

**MASTER**

**A COMPARISON OF SITE EVALUATION METHODS**

**MICHAEL D. ROWE  
AND  
BARBARA L. PIERCE**

**DO NOT MICROFILM  
COVER**

**August 1979**

**DIVISION OF REGIONAL STUDIES  
NATIONAL CENTER FOR ANALYSIS OF ENERGY SYSTEMS  
BROOKHAVEN NATIONAL LABORATORY  
UPTON, NEW YORK 11973**

**Prepared for:  
OFFICE OF NUCLEAR REGULATORY RESEARCH  
U.S. NUCLEAR REGULATORY COMMISSION  
WASHINGTON, D.C. 20555**



## DISCLAIMER

**This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.**

## **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

NUREG/CR-1684  
BNL-NUREG-51203

# A COMPARISON OF SITE EVALUATION METHODS

BY  
**MICHAEL D. ROWE**  
AND  
**BARBARA L. PIERCE**

NUREG/CR--1684  
TI86 002672

August 1979

DIVISION OF REGIONAL STUDIES  
NATIONAL CENTER FOR ANALYSIS OF ENERGY SYSTEMS  
BROOKHAVEN NATIONAL LABORATORY  
UPTON, NEW YORK 11973

Prepared for  
OFFICE OF NUCLEAR REGULATORY RESEARCH  
U.S. NUCLEAR REGULATORY COMMISSION  
WASHINGTON, D.C 20555  
CONTRACT DE-AC02-76CH00016  
NRC FIN NO. A-3076

**MASTER**

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

## NOTICE

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

The views expressed in this report are not necessarily those of the U.S. Nuclear Regulatory Commission.

Available from  
GPO Sales Program  
Division of Technical Information and Document Control  
U.S. Nuclear Regulatory Commission  
Washington, D.C. 20555  
and  
National Technical Information Service  
Springfield, Virginia 22161

THIS PAGE  
WAS INTENTIONALLY  
LEFT BLANK

## PREFACE

This report is one of a series on quantitative methods for nuclear power plant siting prepared by the BNL Division of Regional Studies for the Site Standards Designation Branch of the U.S. Nuclear Regulatory Commission. The other reports in this series are:

- Hobbs, Benjamin F., Analytical Multiobjective Decision Methods for Power Plant Siting: A Review of Theory and Applications, BNL-NUREG-51204, Division of Regional Studies, Brookhaven National Laboratory, Upton, N.Y., September 1979.
- Hobbs, Benjamin F., and Michael D. Rowe, A Comparison of Regional Screening Methods, BNL-NUREG-51205, Division of Regional Studies, Brookhaven National Laboratory, Upton, N.Y., September 1979.
- Pierce, Barbara L., and Michael D. Rowe, Quantitative Nuclear Power Plant Siting Methods: A Review of Current Practice, BNL-NUREG-28115, Division of Regional Studies, Brookhaven National Laboratory, Upton, N.Y., February 1979.
- Rowe, Michael D., Benjamin F. Hobbs, Barbara L. Pierce, and Peter M. Meier, An Assessment of Nuclear Power Plant Siting Methods, BNL-NUREG-51206, Division of Regional Studies, Brookhaven National Laboratory, Upton, N.Y., September 1979.

## ACKNOWLEDGMENTS

The authors acknowledge the contributions of the many persons who participated on the various test and advisory panels listed herein, especially Brookhaven colleagues P. Meier, W. Metz., and B. Hobbs, who not only participated on panels, but also assisted in all other phases of the project. S. Lee of the Site Standards Designation Branch, Nuclear Regulatory Commission, served as project monitor; his guidance and support are gratefully acknowledged. We are also indebted to R. Stern for assistance with graphics and to B. Burke, L. McCarthy, and P. Miller for typing of the manuscript.



## TABLE OF CONTENTS

Preface.....	iii
Acknowledgments.....	iv
Abstract.....	viii
Summary.....	ix
1. Introduction.....	1
2. Methods.....	3
2.1 Test Panels.....	3
2.2 Test Data.....	3
2.3 Procedure.....	10
2.4 Site Evaluation Methods.....	10
3. Results.....	15
3.1 BNL Panel.....	15
3.2 Advisory Panel.....	20
3.3 Long Island Panel.....	24
3.4 Decision Complexity.....	30
3.5 Systematic Differences Among Weights.....	34
4. Discussion.....	40
5. Conclusions.....	48
References.....	49
Glossary.....	50
Appendix 1. Representative Site Description Evaluated by the Long Island Panel.....	53
Appendix 2. BNL Panel Mean Weights.....	56
Appendix 3. Advisory Panel Mean Scaled Attribute Values.....	57
Appendix 4. Advisory Panel Mean Weights.....	58
Appendix 5. Long Island Panel Scaled Attribute Values, First Meeting.....	59
Appendix 6. Long Island Panel Mean Weights, First Meeting.....	60
Appendix 7. Long Island Panel Scaled Attribute Values, Second Meeting....	61
Appendix 8. Long Island Panel Mean Weights, Second Meeting.....	62
Appendix 9. Long Island Panel Mean Weights, Third Meeting.....	63
Appendix 10. Long Island Panel Individual Site Ranks, First and Second Meetings.....	64
Appendix 11. Long Island Panel Individual Site Ranks, Third Meeting.....	67

## LIST OF TABLES

A1	Probability of Error Attributable to Various Transformations of Ratio-Scaled Weights	xii
A2	Conclusions	xiv
1	Brookhaven National Laboratory (BNL) Panel	4
2	Advisory Panel	4
3	Long Island Panel	5
4	Summary of Site Descriptions Evaluated by the BNL Panel	6
5	Representative Attribute Value Function Evaluated by the BNL Panel	8
6	Summary of Fully Quantified Site Descriptions Evaluated by the Long Island Panel	9
7	Representative Weighting Questionnaire used by the BNL Panel	11
8	Descriptions of Weighting Methods used by the Advisory Panel	13
9	Mean Between-Methods Correlations of Weights, BNL Panel	16
10	Mean Between-Persons Correlations of Weights, BNL Panel	16
11	Site Ranks by Weight Estimation and Amalgamation Method, BNL Panel	17
12	Mean Between-Methods Correlations of Attribute Values and Weights, Advisory Panel	20
13	Site Ranks From Categorization (I), Rating (II), and Metfessel Allocation (III) Methods, Advisory Panel	24
14	Mean Between-Methods Correlations of Attribute Values and Weights, Long Island Panel, First Meeting	26
15	Mean Between-Methods Correlations of Weights, Long Island Panel, Second and Third Meetings	26
16	Mean Between-Method Correlations of Total Suitability, Long Island Panel, All Meetings	29
17	Calculation of Decision Complexity	33
18	Decision Complexity Normalizing Coefficients	35
19	Probability of Error Attributable to Various Transformations of Ratio-Scaled Weights	39

## LIST OF FIGURES

1	Discontinuities Delineating "Significant" Clusters of Site Evaluation Methodology Results, BNL Panel	18
2	Major Clusters of Site Evaluation Methodology Results BNL Panel	19
3	Comparison of Categorization and Rating Attribute Value Estimates, 5 of 17 Advisory Panel Members	21
4	Comparison of Categorization and Rating Weights, Advisory Panel	21
5	Comparison of Metfessel Allocation and Rating Weights, Advisory Panel	22
6	Frequency Distribution of Weighting Summation Impact Estimates Made With Categorization Attribute Values and Weights, Advisory Panel	22
7	Comparison of Weighting Summation Site Ranks from Categorization and Rating Attribute Values and Weights, Advisory Panel	25
8	Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, First Meeting	27
9	Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, Second Meeting	28
10	Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, Third Meeting	28
11	Comparisons of Rating Weights Estimated With and Without Global Knowledge, Long Island Panel, Second Meeting	31
12	Comparisons of Indifference Tradeoff and Rating Weights, Long Island Panel, Third Meeting	31
13	Weight Transformations Attributable to Weighting Methods	36
14	Probability of Decision Error Due to Logarithmic Perception as a Function of Decision Complexity	38

## ABSTRACT

This report presents results of tests of different final site selection methods used for siting large-scale facilities such as nuclear power plants. Test data are adapted from a nuclear power plant siting study conducted on Long Island, New York. The purpose of the tests is to determine whether or not different final site selection methods produce different results, and to obtain some understanding of the nature of any differences found. Decision rules and weighting methods are included. Decision rules tested are Weighting Summation, Power Law, Decision Analysis, Goal Programming, and Goal Attainment; weighting methods tested are Categorization, Ranking, Rating Ratio Estimation, Metfessel Allocation, Indifferent Tradeoff, Decision Analysis lottery, and Global Evaluation.

Results show that different methods can, indeed, produce different results, but that the probability that they will do so is controlled by the structure of differences among the sites being evaluated. Differences in weights and suitability scores attributable to methods have reduced significance if the alternatives include one or two sites that are superior to all others in many attributes. The more tradeoffs there are among good and bad levels of different attributes at different sites, the more important are the specifics of methods to the final decision.

## SUMMARY

The Nuclear Regulatory Commission (NRC) is currently developing criteria for evaluating alternative sites for nuclear power plants and establishing procedures for judging quality of alternative site studies in Environmental Reports submitted at the construction permit stage of licensing. In support of that development, this report examines sources of variability and error in the final site evaluation stage of the site selection process to determine the extent to which different site evaluation methods can affect results.

Three test panels having different levels of siting experience applied eight weight estimation methods and five decision rules (amalgamation methods) to hypothetical site descriptions based on data from Long Island, New York. Weight estimation methods included Categorization, Ranking, Rating, Ratio Estimation, Metfessel Allocation, Indifferent Tradeoff, Decision Analysis lottery, and Global Evaluation; amalgamation methods included Weighting Summation, Power Law, Decision Analysis, Goal Programming, and Goal Attainment. In addition, special studies were conducted of sensitivity of Weighting Summation, the most commonly used amalgamation method, to conceptual errors and theoretical problems associated with the various weight estimation methods.

Among the weight estimation and decision rules chosen for study, there is a range of theoretical validity and difficulty of application. In general, the more theoretically valid methods are also more difficult. There is, therefore, a tradeoff between potential for error due to theoretical problems and potential for error due to difficulty in providing the required responses. The simplest method tested, Global Evaluation, produces results which are obviously different from those of more valid methods. Two more complex methods, Goal Programming and Goal Attainment, also produce obviously different results. There are several reasons why these methods are expected to yield results that do not agree with those obtained by other methods, and most probably they do not accomplish the desired evaluation process.

The most theoretically valid method tested, Decision Analysis, is also the most difficult. Results from Decision Analysis lotteries were mostly so inconsistent among themselves that no weights could be calculated with which to estimate site suitability. This method is clearly inappropriate for application by inexperienced persons without extensive training and consistency checking. We have no results with which to compare application of Decision Analysis by persons more experienced in use of the method.

The remaining methods all elicit weights for application in the Weighting Summation, Power Law, Goal Programming, and Goal Attainment decision rules. Indifference Tradeoff is the most theoretically valid among these. Weights elicited by this method were consistently different from those of obtained by other methods. Since this method requires direct assessment of tradeoffs between a unit of one impact against a unit of another impact, and since at least one panel preferred this method over all others because of that direct comparison, it may be that the weights elicited more closely represent the value systems of the individuals involved. There is some question, however, of the significance of lack of value and preference independence as required by most decision rules. It may be that other weighting methods are better able to capture relative preferences under suboptimal conditions.

Ranking, Categorization, Rating, Ratio Estimation, and Metfessel Allocation all require roughly equivalent responses, and their results are correspondingly similar. These methods are once-removed from Indifference Tradeoff in that they all require implicit rather than explicit tradeoffs; the same tradeoffs are involved, but they be made based on more general information and without mechanisms forcing all tradeoffs to be considered. The tradeoffs are diluted to general impressions about relative importance. Under this condition, assumptions of value and preference independence appear to be incorrect, yet these assumptions are fundamental to all decision rules using weights.

