AMDA, Automated Multi-Dataset Analysis: A web-based service provided by the CDPP

Jacquey[1] C., V. Génot[1], E. Budnik[2], R. Hitier[3], M. Bouchemit[1], M. Gangloff[1], A. Fedorov[1], B. Cecconi[4], N. André[1], B. Lavraud[1], C. Harvey[1], F. Dériot[5], D. Heulet[5], E. Pallier[1], E. Penou[1] and J.L. Pinçon[6]

[1] CDPP/CESR, CNRS/Université Paul Sabatier
9, avenue du Colonel Roche, BP44346, 31028 Toulouse cedex04, France
EMail: Christian.Jacquey@cesr.fr

[2] Noveltis
2 Avenue Europe, 31520 Ramonville Saint Agne, France

[3] Co-Libri
Cremefer 11290 Montréal, France

[4] LESIA
Observatoire de Paris, Section de Meudon, 5, place Jules Janssen, 92195 Meudon cedex, France

[5] CNES
Centre spatial de Toulouse, 18 avenue E. Belin, 31401 Toulouse, France

[6] LPC2E
Laboratoire de Physique et Chimie de l'Environnement, 45071 Orléans, France

ABSTRACT

We present AMDA (Automated Multi-Dataset Analysis), a new service recently opened at CDPP. AMDA is a web-based facility for on line analysis of space physics data coming from either its local database or distant ones. This tool allows the user to perform on line classical manipulations such as data visualization, parameter computation or data extraction. AMDA also offers innovative functionalities such as event search on the content of the data in either visual or automated way. These functionalities extendable for automated recognition of specific signatures can be used for performing classification of events and for generating time-tables and catalogues. These time-tables/catalogues can be seen as a brick of up-coming virtual observatories in space physics, space weather study and planetology.

INTRODUCTION

For in depth studies of plasma objects such as the Earth's magnetosphere or the solar wind, it is necessary to analyse multi-points and multi-instruments measurements. In practice, that means that researchers have to exploit together data coming from many sources which can initially be heterogeneous in their internal organisation, their description or their coding format. At the scale of the individual researcher, or even of a laboratory, this requires an important investment and consumes a large amount of time and energy.

The CDPP (http://cdpp.cesr.fr/, Centre de Données de Physique des Plasmas) is the french national centre for space physics data. It was jointly created by CNES (Centre National d'Etudes Spatiales) and CNRS (Centre National de la Recherche Scientifique) in 1998. Recently, the CDPP has opened a new web-based service called AMDA (Automated Multi-Dataset Analysis, http://cdpp-amda.cesr.fr/). This service offers "classical" functionalities such as data extraction, merging or visualisation but also innovative ones such as user-edited automated search or computation on the content of the data as well as the creation and the use of time-tables. The aim of this paper is to present this service. More detailed description have been previously published [1,2].

WHAT ARE THE BASIC NEEDS FOR DATA EXPLOITATION IN SPACE PHYSICS?

Case studies or statistical studies in space physics are generally performed as follows: (1) The first step is generally the search of cases of interest. This search can be performed through examination of quicklooks or using quantitative criteria on the content of the data. (2) When an event or a set of events has been identified, the second step consists of retrieving and extracting the numerical data. (3) Before analysis with the user preferred software, the data may need to be formatted depending on their original format. This step benefits of the more and more common use of standard formats as CDF, CEF or NetCDF. (4) Then, basic treatments generally need to be applied e.g. bad value filtering, gap interpolation or merging data at the same baseline, … (5) In this step, the user performs various manipulations on the data. Many of them are specific but commonly include visualisation, parameter computation and comparison to standard models such as magnetic field models.

It is only when these steps have been accomplished that the interesting work, i.e., the interpretation of the observations, starts. Because there are many datasets available, most of which coming from different origins, formatted in various ways, the work required before interpreting the observations often consumes large amount of time and energy from both researchers and software engineers. This has direct impact on the data exploitation, i.e. on the scientific return of the experimental investment.

"PHILOSOPHY" AND FUNCTIONNALITIES OF AMDA.

The main objective of AMDA is to help the user in performing the different steps described above. In order to enable complex operations on data, AMDA has been built around elementary objects: the *parameter* and the *time table*. A first and immediate difference relatively to most databases is that the concept of data files does not exist in AMDA at the user level. The user manipulates parameters without taking care of the files where they come from. Inside the AMDA system, the *parameter* is associated with properties (scalar, vector, tensor, units, …) and with corresponding options (decomposition, frame of reference, …). The second useful object in AMDA is the *time table* which is basically a collection of time intervals.

The functionalities of AMDA allow to use and to couple these two classes of objects. They are interactively activated through web-interfaces by the user who edits requests and collects the results in his/her workspace. All the requests and results may be saved for further sessions as soon as the user has registered. We now briefly describe these functionalities.

