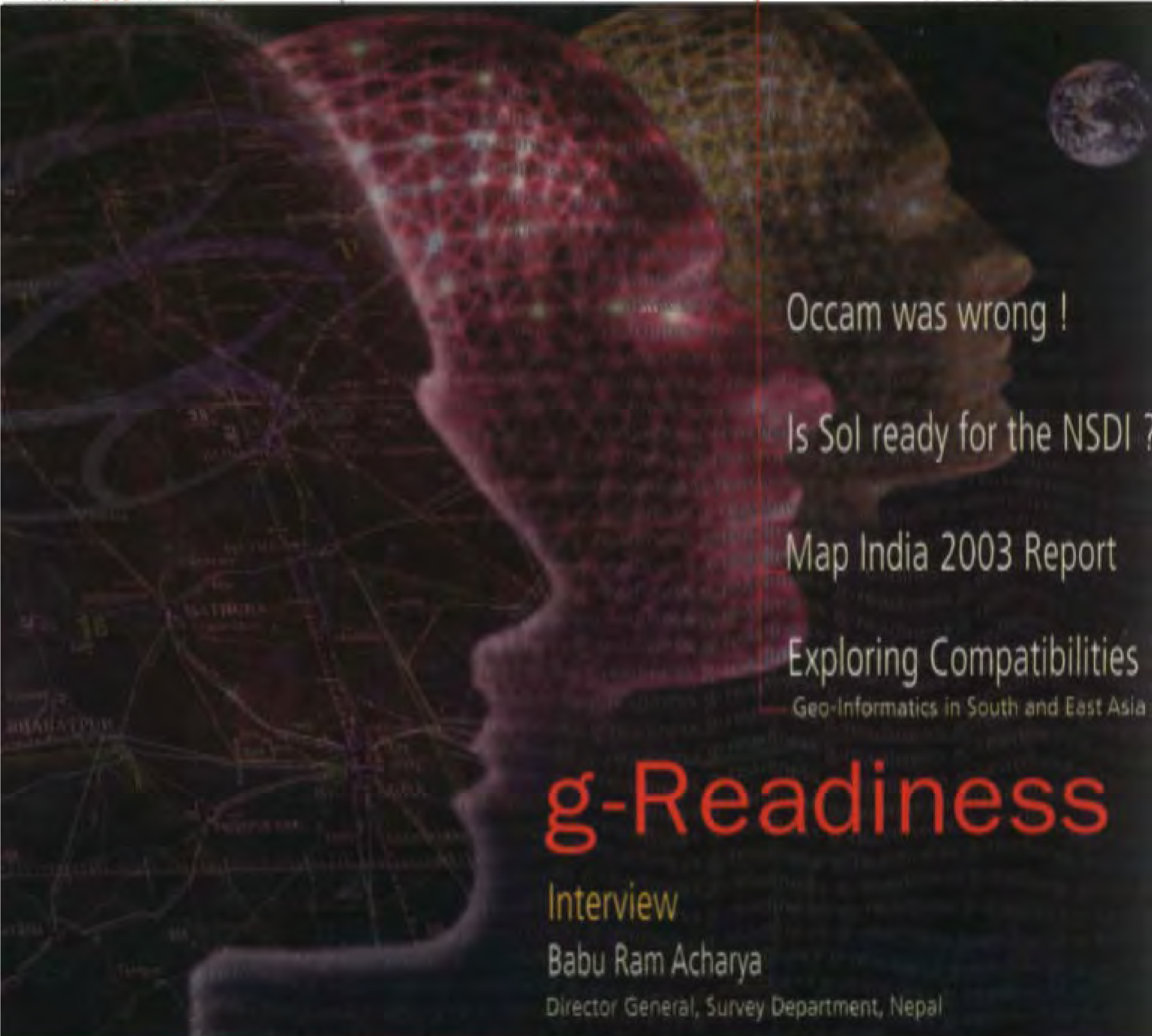
GIS DEVELOPMENT

FEBRUARY 2003 VOL. 7 ISSUE 2 | GIS@development | ISSN 0971 9377



Occam was wrong !

