Assessing Quality Improvement Initiatives when Expert Judgments are Uncertain

Eugene D. Hahn
Department of Information
and Decision Sciences
Salisbury University
Salisbury, MD 21801

Cynthia L. Knott
Department of Information
Management and Management Science
Marymount University
Arlington, VA 22207

Abstract

A new approach for examining quality improvement initiatives regarding errors in the U.S. Census Bureau's Master Address File (MAF) and the Topologically Integrated Geographic and Referencing System (TIGER) databases is presented. A stochastic multi-criteria decision-making method involving Bayesian weighted hierarchical multinomial logit models is used to conduct inference on the priorities in a multiple-expert decision scenario. Quality initiatives regarding basic street address-level address matching, geocoding completeness, and geocoding quality were judged to be the most important for MAF/TIGER improvement at the 95% probability level. The approach allows managers to go one step further in understanding the relative impact of various types of errors on overall quality and thus be better prepared to select approaches to reduce these errors.

Keywords: decision support systems; stochastic multi-criteria decision making; government; geographic information systems; group decision making

Introduction

The requirement of a decennial census is specified in the Constitution of the United States (Article I, §2) so that the seats in the U.S. House of Representatives may be apportioned among the states. To ensure this apportionment is fair and equitable, the U.S. Census Bureau has the following key responsibility: to count each person within the United States in the correct location. The information derived from the decennial census and supplementary surveys is used in an extremely broad context, supporting not only policy decision making at the federal, state, and local levels, but also diverse activities such as academic social research, business development, transportation planning, school and housing development, public health initiatives, law enforcement, and emergency planning (National Research Council, 1995). Census information also plays an important role in national elections as the composition of the U.S. House of Representatives (derived from Census results on population) is used to allocate the composition of the Electoral College (e.g., U.S. Census Bureau, 2001). Clearly it is important for Census data to be as accurate as possible and this fact has not been lost on the executive and legislative branches of the government. Recently the Office of Management and Budget (OMB), in its capacity of assisting the President, has issued guidelines to all federal agencies with respect to maximizing data quality and ensuring the accuracy of statistical information. These guidelines, part of the Federal Data Quality Act (Public Law 106-554; H.R. 5658, §515) passed in December 2000, charge federal agencies with the responsibility to "issue guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by the agency" (OMB, 2001).

In order to be able to count each person in the correct location, the Census Bureau must be able to accurately determine address location. To do this, the Bureau relies on two important sources of information. One of these is TIGER, the Topologically Integrated Geographic and Referencing System. This geospatial database, which can be likened to a digital map, contains visible physical features such as roads, railway lines and bodies of water as well as numerous invisible administrative and political boundaries such as state and county lines, ZIP code regions and Congressional district borders. The second source of information is the Master Address File (MAF). The MAF is a database containing all of the addresses used in the decennial census and other Bureau surveys. An estimated 120 million household residences and 60 million business and other addresses are in the MAF (U.S. Department of Commerce, Office of Inspector General, 2003). Taken together, MAF/TIGER is an extensive geographic information system.

It is well known that statistics obtained from surveys are subject to sampling error. Additionally, an array of different types of non-sampling errors is known to affect surveys and censuses in general. For example, if an address is not included in the MAF, then no data can be obtained from the person(s) at that address. Or, if TIGER's representation of the location of an address is not correct, then there is a (small) possibility that the true address is on the other side of a census block boundary. If this happens, the data will be tabulated in the wrong census block. In either of the two above cases, the Bureau will be unable to count such persons in the correct location.

Thus, the question naturally arises as to how to reduce or possibly eliminate these different types of error. In organizations, resources are always finite. So, after ways are found to reduce or eliminate errors, a prioritization initiative needs to be undertaken to determine which remediating steps are to be taken first. Multi-criteria decision-making methods (MCDMs) have been shown to be helpful in addressing such problems. In using an MCDM, a decision maker is able to determine which courses of action are most important or most preferable by using his or her expert judgment. However, classical deterministic MCDMs can be shown to be predicated on an assumption that a decision maker's judgments are so accurate and well-formed that they can be taken to be completely certain (Hahn, 2003). While this may be true in some decision-making scenarios, the complexity of MAF/TIGER ensures that this is quite unlikely to be the case. It is thus preferable to approach the current problem using stochastic MCDMs (e.g., Ramanathan, 1997; Laininen & Hämäläinen, 2003; Mustajoki *et al.*, 2005). A key advantage of stochastic approaches to multi-criteria decision making is that one can conduct inferential hypothesis tests regarding the alternatives under consideration in a decision problem. Phrased differently, one can examine whether one can be, say, 95% confident that one alternative is judged to be superior to one another given that error in the judgments is present.