Of the five methods above, only Categorization stands out as producing obviously different results. Because of the potential for large ambiguities in definitions of categories, this method has significantly reduced capability for dealing rigorously with the level of quantification required of the Weighting Summation method.

Since it is difficult in these results to separate effects of differences among methods from effects of individual uncertainty and experimental methodology, a series of tests was made of differences in results produced by differences in weights attributable to known characteristics of each method. Categorization, for example, in an ideal case, might cause truncation of correctly ratio-scaled weights to integers. The truncation causes loss of resolution of weights falling between category boundaries. In addition, if apparently linear categories (low, medium, high) are perceived in a nonlinear manner, so that the magnitude of the difference between low and medium is different from that between medium and high (a common response), then

truncation will also be nonlinear. Some categorization schemes start at zero, some at one.

Rating methods produce interval-scaled weights having an arbitrary zero point; the least important attribute may even be incorrectly assigned to have zero weight when it does not have zero importance. Even correctly applied rating procedures usually have no mechanism for fixing the values and ratios of the weights so that a nonarbitrary zero point is assured. Both of these cause an arbitrary shift of scale.

All weighting methods contain individual uncertainty, the magnitude of which can depend on method.

Repeated tests of the above classes of differences among weighting methods show that potential for decision errors attributable to systematic differences among methods is roughly equal to that attributable to individual uncertainty (Table A1). Depending on the structure of the decision problem involved, probability of causing a decision error can be as high as 30% through either methodology or individual uncertainty alone. The two types of error are not strictly additive.

Among the weight estimation methods tested there occasionally were large differences in weights, both for individuals and, less so, for group means. Nevertheless, the sites selected by application of these weights to decision rules are all nearly identical among the two or three top-ranked sites. This robustness of outcome arises from the structure of the siting problems under evaluation with respect to numbers and kinds of tradeoffs that were made between good and bad levels of different attributes. If one site is better in all attributes than all other sites, then it is dominant and it will be first-ranked for all possible combinations of positive weights. Such a siting problem is completely insensitive to magnitudes of weights or methods used to elicit them. If, in contrast, there are large numbers of tradeoffs that must be made between relatively good and bad levels of attributes for all sites, so that each has a few good attribute values and a few bad attribute values, then results can be exceedingly sensitive to the specific magnitudes of weights and, therefore, to any differences attributable to methods for eliciting them. This is especially true if there is a large diversity of sites.

A complexity index was derived which quantifies the magnitudes of tradeoffs that must be made in a particular siting problem, including consideration

TABLE A1  
 PROBABILITY OF ERROR ATTRIBUTABLE TO VARIOUS  
 TRANSFORMATIONS OF RATIO-SCALED WEIGHTS

Transformation	Changes in Probability of Error per Unit Complexity
Linear categorization	
2 categories	0.22
3 categories	0.10
5 categories	0.08
Categorization based at 0 vs categorization based at 1	
2 categories	0.09
3 categories	0.05
5 categories	0.00
Nonlinear categorization logarithmic transformation	
2 categories	0.22
3 categories	0.18
5 categories	0.13
Nonlinear transformation	
logarithmic	0.12
exponential	0.13
Ratio to interval scale	
Random zero point	0.04
+2 zero point	0.07
+4 zero point	0.14
+6 zero point	0.20
Random variability (Means = 0-10)	
= 0.5	0.04
= 1.0	0.08
= 2.0	0.15
= 3.0	0.27

of the likelihood that the tradeoffs will be among attributes of high-ranked sites. It is a measure of the complexity of the decision that must be made and, therefore, the relative sensitivity of the decision to magnitudes of weights. Tests of artificial siting problems of known complexity ranging from zero to near the maximum possible show that the influence of complexity on decision is on the order of twice that of other variables associated with estimation of weights, including method and individual uncertainty.



Complexity of a final site evaluation is controlled by the characteristics of the region of interest and the stage in the siting process immediately preceding final selection in which candidate sites are identified from candidate areas and reduced to a manageable number. This is the least formalized stage of the process and the one most likely to be conducted "by the seat of the pants." Usually, experienced persons sit down with a large number of maps and hunt up places that "look good." There is no reason why this method can not yield high quality choices. Depending on the skill and biases of the persons involved, however, it has high potential for introducing error, bias, and sloppy workmanship, none of which is easy to discover after the fact. Strong preconceived notions about characteristics of a "good" power plant site, for example, reduce diversity of the sites finally evaluated, not only reducing the probability that any one site will be dominant or clearly superior, but also reducing likelihood that there will be sites that are "good" under value systems other than those of the persons choosing candidate sites. Political or NRC requirements that there be sites from several different geophysical or jurisdictional areas increase diversity, but can cause sites to enter the slate which might otherwise be considered unsuitable. It is possible for final site selection to be made from among sites which are inferior from the perspectives of persons other than those who have done the selecting, and there is no good way to determine that this is the case other than by finding an obviously superior site. This places a burden of proof on interveners that should rightfully fall on utilities.

In view of the importance of the candidate site selection stage to the complexity and quality of final results, and in view of its relative lack of accountability because of the manner in which it is conducted, more detailed studies should be made to determine what is necessary to ensure that its results are acceptable.

---

TABLE A2  
CONCLUSIONS

---

1. Different scaling and weight estimation methods can produce different attribute values and weights.
    - Categorization lacks capability to quantify with the level of rigor required of the Weighting Summation decision rule and can suffer from unquantified nonlinear perception of relative importance.
    - Global Evaluation of more than three or four attributes is unsuccessful because decision makers can not process that much information at once. Instead, they evaluate an unspecified subset of the information.
    - Indifference Tradeoff is the only method tested which both can assure valid weights and can be applied by inexperienced persons. Its results may be different for this reason. The assumptions underlying the methods may not hold, however, so the significance of these "more valid" weights is unclear.
    - Decision Analysis lotteries produced few feasible weights in this study. This method should not be used by inexperienced persons without extensive training and consistency checking.
    - Central tendency of means reduces differences among methods for group mean weights.
  2. Different weights can produce different results.
    - Results are strongly influenced by the structure of the data to which weights are applied. Some data structures are less sensitive than others to differences in weights attributable to weight estimation methods.
    - A "decision complexity" index is derived which quantifies sensitivity of results to differences in weights.
  3. Different decision rules can produce different results.
    - Goal Programming and Goal Attainment techniques produce obviously different results from other decision rules. This may be because they are insensitive to dominated solutions, or it may be because given lack of value and preference independence, they better represent the satisficing behavior attempted by many of the decision makers in this study.
  4. Selection of Candidate Sites from Candidate Areas may be the most important stage of the siting process.
    - Sensitivity of results to differences attributable to weight estimation method and decision rule is strongly influenced by the diversity and complexity of the sites chosen for evaluation.
    - Candidate Site selection is the least quantified and least formalized stage of the site selection process. Because of its importance, it should be conducted with special care.
-

## 1 INTRODUCTION

Since enactment of the National Environmental Policy Act (NEPA) of 1969, review of alternatives to any major federal action which would significantly affect quality of the human environment has become an integral part of planning and decision making in federal agencies. Nuclear Regulatory Commission (NRC) licensing and regulatory policy and procedures now require that each applicant for construction of a nuclear power plant discuss in an environmental report "appropriate alternatives" to the proposed facility, including alternative locations.

NRC is currently developing criteria for evaluating alternative sites for nuclear power plants and establishing procedures for judging quality of alternative site studies.<sup>1</sup> The Division of Regional Studies of Brookhaven National Laboratory is assisting NRC in this effort by preparing a procedure for use in evaluating alternative site studies with respect to correctness of application and quality of results of different siting methodologies.

In support of development of this procedure, several experimental studies have been conducted to provide information on differences among siting methodologies and to determine what characteristics of those methodologies appear to be important in controlling quality of results. Studies include testing methods for screening large areas to isolate a smaller set of candidate areas likely to contain suitable nuclear power plant sites, and methods for evaluating a set of candidate sites to select one that appears "best" based on pre-specified evaluation criteria. The intermediate stage, candidate site selection, is discussed but not studied.

This report presents results of studies of application of final site selection methods by three panels of experts. The panels were specifically selected to test different aspects of the methods and influences of different levels of knowledge of the specifics of power plant siting; the project evolved in time based on results obtained from early panels, therefore, the methods tested and the data base used by each panel were not always the same. For this reason, results from the various panels are presented separately. The report also presents results of studies of sensitivity of outcomes of site evaluations to incorrect application of the most frequently used of all final site selection methods, Weighting Summation.

Site evaluation methods were tested by creating data sets representative of those commonly used for siting studies, and having test panels apply each

of several different weight or goal estimation methods and decision rules (amalgamation methods) to the same data. Details of the weight estimation and decision rules used are presented in a companion volume to this report.<sup>3</sup>

## 2 METHODS

### 2.1 TEST PANELS

Three panels of experts were tested. The first (the BNL Panel, Table 1), organized early in the project, was composed of several Brookhaven National Laboratory scientists who specialize in a variety of disciplines, including: Ecology, Water Resources, Health Impact Analysis, Regional Planning, and Meteorology. These persons were selected because of their broad experience in general problems of power plant siting but lack of specific experience with siting methodologies. They represent "intelligent initiates" with respect to siting methodology and, therefore, their responses are useful for determining influences of knowledge and understanding of the specifics of methodologies on consistency of results.

The second panel (the Siting Methodology Advisory Panel, Table 2) was organized not only for testing methodologies, but also for general advice and direction for the project as a whole. The Advisory Panel consisted of 17 members including private consultants having extensive experience in nuclear power plant siting, Nuclear Regulatory Commission personnel, state-level site planners, utility planners, academicians having special knowledge of siting problems, and members of the general public having experience as interveners. This panel, as a group, is extensively experienced in all aspects of power plant siting; the experience is not uniform, however, and individual familiarity and understanding of the specifics of siting methodologies vary.

The third panel (the Long Island Panel, Table 3) was organized specifically to address problems encountered by the Advisory Panel, which felt hampered in its judgments because of lack of detailed knowledge of the environment of Long Island, New York, on which the test data are based. The Long Island Panel, therefore, was composed of persons knowledgeable about siting problems on Long Island, including Brookhaven National Laboratory employees recently transferred from positions in which they were working directly with siting problems in the area, experienced interveners from the Long Island community, and town planners.

### 2.2 TEST DATA

The data chosen for analysis are adapted from those developed for Long Island Lighting Company (LILCO) by Grumman Ecosystems Corporation.<sup>2</sup> These data were chosen for several reasons, primarily because they are representative and well described, and because the Long Island area is familiar to

---

TABLE 1  
BROOKHAVEN NATIONAL LABORATORY (BNL) PANEL

---

Dr. E. Kaplan, Water Resources Scientist, Biomedical and Environmental Assessment Division  
Dr. P. Meier, Regional Scientist and Head, Division of Regional Studies  
Dr. W. Metz, Social Scientist, Division of Regional Studies  
Mr. S. Munson, Geographer, Division of Regional Studies  
Dr. J. Nagy, Radiation Health Effects Analyst, Biomedical and Environmental Assessment Division  
Ms. B. Pierce, Ecologist, Division of Regional Studies  
Mr. G. Raynor, Meteorologist, Atmospheric Sciences Division

---

---

TABLE 2  
ADVISORY PANEL

---

Consultants

Dr. K. Nair, Woodward-Clyde  
Mr. W. Heilman, Gilbert/Commonwealth  
Dr. J. Horst, Stone & Webster  
Mr. M. Hartman, Envirosphere Company  
Mr. J. Halpern, Dames & Moore  
Mr. D. Siefken, Sargent & Lundy

Nuclear Regulatory Commission

Mr. D. Cleary, NRC  
Mr. J. Klein, NRC  
Mr. W. Ott, NRC

Planning Agencies

Mr. R. Kasvinsky, New England River Basin Commission  
Mr. H. Mueller, Maryland Power Plant Siting Program

Electric Utilities

Mr. J. Lagowski, Detroit Edison, Detroit, Michigan  
Mr. E. Hall, Pacific Gas and Electric, San Francisco, CA

Academia

Dr. R. Bolton, Professor of Economics, Williams College,  
Williamstown, MA  
Dr. S. Linke, Professor of Electric Engineering, Cornell University  
Ithaca, NY  
Mr. B. Hobbs, School of Civil and Environmental Engineering,  
Cornell University, Ithaca, NY

Informed Public

Ms. J. Tiedke, League of Woman Voters, New York

---

---

TABLE 3  
LONG ISLAND PANEL

---

<u>Dr. F. Lipfert,</u>	formerly Manager of Air Quality Control, Long Island Lighting Company
<u>Dr. W. Metz,</u>	formerly a siting analyst for Energy Impact Associates, Pittsburgh, PA
<u>Mr. P. Moskowitz,</u>	formerly an environmental analyst for the Long Island Office of the Environmental Defense Fund
<u>Ms. K. Romano,</u>	member of the Southold Town Planning Board, Long Island, NY
<u>Ms. J. Tiedke,</u>	member of the Long Island Chapter of the League of Women Voters (also a member of the Advisory Panel)

---

BNL Panel members. The data are slightly modified to make them more useful for this analysis and scrambled to preserve anonymity of sites and to prevent knowledge of site specifics not included in the analysis from affecting judgments.