Is SoI ready for the NSDI ?

Map India 2003 Report

Exploring Compatibilities

Geo-Informatics in South and East Asia

g-Readiness

Interview

Babu Ram Acharya

Director General, Survey Department, Nepal

Occam Was Wrong !

Harold G Campbell

When faced with complex questions about the interaction of phenomena or while searching for an explanation about the cause and effect of things upon each other, the vast majority of people tend to dissect and interpret how the world is put together from a rather univariate perspective. With the advent of GIS systems it has become "somewhat" easier to construct advanced models that account for not only a linear form of correlation between variables, but also to address the multivariate and multidirectional interrelations that exist between phenomena, provided that researchers remember to envision such potential complexities in order to properly construct an analytical environment inside the GIS software that accounts for such events. The research design strategy presented within this article depicts just how complicated this process of multidirectional hypothesis formulation and spatial correlation modeling can become and further illustrates the degree of comprehensiveness required in order for the researcher to build-in such multidimensional considerations.

PROF HAROLD G CAMPBELL
CALIFORNIA STATE UNIVERSITY - HUMBOLDT
DEPARTMENT OF COMPUTING SCIENCE, CALIFORNIA, USA
Email: hal.campbell@cox.net

Generally speaking, when faced with complex questions about the interaction of phenomena or while searching for an explanation about the cause and effect of things upon each other, the vast majority of people tend to dissect and interpret how the world is put together from a rather univariate perspective. The temptation to oversimplify things and to reduce a complex question to its simplest form is quite understandable really, due largely to the fact that contemplation of multiple interrelationships between variables is difficult to do and as a consequence most people naturally grab hold of the first reasonable explanation that occurs to them regarding how particular phenomenon interact so that they can expediently articulate their conclusion. The problem with this approach to problem solving is that people (once they've decided upon an explanation) tend to cling to their initial argument as though it were a reflection of their personal character, in spite of the introduction of new information that may either invalidate their assertion or further explain the situation.

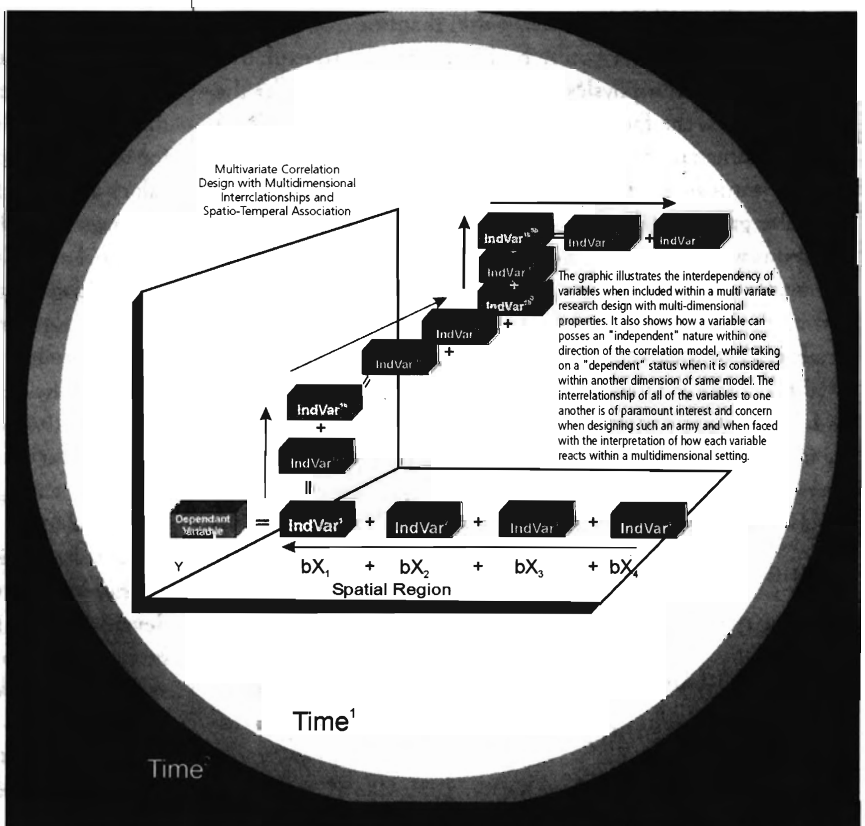
The natural byproduct of such an approach to problem solving (especially if challenged by another during a debate over the issue) is that the dialog typically degenerates to nothing more than a contest of wills, and the truth of the matter is typically never isolated fully by anyone. This approach to problem solving (to reduce the complexity of the world's natural interactions to the most simplistic view possible) is certainly understandable however. After all, it's hard to think up all of the possible reasons that something happens and then prioritize the potentially contributive factors into a coherent argument. It is extremely difficult for people to change who they are, how they think about things, and even more difficult for us to withhold judgment about something until all of the possible alternatives have been examined. We all know that who we are, our cognitive abilities to reason, the methods we employ to arrive at a particular conclusion, and the judgments we make about the world cannot possibly be flawed, because that would suggest that we are somehow flawed and this is simply not acceptable to us.