Multi-criteria decision making methods: deterministic and stochastic approaches

MCDMs are an important set of tools for addressing challenging decisions as they allow individuals to better proceed in the face of uncertainty, complexity and conflicting objectives. In using these methods, a decision maker evaluates various alternatives along different criteria by means of expert judgments. The goal is to determine which of the alternatives best satisfies all of the (possibly conflicting) criteria or objectives. The decision maker will have to supply expert judgments regarding the extent to which the alternatives satisfy the objectives as well as the relative importance of the objectives. If the decision maker is able to do this, he or she need only use a MCDM which is a mathematical procedure for determining the underlying "priority" or weight of each alternative. There are a variety of MCDMs to choose from (e.g., Simpson, 1996; Triantaphyllou, 2000, provides a recent comparative review).

MCDMs are predicated on a series of axioms. As mentioned previously, it can be shown that classical deterministic MCDMs which utilize scalar numeric judgments are predicated on judgmental certainty. There are undoubtedly situations where judgments are completely certain. Nonetheless, we may prefer to use a stochastic MCDM when judgments are uncertain. In such an approach, the uncertainty in judgments is reflected in a probability model. We can see therefore that classical MCDMs are a special case of stochastic MCDMs where the uncertainty has been eliminated (or has been assumed to have been eliminated). A key advantage of using a stochastic MCDM is that a variety of probability statements can be made regarding the alternatives. For example, we would be in a position to determine whether we could be 95% confident that one alternative was superior to another.

Given the complexity of MAF/TIGER, in the current study we utilize a newly-developed stochastic MCDM (Hahn, 2003), which for brevity can be termed a stochastic judgment method or SJM, and furthermore extend the methodology to the group decision-making context. An essential aspect of this method is that the pairwise comparison judgments are assumed to arise as manifestations of underlying stochastic phenomena. Note that a pairwise comparison can be conceptualized as a series of "preference outcomes" for a pair of alternatives. For example, a 2:1 preference for Alternative A over Alternative B indicates that there are two outcomes of preference for A for every outcome of preference for B. From this, one can show that the preference outcomes follow a binomial distribution with underlying priority parameter, p. Briefly, a pairwise comparison ratio C_{ij}, where $i \neq j$, results from a pairwise comparison of two alternatives O_i and O_j with underlying weights w_i and w_j. Clearly, all w must be positive by analogy to physical weights and for the moment we constrain $w \in \mathbb{N}$ and $w_i \geq w_j$, so that $C_{ij} \geq 1$. Then C_{ij} indicates the amount by which O_i is preferred to O_j. That is, for every outcome of preference for O_j, there are C_{ij} outcomes of preference for O_i. This is essentially the same as the ratio of success outcomes and failure outcomes in a binomial process. As such, the pairwise comparison ratios can be used to estimate a binomial process in which w_i successes have been observed in $(w_i + w_j)$ trials subject to an unobserved preference parameter, p_i. Note that this also remains true when $w_i < w_j$, the difference being that in such a case O_j would be preferred to O_i. It is revealing to divide the numerator and the denominator of C_{ij} by the sum of the weights to show

$$C_{ij} = \frac{w_i}{w_j} = \frac{\frac{w_i}{w_i+w_j}}{\frac{w_j}{w_i+w_j}} = \frac{p_i}{1-p_i}.$$

In the above expression, $p_i/(1-p_i)$ is the ratio of preferences and p_i is the stochastically derived priority if we take $w_i \sim \text{Binomial}(w_i+w_j, p_i)$. If there are more than two alternatives, the process can be shown to be multinomial by extension. Thus, estimation of the priority vector, $\mathbf{p}$, may be undertaken by the use of multinomial logit models. Specifically, if i

indexes the rows and k the columns of the judgment matrix, we have

$$w_{ik} \sim \text{Multinomial}\left(\sum_{k=1}^{K} w_{ik}, p_{ik}\right), \tag{1}$$

which indicates that the weights have a multinomial distribution given the total preference outcomes and the underlying priorities of the alternatives. The final priority for each alternative is obtained by averaging over the I rows in the matrix. So, the final priority for the k^{th} alternative is $\frac{1}{I}\sum_{i=1}^{I} p_{ik}$. The collection of the I final priorities is the priority vector, $\mathbf{p}$.