During the course of the project the data evolved from a straightforward presentation of general conditions with specific descriptions of levels of each attribute, through detailed descriptions of each site being evaluated, to highly specific, quantitative descriptions of each attribute. The BNL Panel worked with the data summarized in Table 4. To confine analysis to weight selection and decision rules, each attribute (measured or described characteristic or impact) was prescaled to form an attribute value function expressing the relative suitability of the sites for a nuclear power plant with respect to that particular attribute. Values range from 10.0 (best) to 0.0 (worst). Table 5 shows a representative example of a scaled value function.

The BNL Panel was disturbed by having to make value judgments about attributes which were prescaled, because the scaling itself is a value judgment and not all Panel members agreed with the values expressed therein. For this reason, the same attributes and descriptions were used by the Advisory Panel, but attributes were not prescaled. Instead, the Advisory Panel members made their own subjective judgments of relative suitability.

TABLE 4  
SUMMARY OF SITE DESCRIPTIONS EVALUATED BY THE BNL PANEL

Site	Site Cost	Land use	Transmission	Water quality/ ecosystems	Terrestrial ecosystems	Population	Aesthetics
1	\$13.0 x 10 <sup>6</sup>	8 homes on site	24 miles through farmland	Quality, good offshore shoal, fish traps	10 acres coastal marsh	25 homes within 1/2 mile, 1340 persons within 10 sq miles	Pastoral setting, some natural screening
2	\$14.7 x 10 <sup>6</sup>	8 homes on site, 255 ft height limit, 200 acres farmland	23 miles through farmland and woodland	Deep, cool water, less sensitive ecosystem than others	2 small ponds within marsh	15 homes within 1/2 mile, 3240 persons within 10 miles	Open farmland and beach, low line-of-sight in all directions
3	\$15.7 x 10 <sup>6</sup>		40 miles through land and woodland	Quality good, productive bays nearby, commercial shell fisheries		720 persons within 10 sq miles	Woods, some natural screening
4	\$11.1 x 10 <sup>6</sup>	3 homes on site large land development planned, 40 acres farmland	13 miles through residential	Quality fair, tanker port nearby	28 acres coastal marsh	17 homes within 1/2 mile, 5270 persons within 10 sq mile	Pastoral setting, noise-sensitive area at 2500 ft
5	\$13.2 x 10 <sup>6</sup>	4 homes on site, 160 acres farmland	9 miles through farmland	Quality fair to good		10 homes within 1/2 mile, 1880 persons within 10 sq mile	Pastoral bluff, little screening
6	\$23.0 x 10 <sup>6</sup>	2 homes on site, 160 acres farmland	5 miles through farmland	Quality fair to good, tanker port nearby, offshore shoal		3 homes within 1/2 mile, 2500 persons within 10 sq miles	Pastoral bluff, some natural screening, noise-sensitive area at 1000 ft

1  
9  
1



TABLE 4 SITE DESCRIPTIONS (Cont.)

Site	Site Cost	Land use	Transmission	Water quality/ ecosystems	Terrestrial ecosity stems	Population	Aesthetics
7	\$18.3 x 10 <sup>6</sup>	425 foot height limit, 90 acres farmland	4 miles through farmland	Quality fair to good		1 home within 1/2 mile, 1530 persons within 10 sq mi	Pastoral bluff, elevated, noise- sensitive area at 2500 feet
8	\$9.8 x 10 <sup>6</sup>	425 foot height limit, LILCO-owned	7 miles through fields and wood- land	Quality, fair to good, river out- flow, offshore shoal	12 acres coastal marsh, 1 small point	3 homes within 1/2 mile, 3270 persons, with- in 10 sq mi	Woods, natural screening
9	\$14.5 x 10 <sup>6</sup>	LILCO-owned 85 acres farmland	7 miles through farmland	Quality fair to good, tanker port nearby	1 small pond	2 homes within 1/2 mile, 2500 persons within 10 sq mi	Open bluff scarred by sand pits, no sensitive areas adjacent
10	\$17.8 x 10 <sup>6</sup>	9 homes on site, 190 acres farm- land	14 miles through farmland	Quality fair to good		13 homes within 1/2 mile, 1270 persons within 10 sq mi	Pastoral, low elevation
11	\$12.1 x 10 <sup>6</sup>	200 acres farmland	12 miles through farmland	Quality fair to good		4 homes within 1/2 mile, 1270 persons within 10 sq mi	Pastoral, low elevation

TABLE 5  
 REPRESENTATIVE ATTRIBUTE VALUE FUNCTION  
 EVALUATED BY THE BNL PANEL

Population resistance and potential radiation exposure to nearby population measured in terms of the numbers of homes within a half-mile radius and the numbers of persons in the surrounding 10 square miles (about a 2-mile radius).

SITE DESCRIPTION		SUITABILITY
Homes within 1/2 miles (number)	Persons within 10 sq mi (number)	Scale = 0.0-10.
0	720	10.0
1	1530	8.6
4	1270	8.5
13	1270	7.0
2	2500	6.8
3	2500	6.7
10	1880	6.5
3	3270	5.4
25	1340	5.0
15	3240	4.6
17	5270	0.0

The Advisory Panel was not only hampered by lack of knowledge of the Long Island test area, but also considered the written descriptions of site attributes too simplified and lacking sufficient quantitative information.\* The data used by the Long Island Panel were, therefore, expanded by adding more quantitative information (Appendix 1). The Long Island Panel was still relatively unhappy with the quality of the data, so a new data set was created which is completely quantitative with no qualitative descriptions. Table 6 shows a summary. These data were presented as attributes to be evaluated rather than site-related information.

\*Full site descriptions used by the various Panels are not included here. The Advisory Panel used an information packet, three-quarter inches thick, mailed out for review before the first meeting.

TABLE 6

## SUMMARY OF FULLY QUANTIFIED SITE DESCRIPTIONS EVALUATED BY THE LONG ISLAND PANEL

Site	Population (persons at radius)			Transmission lines (miles)	Social Services (% change in cost)	Coastal wetlands	Farmland	
	2 mi	5 mi	10 mi				Acres	% Total
A	1250	7850	31,420	3	0	10 acres, not fenced	50	1.0
B	942	5890	23,560	20	1	1 acre of 10 destroyed	90	1.8
C	315	1960	7,850	10	-20	10 acres, productivity permanently reduced 10%	0	0
D	30	200	785	5	5	100 acres, in exclusion zone but not affected	100	2.0
E	630	3930	15,710	1	10	10 acres, in exclusion zone but not affected	20	0.4

Site	Clam bed productivity (1000 acres)	Aesthetics	Site Cost		Rate increase (%)
			Total (\$10 <sup>9</sup> )	Incremental (\$10 <sup>9</sup> )	
A	2% permanent reduction	Screening in all direc- tions	2.8	200	5.0
B	No change	Industrial complex, scarred bluff, visible 270°	2.7	100	2.5
C	Permanent destruction of 20 acres	Scenic valley, visible 90° from highway and 20 houses	2.9	300	7.5
D	No reduction, possible increase	Remote beach, no screening, visible 360°	2.6	0	0
E	1% reduction during construction only	High scenic bluff visible from highway, town, and water	3.0	400	10.0

### 2.3 PROCEDURE

For both the BNL and Advisory Panels, tests were preceded by introductory presentations by specialists on important characteristics of the Long Island environment. A brief slide show was included for the Advisory Panel, many members of which had not visited Long Island before. Written descriptions of the study area (more specific than that mailed to the Advisory Panel) and of specifics of the various sites being evaluated were then distributed, followed by questionnaires for attribute suitability scaling and attribute weighting. Throughout, the Panels were cautioned that the purpose of the tests was not to choose a good power plant site, which is not possible given the time and level of information available to them, but to assess their subjective values for attributes of the kind normally used for power plant siting. These subjective judgments can be made at any level of information. Because of time limitations, successive tests of methods were not independent and some learning and carryover occurred among simpler methods. The more complex methods are so different in concept, however, that no carryover is likely.

The Long Island Panel did not require descriptions of the study area. Nevertheless, the detail of the written descriptions of specific sites was greatest for this Panel.

### 2.4 SITE EVALUATION METHODS

The BNL Panel tested six weight selection methods: 1) Ranking; 2) Rating; 3) Ratio Estimation; 4) Indifference Tradeoff; 5) Decision Analysis; and 6) Global Evaluation. The weights were applied as appropriate in five decision rules: 1) Weighting Summation; 2) Power Law; 3) Decision Analysis; 4) Goal Programming; and 5) Goal Attainment.\* Table 7 shows an example of a weighting questionnaire.

Because the mathematics underlying most decision rules presumes that subjective assessments of weights are independent of one another, information on the suitability levels of the site tests was organized so that it is not possible to determine the particular combination of attribute values assigned to each site. In this manner Panel members were forced to consider only the ranges of values for each attribute and not the relative suitability of particular combinations which they might find especially desirable or undesirable. Full site descriptions were given only for the global evaluation method for which they are required.

---

\*See Reference 3 for detailed descriptions of methods.

---

TABLE 7  
REPRESENTATIVE WEIGHTING QUESTIONNAIRE  
USED BY THE BNL PANEL

---

Goal Attainment in Nuclear Power Plant Siting

Examine the descriptions of nuclear power plant siting attributes. Keep in mind that the sites described are the winners among many sites and none of them is really bad with respect to any particular attribute. For each attribute, without regard for the levels of others, choose a level that you consider "sufficient" or "acceptable." The scale of suitability is 0.0-10.0 and the general significance of a unit change in suitability level can be inferred from the site descriptions.

Enter the "acceptable" levels in the spaces provided.

<u>Attribute</u>	<u>Acceptable Level</u>
A. Site Cost	_____
B. Land Use	_____
C. Transmission	_____
D. Water Quality/Ecosystems	_____
E. Terrestrial Ecosystems	_____
F. Population	_____
G. Aesthetics	_____

How long (in minutes) did this evaluation take you? \_\_\_\_\_

Did you have trouble understanding the instructions? \_\_\_\_\_

Comments or suggestions \_\_\_\_\_

\_\_\_\_\_

---

Rating was used twice by the BNL Panel at an interval of three weeks to test for effects of time. Six Decision Analysis lotteries were used to test for effects of choice of lottery basis and lottery form. Four attributes were weighted with the usual form of Decision Analysis lottery in which the attribute being weighted in the "certain" alternative is at its best value and all others are at their worst values. Two of these attributes were weighted again using the opposite lottery form with the attribute being weighted at its worst value and all others at their best value. Hereinafter these forms are post-scripted 1 and 2, respectively.

