An Empirical Approach To Estimating GIS Benefits

By Stephen R. Gillespie

Abstract: *Data on the benefits of using geographic information system (GIS) technology are rare. The U.S. Geological Survey (USGS) has developed a model to predict the benefits of using GIS technology. The USGS model focuses on the complexity of a GIS application as the key factor influencing the level of benefits. Three different aspects of complexity are input to a model consisting of a pair of multiple regression equations. The equations explain from one-half to three-fourths of the measured variation in GIS benefits and present a powerful tool for improving the quality of GIS costs/benefits studies.*

Introduction

All current and potential users of geographic information system (GIS) technology must deal with the issue of the costs and benefits of their activities. It has long been recognized that the only justification for any organization's expenditures on digital data is that the data's benefits exceed their cost (Dickinson & Calkins 1988). Nonetheless, accurate data on benefits generated by GIS technology are rare.

The 1994 Urban and Regional Information Systems Association (URISA) Conference dramatically illustrated that the GIS user community recognizes both the importance and the current paucity of benefits information. The theme of the conference was "Integrating Information and Technology: IT Makes $ense" (Tsui 1994). By this, the conference coordinators meant that the technology must be cost effective. Despite this stated objective, GIS management consultant Rebecca Somers, reviewing the conference, wrote, "A notable absence was that of any real discussion about the actual costs and benefits of GIS ... conference attendees would expect a range of presentations presenting real figures and results, and perhaps even some guidelines-something that we desperately need, but the dearth of information in this area persists" (Somers 1994).

The lack of reliable benefits estimates can have a real cost. Failure to adequately quantify potential benefits can lead to undervaluing GIS technology in costs/benefits studies designed to justify its implementation or expansion. Too conservative an estimate of net benefits can cause the delay or cancellation of investment in a technology that might be seen as highly cost effective if benefits were measured more thoroughly.

The problem is not fundamentally a theoretical one; the issues involved in accurate benefits measurement are well known (Obermeyer 1999). Theoretically, benefits estimates should be based on the societal marginal willingness to pay for GIS-provided improvements over the present system (Peterson & Sorg 1987). In the absence of externalities and monopoly, societal willingness to pay is measured by market prices. However, it is in the nature of government involvement in GIS operation that markets cannot set meaningful prices for many of the changes. There is extensive literature on the theoretical valuation of nonpriced and nonpriceable goods. Contingent valuation studies rely on surveys to determine how much respondents would be willing to pay (Cummings et al. 1986). Travel costs have been used to estimate the value of recreational resources (Clawson & Knetsch 1966). Hedonic models infer values not directly observable from related markets where values are directly observable (Brookshire et al. 1982). Methods include the use of property values and wage differentials in labor markets (Viscusi 1993).

USGS research published in the Fall 1994 *URISA Journal* (Gillespie 1994) demonstrates that there are practical techniques for measuring benefits that might initially appear to be nonquantifiable. The real difficulty in applying such techniques is that they can be time consuming and expensive. Converting from qualitative to quantitative benefits measurement can easily double or triple the cost of a costs/benefits study. It would be very useful to have a relatively quick and inexpensive method for making ballpark estimates

Stephen R. Gillespie *is an economist in the Office of Strategic Planning & Analysis of the U.S. Geological Survey. He received his Ph.D. in economics from George Mason University. He came to the Geological Survey in 1989 following 17 years of work at the Bureau of Labor Statistics. Dr. Gillespie's primary research interest is the economic value of geographically referenced data.*

of the likely benefits an organization would gain from the use of GIS technology.

One way to avoid the expensive process of directly measuring the benefits of using a GIS is to identify and stress factors that contribute to a successful GIS application. Numerous published studies address the question of how to successfully implement GIS technology in an organization. They concentrate on organizational factors, such as "selling" the technology to high-level management, involving users, designing effective pilot projects, and consensually creating a vision for the organization's GIS (Anderson 1992). The social interactionist approach (Campbell 1999) stresses a focus on the traditions, values, and skill bases of individual organizations to ensure successful exploitation of a GIS.