• **Parameter builder.** In the "My parameters" interface, the user can compute new parameters by editing mathematical expression combining existing parameters. Heterogeneous time bases are handled by AMDA transparently to the user. He/she freely chooses the time resolution of his/her final new parameter and can also manage data gaps. Once a new parameter has been created it can be used as any pre-existing parameter.

• **Remote data access.** Following the Virtual Observatory paradigm AMDA goes further in giving a direct access to parameters hold in distant databases (noticeably CDAWeb, CASSINI/MAPSKP, VEX-MAG, THEMIS databases are available through AMDA). In the "External Data" interface, the user can browse through the parameters of the connected distant databases, select the desired ones and finally save them in his/her own *external data tree*. This data tree, with the usual hierarchy mission/instrument/dataset/parameter, will then be added to the existing local parameters. The parameters referenced in the external data tree can be used thereafter like any pre-existing ones.

• **Data download.** In the "Download Data" interface, newly created, remotely accessed or local parameters can be equally downloaded (or exported) by the user in different formats (plain ASCII, CEF, CDF) and with a chosen resolution. Time series may optionally be joined on a common time basis before exporting. The data merging is performed either by averaging or interpolating, depending on the time resolution of the original datasets. As it was mentioned above, the user does not need to care about new data downloads. It is fully automated according to the requested time interval and the parameters needed. Note that this functionality also accepts time tables as input allowing to perform easily massive download.

• **Data visualisation.** In the "Plot Data" interface, the user can edit a figure combining any available parameter with the desired options (reference frame, scaling,…). The request can be saved in such a way that the edited figure can be applied to any time interval chosen by the user. Note that the interface provides the option to apply time shifting for solar wind propagation. Widgets also allow to shift the figure to the following/preceding time interval. This functionality also accepts time tables as input. In this latter case, widgets allow to skip the figure from a time interval to the next/previous one.

• **Visual search.** While browsing the requested figures, the user can retain intervals when a special feature, visually determined, occurs. He can do so by double clicking at the start/stop of the event. Then an interface is provided to store this interval in a list; that is, the user can create a time table by visual inspection.

• **Automated conditional search.** A second way to create time tables is by selecting time intervals when a particular mathematical condition applied on given parameters is fulfilled. In the "Conditional Search" interface, the user can edit his/her condition with mathematical functions and logical operands (>,<,&,|). This condition is applied on a given time window and only sub-intervals fulfilling the condition are retained to populate the time table. The time window may itself be a single time interval or a time table, offering the possibility to perform successive automated searches.

• **Time-table manager.** In the "My Time Tables" interface, time tables may be subsequently edited and it is possible to apply operations on a single time table: extend/shrink/shift by a given duration; or on multiple time tables: union or intersection. The time table can be exported in ASCII format or in VOTables compliant to the IVOA standards.

CONCLUSION AND PERSPECTIVES.

AMDA is a web-based service developed with the aim to help researchers in space physics. AMDA offers functionalities to access and analyse multi-dataset in a transparent fashion and allows perform event search. AMDA is in continuous development. Future developments concern new analysis tools, new products, enhancement of interoperability and the creation/management of catalogues.

AMDA is evolving in the Virtual Observatory paradigm. It gives a direct access to data from distant databases and includes a connection layer compliant with the SPASE (http://www.spase-group.org/) standards. AMDA produces, manipulates and uses time tables. The time tables can be seen as one of the primary brick to be used for the interoperable exchanges in space physics.

The CDPP participates to project of virtual observatories of heliophysics (HELIO, FP7), of planetology (EUROPLANET RI/FP7) and space weather (VISPANET/ESA) and will integrate AMDA into them as an interoperable external resource.

REFERENCES

[1] - C. Jacquey, V. Génot , E. Budnik , R. Hitier , M. Bouchemit , M. Gangloff , A. Fedorov , B. Cecconi, N. André , B. Lavraud , C. Harvey , F. Dériot , D. Heulet, E. Pallier , E. Penou and J.L. Pinçon, AMDA, Automated Multi-Dataset Analysis: A web-based service provided by the CDPP , Studying the Earth's Space Plasma Environment Series: Astrophysics and Space Science Proceedings , 2009

[2] - V. Génot, C. Jacquey, , E. Budnik , R. Hitier , M. Bouchemit , M. Gangloff , A. Fedorov , B. Cecconi, N. André , B. Lavraud , L. Broussillou, C. Harvey , F. Dériot , D. Heulet, E. Pallier , E. Penou and J.L. Pinçon, Space Weather applications with CDPP/AMDA, submitted to Advances in Space Research, 2009