The multinomial logit model continues with the specification that

$$p_{ik} = \frac{\exp(\alpha_k + \beta_{ik})}{\sum_{k=1}^{K}\exp(\alpha_k + \beta_{ik})}.$$

However, it is important to recognize that the judgments made by a particular decision maker for a particular set of alternatives are not independent of one another. As has been recently pointed out by Leung *et al.* (2005), dependency in judgments can be handled in a non-statistical way in MCDMS by, for example, using the ANP extension of AHP. In the SJM, however, it is straightforward to characterize dependence statistically through the incorporation of correlation structures in the model (e.g., Hahn, 2006). As such, a hierarchical model is used here. The hierarchical model provides a means for accounting for judgment dependence which is inherent in multi-criteria decision approaches. The hierarchical component of the model is

$$\beta_{22}, \ldots, \beta_{2K}, \beta_{32}, \ldots, \beta_{IK} \sim \text{Normal}(0, \tau),$$

which indicates that the unrestricted β parameters follow a common normal distribution with mean zero and precision τ. Note the parameterization is in terms of the precision as opposed to the variance (i.e., $\tau = 1/\sigma^2$). A Bayesian approach is adopted for modelling purposes. With the modelling specification given above, the posterior is

$$p(\alpha_k, \beta_{ik}, \tau | w_{ik}) \propto \prod_{i=1}^{I} \prod_{k=1}^{K} p(w_{ik} | \alpha_k, \beta_{ik}) \times \prod_{k=2}^{K} p(\alpha_k) \times \prod_{i=2}^{I} \prod_{k=2}^{K} p(\beta_{ik} | \tau) \times p(\tau). \tag{2}$$

Another factor to be addressed involves the fact that the judgment matrix is reciprocally symmetric across the diagonal. Hence, some of the information in the matrix is redundant with information in other parts of the matrix. It is thus necessary to downweight the information content of the matrix to obtain accurate standard errors. The appropriate weight can be shown to be

$$Q = \frac{I^2 + I - 2}{2I^2}$$

where I is the number of rows. Then, for modelling purposes the adjusted likelihood is

$$w_{ik} \sim \text{Multinomial}\left(Q \sum_{k=1}^{K} w_{ik}, p_{ik} \right) \tag{3}$$

and we use the adjusted likelihood in Expression (3) instead of the original likelihood in Expression (1). In summary, the SJM methodology involves the use of Bayesian weighted hierarchical multinomial logit models. Estimation of the model is straightforward to accomplish by using Markov chain Monte Carlo methods. In the original formulation of the SJM methodology, it was assumed that there was only one expert decision maker providing judgments. Here, we extend the methodology to accommodate the more general case of there being a group of experts. Prior to doing so, however, we review the key components of the model.

Model goal, objectives and alternatives

The information in TIGER was drawn from a variety of available data sources (O'Grady & Godwin, 2000) of varying quality (Liadis, 2000). Like TIGER, the MAF has been compiled from numerous data sources (National Research Council, 2003). Thus, as with TIGER, quality assurance in the MAF is an important and ongoing concern. Previous MAF quality assurance efforts in preparation for Census 2000 have been described by Gbur *et al.* (1994). The goal of the model is therefore to determine how to best prioritize error remediation initiatives.

Housing unit coverage

The goal of a census is to provide a complete count of the entire population at a given point in time. If the whereabouts of every single member of the population is known, then in principle each person may be counted. However, if the whereabouts of a particular person is not known, then he or she is very unlikely to be counted. This leads to what is called an undercount. Because multiple strategies may be used to locate hard-to-find subpopulations, it is also possible for a variety of reasons to count the same individual more than once. This leads to what is called an overcount. Perfect coverage of all of the individuals in the population is the desired goal. As such, this objective is one of high priority for the Bureau.

Block assignment

The Census Bureau tabulates data at a number of geographical levels. These include the familiar state and county levels as well as smaller levels. The smallest geographical level utilized by the Bureau is the census block. If an individual is assigned to the wrong block, then she will be counted but in the wrong location. Thus, a failure to locate an individual in the correct block is of major concern. However, this kind of error is less severe than missing the individual entirely and moreover after aggregation over one or two geographic levels the chance that such an error will disappear is considerable (for example, an individual may be assigned to the wrong block but the chances he or she is assigned to the wrong state are very small).

Field operations

The Bureau relies considerably on mailed surveys to collect information. Collecting information by mail is very inexpensive. However it has long been appreciated that, despite being economical, mail survey results are susceptible to bias due to nonresponse (Hansen & Hurwitz, 1946). To ameliorate this bias, Census field enumerators conduct follow-up interviews with non-responders. Thus, we see that the objective of enhancing field operations is important. However, given an adequate level of field operations, the previously mentioned objectives which pertain to counting each person in the right location are more crucial to the integrity of census results.

