Online Poster paper for the
1st Crisis In Cosmology Conference
Monção, Portugal June 23-25, 2005

Breaking the Logjam: Applying the Observational-Inductive Framework for Science

Timothy E. Eastman, Plasmas International, Silver Spring, Maryland

Summary: The current logjam in the controversy between the Big Bang model and several alternatives has continued for more than half a century and promises to continue for much longer. The standard hypothetical-deductive methodological framework for science, which focuses on logical implication, derives its strength from the consistency, coherence, and testability of deduced consequences resulting from initial hypotheses. The hypotheses are, in part, inspired by observations but may be highly dependent on prior theory as, for example, research on gamma ray bursts or black holes. Recent work in philosophy of science has recognized fundamental limitations to this standard account of scientific process and has introduced hypothetical-inductive inference in addition to hypothetical-deductive inference [*Niiniluoto et al.,* 1973]. Nevertheless, a third methodological framework has begun to emerge, especially in those fields such as geophysics and space science where direct testing of certain initial conditions or core hypotheses is difficult, if not impossible, but where gigabyte to petabyte datasets have emerged. This new *observational-inductive* mode of inference now shows promise in the wake of advances in high-performance computing, rich data sets, high-speed sensor systems, and multi-dimensional, multi-scale modeling. For the first time in the 400 years since Francis Bacon introduced induction, new developments in computer-aided knowledge discovery (data mining, pattern recognition, artificial intelligence) enable observational-inductive inferences, which are observation-driven, focus on causal implication, and can minimize theory-dependent assumptions. The recent discovery of T-dwarf stars by data mining massive datasets demonstrates the potential of such data-driven inductive, knowledge discovery methods.
The hypothetical-deductive/inductive and observational-inductive frameworks are complementary and synergistic; however, reduction in theory dependence through applying observational-inductive inference may be the only way to break the logjam.

Introduction

A continuing controversy between the Big Bang model and several major alternatives is well known but seldom discussed. The Alternative Cosmology Group *(www.cosmology.info)* and the 2005 Monção, Portugal Conference seek to change this by providing an international forum focused on hypotheses that are, in various ways, alternatives to the standard Big Bang model and its variants. For example, research results on electromagnetism and plasmas in astrophysical settings at all scales are rapidly advancing to fill gaps in the "standard view of nature." Whatever the long-term outcome of this controversy, I recommend that all researchers eschew the ideological stance so common in political settings and apply the best ideals of the scientific process, which include systematic examination of presuppositions, framing of testable hypotheses (falsifiable in principle), model development (preferably quantitative), and careful design of observational tests.

A confluence of new observational datasets and data-processing technologies has the potential to resolve this controversy through the marketplace of ideas. A wealth of observations is becoming available, exemplified by the Sloan Digital Sky Survey and Hubble datasets, providing new possibilities for tests of cosmological theories. In addition, new computer and data-intensive Grid systems promise to bring these datasets more effectively to researchers worldwide. Finally, recent developments in Knowledge Discovery in Databases (KDD), including data mining, pattern recognition, and neural network tools, enable new research methodologies that are less theory-dependent. Because many scientific controversies hinge on certain presuppositions or theory-dependent assumptions, the application of KDD tools, when appropriate, can contribute to breaking the logjam. Though KDD tools embody certain assumptions with regard to data relevance, etc., these are transparent. Unlike theory-based assumptions imbedded in most research activities, KDD assumptions must be fully explicated in order to design and use KDD tools.

A new scientific framework is beginning to emerge as a result of the above confluence of observations and technologies. This observational-inductive framework complements the standard hypothetical-deductive model that has often been held up as the sole standard of what is meant by "science." It is important to note that the hypothetical-deductive/inductive schema were developed prior to, and have not significantly changed since the massive growth (by orders of magnitude) in the volume of observational data and power of high performance computing techniques. The strength of the observational-inductive model is its firm foundation on both of these revolutionary developments in the history of science.

Filling in the Gaps

Significant developments have been occurring for theories and models that fill in the gaps left in the standard view of nature, which "jumps eighteen orders of magnitude from atoms to stars" [*Goedbloed and Poedts*, 2004].

In their 613-page major work on the *Principles of Magnetohydrodynamics* these authors contribute substantially to filling in these gaps by documenting in detail how electromagnetism and plasmas are major dynamical players, from planetary and neutron star magnetospheres to galaxies and beyond. They point out that the key law for magnetic fields, $\nabla \bullet B$,

> "is incompatible with spherical symmetry so that the simplest basic geometries of magnetized plasmas are completely different from the ones prevailing on the atomic and gravitational scales. In particular, large scale tubular magnetic structures occur which move with the plasma so that magnetic forces are transmitted with the fluid. One could hardly imagine a bigger contrast with central electrostatic and gravitational forces decaying in vacuum with distance as r^{-2} ! Striking examples are solar flares, the X-ray emitting corona of the Sun, and coronal mass ejections…, the interaction of the solar wind with the planetary magnetic fields, waves and flows in neutron star magneto-spheres, extragalactic jets, spiral arm instabilities, etc." [*Goedbloed and Poedts*, 2004].

