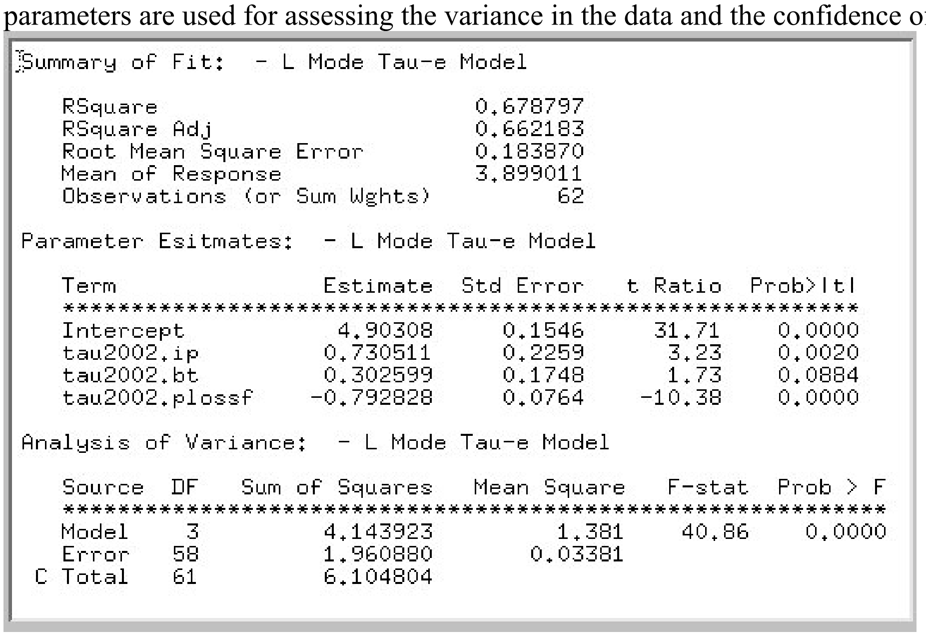


Fig. 3. DbAccess Model Configuration Widget

Clicking on “Run Model” in Fig. 3 pops up two windows. The output shown in Fig. 4 is similar to that produced from Excel® Data Analysis tools and JMP® [8]. These common statistical parameters are used for assessing the variance in the data and the confidence of the fit.


```
Summary of Fit:  - L Mode Tau-e Model

   RSquare                         0.678797
   RSquare Adj                     0.662183
   Root Mean Square Error          0.183870
   Mean of Response                3.899011
   Observations (or Sum Wghts)           62

Parameter Esitmates:  - L Mode Tau-e Model

   Term                Estimate  Std Error   t Ratio  Prob>|t|
   ***********************************************************
   Intercept            4.90308     0.1546     31.71    0.0000
   tau2002.ip          0.730511     0.2259      3.23    0.0020
   tau2002.bt          0.302599     0.1748      1.73    0.0884
   tau2002.plossf     -0.792828     0.0764    -10.38    0.0000

Analysis of Variance:  - L Mode Tau-e Model

   Source  DF   Sum of Squares   Mean Square   F-stat  Prob > F
   *************************************************************
   Model    3         4.143923         1.381    40.86    0.0000
   Error   58         1.960880       0.03381
 C Total   61         6.104804
```
Fig. 4. DbAccess Regression Statistics Example