Software applications development

A key task of the software applications associated with MAF/TIGER is to provide for geocoding. When an address is geocoded, it is mapped or assigned to a specific location so that the data associated with that address may be tabulated appropriately. Some kinds of errors are known to affect this automated geocoding process and thus require additional resources to be utilized for correction. Thus, errors that affect software applications are undesirable and the expectation is that these errors should not exist. However, in many cases workarounds for these errors can be devised.

Alternatives

There are nine quality initiatives that constitute the alternatives under consideration. Initiative 1 is *geocoding completeness.* Geocoding completeness refers to the percentage of addresses that the automated address geocoding system assigned to a set of census geographies. Initiative 2 is *geocoding quality*, which refers to the correctness of the census block assignment. *Address conversions*, Initiative 3, involve mainly the change of a rural route address to a city style address. *Feature coverage*, Initiative 4, refers to the presence of all existing road features in the TIGER database. *Feature ID accuracy*, Initiative 5, refers to the presence of a feature's name and the accuracy of the name. For example, required prefixes (such as 123 *North* Main) and/or suffixes (such as 1600 Pennsylvania Avenue, *N.W.*) must be present and correct. A basic street address (BSA) is the "house" number and street name assigned by local government authorities to a structure or building for identification. Thus, *address matching-BSA level*, Initiative 6, refers to the accuracy of the automated address matching of a new address to a building. *Address matching-within BSA*, Initiative 7, refers to the accuracy of the automated address matching of a new address with an apartment designation within a building. *Group quarters identification*, Initiative 8, refers to the correct identification of a group quarters (e.g., jails, nursing homes) as a group quarters since enumeration procedures for group quarters are different. Finally, *TIGER Map Enhancements*, Initiative 9, are additional locations added to the TIGER database to improve map usability (primarily for enumerators) and map appearance.

Empirical analyses

Examining the initiative categories

The four categories of objectives were examined first (see also the top of Figure 1). Judgments regarding the contribution of the different initiatives towards MAF/TIGER improvement efforts were obtained from four senior Bureau members with MAF/TIGER responsibilities (three with collective Bureau experience of over 70 years and the fourth a Ph.D. geographer). The decision task was administered by having each decision maker make pairwise comparisons among the alternatives with respect to a particular objective. For example, with respect to housing unit coverage, each expert was asked how much more preferable was geocoding completeness as compared to feature ID accuracy. Experts used a numeric scale anchored with descriptive labels. A 1:1 preference for Alternative A over Alternative B denoted equal preference, a 3:1 preference denoted a moderate preference for A, and 5:1, 7:1, and 9:1 preferences denoted strong, very strong, and extreme preferences respectively. Intermediate scale values were used for intermediate judgments and the reciprocal of the above values were used when B was preferred to A.

In the case of four experts, we again note that judgments are conditionally independent given the parameter vector. We can then extend the method described in Expression 2 as follows. The model as described is replicated for each one of the experts. Specifically, if the experts are indexed by n, the posterior is

$$p(\alpha_{kn}, \beta_{ikn}, \tau_n | w_{ikn}) \propto \prod_{n=1}^{N} \prod_{i=1}^{I} \prod_{k=1}^{K} p(w_{ikn} | \alpha_{kn}, \beta_{ikn}) \times \prod_{k=2}^{K} p(\alpha_{kn}) \times \prod_{i=2}^{I} \prod_{k=2}^{K} p(\beta_{ikn} | \tau_n) \times p(\tau_n). \quad (4)$$

Markov chain Monte Carlo integration can then be utilized to obtain aggregate results for the group. In the current study, vague priors were placed on all model parameters. Specifically, the α and β coefficients were given vague independent normal priors (means of zero and variances of 100,000) and τ was given a vague gamma prior (shape and scale parameters of 0.05). The first 10,000 iterations of the run were discarded as a burn-in period. The chain was then allowed to run for an additional 50,000 iterations. In general, this class of models is well-behaved and convergence to the posterior is rapid.

———————— Figure 1 About Here ————————

Among the categories of objectives, in Table 1 we see that housing unit coverage is of the highest concern overall with a priority of almost 50%. It was followed by block assignment with an appreciably smaller contribution than housing unit coverage (priority of 35%). By contrast, field operations and software applications development had priorities of less than

10%. As a point of reference, we compare the results of the current method to a deterministic MCDM based on pairwise comparisons, the Analytic Hierarchy Process (Saaty, 1977). The priority estimates from both of these MCDMs appear in Table 1. AHP gives more weight to the two middle-ranked categories of block assignment and field operations. SJM by contrast gives more weight to housing unit coverage. However, at a broader level the priorities are fairly consonant, especially considering the priorities in both cases were generated by distinctly different mathematical approaches. For example, the mean absolute deviation of one set of estimates from the other was only 0.004.

