Intelligent Archive Technologies for NASA/IMAGE Radio Plasma Imager Data

Ivan A. Galkin, Grigori M. Khmyrov, Alexander P. Kozlov, and Bodo W. Reinisch
University of Massachusetts Lowell, Center for Atmospheric Research
Robert F. Benson and Shing F. Fung
NASA Goddard Space Flight Center

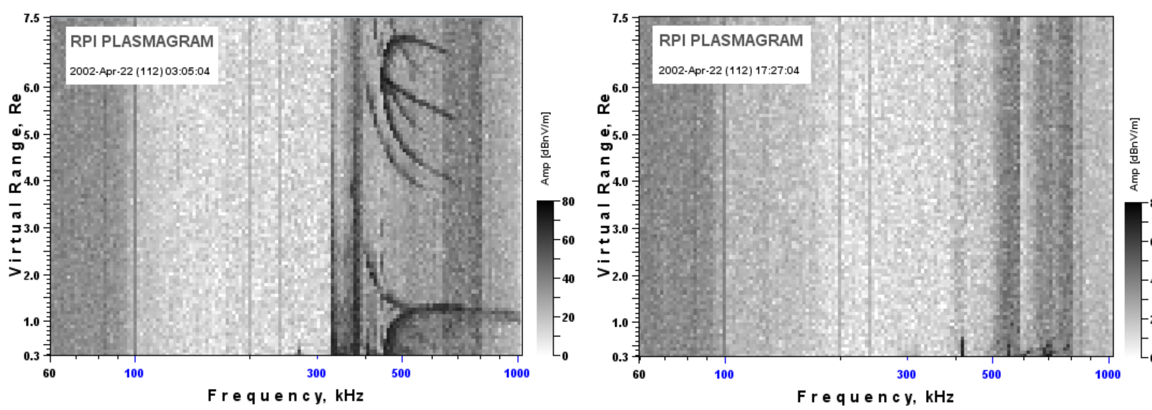
ABSTRACT

The Radio Plasma Imager (RPI) on NASA's IMAGE spacecraft has acquired over 1.2 million plasmagrams, images of remote and in-situ sensing of the Earth's magnetosphere. The RPI plasmagram archive is a classic example of a dataset posing unrealistic demands of manual labor in order to interpret each image or even to simply tell if it contains useful information. We present a pilot project implementing intelligent archive (IA) technologies to assist in the manipulation of the RPI plasmagram collection for the purpose of knowledge discovery. The key elements of the IA project are (1) CORPRAL, a tool for automated plasmagram data prospecting, (2) BinBrowser, a data visualization platform with automated resonance signature identification, and (3) a commercial-strength database, with read-write access over the Internet, for RPI telemetry and derived data products. We discuss our progress to date and lessons learned from project operations during 2002-2006.

1. INTRODUCTION

The growing information capacities of space-borne instrumentation put data managing facilities under pressure of the sheer volume of collected data. The need for knowledge clearinghouses has been recognized since the 1940s, but never before has this need been so clear and present. Our study was inspired by a wish to get an insight into the depths of the NASA data archive holding 1.2 million plasmagram images acquired by the Radio Plasma Imager (RPI) [1] on the IMAGE spacecraft [2]. Even on this small scale of single instrument, the problem of the information avalanche has been difficult to deal with.

The RPI on IMAGE collects radio remote-sensing data about the density distribution of magnetospheric plasmas. RPI's chief product is the plasmagram, as shown in Figures 1-a and 1-b. The plasmagram display shows received signal strength (image intensity) as a function of virtual range (or echo delay; vertical scale) and radio-sounder frequency (horizontal scale) of the radar echo pulses.