Perhaps the most demonstrative difference between people who are trained in the scientific approach to problem solving and those practices employed by "normal people" is the ability of the former to recognize the innate complexities and interrelationships of the world and to employ a methodological structure to the problem solving effort which roots out the cause and effects of any situation under study (or at least we scientists like to think so). The byproduct of these analyses are the development of an awareness and the construction of quantitative formulas that can be employed to describe the relationships discovered and which can extend the analysis to the formulation of strategic policies that are based on this understanding and that hopefully provide effective control over the outcome of efforts to manipulate the environment in which we live.

This technique is not hard to master, but it does require practice and self-discipline. The pace of today's world exacerbates the temptation to cling to the univariate model of problem solving because decisions must be made quickly. However, those people who are effective at policy formulation, I believe, tend to realize that decisions about how forces interact, what actions would be most effective at achieving the desired results, and the consequences of such actions, do so from an informed perspective rather than a quickly acquired univariate determination. This means that they do not make decisions in haste. Rather, they take time to examine the complexities of the issues under study and render decisions based on their assessments of what is best for all concerned based on the volumes of information that they have painstakingly assembled and analyzed (or at least that's the way in which it is supposed to happen).

As many researchers have discovered, using computer programs such as statistical analysis software or spatial analysis systems can greatly aid in the development of complex analytical models. The inherent design of these types of software, as applied to the analytical process, is extremely conducive to replicating the scientific approach to problem solving. Within GIS systems specifically, data are stored into separate relational tables according to some logical design structure and represent information in a manner that considers the value of these attributes, along with anchor points that depict space and time considerations. Similarly, the methods employed to represent geographical areas themselves inside GIS, also force the process toward a layered approach, which necessarily subcategorizes each individual layer into its own unique file, while all the time paying attention to the spatial integrity of each layer independently and when aggregated. When combined with the ability to develop horizontal and vertical oriented SQL queries about data values relative to one another and collectively relative to space and time, GIS become a most effective tool for developing complex scientifically oriented examinations of the world and its associated phenomena. The key to using these tools to facilitate such scientific examinations rests with the awareness (on the part of the researcher) that things are not as simple as they may appear to be at first glance. It is normally the combined effect of multiple factors (in multiple directions) that is responsible for fluctuations in the observed behavior of a dependent variable and when using such systems, the identification of these causative factors must be achieved prior to rendering a judgment if the researcher has any hope of isolating those variables which are not only causative in nature, but which also can be controlled or manipulated to achieve the desired changes in the dependent variable. Even those people with more advanced knowledge and skill about how to construct scientifically structured methodologies for problem solving, I find, tend to fall victim to the temptation of limiting their examination of phenomena to a very linear form of multiple correlation equation or discriminant analysis that often fails to fully illustrate the multidirectional intricacies and associations that also exist between phenomena. Because of our training, we rely on the standard quantitative design methodologies we learned in graduate school to help us in developing such models, but often we fail to extend those practices in order to attain the most complete explanation possible. This again is certainly understandable because of the difficulty associated with building complex models and the time required to construct such predictive equations.