——————— Table 1 About Here ———————

Table 1 shows that the priorities of some categories of initiatives seem to be much different from others. For example, the priority of housing unit coverage is over six times greater than the priority of software applications development. However, a user of a deterministic MCDM could only make qualitative statements regarding any differences in the priorities. By contrast, a user of a stochastic MCDM can conduct inference on the differences in priorities. That is, he or she will be able to determine whether the priorities of two alternatives are significantly different at the 95% probability level (or some other level of his or her choosing). This enables the user of a stochastic approach to put his or her conclusions on a more robust footing. By the same token, if two alternatives *cannot* be found to be significantly different from one another, again some valuable information has been obtained regarding the situation at hand. The second-to-last column of Table 1 lists the 95% posterior probability intervals for the priority parameters. We see that the probability interval for housing unit coverage does not overlap with those of field operations and software applications development. Thus, the priority for housing unit coverage is significantly different from the latter two. By contrast, we can see there is some overlap between the intervals for housing unit coverage and block assignment. Not surprisingly, the two priorities are not significantly different. In particular the 95% probability interval for the difference score of these two priorities contains the value zero. Hence we cannot conclude that these two categories' priorities are different from one another at the 95% probability level. Along similar lines, block assignment differs from field operations and software applications development at the 95% level. Field operations and software applications development, however, cannot be distinguished at the 95% level.

Examining the initiatives

Having examined the priorities of the criteria, we now turn to an examination of the overall model and the quality improvement initiatives themselves. The four categories and the nine proposed initiatives were arranged into the structure that appears in Figure 1. All other things being equal, we might expect initiatives which address the highly important category of housing unit coverage to be the most preferred. Yet we would also want to simultaneously consider the impact of the various initiatives themselves. For example, it might make sense to prioritize a very impactful initiative for a moderate-priority category ahead of a minor initiative for a high-priority category. This can be accomplished by separating the initiatives for each particular error category and calculating their individual priorities. Then, one may calculate an overall or global priority for each initiative by multiplying its individual priority by the priority of the error category with which it is associated. The stochastic modelling proceeds in a likewise fashion. Note that initiatives which do not belong to the same error category are independent of one another conditional on the parameter vector. So, five multinomial logit models as described in Expression 4 can be run essentially in parallel: one model for the error categories and one model for the set of initiatives under each of the four error categories. Specifically, if the models are indexed by m and the experts by n, the final posterior is

$$\begin{aligned} p(\alpha_{kmn}, \beta_{ikmn}, \tau_{mn} | w_{ikmn}) \propto & \prod_{m=1}^{M} \prod_{n=1}^{N} \prod_{i=1}^{I} \prod_{k=1}^{K} p(w_{ikmn} | \alpha_{kmn}, \beta_{ikmn}) \\ & \times \prod_{k=2}^{K} p(\alpha_{kmn}) \times \prod_{i=2}^{I} \prod_{k=2}^{K} p(\beta_{ikmn} | \tau_{mn}) \times p(\tau_{mn}). \end{aligned} \tag{5}$$

The global priorities for the initiatives are then obtained by MCMC methods by multiplying each initiative's priority by the priority of its error category and summing across categories. Vague priors were again used for all parameters and estimation was based on 50,000 iterations after a 10,000-iteration burn-in.

Table 2 contains the summary statistics of the initiatives' aggregate global priorities. More tellingly, Figure 2 displays a box plot of the global priorities of the 9 initiatives sorted in descending order of priority. The whiskers denote the beginning and the end of the 95% posterior credible interval for each initiative's global priority. We can see that the initiative that was judged to contribute most to the overall goal was the BSA-level address matching initiative (I6). Following our previous development, for inferential purposes we may construct the difference of the priorities of initiatives i and j, which may be denoted δ_{ij}. If the 95% posterior interval for δ_{ij} excludes zero, then evidence that $p_i \neq p_j$ is obtained. Returning to the initiatives, I6 was significantly greater than (or significantly preferable to) the remaining initiatives at the 95% level with the exception of geocoding completeness (I1). Thus, the BSA-level address matching initiative is judged to be one of the most important initiatives to implement. The initiative with the second highest priority was geocoding completeness (I1). This initiative was significantly preferable to the remaining lower-ranked initiatives with the exception of geocoding quality (I2). In turn, geocoding quality (I2) was found to be significantly preferable to the remaining alternatives below it in Figure 2. Hence, I6, I1, and I2 appeared to form a top tier of alternatives in terms of preferability or priority. Regarding the next set of initiatives, address conversions (I3) was found to be significantly preferable to the initiatives below it save feature coverage (I4). In turn, both feature coverage (I4) and address matching within BSA (I7) could only be differentiated from lower-ranked group quarters identification (I8) as well as TIGER map enhancements (I9). Lastly, both feature ID accuracy (I5) and group quarters identification (I8) were judged significantly preferable only to TIGER map enhancements (I9). A complete listing of the 36 difference score summary statistics can be found in the appendix.