(a) (b)

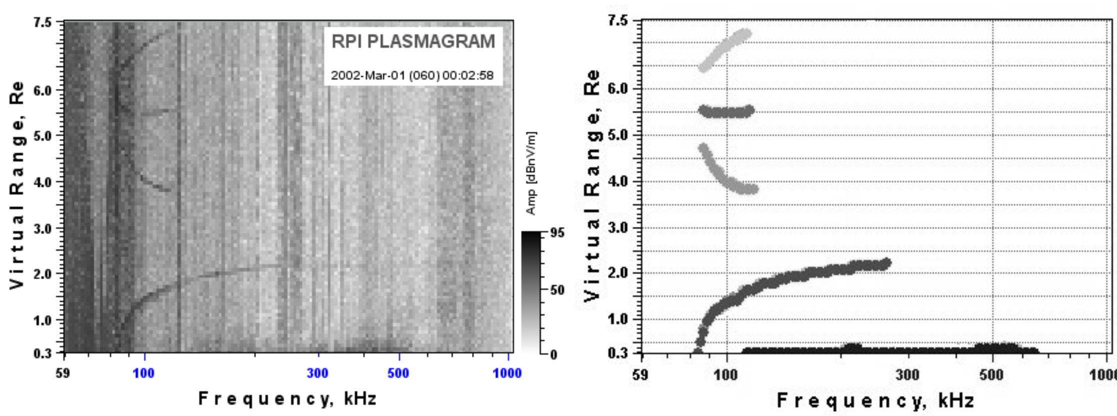
Figure 1. (a) IMAGE RPI plasmagram recorded at 03:06:04 UT on April 22, 2002, showing signal reflections from remote plasma locations (dark traces) intermixed with stimulated resonances in the local plasma (vertical traces typically showing decreasing intensity with increasing virtual range) and natural radio emissions (vertical bands extending over all virtual ranges). (b) Same as (a) except 14 hours later, when IMAGE was in a similar orbital position: in this case remote reflections are not observed.

The thin black traces observed above 400 kHz in Figure 1-a are attributed to radar echoes from signals guided along magnetic field lines in the magnetosphere [3,4]. These echo traces are intermixed with vertical signatures corresponding to locally excited plasma resonances (e.g., the intensification near 320 and 380 kHz in Figure 1-a) and various natural emissions detected by RPI [5]. Less than 20% of all plasmagrams contain echo traces because RPI is a radar of opportunity: for its 10 Watt signal to reflect at a remote location as far as 40,000 km away, return to the spacecraft location, and appear above the noise level to be detected, a number of conditions need to be satisfied. Figure 1b shows a plasmagram, without traces of remote signal reflections, that was taken during the same day at a similar orbital location; this time the conditions for echo detection were not met.

Locating and processing plasmagrams with echo traces is a major exercise requiring significant manual labor. Typically plasmagram scalers analyze 300-400 plasmagrams a day, which roughly translates to 12 years of non-stop work to process the available 1,200,000 images. We estimate that scientists will inspect < 10% of all the collected plasmagrams.

2. "CORPRAL" DATA PROSPECTOR

An automated pre-classification of RPI plasmagrams into categories will improve the chances of extracting useful data from these records. Cognitive Online Rpi Plasmagram Ranking ALgorithm (CORPRAL) [6] is an automated data *prospector* designed to find meaningful data among the large database in order to draw attention of human analysts to possible nuggets of information. Building an imagery data prospector has been a significant task that involved modeling of the visual data perception by humans. Our choice of a feature-binding approach for plasmagrams is based on a bio-plausible model of pre-attentive vision that adheres to a "bottom-up" analysis strategy. Bottom-up feature-extraction algorithms first seek detectable low-level image features (dots, bars) that can be grouped together in salient contours under Gestalt restrictions of proximity, good continuation, and smoothness [7]. They assume no prior knowledge of the features to be discovered in the analyzed image. The bottom-up model is only aware of a general perceptual quality of the image features that makes them stand out against the background. Figure 2 presents the results of applying the CORPRAL analysis to an RPI plasmagram.



(a) (b)

Figure 2. (a)RPI plasmagram recorded at 00:02:58 UT on March 01, 2002. (b) Five automatically-scaled traces resulting from applying the CORPRAL plasmagram prospector.

All of the available 1.2 million plasmagrams were processed to detect traces; over 200,000 of them are now labeled in the RPI mission database as containing echo signatures. We tested the CORPRAL performance on a test set of 25,000 manually-inspected plasmagrams with an 8% prevalence of remote echoes. CORPRAL was able to locate 85% of all plasmagrams with remote echoes; most of the false negative decisions were attributed to plasmagrams with faint traces barely visible in the noise. The overall accuracy of plasmagram prospecting for both positive and negative decisions was 94%. CORPRAL's positive predictive value (PPV), a measure of probability that automatically tagged plasmagrams actually have traces, varied substantially over different periods of mission, reflecting changes in the trace prevalence as the satellite orbit evolved. Overall value of 56% PPV for our 8% prevalence dataset is consistent with our choice to bias prospecting toward the false positive errors in order to increase its sensitivity to faint plasmagram traces.