Less common is the identification of factors that influence the success of particular GIS applications. Aronoff (1989) discusses how the usefulness of existing spatial data (factors such as correctness, comparability, and consistency) affects the success of a GIS. The more useful the existing data, the greater the likelihood that the GIS can be successfully used. The causative link is cost avoidance; that is, when the existing data are good, the user does not have to spend as much to provide good data input for the GIS. Ripple (1987) identifies the rate of change for existing data and the likelihood of legal challenges to decisions in which GIS applications were important factors. The faster the rate of change, the greater the value of the GIS, the value stemming from the relative ease of updating computer files. The greater the likelihood of legal challenge, the greater the value of the GIS, the value coming from the appearance of professionalism and rigor of GIS outputs. The Bureau of Indian Affairs (1988) identifies the existence of repetitive work as a key to a successful GIS application. USGS research has extended this early work by creating and applying a comprehensive framework for analyzing the factors that influence the value of GIS technology for particular applications. The resulting model greatly simplifies the task of quantifying benefits for a broad range of Federal GIS applications.

General Framework for GIS Benefits

Many different taxonomies of benefits are available for use in costs-benefits studies (Smith & Tomlinson 1992). A particularly useful distinction for measuring benefits from the use of GIS technology is between efficiency benefits and effectiveness benefits. Efficiency benefits result when a GIS is used to do a task previously done without a GIS; the same quality of output is produced, but at lower cost. For example, cut and fill calculations can be made by applying planimetric techniques to contour lines on a graphic map or by manipulating digital elevation data in a GIS. Both methods yield the same results, but a GIS is much faster and easier.

Effectiveness benefits result when a GIS is used to improve the quality of a current output or to produce an output not previously available; that is, the GIS is used to do something that could not or would not be done without it. For example, a GIS can quickly and easily produce maps showing how the proposed route for a new road would affect a series of environmentally sensitive resources. Such maps could be manually drafted, but the process would be so expensive that they probably would not be prepared. A GIS also can overlay a large number of separate environmental themes and calculate an overall impact. When there are more than just a few overlays, this task simply is not feasible using non-GIS techniques.

The level of benefits realized when using a GIS to run an application is determined by comparing the cost of using the GIS method with the cost of using the non-GIS method, and by comparing the value of the outputs produced by the two methods.

$$\text{Benefits of GIS} = (\text{Value of output}_{\text{GIS}} - \text{Value of output}_{\text{NON-GIS}}) + (\text{Cost}_{\text{NON-GIS}} - \text{Cost}_{\text{GIS}})$$

Pure efficiency benefits and pure effectiveness benefits can be seen as special cases of this general formula. When the GIS outputs are equivalent to the non-GIS outputs, the first term vanishes, leaving Benefits = $(\text{Cost}_{\text{NON-GIS}} - \text{Cost}_{\text{GIS}})$, or pure efficiency benefits. When the costs of the two methods are the same, the second term vanishes, leaving Benefits = $(\text{Value of output}_{\text{GIS}} - \text{Value of output}_{\text{NON-GIS}})$, or pure effectiveness benefits.

The general formula shows why benefit measurement of a proposed GIS application is expensive. Of the four terms in the formula, $\text{Cost}_{\text{NON-GIS}}$ is the only one for which a government agency is likely to have reasonably accurate information. Estimation of Cost_{GIS} could require an extensive pilot test. Estimation of the value of outputs requires identification of users and uses of the outputs, impacts of changes in outputs on the users and uses, and dollar valuations of the impacts, none of which is likely to be easy. Because this is such a daunting task, it is not surprising that quantitative measurement of GIS benefits is so rare.

Factors Influencing Level of GIS Benefits

USGS research focuses on the complexity of a GIS application as the key factor influencing the level of benefits realized from the application. Complexity was chosen as the key factor because (1) the theoretical direction of its influence on efficiency and effectiveness benefits is clear, (2) it is identified as being an important GIS success factor in the previously cited studies, and (3) it proved to be a useful and measurable concept in a series of USGS case studies of specific Federal GIS applications.