——————— Table 2 About Here ———————

——————— Figure 2 About Here ———————

Discussion

The current paper extends the literature on stochastic MCDMs by providing a multiple-expert generalization of the stochastic judgment method. It also appears to be the first paper to describe the use of multiple-expert MCDMs in the context of GIS, as a recent review of the literature on the intersection of decision making and GIS only describes single-expert decision making (Huerta *et al.*, 2005). In terms of governmental decision making, the method allows decision makers to conduct mult-criteria decision making in the context of uncertain judgments which are often prevelant in the highly complex issues faced by government agencies. Here, the results show that in terms of program enactment there is a top-tier grouping of initiatives that are judged to be significantly most important in terms of impact on MAF/TIGER improvement efforts. These initiatives are BSA-level address matching (I6), geocoding completeness (I1), and geocoding conversions (I2) respectively. Additional attention should be focused on these three initiatives. Among this top tier, further delineation is possible in that the first-ranked initiative was significantly differentiable from the third-ranked initiative. A second tier of alternatives also is present in the results. This second tier consists of address conversions (I3). Note that within a tier some initiatives are not significantly different from one another. This indicates that the evidence for distinguishing these initiatives from one another may be insufficient given the presence of error in the judgments. Thus, for practical purposes one could consider treating them as a grouping or a block to be implemented simultaneously if resources were available.

It is worth pointing out that while a user of a stochastic MCDM such as the one described here would know which initiatives are significantly different from one another, a user of a deterministic MCDM would not. Thus, the practical usefulness of the approach is twofold.

First, knowledge about which alternatives are significantly preferable allows for more robust conclusions to be drawn. Second, this knowledge also permits suspect conclusions to be avoided. Consider feature coverage (I4) and feature ID accuracy (I5). A user of a deterministic MCDM might be very "confident" that feature coverage was preferable to I5 because the priority for feature coverage is over 20% larger than that of feature ID accuracy in Table 2. However, we can see from Figure 2 and the δ statistics in the appendix that such a conclusion is unwarranted. In the current scenario the decision makers could look at Figure 2 and the δ values and see that they could not be very confident that feature coverage was preferable to feature ID accuracy. In terms of governmental policy recommendations, these alternatives cannot be differentiated at the 95% level given the current state of information.

References

Gbur, P. M., Kindred, M., & Mersch, M. L. (1994). Master Address File: Update methodology and quality improvement program. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association: Alexandria, VA. 748–753.

Hahn, E. D. (2003). Decision making with uncertain judgments: A stochastic formulation of the Analytic Hierarchy Process. *Decision Sciences*, *34*(3), 443–466.

Hahn, E. D. (2006). Link function selection in stochastic multicriteria decision making models. *European Journal of Operational Research*, in press.

Hansen, M. H. & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, *41*, 517–529.

Huerta, E., Navarrete, C., & Ryan, T. (2005). GIS and decision-making in business: A literature review. In J. B. Pick (Ed.), *Geographic Information Systems in Business*, Idea Group: Hershey, PA. 20–35.

Laininen, P. & Hämäläinen, R. P. (2003). Analyzing AHP-matrices by regression. *European Journal of Operational Research*, *148*(3), 514–524.

Leung, L., Lam, K., & Cao, D. (2005). Implementing the balanced scorecard using the Analytic Hierarchy Process & the Analytic Network Process. *Journal of the Operational Research Society*, in press.

Liadis, J. S. (2000). GPS TIGER accuracy analysis tools (GTAAT) evaluation and test results. Technical report, Washington, DC, U.S. Census Bureau, http://www.census.gov/geo/ mod/gtaat2000.pdf accessed 7 Nov 2004.

Mustajoki, J., Hämäläinen, R., & Salo, A. (2005). Decision support by interval SMART/SWING—Incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, *36*(2), 317–339.

National Research Council (1995). In B. Edmonston & C. Schultze (Eds.), *Modernizing the U.S. Census*, National Academy Press: Washington D.C.