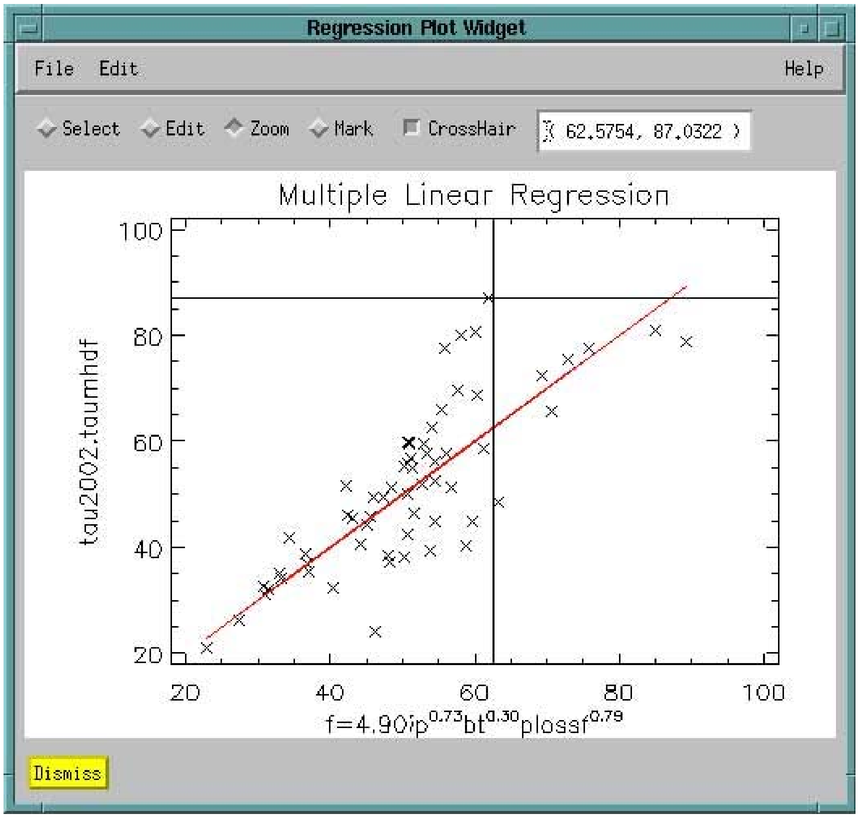


Fig. 5. DbAccess Regression Plot Example

The power law fit equation, $f = 4.0\ ip^{0.73}\ bt^{0.30}\ plossf^{0.79}$, the label on the X-axis in Fig. 5, is plotted against taumhdf from the database. The exponents derived indicate how important the respective quantities are for energy confinement. These values may be compared to those from other fusion devices.

The plot in Fig. 5 was produced using GA Plot Objects in IDL. Some coding changes were necessary, but the net gain in features and "friendliness" was enormous. For example, crosshairs may be used to select a point for which the x and y values are displayed in the white box at the top. The plot can be zoomed or resized with the mouse. After checking the Mark box, clicking on a data point will cause the corresponding row in the Selection table widget (not shown) to be highlighted. Virtually all IDL plotting options can be changed from various dialog boxes by selecting "Set Plot Appearance" under the Edit menu. See the DbAccess manual [13], or documentation on General Atomics' ReviewPlus [14], for more information.

6. Future Plans

We plan to remove the need for IDL's Dataminer when using ODBC drivers to connect to the database and possibly write our own IDL ODBC interface. We also hope to make the data-loading function more user-friendly so as to reduce the need for programmer assistance. We may add security features to prevent inadvertent changes to database tables.

7. Summary

Summary databases were used more frequently on early fusion experiments at PPPL than they are today. We hope that an easy-to-use tool that automates complex, but regularly used, operations, such as power law scaling, will empower our physicists to find insights into their data more readily. We also hope this work can benefit other laboratories, just as our use of the GA Plot Objects has benefited this project.

Acknowledgements

The authors thank Stan Kaye for specifying requirements for this application and for testing and providing valuable feedback. We thank Qian Peng from General Atomics for installing the GA Plot Objects at PPPL. This work was supported by DOE Contract DE-AC02-CHO3073.

References

[1] Kaye et al., The Physics Design of the National Spherical Torus Experiment. *Fusion Technology*, **36**, July 1999, p. 16, or http://nstx.pppl.gov/index.shtml.
[2] J.A. Stillerman, T.W. Fredian, K.A. Klare, G. Manduchi, MDSplus Data Acquisition System. *Review of Scientific Instruments* **68** (1) January 1997, p. 939.
[3] W. Davis, P. Roney, T. Carroll, T. Gibney, D. Mastrovito, The Use of MDSplus on NSTX at PPPL, *Fusion Eng. Des*. **60** (2002), 247-251.
[4] J.P. Christiansen, et al., Global Energy Confinement Database for ITER, *Nuclear Fusion* **32** (1992) 291-338.
[5] G.T. Hoang, C. Bourdelle, X. Garbet, G. Giruzzi, T. Aniel, M. Ottaviani, W. Horton, P. Zhu, and R.V. Budny, Experimental determination of critical threshold in electron transport on Tore Supra, *Phys. Rev. Lett.* **87** (2001) 125001.
[6] J.A. Murphy, R.M. Wieland, LOCUS Database System. *Review of Scientific Instruments* **59** (8) August, 1988.
[7] R.M. Wieland, et al., MINGL Database System for TFTR. *Review of Scientific Instruments* **59** (8) August, 1988.
[8] *JMP*® Statistics and Graphics Guide, SAS Institute Inc., Cary, NC, 1995, pp. 50-53.
[9] IDL® – The Interactive Data Language http://www.rsinc.com/idl/.
[10] J. Schachter, Q. Peng, D.P. Schissel, Data analysis software tools for enhanced collaboration at the DIII-D National Fusion Facility, *Fusion Engineering Design* **48** (2000) 91-98.
[11] S.A. Sabbagh, S.M. Kaye, J. Menard, et al., Equilibrium Properties of Spherical Torus Plasmas in NSTX, *Nuclear Fusion* **41** (2001) 1601.
[12] K. Thomsen, et al., The international global H-mode confinement database: storage and distribution, *Fusion Eng. Des*. **60** (2002), 347-352.
[13] The DbAccess Manual, http://nstx.pppl.gov/nstx/Software/pdf_files/DbAccess.pdf.
[14] The ReviewPlus Manual, http://web.gat.com/comp/analysis/uwpc/reviewplus/manual/.