There are three aspects to the complexity of an application:

1. Input complexity concerns the data themes needed to perform the application. It involves such things as the number and diversity of data themes, the total volume of input data, and the areal extent of the application.
2. Analysis complexity concerns how the data themes are manipulated inside the application. It involves such things as the maximum number of concurrent overlays, the number of steps in the analysis, the number of intermediate data

themes created, and the number of potential interactions between data themes.

3. Output complexity concerns the products of the application. It involves such things as the number of distinct uses for the outputs and the likelihood that the outputs will be used in adversarial hearings.

Each complexity factor can be expected to influence the level of efficiency and effectiveness benefits in a predictable way.

Among the measures for input complexity, both the areal extent of the application and the volume of input data are expected to be positively related to the level of efficiency benefits. All other things being equal, the larger the study area or the greater the amount of physical data, the greater the manual inputs required. Having greater manual inputs implies a larger potential for efficiency benefits by using GIS technology. Both also are expected to be positively related to the level of effectiveness benefits. All other things being equal, the larger the study area, the greater the value of outputs. A greater volume of input data implies a larger information content in the outputs; a greater value of outputs implies a larger potential for effectiveness benefits.

In fact, it is expected that both of these input complexity measures would have a log linear relationship to the level of GIS benefits. This is because there are economies of scale in dealing with inputs, so that a doubling of the volume of inputs does not double the complexity of the application. There are two general types of economies of scale that operate with input complexity. Both types are illustrated by an application to find an optimal route.

For example, assume that the best route must be found for shipping something from point A to point B. The road network would be an input to this application. The complexity of the input would be affected by the level of detail sought about the road network. If only interstate highways are relevant to the analysis, then the input is not very complex. As more levels of detail are needed (for example, primary roads, secondary roads, and unpaved roads) the input becomes more complex. However, the complexity does not increase as quickly as does the volume of input data. There may be five times as many miles of secondary roads as primary roads, but their inclusion only raises the input complexity by one level.

Another economy of scale comes about because much of the input data is not relevant to the problem. For example, most secondary roads are clearly not on the optimal route and quickly can be eliminated from further consideration. Adding the entire secondary road network could double or triple the volume of input data but probably would add only slightly to the volume that must be seriously considered.

Among the measures for analysis complexity, both the number of concurrent overlays and the number of potential interactions between data themes are expected to be positively related to the level of benefits. All other things being equal, the greater the number of themes overlaid, the greater the manual inputs required, and the larger the potential for efficiency benefits. Similarly, the greater the information content in the outputs, the larger the potential for effectiveness benefits.

The number of data themes overlaid is expected to have a linear relationship with the level of efficiency benefits. There are no economies of scale with analysis complexity, however, it is expected to have a curvilinear relationship to the level of effectiveness benefits. It is true that diminishing returns apply to the simple addition of data themes. For example, assume that it is necessary to predict what effect increased logging in a national forest would have on an endangered species. Expanded logging would create various environmental stresses that could affect the endangered species. To find the single most dangerous stress, one would examine each stress independently. As more and more separate stresses were examined, diminishing returns would quickly set in.

However, the concurrent examination of multiple data themes also involves the ever-increasing complexity of interaction effects. Interaction effects can be very important. For example, perhaps no one environmental stress would have a serious effect on the endangered species, but the cumulative effect of many stresses would be fatal. The interaction effects created as more themes are overlaid could make a major contribution to the value of the output. The number of interactions between data themes increases geometrically as the number of data themes increases arithmetically.

The measures for output complexity are expected to have a linear relationship to the level of both efficiency benefits and effectiveness benefits. There are no economies of scale with these measures. Increasing any of them is likely to result in a proportionate increase in the complexity of the application. Likewise, there are no significant diminishing returns to the number of different uses for the output or to the probability of the output being used in adversarial hearings. Increasing either of these measures is likely to result in a proportionate increase in the value of the output.

It also is likely that there are interaction effects between the different aspects of complexity. An application's overall complexity is more than just the sum of its input, analysis, and output complexity; these aspects are more likely to be multiplicative than additive. The impact of overall complexity on the level of GIS benefits is expected to vary depending on the relative strengths of the three different aspects of complexity.