National Research Council (2003). In D. L. Cork, M. L. Cohen, & B. F. King (Eds.), *Planning the 2010 Census: Second Interim Report*, National Academy Press: Washington D.C.

Office of Management and Budget (2001). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Web page, Washington, DC, Office of Management and Budget, http://www.whitehouse.gov/omb/fedreg/reproducible.html accessed 22 Feb 2004.

O'Grady, K. & Godwin, L. (2000). The positional accuracy of MAF/TIGER. Technical report, Washington, DC, U.S. Census Bureau, http://www.census.gov/geo/mod/positional_accuracy.pdf accessed 19 Sep 2004.

Ramanathan, R. (1997). Stochastic decision making using multiplicative AHP. *European Journal of Operational Research*, *97*(3), 543–549.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234–281.

Simpson, L. (1996). Do decision makers know what they prefer?: MAVT and ELECTRE II. *Journal of the Operational Research Society*, *47*(7), 919–929.

Triantaphyllou, E. (2000). *Multi-criteria Decision Making Methods: A Comparative Study*. Kluwer: Dordrecht.

U.S. Census Bureau (2001). Congressional apportionment. C2KBR/01-7, PDF file, Washington, DC, U.S. Census Bureau, http://www.census.gov/prod/2001pubs/c2kbr01-7.pdf accessed 16 Nov 2004.

U.S. Department of Commerce, Office of Inspector General (2003). MAF/TIGER redesign project needs management improvements to meet its decennial goals and cost objective. Final Inspection Report No. OSE-15725, U.S. Department of Commerce: Washington, DC.

Appendix – Difference Scores for Aggregate Global Priorities

Parameter	*Mean*	*S.D.*	*2.5% Quantile*	*97.5% Quantile*
δ_{12}	0.029	0.018	-0.006	0.063
δ_{13}	0.059	0.014	0.031	0.087
δ_{14}	0.080	0.014	0.053	0.108
δ_{15}	0.098	0.013	0.072	0.124
δ_{16}	-0.014	0.016	-0.044	0.018
δ_{17}	0.084	0.014	0.057	0.110
δ_{18}	0.104	0.013	0.080	0.130
δ_{19}	0.157	0.012	0.134	0.180
δ_{23}	0.030	0.015	0.002	0.059
δ_{24}	0.051	0.017	0.019	0.084
δ_{25}	0.069	0.014	0.041	0.097
δ_{26}	-0.043	0.017	-0.076	-0.009
δ_{27}	0.055	0.015	0.025	0.085
δ_{28}	0.075	0.013	0.050	0.102
δ_{29}	0.128	0.012	0.104	0.153
δ_{34}	0.021	0.013	-0.004	0.046
δ_{35}	0.039	0.011	0.017	0.060
δ_{36}	-0.073	0.014	-0.100	-0.046
δ_{37}	0.025	0.012	0.002	0.047
δ_{38}	0.045	0.010	0.025	0.066
δ_{39}	0.098	0.009	0.081	0.116
δ_{45}	0.018	0.011	-0.003	0.039
δ_{46}	-0.094	0.014	-0.122	-0.065
δ_{47}	0.004	0.011	-0.018	0.026
δ_{48}	0.024	0.010	0.004	0.045
δ_{49}	0.077	0.009	0.060	0.095
δ_{56}	-0.111	0.013	-0.137	-0.086
δ_{57}	-0.014	0.010	-0.034	0.005
δ_{58}	0.007	0.009	-0.011	0.024
δ_{59}	0.059	0.008	0.044	0.074
δ_{67}	0.097	0.014	0.071	0.124
δ_{68}	0.118	0.013	0.093	0.143
δ_{69}	0.170	0.012	0.148	0.193
δ_{78}	0.021	0.010	0.001	0.040
δ_{79}	0.073	0.008	0.057	0.090
δ_{89}	0.052	0.007	0.039	0.066