The Advisory Panel tested three weighting methods: 1) Categorization; 2) Rating; and 3) Metfessel Allocation, and the weights were applied to the Weighting Summation method only. Table 8 shows the methodology descriptions used. Categorization was substituted for Ranking because the former is commonly used while the latter is not; Metfessel Allocation was substituted for Ratio Estimation because the former is more rigorously defined than the latter while yielding the same kind of weights. Note that the Rating instructions require that the least important attribute arbitrarily be assigned zero weight. This produces interval-scaled weights which are not valid. Because of difficulties encountered by the Advisory Panel in scaling and weighting attributes, no attempt was made to test a large number of amalgamation methods using their results.

As mentioned above, the Long Island Panel was created specifically to address problems encountered by the Advisory Panel. The Long Island Panel is knowledgeable about environmental problems in the study area, which the Advisory Panel is not, and the detail of the information provided to the Long Island Panel was greater than that provided to the Advisory Panel. In addition, the Long Island Panel was tested for effects of making independent assessments of weights as opposed to making them with full knowledge of the specific combinations of attribute values represented at each site. Site attributes were first scaled and weighted with independent information using Categorization, Rating, and Metfessel Allocation. The attribute values were then combined into "site descriptions" with a random number table and the Panel made a global evaluation of each "site" by considering all impacts together. The attribute values were then reweighted using Categorization, Rating and Metfessel Allocation with the full descriptions of the particular combination of values at each "site."

---

TABLE 8  
DESCRIPTIONS OF WEIGHTING METHODS USED BY  
THE ADVISORY PANEL

---

WEIGHTING METHODOLOGY

Categorization. Estimate the relative importance you think each site evaluation attribute should have in a siting decision using the following categories.

- 0 - No importance
- 1 - Low importance
- 2 - Moderate importance
- 3 - High importance
- 4 - Very high importance

Rating. Rate the relative importance you think each site evaluation attribute should have in a siting decision using a scale of 0.0 (least importance) to 10.0 (most importance). Scan the list and assign the most important attribute a 10.0 and then estimate the distances the other attributes are in importance from the most important.

- 0.0 - Least important
- 10.0 - Most important

Metfessel Allocation. Estimate the ratios of importances you think each site evaluation attribute should have in a siting decision by allocating a total of 100 points to all attributes in proportion to your perception of their relative importance. Watch the ratios among the attributes carefully as they develop and correct them when you find inconsistencies among two or more ratios. Remember that the total number of points must equal 100.

---

A third set of tests was made with the Long Island Panel related to the specificity of quantification of attribute levels. Many of the attributes normally used are described in qualitative terms; of these, many (aesthetics, for example) can only be described in these terms. The fully quantified attributes created for the Long Island Panel (summarized in Table 6, above) are not sufficiently complete to use for selecting a power plant site, so they were presented as attributes to be evaluated rather than as descriptions of potential sites. These attributes were weighted twice, at an interval of one month, to test for individual variability.\* In addition, two more weighting methods, Indifference Tradeoff and Decision Analysis, were applied during the second test using these fully quantified attributes.

---

\*The Three-Mile Island nuclear accident occurred in the interim between these two tests which may have changed perception of some siting attributes.



### 3 RESULTS

#### 3.1 BNL PANEL

Mean weights for six weight estimation methods are shown in Appendix 2. Few weights were calculated for the Decision Analysis method; therefore, the mean weight shown is derived from a mean lottery probability rather than individual weights.

The Decision Analysis lottery is difficult and time consuming. It is therefore normally applied to one attribute, and weights for other attributes are calculated from Indifference Tradeoff ratios using the lottery weight as a base. We conducted several lotteries to determine consistency among trials. Nearly all of the results from lottery questions formed in the normal manner, with the attribute being weighted in the "certain" alternative at its best value and all others at their worst value, were so inconsistent with Indifference Tradeoff ratios that no feasible set of weights (all weights  $\leq 1.0$ ) could be calculated. Those that could be calculated were often inconsistent among the different lottery questions. Only the lottery questions formed in the opposite manner from normal, with the attribute being weighted in the "certain" alternative at its worst value and all others at their best values, yielded enough feasible weights to be considered meaningful. Nevertheless, all weights based on mean probabilities are reported, in spite of the fact that they may be meaningless.

Mean between-methods correlations for all BNL Panel members are shown in Table 9. A decrease in correlation from simpler to more complex methods is obvious. Decision Analysis weights are not included because of the poor quality of the results and lack of representation of all individuals.

Mean between-persons correlations for all Panel members are shown in Table 10. As is expected for such a diverse group, correlations are low and no consistent pattern is apparent.

Group mean weights for each method are applied to the appropriate decision rules to obtain estimates of total site suitability and site suitability ranks. All but 8 of the 276 between-methods correlations of total site suitability are significant at the 5% level. This is caused by the central tendency of means which makes the means more similar among methods than are the weights from which they are calculated. Nevertheless, 64% of these combinations do not agree on the first-ranked site, and of these, 76% do not even agree that the first-ranked site from one method is among the top three sites for the other method (Table 11).

TABLE 9  
MEAN BETWEEN-METHODS CORRELATIONS OF WEIGHTS, BNL PANEL

	Rating 1	Rating 2	Ratio Estimation	Indifference Tradeoff	Global Evaluation
Ranking	0.817	0.705	0.732	0.440	0.370
Rating 1		0.771	0.766	0.450	0.256
Rating 2			0.865	0.502	0.484
Ratio Estimation				0.493	0.343
Indifference Tradeoff					0.125

In spite of these differences, however, the various methods considered tend to fall into a few groups which produce similar results among themselves. Figures 1 and 2 show results of cluster analysis of 29 combinations of weight estimation methods (group mean weights) and decision rules. (Table 11 has more understandable descriptions of the Goal Programming and Goal Attainment Methods.) There is no particular level of amalgamation distance in Figure 1 that is significant by definition; therefore, in Figure 2, methods are combined into groups formed by the break points shown in Figure 1. These are points at which there are relatively large discontinuities in the amalgamation distance-group number curve.

In Figure 2, Level 1 clusters have identical results. This level includes most applications of the Decision Analysis method and Indifference Tradeoff. Since Decision Analysis weights are calculated from Indifference Tradeoff results, it is not surprising that these two methods are similar. At

TABLE 10  
MEAN BETWEEN-PERSONS CORRELATIONS OF WEIGHTS, BNL PANEL

Person	A	B	C	D	E	F	G
A		0.500	0.369	0.283	0.358	0.318	0.648
B			0.572	0.130	0.344	0.552	0.387
C				0.447	0.320	0.556	0.277
D					0.102	0.213	0.280
E						0.174	0.219
F							0.415

TABLE 11

SITE RANKS BY WEIGHT ESTIMATION AND METHOD AND DECISION RULE, BNL PANEL<sup>a</sup>

Methods	Sites	1	2	3	4	5	6	7	8	9	10	11
Ranking		9	10	8	11	4	7	3	5	1	6	2
Rating 1		9	10	8	11	4	7	3	6	1	5	2
Rating 2		10	9	8	11	4	7	3	6	1	5	2
Ratio												
Estimation		10	9	8	11	4	7	3	6	2	5	2
Indifference												
Tradeoff		9	10	7	11	4	8	3	6	1	5	2
Global												
Evaluation		10	8	9	11	3	5	1	7	4	6	2
POWERLAW												
all weights		7	9.5	9.5	9.5	3	9.5	4	6	1	5	2
DECISION ANALYSIS												
all weights		10	9	7	11	4	8	3	6	1	5	2
Site cost 1		--	--	--	--	--	--	--	--	--	--	--
Population 1		10	9	7	11	4	8	3	6	1	5	2
Water quality		10	9	7	11	4	8	3	6	1	5	2
Aesthetics		10	9	7	11	4	8	3	6	1	5	2
Site cost 2		9	10	7	11	4	8	3	6	1	5	2
Population 2		9	10	7	11	3	7	4	5	1	6	2
1. Goal Programming (Rating weights)												
a. Absolute value of deviations		8	7	10	11	1	3	5	9	6	2	4
b. Squared deviations		7	8	10	11	2	3	4	9	6	1	5
c. Squared devia- tions from maximum		8	10	9	11	3	7	4	6	1	5	2
2. Goal Programming (Indifference Trade- off weights)												
a. Absolute value of deviations		8	7	9	11	2	3	4	10	6	1	5
b. Squared deviations		7	8	10	11	4	2	3	9	6	1	5
c. Square devia- tions from maximum		7	10	9	11	3	8	4	6	1	5	2
3. Goal attainment (rating weights)												
a. Minimax		8	9	10	11	3	4	5	7	6	1	2
b. Maximax		8	9	11	10	3.5	6.5	2	6.5	1	5	3.5
4. Goal Attainment (Indifference Trade- off weights)												
a. Minimax		8	9	10	11	2	4	5	7	6	1	3
b. Maximax		8	9	10	11	3.5	5.5	2	5.5	1	7	3.5

<sup>a</sup>Using group mean weights.

cluster Level 2, the Goal Programming and Goal Attainment methods as well as Global Evaluation are separated from all others except Ratio Estimation and Power Law amalgamation using Decision Analysis. The first cluster of Level 2 contains the simpler methods along with Decision Analysis and Indifference Tradeoff. Considering that the BNL Panel was very uncomfortable with Indifference Tradeoff and Decision Analysis, and considering the inconsistencies and generally poor quality of the Decision Analysis results as mentioned above, it is surprising that the Panel mean weights are so similar to those of the simpler methods. Level 3 is the highest level at which Ratio Estimation and Indifference Tradeoff are grouped. These two methods elicit similar information, but Panel members preferred Ratio Estimation by far, grouping it with Ranking and Rating with respect to difficulty. At Level 4, many of the Goal Programming methods are still so different from others that their amalgamation distance is 3 to 10 times that between any other clusters. These methods are clearly "out in left field" and probably do not correctly accomplish the desired evaluation process.

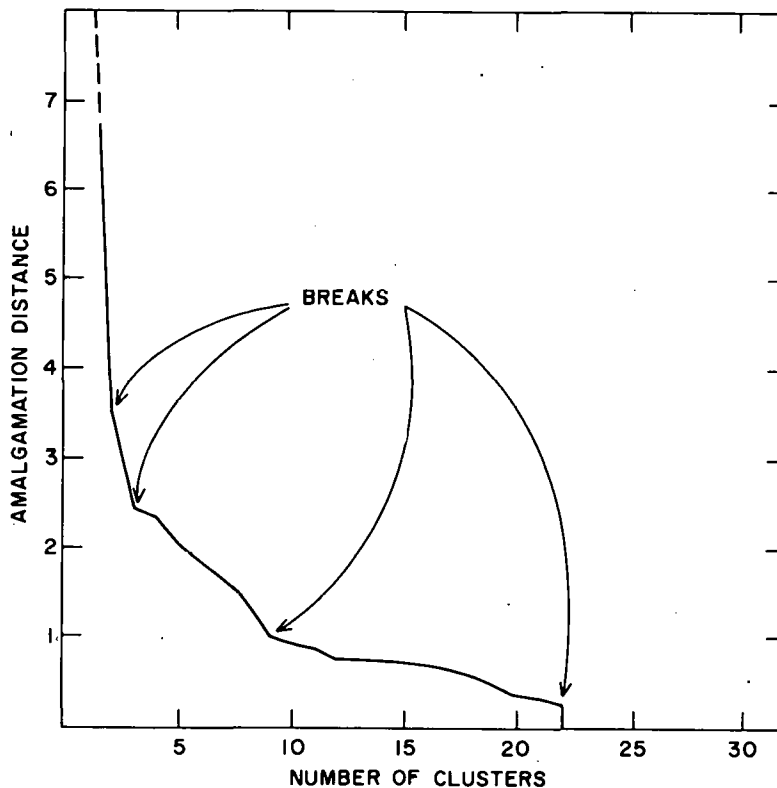


Figure 1. Discontinuities delineating "Significant" clusters of site evaluation methodology results, BNL Panel.