A Model to Estimate Benefits

The USGS has linked the complexity factors discussed in the previous section from a theoretical perspective to the general framework for GIS benefits to produce a quantitative model for estimating these benefits. The model is specified as a pair of ordinary least squares multiple regression equations. Input to the model is provided by a series of 62 case studies of Federal GIS applications conducted by the USGS in 1990 and 1991 (Gillespie 1991). The model estimates efficiency and effectiveness benefits independently.

Pure effectiveness benefits (that is, where $\text{Cost}_{\text{NON-GIS}} = \text{Cost}_{\text{GIS}}$) are estimated by the equation

$$\text{LT} = \underset{(3.5)}{3.752} + \underset{(5.7)}{0.673}\ \text{INPLEX1} + \underset{(1.6)}{0.045}\ \text{INTERACT} + \underset{(2.3)}{0.429}\ \text{OUTPLEX} + \underset{(2.8)}{3.147}\ \text{SMALL} + \text{residual}$$

where LT = Natural log of the dollar value of the pure effectiveness benefits
INPLEX1 = Measure of input complexity
INTERACT = Measure of analysis complexity
OUTPLEX = Measure of output complexity
SMALL = Dummy variable reflecting overall complexity of application

The equation has an R^2 of 0.592, an F value of 11.250, and is based on 36 observations. The t statistics are in parentheses below each coefficient.

The R^2 value means that the equation explains about three-fifths of the measured variation in the level of effectiveness benefits across the 36 applications studied. The F statistic tests the hypothesis that all of the coefficients except the intercept are 0. There is less than 1 chance in 10,000 of obtaining an F value this high if all of the coefficients are 0. The t statistics test if each coefficient individually is equal to 0. All of the variables except INTERACT are significant at the 99 percent confidence level. This means that there is less than 1 chance in 100 that the coefficient is 0. INTERACT is significant at the 80 percent level.

The equation predicts the natural log of pure effectiveness benefits, implying that unit changes in the independent variables cause percentage changes in the level of effectiveness benefits. For example, consider the effect of a one-unit increase in the output complexity factor of the number of distinct uses. The level of effectiveness benefits increases by the value of the outputs to the new class of users. Lacking other information, the best estimate of the value to the new class of users is the mean value to the previous classes of users. The increase in the level of effectiveness benefits depends on the previous level; the increase is a constant percentage, not a constant dollar amount. That is, the marginal effect of each of the independent variables on the dollar amount of effectiveness benefits increases with the level of effectiveness benefits.

The dollar estimate of pure effectiveness benefits is found by taking the antilog of the estimated natural log. For example,

when LT = 5, the dollar value = \$148;
when LT = 7, the dollar value = \$1,097;
when LT = 9, the dollar value = \$8,103.

If the effectiveness benefits are not pure (that is, if $\text{Cost}_{\text{GIS}} \neq \text{Cost}_{\text{NON-GIS}}$), then the difference between Cost_{GIS} and $\text{Cost}_{\text{NON-GIS}}$ must be subtracted from the estimated total. For example, if estimated pure effectiveness benefits = \$5,000, Cost_{GIS} = \$2,000, and $\text{Cost}_{\text{NON-GIS}}$ = \$500, then estimated net effectiveness benefits are \$5,000 - (\$2,000 - \$500) = \$3,500.

Pure efficiency benefits are estimated by the equation

$$\text{RATIO} = \underset{(7.9)}{0.477} + \underset{(6.5)}{0.100}\ \text{INPLEX2} - \underset{(-0.4)}{0.001}\ \text{INTERACT} + \underset{(4.3)}{0.051}\ \text{OUTPLEX} + \underset{(6.2)}{0.377}\ \text{SMALL}$$

$$+ \underset{(4.4)}{0.232}\ \text{COST} - \underset{(-4.1)}{0.186}\ \text{LAND} + \text{residual}$$

where RATIO = Ratio of efficiency benefits from GIS to manual cost of running the application
INPLEX2 = Measure of input complexity
INTERACT = Measure of analysis complexity
OUTPLEX = Measure of output complexity
SMALL = Dummy variable reflecting overall complexity of application
COST = Dummy variable reflecting cost of performing application with manual methods
LAND = Dummy variable reflecting subject area of application

The equation has an R^2 of 0.742, an F value of 11.531, and is based on 31 observations. The t statistics are in parentheses below each coefficient. The equation explains about three-quarters of the measured variation in the ratio of efficiency benefits to manual cost across the 31 applications studied.