External Distribution

Plasma Research Laboratory, Australian National University, Australia
Professor I.R. Jones, Flinders University, Australia
Professor João Canalle, Instituto de Fisica DEQ/IF - UERJ, Brazil
Mr. Gerson O. Ludwig, Instituto Nacional de Pesquisas, Brazil
Dr. P.H. Sakanaka, Instituto Fisica, Brazil
The Librarian, Culham Laboratory, England
Mrs. S.A. Hutchinson, JET Library, England
Professor M.N. Bussac, Ecole Polytechnique, France
Librarian, Max-Planck-Institut für Plasmaphysik, Germany
Jolan Moldvai, Reports Library, Hungarian Academy of Sciences, Central Research Institute for Physics, Hungary
Dr. P. Kaw, Institute for Plasma Research, India
Ms. P.J. Pathak, Librarian, Institute for Plasma Research, India
Ms. Clelia De Palo, Associazione EURATOM-ENEA, Italy
Dr. G. Grosso, Instituto di Fisica del Plasma, Italy
Librarian, Naka Fusion Research Establishment, JAERI, Japan
Library, Laboratory for Complex Energy Processes, Institute for Advanced Study, Kyoto University, Japan
Research Information Center, National Institute for Fusion Science, Japan
Dr. O. Mitarai, Kyushu Tokai University, Japan
Dr. Jiangang Li, Institute of Plasma Physics, Chinese Academy of Sciences, People's Republic of China
Professor Yuping Huo, School of Physical Science and Technology, People's Republic of China
Library, Academia Sinica, Institute of Plasma Physics, People's Republic of China
Librarian, Institute of Physics, Chinese Academy of Sciences, People's Republic of China
Dr. S. Mirnov, TRINITI, Troitsk, Russian Federation, Russia
Dr. V.S. Strelkov, Kurchatov Institute, Russian Federation, Russia
Professor Peter Lukac, Katedra Fyziky Plazmy MFF UK, Mlynska dolina F-2, Komenskeho Univerzita, SK-842 15 Bratislava, Slovakia
Dr. G.S. Lee, Korea Basic Science Institute, South Korea
Institute for Plasma Research, University of Maryland, USA
Librarian, Fusion Energy Division, Oak Ridge National Laboratory, USA
Librarian, Institute of Fusion Studies, University of Texas, USA
Librarian, Magnetic Fusion Program, Lawrence Livermore National Laboratory, USA
Library, General Atomics, USA
Plasma Physics Group, Fusion Energy Research Program, University of California at San Diego, USA
Plasma Physics Library, Columbia University, USA
Alkesh Punjabi, Center for Fusion Research and Training, Hampton University, USA
Dr. W.M. Stacey, Fusion Research Center, Georgia Institute of Technology, USA
Dr. John Willis, U.S. Department of Energy, Office of Fusion Energy Sciences, USA
Mr. Paul H. Wright, Indianapolis, Indiana, USA

The Princeton Plasma Physics Laboratory is operated
by Princeton University under contract
with the U.S. Department of Energy.

Information Services
Princeton Plasma Physics Laboratory
P.O. Box 451
Princeton, NJ 08543

Phone: 609-243-2750
Fax: 609-243-2751
e-mail: pppl_info@pppl.gov
Internet Address: http://www.pppl.gov