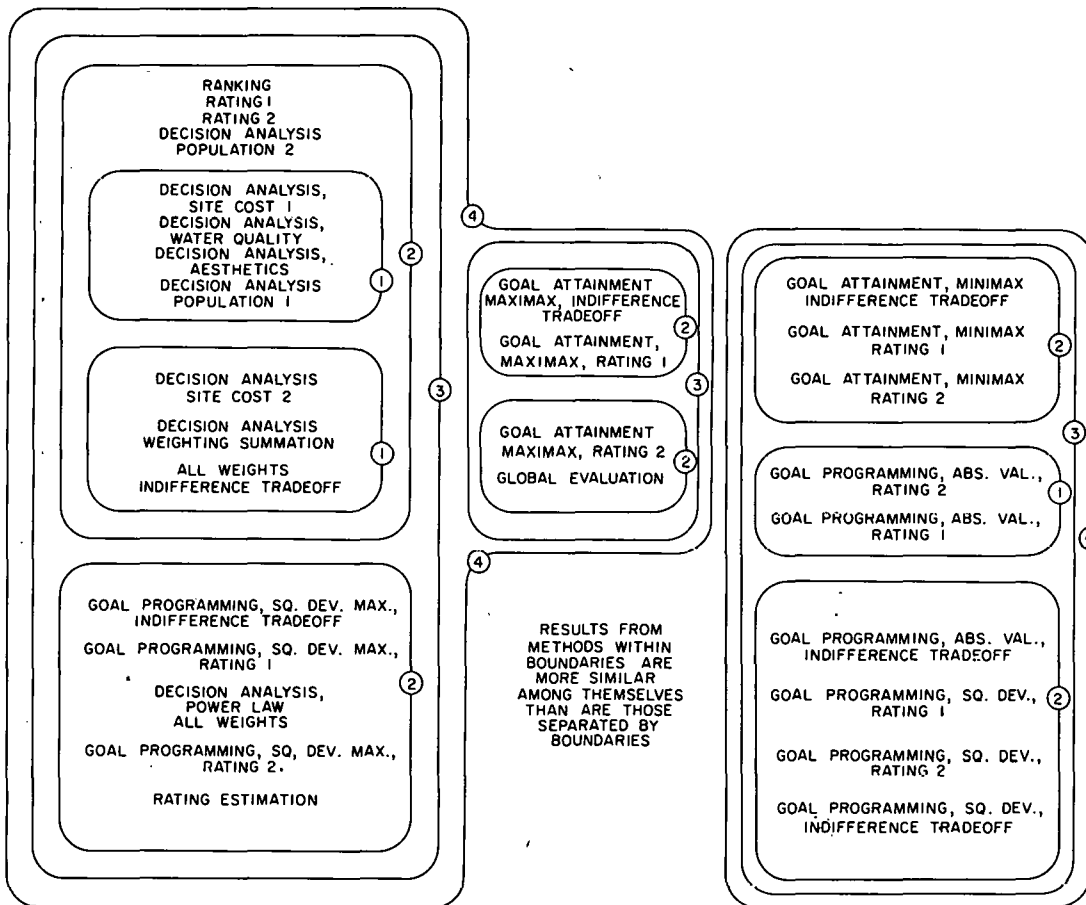


Figure 2. Major clusters of site evaluation methodology results, BNL Panel.  
 (Results from methods within boundaries are more similar among themselves than are those separated by boundaries.)

### 3.2 ADVISORY PANEL

Mean scaled attributes and weights for three value elicitation methods are shown in Appendices 3 and 4. Note that attribute values for the Advisory Panel are expressed as impacts rather than suitabilities, making a high number bad, whereas for the BNL Panel a high number is good. Correlations among methods for attribute values, weights, and total impact are shown in Table 12.

In general, the correlations are moderate and no obvious pattern can be seen among methods. Figures 3 through 5 also show some of the relationships. Figure 3 shows one of the reasons why correlations between scaling by Categorization and Rating are poor. Categorization uses ambiguously defined categories (no impact, moderate impact, unacceptably high impact, etc.), while Rating requires that the highest impact be arbitrarily assigned a value of 10.0, the lowest impact be assigned a value of 0.0, and all other impacts be appropriately scaled between the extremes. Although a group might agree reasonably well on what are the relative positions of the various impacts, it clearly does not agree on whether or not these impacts are "unacceptable." Note that few persons from the subset of the total Advisory Panel represented in Figure 3 considered the highest impact to be "unacceptable" (category 5) while most agreed that the lowest impacts are "low" or "no impact" (categories 2 and 1).

Not only are correlations among methods low, but also the spread of estimates within a single method is high. Figure 6 shows the distribution of Categorization Weighting Summation impact estimates for each site for Advisory Panel members. Ignoring covariance among persons, none of the sites appears to be significantly different from any of the other sites.

TABLE 12  
MEAN BETWEEN-METHODS CORRELATIONS ADVISORY PANEL  
OF ATTRIBUTE VALUES AND WEIGHTS

Method	Attribute value		Weights		Total impact	
	Rating	Metfessel Allocation	Rating	Metfessel Allocation	Rating	Metfessel Allocation
Categorization	0.66	0.60	0.85	0.76	0.78	0.69
Rating		0.77		0.81		0.77

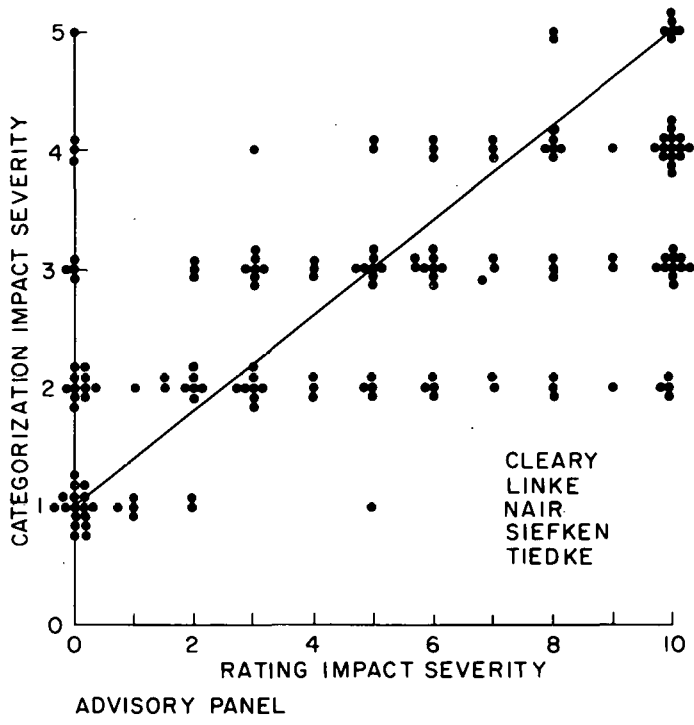


Figure 3. Comparison of Categorization and rating attribute value estimates, 5 of 17 Advisory Panel Members.

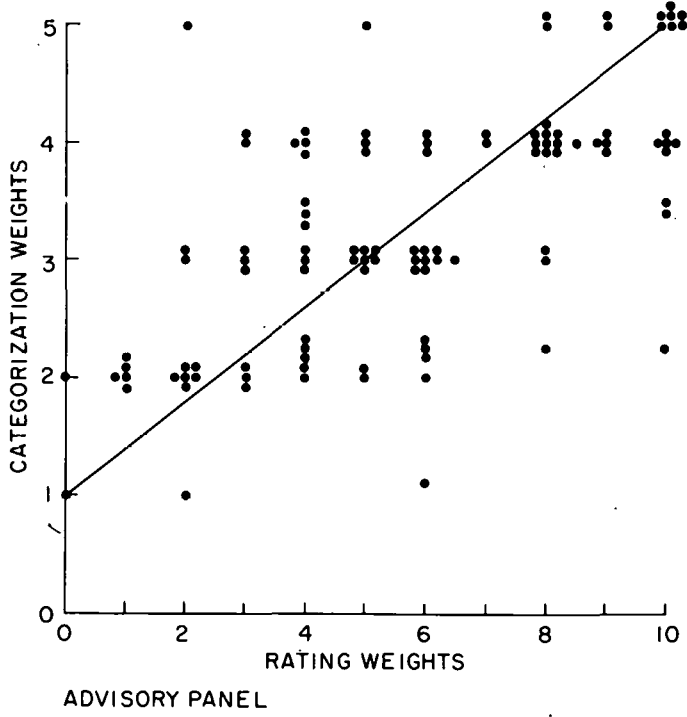


Figure 4. Comparison of Categorization and rating weights, Advisory Panel.

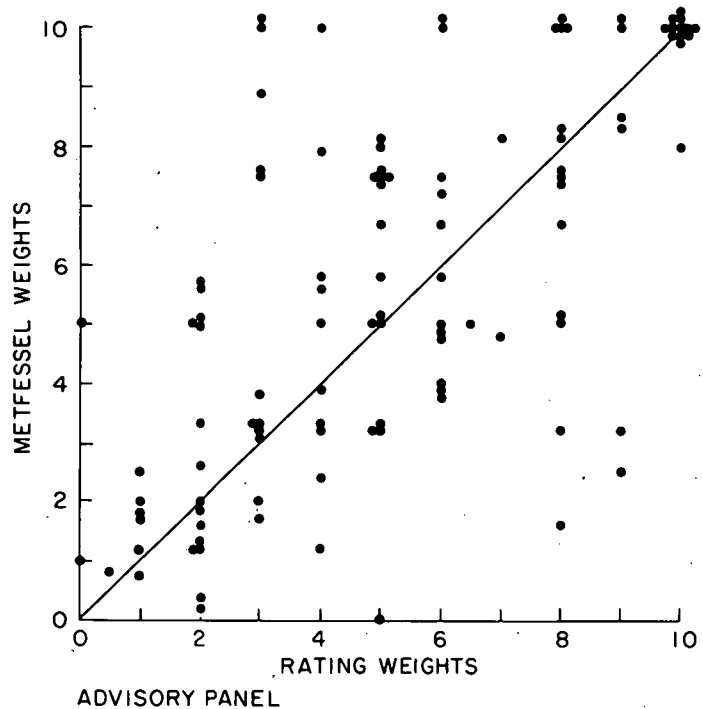


Figure 5. Comparison of Metfessel Allocation and rating weights, Advisory Panel.

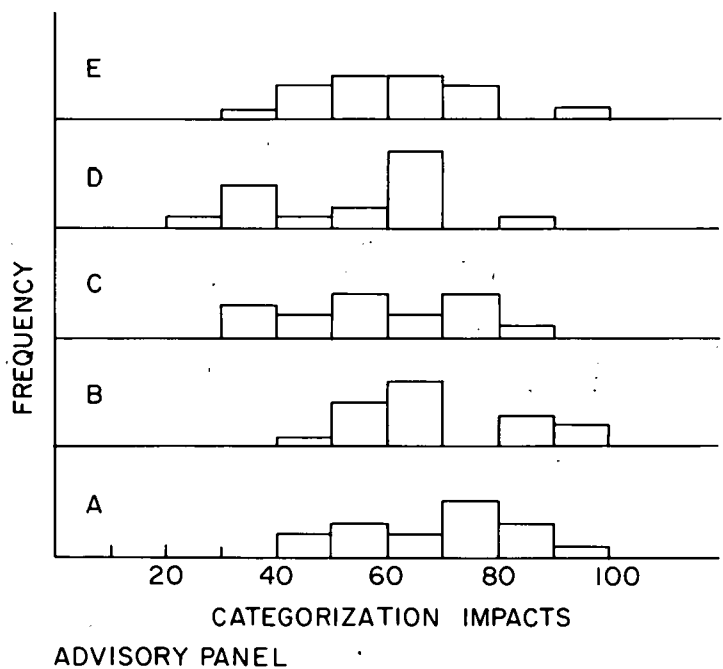


Figure 6. Frequency distribution of Weighting Summation impact estimates made with Categorization attribute values and weights, Advisory Panel.



There are several possible explanations for the general lack of agreement among scaling and weighting methods. First, the Advisory Panel was particularly dismayed at the low level of detail of the site descriptions made necessary by time constraints. Under normal circumstances, Panel members would have many weeks to study site characteristics and to evaluate relative impacts. Although this project is an academic exercise in which there is no attempt to actually site a nuclear power plant, and although it should be possible to make the necessary value judgments on the basis of whatever level of information is provided (specific judgments will, of course, be a function of the level of information), nevertheless, Panel members were uncomfortable and reluctant to make any judgments at all. The information was inadequate for siting a real power plant, which was what many thought they should be doing.