One could describe much of mainstream astrophysics as providing a lot of gap-filling for the standard view of nature, at least for those researchers who prefer to remain in domains where testability and in-principle falsifiability remain on reasonably firm grounds. Wolfgang Kundt's excellent work *Astrophysics: A New Approach* [2005] exemplifies this orientation.

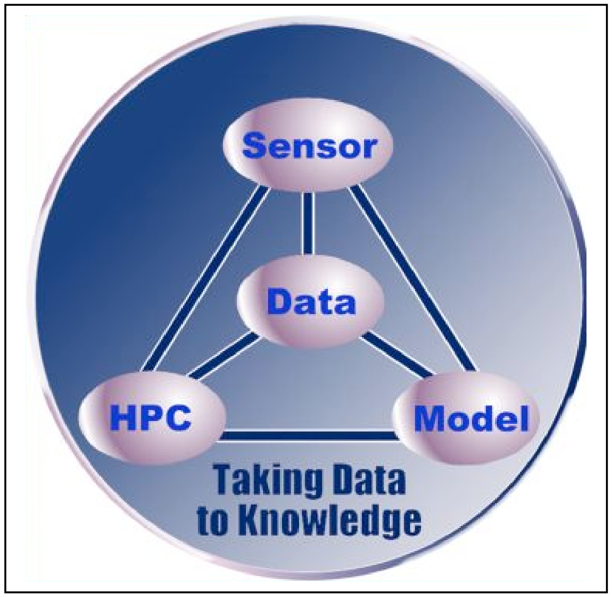
The importance of electromagnetism and plasmas processes in astrophysics and cosmology is beginning to be appreciated by many more space scientists as demonstrated by the rapid growth of plasma astrophysics, especially within the past decade [*Rudiger and Hollerbach*, 2004; *Goedbloed and Poedts*, 2004; *Tajima,* 2004; *Benz,* 2002, *Coppi et al.*, 2000; *Buchner,* 1999; *Tajima and Shibata*, 1997; *Zheleznyakov*, 1996; *Peratt,* 1992, 1995; *Kirk, Melrose, and Priest*, 1994; *Michel,* 1991 – see *www.plasmas.org* for these and many more references in space and astrophysical plasmas].

Taking Data to Knowledge

The need for data grids such as the International Virtual Observatory (IVO) for astronomy derives from the following set of drivers (problems). There is no single solution for these complex problems but many promising solutions depend on technologies that have emerged within the past 1-2 decades [*Eastman et al.*, 2005].

PROBLEMS (all associated with links in the figure)

- Information Explosion
- Understanding Multiscale Physical Systems
- Solving Complex, Nonlinear Systems
- New High Data Rate Sensors
- Distributed, Intelligent Sensor Networks



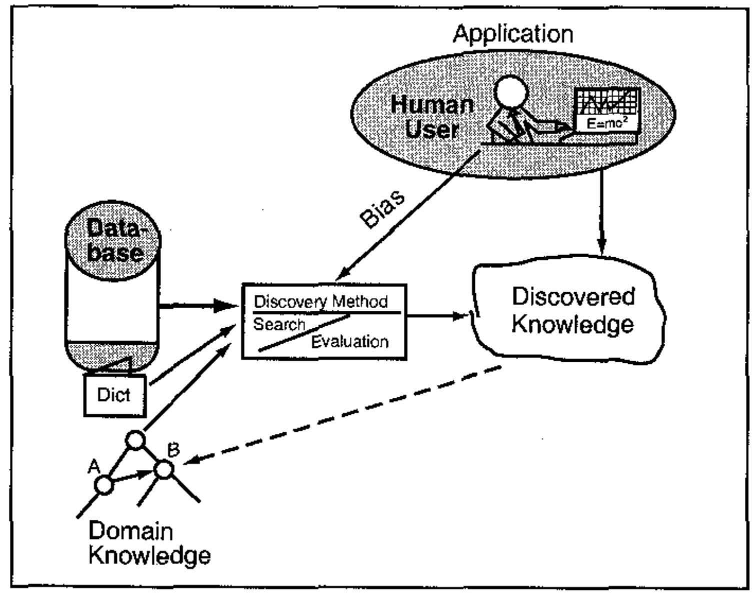
SOLUTIONS (all associated with links in the figure)