3. RESONANCE SIGNATURE IDENTIFICATION

Plasma resonances seen in RPI plasmagrams can be used for highly accurate determination of local plasma density and magnetic field strength [5]. A precise interpretation of the resonances, however, requires skill and patience. Both semi-automatic and automatic algorithms have been developed to assist this identification process. Figure 3 shows an RPI plasmagram where plasma resonances and cutoff frequencies have been automatically determined and labeled [8].

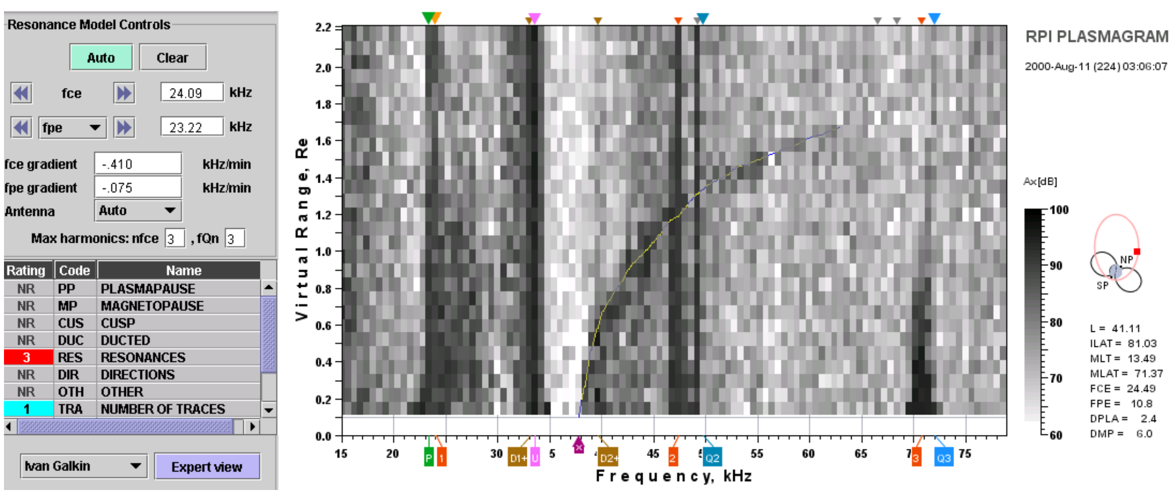


Figure 3. Automatic identification of resonance frequencies in RPI plasmagram recorded at 03:06:07 UT on August 11, 2000. Scaled values are show in the upper left panel as the electron cyclotron frequency fce = 24.09 kHz and the electron plasma frequency fpe = 23.22 kHz. Color labels on the frequency axis are: red 1, 2, 3 = fce and harmonics, green P = fpe, magenta U = upper hybrid frequency, violet X = X cutoff frequency, blue Q = Bernstein-mode resonances, brown D = D-type resonances. The rating table on the left indicates the "presentation quality" of the resonance identification, and one detected trace of remote reflections. (Adapted from [5].)

The identification algorithm uses a theoretical model of dependencies between characteristic frequencies in the plasma driven by four unknown parameters, the electron cyclotron and plasma frequencies, f_{ce} and f_{pe}, and their gradients with time along the orbit. The automated model fitting process uses resonance signature-detection techniques to improve its robustness to the background emissions and instrument noise. Still, whereas f_{ce} is determined reliably by the automated analysis, perfect identification of f_{pe} is currently possible only for picture-perfect plasmagrams like the example in Figure 3. In other cases, human intervention is required to make correction to the autoscaled values. Such corrections are introduced by adjusting values of the four resonance model drivers (the upper left panel in Figure 3) without manual recalculation of the complete suite of resonance frequencies.