A second source of variability, related to the first, was lack of knowledge of the Long Island environment and therefore uncertainty about meanings of the words used in site descriptions. Panel members had limited understanding of the significance of a particular impact description relative to the whole Long Island system.

A third source of variability, mentioned above, arose from the manner in which site impact information was presented. To ensure that judgments of the relative importance of each attribute were made in isolation, only ranges of impacts were presented without full site descriptions showing the specific combination of impacts at each site (global knowledge). This seldom occurs in real siting studies, and several Panel members felt constrained by lack of full information.

A fourth source of variability, common to most site evaluations, arose from ambiguities in definitions of attributes and impacts. Attributes are usually broad and general categories, like "impacts on wetlands," with elements that are diverse and often not well specified. Evaluators are required to bring considerable outside knowledge and experience to bear on an evaluation of these impacts. This problem is exacerbated by a conflict between specificity of attributes and the numbers of such attributes that must be included to cover all aspects of power plant siting. It is clearly not possible to deal with large amounts of specific information, so it must be generalized and aggregated to a manageable size.

Table 13 shows individual site ranks for the Advisory Panel. In spite of the lack of correlation among methods and high variability of the individual

TABLE 13  
SITE RANKS FROM CATEGORIZATION (I), RATING (II),  
AND METFESSEL ALLOCATION (III) METHODS,  
ADVISORY PANEL

Person	Site														
	A			B			C			D			E		
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
JT	3	5	5	5	2	2	4	3	4	1	4	3	2	1	1
SL	3	5	5	4	4	4	1	2	3	2	1	1	5	3	2
KN	5	5	5	4	4	4	2	3	4	3	1	1	1	2	2
DS	5	5	4	4	4	5	2	3	2	1	1	1	3	2	3
DC	3	4	4	5	5	5	2	2	2 $\frac{1}{2}$	1	1	1	4	3	2 $\frac{1}{2}$
WO	4	4	4	5	5	5	3	3	2	1	2	1	2	1	3
BH	3 $\frac{1}{2}$	4	4	3 $\frac{1}{2}$	5	3	2	1	2	1	2	1	5	3	5
HM	4	5	4	5	4	5	2	3	3	1	1	1	3	2	2
EH	4	4	3	5	5	4	3	2	1	1	1	2	2	3	5
JL	4	4	4	5	5	5	3	2	1	1	1	2	2	3	5
TH	5	5	5	3	1	2	1	4	4	2	3	1	4	2	3
MH	5	5	4	3	3	2	4	4	5	2	1	1	1	3	3
FL	4	4	2	5	5	5	1	3	4	2	1	1	3	2	3
PM	5	5	4	2 $\frac{1}{2}$	3	1	4	4	5	1	1	3	2	2	2
WH	4	5	4	5	4	5	1	2	3	3	1	1	2 $\frac{1}{2}$	2	3
JH	5	5	3	4	4	5	1	1	1	2	1	2	3	3	4
Group rank based on mean weights	5	4	4	4	5	5	2	2	2	1	1	1	3	3	3

weights, results appear similar among methods even at the individual level; for group mean weights, results are identical in the top three sites. Figure 7 shows that individual results are not quite so similar as they appear in the table.

Part of the reason for the robustness of these results despite high variability of value estimates can be seen in the structure of the attribute levels of the alternatives (Appendix 3). The top-ranked site is best in only one of seven attributes, but it is not particularly bad for any attribute; it wins, therefore, not because it is good but because it is not very bad, so there are few tradeoffs to be made between good and bad levels of different attributes.

### 3.3 LONG ISLAND PANEL

Mean scaled attribute values and weights for the three meetings of the Long Island Panel are shown in Appendices 5 through 9. Readers should keep in

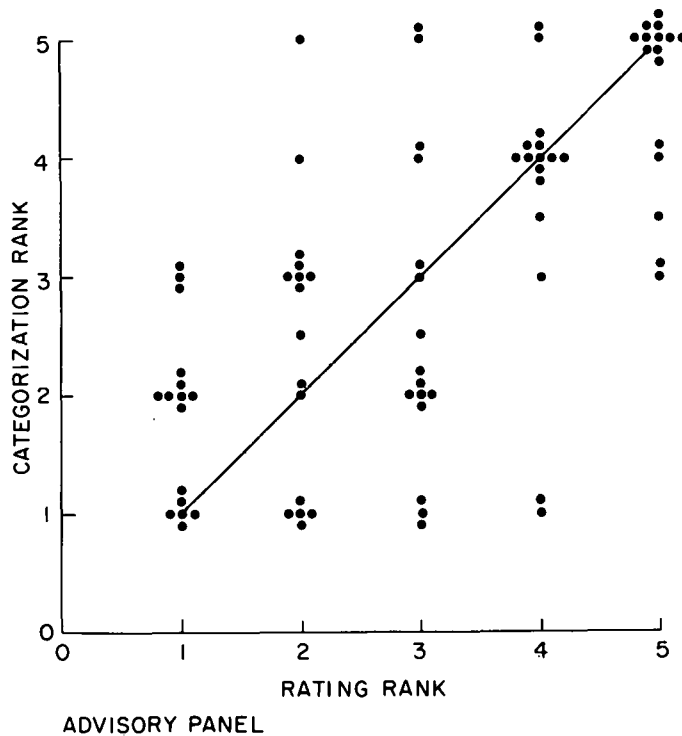


Figure 7. Comparison of Weighting Summation Site Ranks from Categorization and Rating Attribute values and Weights, Advisory Panel. Each point represents one site ranked by one panel member.

mind that the data generated in the first meeting are for a different siting circumstance from that in the second and third meetings. In addition, attribute values were prescaled in the third meeting using group mean values from the second meeting. Global Evaluation questionnaires had insufficient degrees of freedom to permit estimation of weights by linear regression.

Tables 14 and 15 show correlations among methods for attribute impact severity, weights, and Weighting Summation impact severity by three methods. These were repeated once with and once without full knowledge of the particular combination of attributes represented at each site (trials 1 and 2) and for replicates of these data and methods after a time interval of one month. Two additional methods were added at the final meeting of this panel (Table 15). There is a large difference in the level of quantification of the data between the first meeting and the following meetings. Also, because Metfessel Allocation proved particularly difficult to apply to attribute value scaling, and because the mathematical theory underlying the weighting summation method

TABLE 14  
MEAN BETWEEN-METHODS CORRELATIONS OF ATTRIBUTE VALUES AND WEIGHTS,  
LONG ISLAND PANEL, FIRST MEETING

Method	Attribute Value	Weights				
	Rating	RATE 1	METF 1	CAT 2	RATE 2	METF 2
Categorization	0.55	0.89	0.87	0.76	0.82	0.85
Rating 1			0.92	0.87	0.91	0.90
Metfessel Allocation 1				0.82	0.90	0.93
Categorization 2					0.92	0.84
Rating 2						0.93

TABLE 15  
MEAN BETWEEN-METHODS CORRELATIONS OF WEIGHTS, LONG ISLAND PANEL,  
SECOND AND THIRD MEETINGS

A. Second Meeting, Between Methods

Method	Without global knowledge		With global knowledge	
	Rating	Metfessel Allocation	Rating	Metfessel Allocation
Categorization	0.89	0.86	0.93	0.93
Rating		0.98		0.98

B. Second Meeting, With and Without Global Knowledge

Method used without global knowledge	Method used with global knowledge		
	Categorization	Rating	Metfessel Allocation
Categorization	0.78	0.91	0.79
Rating		0.95	0.93
Metfessel Allocation			0.92

C. Third Meeting, Between Methods With Global Knowledge

Method	Rating	Metfessel Allocation	Indifference Tradeoff	Decision Analysis
Categorization	0.89	0.85	0.46	--
Rating		0.92	0.37	--
Metfessel Allocation			0.32	

D. Effect of Time, 2 April to 7 June, 1979

Methods applied in April	Methods applied in June		
	Categorization	Rating	Metfessel Allocation
Categorization	0.79	0.79	0.79
Rating	0.75	0.81	0.84
Metfessel Allocation	0.76	0.81	0.83

does not require the ratio-scaled values yielded by this method, it was not included among the scaling methods used. Metfessel Allocation weights are therefore applied to Rating attribute values. For comparison, Categorization weights are also applied to Rating attribute values.

Correlations among weighting methods for the Long Island Panel are slightly better than those for the Advisory Panel, and those from the second and third meetings are better than those from the first. This is well illustrated in Figures 8 through 10. The differences are not striking, however, and there is no obvious pattern to the results. Correlations from the third meeting are similar to those from the second, except that results from Indifference Tradeoff are clearly different from those of the other methods. Also, Figure 10 shows one of the few examples of apparent nonlinear differences in perception of attribute value between the Rating method and a ratio-scaled method found in this study. Decision Analysis results are so inconsistent that no weights can be calculated. This agrees well with the results from the BNL Panel in which Decision Analysis and Indifference Tradeoff are obviously different from other methods.

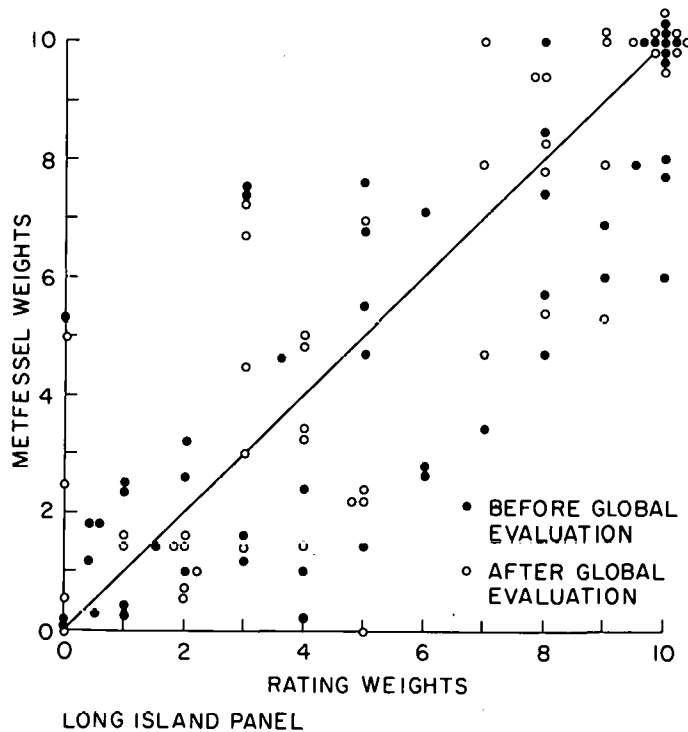


Figure 8. Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, First Meeting.

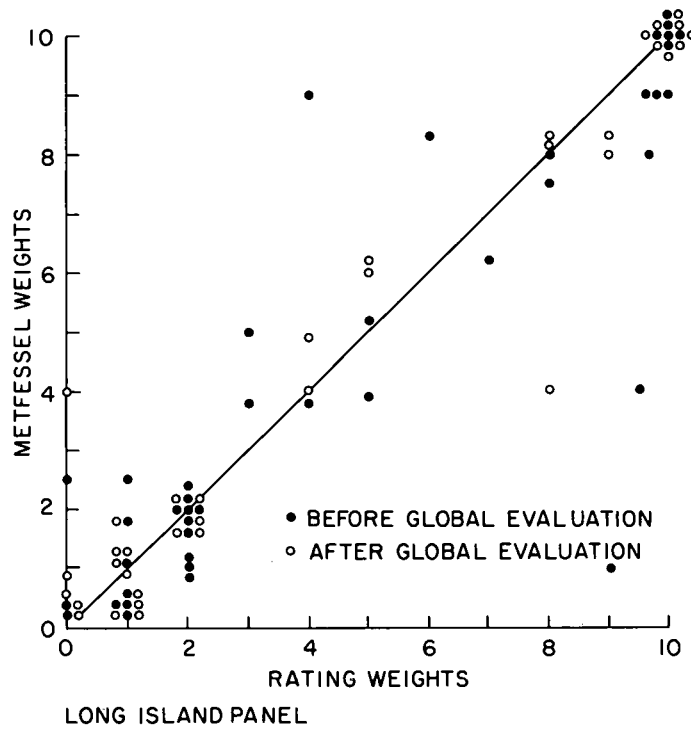


Figure 9. Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, Second Meeting.