Figure 1: Model structure

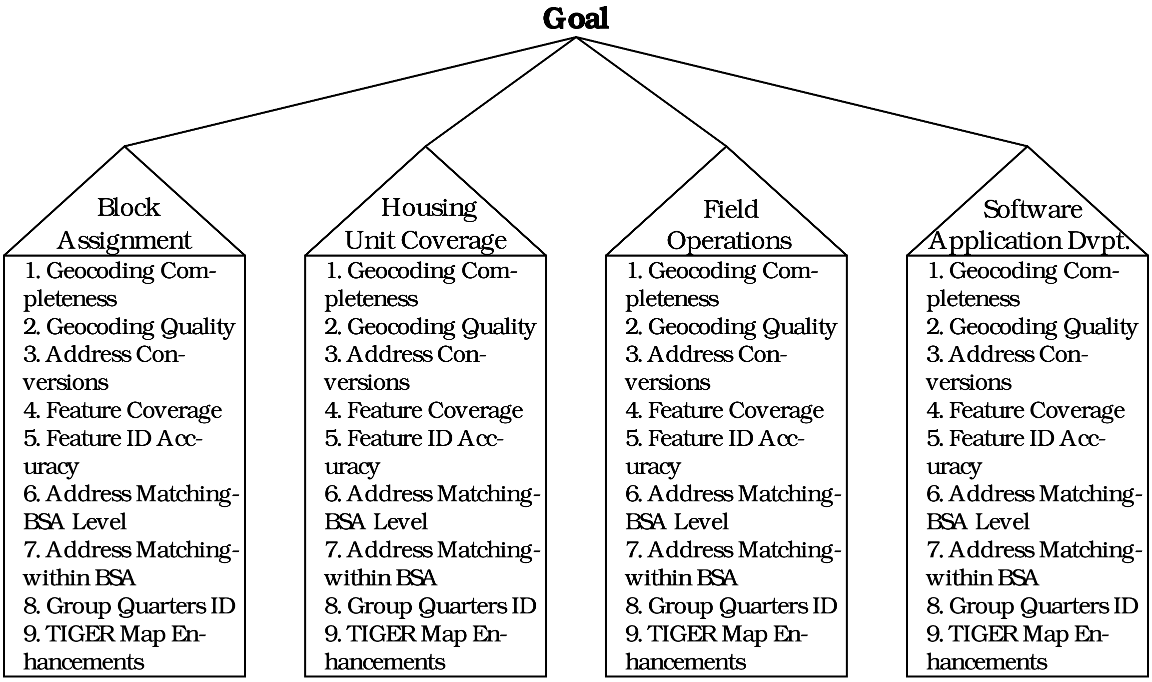

Figure 2: Aggregate global priorities and 95% posterior intervals for quality initiatives

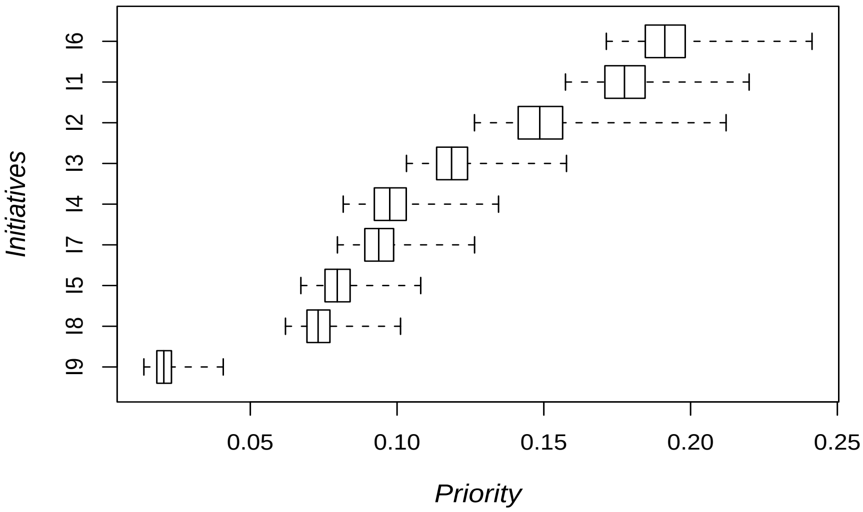

Table 1: Priority estimates for initiative categories

Category	*SJM Priority*	*SJM 95% Interval*	*Deterministic AHP Priority*
Housing Unit Coverage	0.492	(0.380, 0.606)	0.484
Block Assignment	0.353	(0.251, 0.465)	0.359
Field Operations	0.080	(0.028, 0.152)	0.082
Software Applications Development	0.075	(0.024, 0.145)	0.075

Table 2: Aggregate global priority estimates for initiatives

Initiative	*Stochastic Global Priority*	*Parameter S.D.*	*95% Interval*
I1	0.177	0.010	0.157 – 0.198
I2	0.149	0.012	0.126 – 0.172
I3	0.119	0.008	0.103 – 0.135
I4	0.097	0.008	0.082 – 0.115
I5	0.080	0.007	0.067 – 0.093
I6	0.191	0.010	0.171 – 0.212
I7	0.094	0.007	0.080 – 0.109
I8	0.073	0.006	0.062 – 0.085
I9	0.021	0.004	0.014 – 0.029

Full names of the initiatives appear in Figure 1 (e.g., I1 is geocoding completeness.).