The efficiency equation has some structural differences from the effectiveness equation. Rather than estimating the absolute level of efficiency benefits, the equation estimates the fraction of the manual cost of running the application that is saved by the use of GIS technology. Because the manual cost restricts the efficiency benefits to a maximum value, manual cost is an important factor to include in any model. Incorporating the manual cost into the dependent variable eliminates the need to include it as an independent variable. This brings the influences of the other variables into clearer view.

The two additional dummy variables are included because of the above change in the dependent variable. COST flags applications that are neither very expensive nor very inexpensive to run manually. It is expected that applications in the midrange of manual cost will tend to save a larger percentage of their manual cost than would be estimated solely on the basis of the values of the other variables. This is due to the frequency with which this type of application is run; more expensive applications tend to be run less frequently. An agency wouldn't have to save a very large percentage of the manual cost of a less expensive application to make it valuable to use a GIS. Less expensive applications are run very frequently, and the sheer volume makes the total efficiency benefits large. An agency wouldn't have to save a very large percentage of the manual cost of an expensive application to make the use of a GIS valuable. Such applications are so expensive that the efficiency benefits are large in absolute terms anyway. However, an agency does have to save a large percentage of the manual cost of a moderately expensive application to make the use of a GIS valuable. Such applications cannot be justified on the basis of volume (because they are not run very frequently) or on the basis of large absolute savings (because the manual costs are not extremely large), and so they require a larger percentage of savings.

LAND flags applications that are primarily concerned with the economic value of the land (for example, forestry, soils, water resources) rather than with the land as the location of other human activity (for example, transportation, emergency preparedness, urban planning). It is expected that such applications will tend to save a smaller percentage of their manual cost than would be estimated solely on the basis of the values of the other variables. This is because LAND applications are more expensive to run (both manually and with a GIS) than are non-LAND applications because they are more likely to involve continuous variables (for example, soil conditions change incrementally over a geographic area), and non-LAND applications are more likely to involve discrete variables (for example, political units change abruptly at defined boundaries). The fuzziness of continuous variables can increase the difficulty of both processing and analysis and thus raise the cost of running an application. The higher level of both types of costs reduces the ratio of efficiency benefits to manual cost.

All of the variables are significant at the 99 percent level except for INTERACT. The low significance (and negative coefficient) for INTERACT also is due to estimating the fraction of savings from using a GIS. Because a GIS can handle additional concurrent overlays very easily, it was expected that Cost_{GIS} would increase very little when analysis complexity increased. This in turn would lead to an increase in the fraction of the manual cost saved by using a GIS. The equation contradicts this expectation. It appears that there is a significant increase in Cost_{GIS} associated with an increase in analysis complexity. The explanation for this is probably that, even though the marginal cost of physically overlaying another data theme is trivial with a GIS, the marginal cost of interpreting the results is not trivial. Whether the overlays are done manually or with a GIS, it is considerably more difficult to interpret the results of overlaying a larger number of themes.

This does not mean that a GIS is not valuable for handling increased analysis complexity; all other things being equal, the level of efficiency benefits will increase when the analysis complexity of the application increases. However, the effect of increased analysis complexity on the ratio of efficiency benefits to manual cost is indeterminate. That is, there is no firm theoretical expectation as to the direction of the effect; the direction becomes an empirical question.

The dollar estimate of pure efficiency benefits is found by multiplying the estimated ratio times the manual cost of running the application. For example,

when RATIO = 75.0 and $\text{Cost}_{\text{NON-GIS}}$ = \$1,000, the dollar value = \$750;

when RATIO = 80.0 and $\text{Cost}_{\text{NON-GIS}}$ = \$200, the dollar value = \$160.