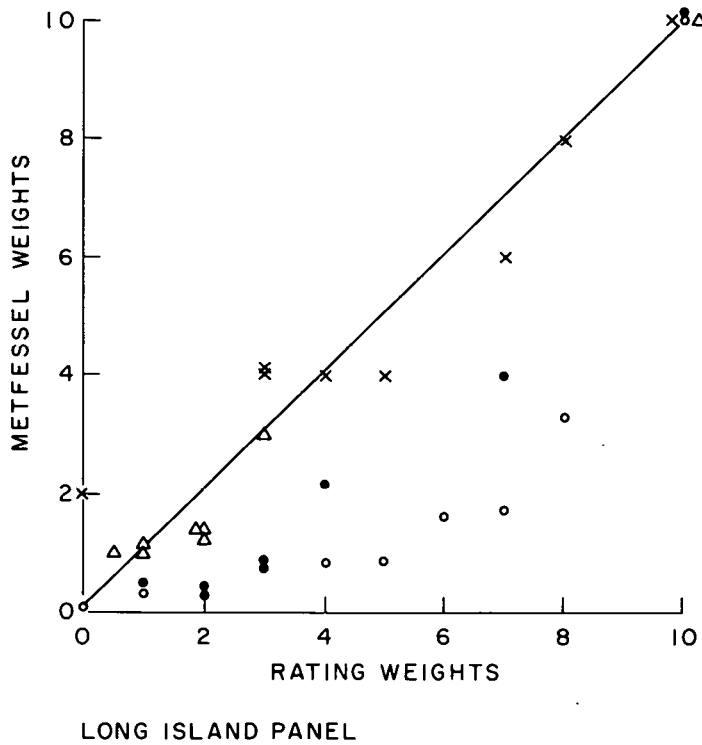


Figure 10. Comparison of Metfessel Allocation and Rating Weights, Long Island Panel, Third Meeting. Symbols represent different persons.

Weighting Summation estimates of total impact (Table 16), however, show several significant patterns. First, it is clear that Categorization results are much different from all the other results from the first meeting. The difference is primarily attributable to a low correlation (0.55) between Categorization and Rating attribute value scaling. When Categorization weights are applied to Rating attribute values, correlations become equal to or better than those among other methods. The problem with Categorization appears, therefore, to be related more to difficulties in using the technique for scaling attribute values than to weighting. This interpretation is supported by an improvement in correlations between Categorization and other methods in the second and third meetings in which the highly simplified attributes were much easier to evaluate.

TABLE 16  
MEAN BETWEEN-METHOD CORRELATIONS OF TOTAL SUITABILITY,  
LONG ISLAND PANEL, ALL MEETINGS<sup>a</sup>

Method	Cat 1*	Rating 1	Metf 1	Cat 2	Cat 2*	Rating 2	Metf 2	Global
<u>31 MARCH MEETING</u>								
Categorization 1	0.48	0.37	0.31	0.98	0.42	0.39	0.38	0.46
Categorization 1*		0.93	0.88	0.47	0.95	0.94	0.92	0.54
Rating 1			0.96	0.39	0.91	0.96	0.95	0.53
Metfessel 1				0.33	0.89	0.97	0.97	0.46
Categorization 2					0.44	0.41	0.39	0.51
Categorization 2*						0.95	0.91	0.61
Rating 2							0.96	0.56
Metfessel 2								0.51
<u>2 APRIL MEETING</u>								
Categorization 1	0.87	0.81	0.81	0.92	0.90	0.88	0.89	0.48
Categorization 2		0.88	0.90	0.75	0.91	0.87	0.85	0.50
Rating 1			0.98	0.79	0.88	0.96	0.94	0.50
Metfessel 1				0.78	0.91	0.94	0.93	0.44
Categorization 2					0.91	0.88	0.92	0.52
Categorization 2*						0.94	0.96	0.54
Rating 2							0.99	0.52
Metfessel 2								0.52
<u>7 JUNE MEETING</u>								
Categorization 2*						0.95	0.88	
Rating 2							0.92	
Indifference Tradeoff					0.80	0.72	0.61	

<sup>a</sup>Suffixes 1 and 2 refer to results with and without full knowledge of specific combinations of impacts, respectively.

\*Categorization weights applied to Rating attribute values.

A second observation on these results is that Global Evaluation appears to be significantly different from all other methods under all conditions tested. Again, this agrees well with results from the BNL Panel.

A third observation is that, although the Long Island Panel felt much more comfortable making value judgments after the Global Evaluation exercise and therefore with full knowledge of the specific combinations of impacts at each site, this knowledge did not significantly improve consistency of results. Two Panel members did, however, say that their relative weights changed after Global Evaluation; one member changed a weight by a large amount and with great reluctance because he felt he had to reduce the influence of something that would have had high value under other circumstances. This suggests that he did not consider the range of attribute values for the first weighting exercise. Most weights changed very little (Figure 11).

Unlike the BNL Panel, Long Island Panel members felt relatively comfortable with the Indifference Tradeoff weighting method which they thought better able to capture specific tradeoffs among attributes. As can be seen from Table 15 and Figure 12, this method yields results which have obviously different correlations with other methods tested by this Panel. Whether this indicates that the well-correlated but different methods are significantly less good at capturing participants' preferences or vice versa is not clear. Indifference Tradeoff is the only method shown which in theory produces valid weights.

Appendices 10 and 11 show site ranks based on Weighting Summation suitability for each member and each method tested by the Long Island Panel. Again in spite of high variability, the results are similar. Using group mean weights, only Global Evaluation, in one instance, produced different results. Unlike that for the Advisory Panel, the structure of the attribute values in this case is complex, and there are many tradeoffs to be made.

### 3.4 DECISION COMPLEXITY

The robustness of the siting problems tested despite high variability among weights is surprising. Given some thought, however, it becomes obvious that the numbers of tradeoffs and magnitudes of differences among alternatives can in some circumstances have more influence on results than weights and, by extension, weighting methods. Consider, for example, a set of alternative sites in which the best site is best in each attribute, the second best site is secondbest in each attribute, and so on for all sites. Since there are no



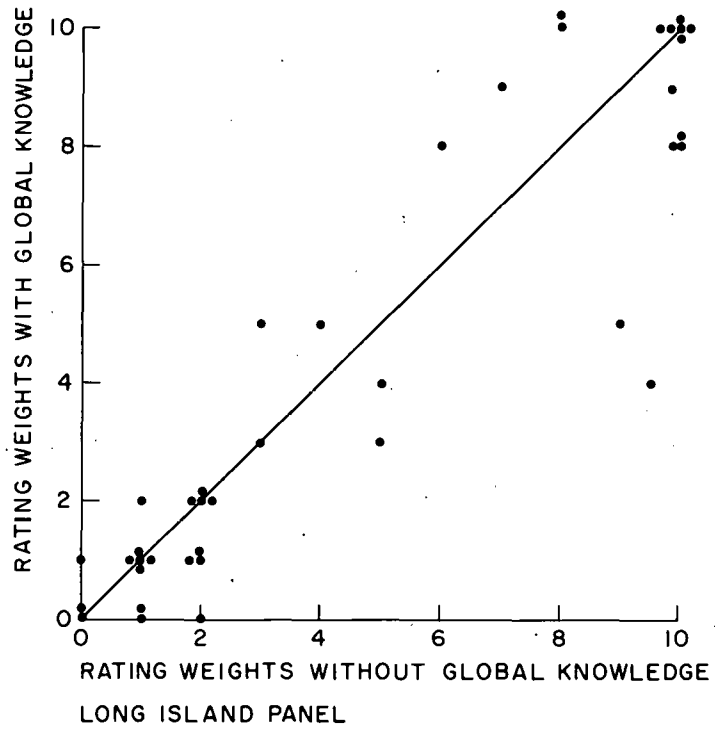


Figure 11. Comparisons of Rating Weights estimated with and without global knowledge, Long Island Panel, Second Meeting.

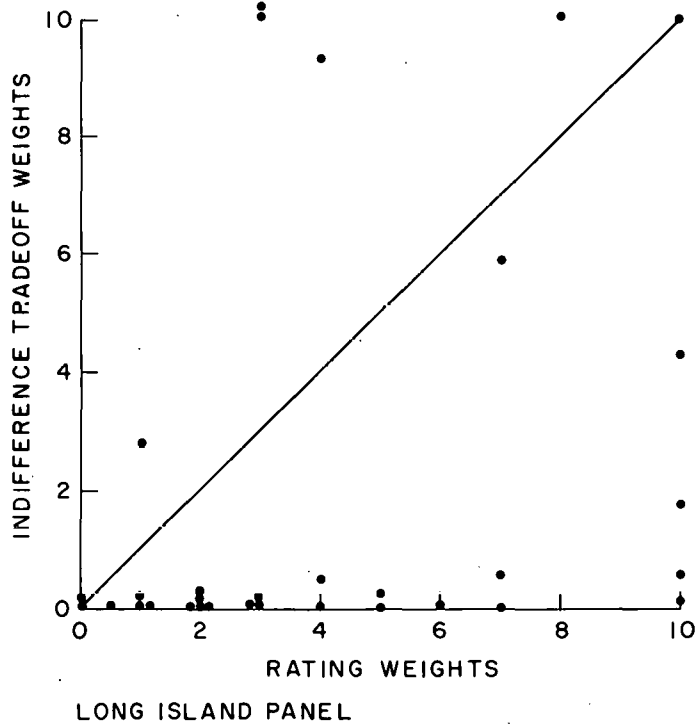


Figure 12. Comparisons of Indifference Tradeoff and Ratings Weights, Long Island Panel, Third Meeting.

tradeoffs among alternatives, there is complete dominance. Under this condition the rank of each alternative is independent of weights or amalgamation method.\* Such a siting problem has zero complexity. With respect to the final decision, it is only necessary that the first-ranked site be best in all attributes to make that decision independent of weights.

At the other extreme, one can envision a situation in which attribute values are systematically organized so as to create the maximum possible number of tradeoffs among very high and very low values of different attributes. Such a siting problem would be highly complex, very difficult, and very sensitive to magnitudes of weights for each attribute. Between these two extremes there is a continuum of possibilities of increasing complexity from no tradeoffs, through a random distribution of tradeoffs, to systematic increases in tradeoffs among alternatives.

To quantify this "complexity" and sensitivity to magnitudes of weights we have derived an index that measures the extent to which a set of attribute values contains tradeoffs which must be resolved by subjective judgments of relative importance - weights. The derivation begins by considering what decision one would make if all attributes have equal weight. Under this condition the best alternative would be that having the highest sum of attribute values. Rank-ordering the alternatives with respect to this total (grand rank) provides a resolution of all tradeoffs in the absence of differences in weights. Examining the magnitudes of these tradeoffs for each attribute in some quantified manner then provides an estimate of the potential that differences in attribute importance (weights) have to change the decision based on attribute values alone. The procedure is as follows (Table 17). Alternatives are rank-ordered by the sum of all attribute values (grand rank). Within each attribute, alternatives are rank-ordered by the magnitude of that particular attribute value (local rank). By ascending rank,  $i$ , each  $i$ -th local-ranked attribute value is subtracted from the corresponding  $i$ -th grand-ranked attribute value and the difference is divided by the range of attribute values across all alternatives. This yields a measure of magnitude of tradeoff relative to the total range of tradeoffs possible. Were there complete dominance, the local rank-order and grand rank-order of the alternatives would be the same -- the best alternative is best in each attribute, etc. If the local rank-order is not the same as the grand-rank order, then there are tradeoffs of known magnitude.