- Distributed Data Environments
- Grid Services (interoperability; semantic web)
- Virtual Observatories, Data Grids, eScience
- Knowledge Discovery, Data Mining
- Data Archive Standards
- Sensor Web
- Sensor Development
- Scientific Modeling
- Advanced Visualization

In the late 20th century, breakthroughs in nonlinear dynamics emerged from the HPC-Model linkage with the advent of new supercomputer resources. Similarly, multiple linkages within the Data-Sensor-HPC-Model framework have the potential of radically transforming science in the 21st century. Of course, any element of this tetrahedron can be placed in the center to yield different perspectives on grid systems. Examples of such perspectives are Grid Computing (HPC as central), Sensor Webs (sensors as central), and Data Grids (data and data systems as central). Considered within this broad conceptual framework, grid systems and virtual observatories may have a transformative impact well beyond their initial role to broaden access to data and computing resources and enhance analysis tools across distributed databases worldwide.

Knowledge Discovery in Databases (KDD)

Within the past decade there have been major advances in data mining, neural networks, pattern recognition, clustering, principal component analysis, Bayesion networks, Markov models and other tools, which are here referred to collectively as KDD tools. KDD is particularly useful for the discovery of hidden relationships in large, complex databases where human means of pattern recognition or even model application may fail. Knowledge discovery denotes "the nontrivial extraction of implicit, previously unknown, and potentially useful information" [quote and figure from *Frawley et al*., 1992].



Data selection, automating access through registries, translation and formatting are just a few of the many data preparation steps that are essential for successful KDD applications and which can consume up to 80% of a data-mining project [*Pyle,* 1999]. With this investment and with sufficiently robust data sets, however, previously hidden facts can be discovered such as specific rare events, anomaly detection, patterns, correlations, linkages, complex multi-variable interdependencies and more [*Borne,* 2003].

Emergence of the International Virtual Observatory (IVO) (*www.ivoa.net*) and other new venues for data access provide important new opportunities for applying KDD tools.

* For this section, I gratefully acknowledge discussions with Dr. Kirk Borne of George Mason University.

Frameworks for Knowledge Discovery

There have been many ways to represent the scientific method (hypothesis formation, experimental preparation, test, model refinement…). Most scientists agree that there is no one single method and that simplistic reference to the scientific method is insufficient; however, it remains unclear what these methods are when more fully evaluated and expressed. One thoughtful attempt to capture the essence of the scientific process led to seeing the "scientific method as information-seeking by questioning" and "problem-solving power [keeping in mind] the basic theoretical presupposition… of one's questioning procedure" [*Hintikka,* 1985]. An "inductive logic of theories" remains incomplete in philosophy of science.

Framework	hypothetical-deductive	observational-inductive	hypothetical-inductive*
Levels	Top-down	Bottom-up	Interplay of levels
Focus	Logical implication	Causal implication	Causal implication
Driver	Theory	Observations	Theory/observation balance

*Hypothetical-inductive inference is discussed by [*Niiniluoto and Tuomela*, 1973].
Current science = hypothetical-deductive plus observational-inductive with hypothetical-deductive as "ideal."
Complementary frameworks are hypothetical-deductive and observational-inductive with hypothetical-inductive as middle ground.
Observational-inductive framework is just now emerging due to the power of new KDD tools and the synergism of Data-Model-HPC-Sensor.
Note: Here "induction" is used broadly to include Peircean abduction as well as induction [*Denecker et al*, 1996].

Early problems with the concept of induction contributed to the success of the hypothetical-deductive framework. However, both inductive and deductive inference remain as irreducible elements of the scientific process [*Niiniluoto and Tuomela*, 1973]. This has led to promising new research in inductive inference [e.g., see *Han et al.,* 1993, and references in "Computational Scientific Discovery" (see *www.isle.org/~langley/discovery.html*)].

Breaking the Logjam

For the first time in the 400 years since Francis Bacon introduced induction, a confluence of new technologies [high-performance computing, large, robust data sets, high-speed sensors, and sophisticated linear and nonlinear models – the *Data-Sensor-HPC-Model* framework – in combination with new Knowledge Discovery in Database (KDD) techniques] are effectively creating a new *observational-inductive* approach to scientific inference that is complementary to the standard hypothetical-deductive approach. The standard approach has been marvelously successful in high-energy physics and certain other fields where highly-refined theories can provide well-defined, falsifiable predictions that can be directly tested in controlled, laboratory experiments. However, for complex, interrelated multi-scale systems like ecology, space physics (my field) or cosmology [especially without a focus on the (in principle) falsifiability of hypotheses], it can lead to circumstances in which a particular research program becomes an arbiter of acceptable data [*Lakatos,* 1970], which then shields the preferred theoretical framework from falsification (as happened with the uniformitarian doctrine in geology prior to plate tectonics). This tendency towards theory-dependence is a key weakness of the hypothetical-deductive approach, a weakness which can be offset by the observations-driven approach of the newly emerging observational-inductive framework. The recent discovery of T-dwarf stars by the data mining of massive datasets [*Burgasser et al.,* 2000] demonstrates the potential of this new methodological framework through the application of data-driven, inductive, knowledge discovery methods.