When an application generates both effectiveness benefits and efficiency benefits, then the estimate of GIS benefits is the sum of the estimates from the two equations.

How to Use the Model

The GIS benefits estimation model can be a powerful tool for improving GIS costs/benefits studies. The model can produce reasonable estimates of the likely level of benefits for a fraction of the cost of direct benefits measurement. There are 10 steps to follow:

1. Identify the different types of applications that will be run using a GIS.

For each type of GIS application:

2. Identify the major source of benefits:
   a. Efficiency benefits: that is, lower cost to run the application
   b. Effectiveness benefits: that is, higher value output from the application

    c. Both types of benefits are important.
3. Estimate how frequently the application will be run.

For each application where efficiency benefits are expected to be important:

4. Estimate the information needed to run the equation:
    a. The values of the complexity variables used in the efficiency equation. (Details on the construction of these variables are provided in the appendix.)
    b. The cost of running the application using the existing (non-GIS) method.
5. Enter the estimated values for the variables into the equation.
    - The result is an estimate of the fraction of $\text{Cost}_{\text{NON-GIS}}$ that will be saved by the use of GIS technology.
6. Convert the fraction to dollars, and aggregate across applications.
    a. Multiply the fraction by the estimated manual cost.
        - The result is an estimate of the dollar value of GIS efficiency benefits for running the application.
    b. Multiply the estimated efficiency benefits times the frequency with which the application will be run.
        - The result is the total dollar value of efficiency savings for the application.
    c. Sum these totals across all efficiency applications.
        - The result is the total dollar value of efficiency benefits for the use of GIS technology.

For each application where effectiveness benefits are expected to be important:

7. Estimate the information needed to run the equation.
    a. The values of the complexity variables used in the effectiveness equation.
    b. If it is more expensive to run the application using a GIS, the amount by which Cost_{GIS} is greater than $\text{Cost}_{\text{NON-GIS}}$.
8. Enter the estimated values for the variables into the equation.
    - The result is an estimate of the natural log of the dollar value of the new or improved outputs the GIS will produce.
9. Convert the estimate to dollars and aggregate across applications.
    a. Take the antilog of the estimated natural log.
        - The result is the dollar value of pure effectiveness benefits for running the application.
    b. Subtract the excess of Cost_{GIS} over $\text{Cost}_{\text{NON-GIS}}$.
        - The result is the dollar value of the net effectiveness benefits for running the application.
    c. Multiply the net effectiveness benefits times the frequency with which the application will be run.
        - The result is the total dollar value of the net effectiveness benefits for the application.
    d. Sum these totals across all effectiveness applications.
        - The result is the total dollar value of the effectiveness benefits for the use of GIS technology.
10. Verify the reasonableness of the benefit estimates by selecting a small number of applications and performing a traditional benefit measurement on them.

The ten-step process produces a suite of outputs that together tell a compelling story about the potential value of GIS technology.

1. Quantitative estimates of GIS benefits. Impressive on their own, they can be combined with cost data to produce costs/benefits ratios, net present values, internal rates of return, and project breakeven dates.
2. Case studies of selected applications. These demonstrate in concrete terms that the estimated benefits are real.
3. Ratio of the dollar values of effectiveness to efficiency benefits. This dramatically demonstrates where the value of a GIS truly lies. Typically the ratio will be large, making it clear that GIS is an enabling technology, primarily important because it helps agencies work better, not because it helps them work cheaper.

The Montana Geographic Information Council (MGIC) followed these ten steps to analyze GIS implementations in state and county governments of Montana (McInnis & Blundell 1998). They report that "the model is an excellent tool for assessing the benefits of GIS installations." The full report of the MGIC is available online at: http://www.mt.gov/isd/groups/eacba/eacba_cba.htm.