4. MISSION DATABASE WITH WRITE ACCESS

The intelligent part of archiving RPI plasmagram data consists in providing read-write access to all automatically- and manually-derived data products (traces and resonances) together with the expert ratings, comments, optimal settings of plasmagram visualization, orbital information, and geospace model predictions. For our pilot project, we use the relational database management system Firebird with Interclient middleware that provides access to data over the Internet to any user. Using BinBrowser visualization and editing software [9,10], it is possible to search and download science mission data by content, not just by time interval, and then process them interactively, evaluating their significance by means of ratings, and upload results of processing and evaluation directly to the database. Thus the database serves not only as a raw telemetry archive, but also as a searchable repository of the expert knowledge provided by both automated intelligent algorithms and affiliated scientists. Without capturing that knowledge, RPI plasmagram archive would appear, to a novel user, only as a large, unfathomable collection of images.

5. DISCUSSION

Since the introduction of the database project in 2002, several scenarios have emerged for “assisted” knowledge discovery in the RPI data. One of the popular data-exploration approaches is to search for plasmagrams with traces, in particular or unusual locations on orbit, or to search for echo signatures in unexpected frequency bands. Prevalence plots (ratios of plasmagrams with traces to total plasmagrams) provide a unified view of the mission and pose questions as to the reasons for higher trace prevalence in particular locations and times. There is an interest in the plasmagrams where automated recognition identified resonance signatures that could not be matched to their theoretical counterparts, which may indicate imperfection of the model, non-linearity of gradients in the medium, or new stimulated plasma wave emissions. In future, plasmagram ratings resulting from this work can be mined for hidden associations with other geophysical events or characteristics. It is our strong belief that the RPI data mine has more knowledge nuggets waiting to be discovered by intelligent algorithms.

6. ACKNOWLEDGEMENTS

The work at the University of Massachusetts Lowell was supported by NASA under subcontract 83822 to the Southwest Research Institute and through the Intelligent Systems (IS) Grant NNG04GO35G. The authors thank Xueqin Huang, Jim Tilton, Antonio Plaza, and Jordan Grinstein for contributions to the project.

REFERENCES

[1] Reinisch, B.W., *et al*., "The Radio Plasma Imager investigation on the IMAGE spacecraft", *Space Science Reviews*, Vol. 91, 2000, pp. 319-35.

[2] Burch, J.L (ed), *The IMAGE Mission*. Dordrecht: Kluwer Acad. Pub., 2000.

[3] Fung, S. F., R. F. Benson, D. L. Carpenter, J. L. Green, V. Jayanti, I. A. Galkin, and B. W. Reinisch, "Guided Echoes in the Magnetosphere: Observations by Radio Plasma Imager on IMAGE," *Geophysical Research Letters*, 30(11), 1589, doi:10.1029/2002GL016531, 2003.

[4] Fung, S.F., and J.L. Green, Modeling of Field-Aligned Guided Echoes in the Plasmasphere, *J. Geophys. Res*., 110, A01210, doi:10.1029/2004JA010658, 2005.

[5] Benson, R.F., V.A. Osherovich, J. Fainberg, and B.W. Reinisch, Classification of IMAGE/RPI-stimulated plasma resonances for the accurate determination of magnetospheric electron-density and magnetic field values, *J. Geophys. Res.*, *108(A5)*, 1207, doi:10.1029/2002JA009589, 2003.

[6] Galkin, I.A., B.W. Reinisch, G. Grinstein, G. Khmyrov, A. Kozlov, and S. F. Fung, "Automated exploration of the Radio Plasma Imager data", *Journal of Geophysical Research*, Vol. 109, 2004.

[7] Rock, I., and S. Palmer, "The legacy of Gestalt psychology", *Scientific American*, Vol. SCA9012, 1990, pp. 84-90.

[8] Galkin, I.A., B. W. Reinisch, H. Xueqin, R. F. Benson, and S. F. Fung, Automated diagnostics for resonance signature recognition on IMAGE/RPI plasmagrams, *Radio Sci.*, 39, 1, RS1015, 10.1029/2003RS002921, 24 February 2004.

[9] Galkin, I.A., G. M. Khmyrov, A. Kozlov, B.W. Reinisch, X. Huang, and G. Sales, New tools for analysis of space-borne sounding data, *Proc. 2001 USNC/URSI Nat. Radio Sci. Meeting*, Boston, MA, 2001, p. 304.

[10] *UMASS Lowell Center for Atmospheric Research: RPI Software Installations* (n.d.), retrieved January 30, 2006 from http://car.uml.edu/Installation/install_all.htm