With the advent of GIS systems however it has become "somewhat" easier to construct advanced models that account



for not only a linear form of correlation but also which address the multidirectional relationships and expanded hypothetical interrelations that exist. It still comes down to the researchers ability to envision such potential complexities however, in order to properly build an analytical environment inside the GIS software that accounts for such events.

The research design strategy presented within Graphic 1 depicts just how complicated this process of multidirectional hypothesis formulation and spatial correlation modeling can become and also depicts the degree of comprehensiveness required in order for the researcher to build-in such multidimensional considerations (which probably explains why most of us never really build these types of behemoth equations in the first place). As can be observed, the relative spatial position of each variable, along with its defined hypothetical interrelation to all of the variables contained within the multidirectional correlation matrix, is prescribed under such a design as well as the relative value of each of these variables within the ever present spatio-temporal envelope.

From a practical perspective, in order for us to use spatial analysis systems as a mechanism to study such complex interrelations, it is imperative that we subscribe to the rules that govern interaction and relation, which are simply that (1) nothing exists separate and apart from everything else in the universe and (2) everything occupies space and time (even variables). In order for us to fully examine the complexities of the universe, we must build our computer models according to the same manner in which they really exist within that universe. It really doesn't matter whether you're building a model to explain a complex equation in physics, biology, natural resources, or the social sciences, the fact remains that everything exists within space and time and everything has either a dependent or independent status depending upon which way you look at it. This particular graphic shows the relational nature of things to one another as well as to space and time and offers a unique look at how such interrelationships can be envisioned by researchers who are attempting to fully account for the interactivity of variables upon one another. As presented, each independent variable contained within the traditional multiple correlation model (the lower equation in the graphic) possesses a contributive influence over the value of the dependent variable, but each of these IndVars is also subject to relative influences itself by other more subtle factors, and in essence becomes a dependent variable itself when considered in a more fully articulated model design. Those subtle variables contained in the perpendicular equations, in turn take on a dependent relationship as you consider even more perpendicular dimensions to the same equation. The really cool part of all of this however is not simply the infinite number of perpendicular relationships that may exist in such an equation (although that is entertaining to consider as well), but rather the degree of subtlety that one can achieve in identifying and controlling first, second, and third order variables. After we have identified the perpendicular relationships that exist between multiple correlation equations and the interrelation that exists between perpendicular equations, we are free to further examine the controllable and non-controllable properties of each IndVar and use this recognition to experiment with the impact that subtle manipulations of each variable may have on the greater principle equation. This process of recognizing, identifying, measuring, and manipulating variables, in multiple directions, can be highly effective at disclosing the subtleties that exist between variables and at providing researchers with a tool for creating models and subsequent policies based on these models, that maximize the degree of prerogative that exists in manipulating our universe.

From the strategist's perspective, such research designs not only offer a comprehensive examination of phenomena, but also further provide for the ability to construct advanced equations that offer a highly refined degree of subtlety over potential courses of action to manipulate the environment. This approach can not only maximize effectiveness in achieving the desired results to control phenomena, but can also provide for the expanded consideration that minor adjustments in the values of the IndVars contained in the perpendicular equations will have upon the more principle aspects of the model.

Spatial analysis systems such as GIS, makes it much easier for us to engage in this form of complex empirical modeling and afford researchers with a relatively inexpensive tool that can be used to produce extremely complex exploratory and predictive models. The key to success in these endeavors is to simply remember that we cannot fall victim to the perils of oversimplification, nor can we forsake or ignore the principles of good science when engaged in such endeavors. In order to create the real world inside of a computer, it is necessary to pay attention to the rules under which it exists outside of the computer.

The world in which we live is complex and in order for us to develop quantitative models that accurately describe the interrelations of the world and the prospects that exist for securing positive changes, we must build our equations with this complexity in mind and develop models that afford the ability to exploit the subtleties that exist within the universe. In other words, Occam was wrong.■

With the advent of GIS it has become "somewhat" easier to construct models that account for not only a linear form of correlation but also which address the multidirectional relationships and expanded hypothetical interrelations that exist