Conclusion

The general framework for GIS benefits is broad enough to support many different models for estimating those benefits. The USGS tactic of concentrating on complexity factors is not the only possible approach, but it has proven to be a fruitful one. Within the broad categories of input, analysis, and output complexity, there is room for much experimentation concerning which variables to include and how to combine them. The specific forms of the complexity variable used in the USGS model work well for the particular set of highly diverse Federal GIS applications studied. Alternative formulations of the variables might be more appropriate for specific types of applications or for applications run by non-Federal agencies. There is much useful work still to be done. The USGS research provides a firm foundation upon which to build a better knowledge of where and why GIS technology is valuable.

Appendix: Construction of Independent Variables in USGS Model

Input Complexity

Input complexity concerns the data themes needed to perform the application. It is modeled as a combination of the total volume of input data and the areal extent of the application.

The total volume of input data (MB) is measured as the number of megabytes of computer memory required to hold the data used during a single occurrence of the GIS application. The areal extent of the application (MU) is measured in map units. One map unit is the area that can be represented on a map sheet with the physical dimensions of a USGS 1:24,000-scale quadrangle (18 inches by 23 inches). The actual square miles of area included in a map unit varies according to the map scale used.

Both MB and MU have a log-linear relationship to GIS benefits. The log forms of these variables are highly correlated and so subject to multicollinearity. The presence of multicollinearity reduces the precision of the estimates, making it difficult to disentangle the relative influences of the affected variables. This problem can be eliminated either by combining the two variables into a single measure of input complexity, or by dropping one of the variables from the model.

INPLEX1 (used in the effectiveness equation) is defined as the natural log of MB plus the natural log of MU. INPLEX2 (used in the efficiency equation) is defined as the natural log of MU.

The alternative treatment of input complexity in the efficiency equation is required by the choice of the dependent variable. Although MB is positively related to the absolute level of efficiency benefits, the effect of MB on the ratio of efficiency benefits to manual cost is theoretically indeterminant. Empirical results show a slight negative relationship. The best solution to the multicollinearity problem, therefore, simply is to drop MB from the efficiency equation.

Analysis Complexity

Analysis complexity concerns how the data themes are manipulated inside the application. It is modeled by the number of potential interactions between data themes. If MAX is the maximum number of distinct data themes overlaid at any one step in the analysis, then the number of potential interactions is given by the formula $(MAX^2 - MAX) / 2$.

Output Complexity

Output complexity concerns the products of the application. It is modeled as a combination of the number of distinct uses for the outputs and the likelihood that the outputs will be used in adversarial hearings.

Variety of uses (VU) is measured as the total number of separate concerns that must be kept in mind by the GIS staff when running the application, and that must be addressed by the output of the application. It is closely related to the number of special interest groups that will be watching the agency's actions.

The likelihood of use in adversarial hearings (AH) is measured by the probability that the output from a typical single occurrence of the GIS application will be used to support the position of one (or both) sides in a formal adversarial setting. This could be in a legal setting (before a jury, judge, or regulatory commission) where the decision reached is legally binding. It could be in an administrative setting (at a public hearing of some sort) where the expressed opinions are advisory only. "Adversarial hearings" definitely refers to something more structured than internal agency disputes over the most appropriate management policy. It refers to situations where, after the agency has reached a decision, some group external to the agency challenges that decision in some structured setting. "Adversarial" implies controversy; there are winners and losers. Tempers are likely to rise. This does not refer to mere informational presentations, but to meetings where at least some of the participants have staked out opposing positions.

Both VU and AH have a linear relationship to GIS benefits. The two variables are highly correlated and so subject to multicollinearity. The problem is eliminated by combining the two variables into a single measure of output complexity.

OUTPLEX = (VU / 3) + (AH / 25). The values for the denominators weight the two variables approximately equally.

Overall Complexity

The overall complexity of the application (that is, interaction effects between the three aspects of complexity) is measured by the dummy variable SMALL. SMALL has a value of 1 when all aspects of complexity are low, and a value of 0 otherwise.

Operationally, SMALL is determined by appropriately weighting and then summing the five factors used to model the complexity factors.