The hypothetical-deductive/inductive models have not changed significantly in response to the massive growth in the volume of observational astrophysical data and the power of high performance computation. The observational-inductive model of scientific inference stands firmly on both of these revolutionary developments in the history of science.

The hypothetical-deductive/inductive and observational-inductive frameworks are complementary and synergistic; however, reduction in theory dependence through applying observational-inductive inference may be the only way to break the logjam.

References

Benz, Arnoldo O., *Plasma Astrophysics: Kinetic Processes in Solar and Stellar Coronae*, Dordrecht: Kluwer Academic Publ., 2002.

Borne, K. D., Distributed data mining in the National Virtual Observatory, in the conference *SPIE Data Mining & Knowledge Discovery V*, vol. 5098, p. 211, 2003.

Buchner, Jorg, *Plasma Astrophysics and Space Physics*, Dordrecht: Kluwer Academic Publ., 1999.

Burgasser, Adam, et al., Discovery of a brown dwarf companion to Gliese 570ABC, *ApJ,* Vol. 531, part 2, pp. L57-L60, 2000.

Coppi, Bruno, et al., eds, *Plasma Astrophysics*, Vol. 142, Int'l School of Physics Enrico Fermi, 2000.

Denecker, Marc, Bern Martens, and Luc De Raedt, On the difference between abduction and induction, in *12th European Conference on Artificial Intelligence,* W. Wahlster, ed., ECAI 96, Wiley & Sons, Ltd., 1996.

Eastman et al., eScience and Archiving for Space Science, submitted to *Data Science Journal*, 2005.

Frawley, William J., G. Piatetsky-Shapiro, and C. Matheus, Knowledge discovery in databases: An overview, in *Knowledge Discovery in Databases*, ed. G. Piatetsky-Shapiro and W. Frawley, Menlo Park, CA: AAAI Press, 1991.

Goedbloed, J., and S. Poedts, *Principles of Magnetohydrodynamics: With Applications to Laboratory and Astrophysical Plasmas*, Cambridge University Press, 2004.

Han, J., Y. Cai, and N. Cercone, Data-driven discovery of quantitative rules in relational databases, *IEEE Trans. On Knowledge and Data Engineering*, Vol. 5, No. 1, 29-40, Feb. 1993.

Hintikka, Jaakko, True and false logics of scientific discovery, in *Logic of Discovery and Logic of Discourse*, J. Hintikka and F. Vandamme, eds, New York: Plenum Press, 1985.

Kirk, J. G., D. B. Melrose, and E. R. Priest, *Plasma Astrophysics*, Berlin: Springer-Verlag, 1994.

Kundt, Wolfgang, *Astrophysics: A New Approach,* Berlin: Springer-Verlag, 2005.

Lakatos, Imre, *Criticism and the Growth of Knowledge*, Cambridge University Press, 1970.

Michel, F. Curtis, *Theory of Neutron Star Magnetospheres*, Chicago: University of Chicago Press, 1991.

Niiniluoto, Ilkka, and Raimo Tuomela, *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht: D. Reidel, 1973.

Pyle, Dorian, *Data Preparation for Data Mining*, San Francisco: Morgan Kaufmann Publishers, 1999.

Peratt, Anthony L., *Physics of the Plasma Universe,* Berlin: Springer-Verlag, 1992.

Peratt, Anthony L., ed., *Plasma Astrophysics and Cosmology,* Dordrecht: Kluwer Academic Publ*., 1995*.

Rudiger, Gunther and Rainer Hollerbach, *The Magnetic Universe: Geophysical and Astrophysical Dynamo Theory ,* Wiley-VCH, 2004.

Tajima, Toshi, and K. Shibata*, Plasma Astrophysics,* Addison-Wesley, 1997.

Tajima, Toshi, *Computational Plasma Physics: With Applications to Fusion & Astrophysics,* Perseus, 2004.

Zheleznyakov, V. V., *Radiation in Astrophysical Plasmas,* Dordrecht: Kluwer Academic Publ., 1996.

Note: I gratefully acknowledge suggestions and corrections provided by Drs. Farzad Mahootian and Carolyn Brown.