- MU and MB are weighted by taking the natural log of the raw value, and rounding up to 0 when the natural log is less than 0.
- MAX and VU are weighted by dividing the raw value by 3 and rounding up to the next highest integer, with a maximum value of 4.
- AH is weighted as follows:

Raw percentage = 0 (never used): weighted value = 0
Raw percentage = 1 to 49 (sometimes used): weighted value = 1
Raw percentage = 50 to 99 (often used): weighted value = 2
Raw percentage = 100 (always used): weighted value = 3

In the effectiveness equation, SMALL=1 when the sum of the weighted factors is less than 6. In the efficiency equation, SMALL=1 when the sum of the weighted factors is less than 5. There are natural breaks in the data from the USGS case studies at these points. In addition, these values make sense intuitively as reasonable definitions of a "simple" GIS application generating each type of benefit.

Level of Manual Cost

The dummy variable COST has a value of 1 when the cost of performing a single occurrence of a GIS application using non-GIS methods is between $20,000 and $50,000, and a value of 0 otherwise.

Subject Area of Application

The dummy variable LAND has a value of 1 when the primary function of the GIS application is in the areas of agriculture, fish and wildlife management, forestry, land management, soils, or water resources. LAND has a value of 0 when the primary function of the GIS application is in the areas of commerce and economic development, defense law enforcement and emergency preparedness, energy and mineral management, environmental protection, geological surveys, library and academic research, parks and recreation, taxation and revenue, transportation, and urban and regional planning.

References

Anderson, Carrie S. 1992. "GIS Development Process: A Proactive Approach to the Introduction of GIS Technology." *GIS/LIS Proceedings*, pp. 1-10.

Aronoff, Stan. 1989. *Geographic Information Systems: A Management Perspective.* WDL Publications, Ottawa, Canada.

Brookshire, D.S., M.S. Thayer, W.D. Schulze, and R.C. d'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches." *The American Economic Review*, 72, pp. 165-177.

Bureau of Indian Affairs. 1988. *Final Report on Cost and Benefit Analysis of Geographic Information System Implementation to Bureau of Indian Affairs.*

Campbell, H.J. 1999. "Institutional Consequences of the Use of GIS." Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind, eds. *Geographical Information Systems*, pp. 621-631.

Clawson, Marion and Jack Knetsch. 1966. *Economics of Outdoor Recreation.* Johns Hopkins University Press for Resources for the Future, Baltimore, MD.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Rowman and Littlefield, Totowa, NJ.

Dickinson, Holly J. and Hugh W. Calkins. 1988. "The Economic Evaluation of Implementing a GIS." *International Journal of Geographical Information Systems*, 2:4, pp. 307-327.

Gillespie, S. 1991. "Measuring the Benefits of GIS Use." *Technical Papers, 1991 ACSM-ASPRS Fall Convention*, pp. 84-94.

Gillespie, S. 1994. "Measuring the Benefits of GIS Use: Two Transportation Case Studies." *URISA Journal*, 6:2, pp. 62-67.

McInnis, Logan and Stuart Blundell.1998. *Analysis of Geographic Information Systems (GIS) Implementations in State and County Governments of Montana,* http://www.mt.gov/isd/groups/eacba/eacba_cba.htm.

Obermeyer, N.J. 1999. "Measuring the Benefits and Costs of GIS." Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind, eds. *Geographical Information Systems*, pp. 601-610.

Peterson, George L. and Cindy F. Sorg. 1987. *Toward the Measurement of Total Economic Value.* USDA Forest Service General Technical Report RM-148.

Ripple, William J., editor. 1987. *Geographic Information Systems for Resource Management.* American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping.

Smith, Douglas A. and Roger F. Tomlinson. 1992. "Assessing Costs and Benefits of Geographical Information Systems: Methodological and Implementation Issues." *International Journal of Geographical Information Systems*, 6:3, pp. 247-256.

Somers, Rebecca. 1994. "URISA '94: Trying To Keep Up." *GeoInfoSystems*, October 1994, pp. 20-22.

Tsui, Mary. 1994. "URISA '94: A Conference Overview." *GeoInfoSystems*, July 1994, pp. 42-45.

Viscusi, W. Kip. 1993. "The Value of Risks to Life and Health." *Journal of Economic Literature*, 31:4, pp. 1912-1946.