---

\*Goal Programming and Goal Attainment methods may produce different results because they can select dominated alternatives.

TABLE 17  
CALCULATION OF DECISION COMPLEXITY

Alternatives	Attribute Value		Local Rank		Sum of Attribute Values	Grand Rank
	I	II	I	II		
A	2	0	4	5	2	5
B	0	9	5	2	9	4
C	6	10	3	1	16	1
D	10	4	1	4	14	3
E	7	8	2	3	15	2

Rank	Attribute Value Listed by Local Rank		Attribute Value Listed by Grand Rank		Absolute Value of Difference Divided by Range		Relative Difference Divided by Grand Rank	
	I	II	I	II	I	II	I	II
1	10	10	6	10	4/10	0/10	0.4/1	0.0/1
2	7	9	7	8	0/10	1/10	0.0/2	0.1/2
3	6	8	10	4	4/10	4/10	0.4/3	0.4/3
4	2	4	4	9	2/10	5/10	0.2/4	0.5/4
5	0	0	2	0	2/10	0/10	0.2/5	0.0/5
Subtotal							0.623	0.308
TOTAL							0.931	

The importance of tradeoffs to a decision depends on their position in the rank-order. Tradeoffs in the first-ranked alternative means that the best alternative is not best in everything, and if the attributes for which it is not best are important, some other alternative may become first-ranked after application of weights. Tradeoffs in low-ranked sites have less significance to a decision because low-ranked alternatives are unlikely to become high-ranked whatever weights are applied. This difference in relative importance of tradeoffs can be approximately accounted for by weighting each tradeoff in inverse proportion to its grand rank. In this manner, a tradeoff in the second-ranked alternative is half as important as one in the first-ranked alternative, and a tradeoff in the fourth-ranked alternative is one-fourth as important as one in the first-ranked alternative, etc. The correction factor is approximately to the extent that weighted ranks are different from the grand rank.

The absolute value of resulting differences between attribute values listed by local rank and attribute values listed by grand rank divided by

grand rank are added over all attributes and all alternatives to form an index of total complexity of a multiobjective decision. In order to compare complexities among problems having different numbers of attributes and alternatives, the index can be normalized with the coefficients shown in Table 18. These coefficients are approximate expected values of complexity for a random distribution of tradeoffs. They were generated by Monte Carlo simulation of random tradeoffs among interval-scaled attribute values. Using these coefficients, the expected complexity of a random combination of attribute values is 1.0; the range of the index is 0 to about 1.6.

### 3.5 SYSTEMATIC DIFFERENCES AMONG WEIGHTS

Lacking specific information on systematic differences among weight estimation methods, we can hypothesize differences based on the manner in which the methods work. This can yield information on the potential each known kind of methodological difference has to affect decisions in the absence of individual and methodological uncertainty. We have identified five classes of transformations attributable to weighting methods that can produce weights which are not correctly ratio scaled (Figure 13). The first is linear categorization, in which ratio-scaled weights are grouped into 2, 3, 5, etc. categories of equal range. In this class of transformation, ratio-scaled weights ranging from 0-10 might be grouped into three categories -- low, medium, and high -- such that all weights from 0 to 3.33 are categorized as low, those between 3.34 and 6.66 are categorized as medium, and those from 6.67 through 10.0 are categorized as high. The second class is nonlinear categorization in which categories are not equally allocated throughout the range of the original weights. Categories having "linear" descriptions, such as low, medium, and high, might be perceived in such a manner that the difference between low and medium is less than the difference between medium and high or vice versa. The particular transformation considered here is logarithmic perception, in which intervals are interpreted proportionally so that logarithms are perceived to be linear. The third transformation is application of logarithmic (or exponential) perception directly without classifying the result. A logarithmic transformation shifts weights toward the high end of the scale (see Figure 10); an exponential transformation shifts weights toward the low end of the scale. The fourth class of transformation is a shift of scale. Such a shift can occur in the rating method of weight estimation if the attribute having the lowest importance is arbitrarily assigned a weight of 0.0.

TABLE 18

DECISION COMPLEXITY NORMALIZING COEFFICIENTS<sup>a</sup>

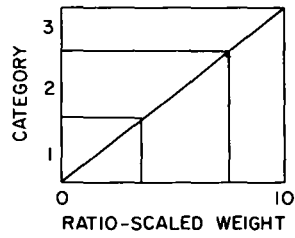
Number of Alternatives	Normalizing Coefficient (M = number of attributes)
3	0.76 (M - 1.33)
4	0.83 (M - 1.33)
5	0.88 (M - 1.33)
6	0.92 (M - 1.33)
7	0.96 (M - 1.33)
8	0.99 (M - 1.33)
9	1.02 (M - 1.33)
10	1.05 (M - 1.33)
11	1.07 (M - 1.33)
12	1.09 (M - 1.33)
13	1.11 (M - 1.33)
14	1.13 (M - 1.33)
15	1.15 (M - 1.33)

<sup>a</sup>Derived from Monte Carlo simulation of random combinations of attribute values.

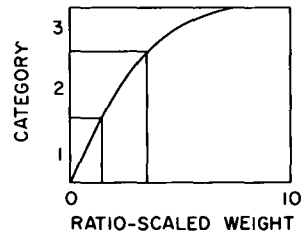
Interval-scaled weights of this kind almost always have a zero point that does not represent zero importance. Two such scale shifts are examined, one in which correctly ratio-scaled weights are converted to interval-scaled weights by assigning the lowest weight to be 0.0 and rescaling the remaining weights on the interval 0-10 (the scale shift is different for each weight set), and another in which the lowest weight is set at a specific distance from "true" zero, and several different distances are tested.

The fifth class of transformation is random variation without systematic shift or bias. Normally distributed random variation with standard deviation 0.5, 1.0, and 2.0 weight units are tested to compare potential effects of uncertainty with those of systematic differences. Uncertainty of this kind can be interpreted as individual uncertainty about the magnitude of weights, as an effect of increase in uncertainty produced by difficulty in applying a particular weighting method, or as differences in definition of attribute "importance."

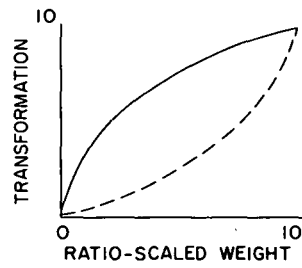
Test data containing 48 different siting problems having equal numbers of attributes, and alternatives were used to examine effects of the five classes of weight transformations attributable to weight estimation methods. Complexity ranged from 0.0 to close to the maximum possible. In each case results of



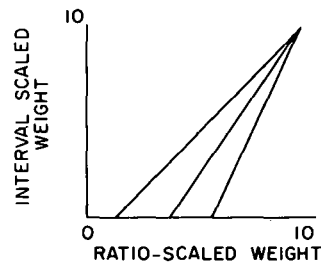
a. LINEAR CATEGORIZATION



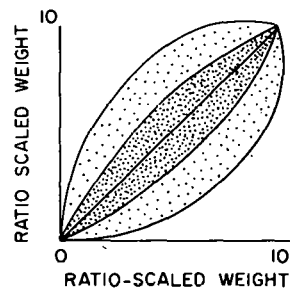
b. NONLINEAR CATEGORIZATION



c. NONLINEAR PERCEPTION



d. SCALE SHIFT WITH ARBITRARY ZERO



e. RANDOM VARIATION

Figure 13. Weight transformations attributable to weighting methods.

applying 10 randomly generated ratio-scaled weight sets were compared with those from the same weights transformed in the specified manner. The measure of difference in results was the probability of a decision error (a change in the first-ranked site). Figure 14a is an example of the results. The particular shape of the distribution, with high variability for high complexities, is explained by Figure 14b, which shows the expected probability density functions for 10 such trials. Density functions are based on an assumption of 10 random Bernoulli trials having an expected value equal to that shown by the regression line in Figure 14. Figure 14 shows that, although the expected probability of a decision error for a logarithmic transformation is 0.12 at a complexity of 1.0, there is a 10% probability that such a transformation will yield a 30% probability of error.

To generalize results of the various transformation tested, a regression line was calculated for each and the slope of that line reported as the average increase in probability of error per unit decision complexity. The probit or logit transformation should properly be used for probabilities of this kind<sup>5</sup>; however, these methods do not easily accept observations of zero probability and the resulting slope constants are not so easily interpreted as those of linear regression. A test of the logit transformation showed the two lines to be nearly identical, so linear regression was retained for its ease of interpretation and use. Regression coefficients for all transformations are shown in Table 19.

Since the complexity index is normalized using values for a random distribution of tradeoffs among attribute values, the slope of the error probability line (Figure 14) is equal to the probability of error for that random distribution. A two-level linear categorization scheme (acceptable, unacceptable), for example, would, on average, introduce a 22% probability of error for a random set of attribute values. The average probability would be higher for problems of higher complexity, up to 33% for a maximum complexity of about 1.5 units, with an 8% chance that the probability exceeds 50% at this complexity.

The results shown in Table 19 suggest first that potential for error introduced by individual uncertainty (random variability) is less than or equal to that of systematic error produced by incorrect methodologies. Categorization, in particular, has high potential for introducing error unless there are a sufficient number of categories. Even five categories are not sufficient to

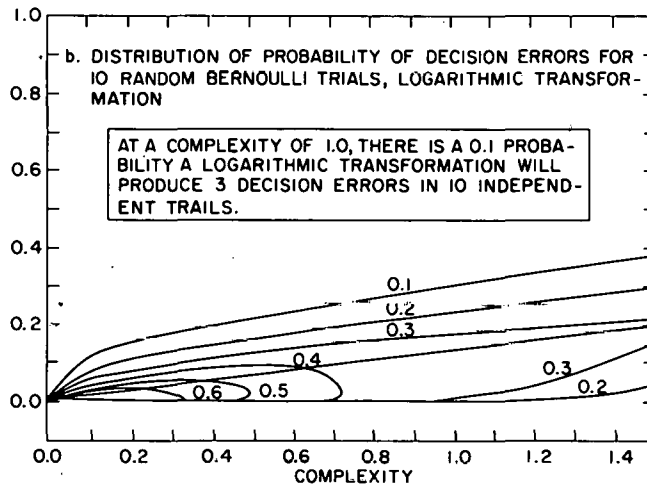
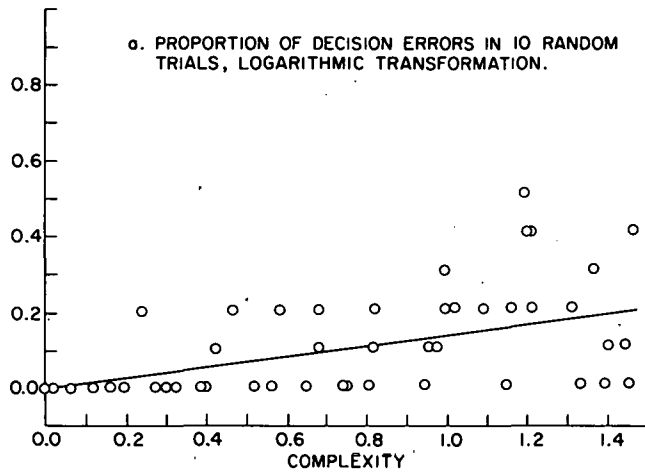


Figure 14. Probability of decision error due to logarithmic perception as a function of decision complexity.



TABLE 19  
 PROBABILITY OF ERROR ATTRIBUTABLE TO VARIOUS  
 TRANSFORMATIONS OF RATIO-SCALED WEIGHTS

Transformation	Changes in Probability of Error per Unit Complexity
Linear categorization	
2 categories	0.22
3 categories	0.10
5 categories	0.08
Categorization based at 0 vs categorization based at 1	
2 categories	0.09
3 categories	0.05
5 categories	0.00
Nonlinear categorization logarithmic transformation	
2 categories	0.22
3 categories	0.18
5 categories	0.13
Nonlinear transformation	
logarithmic	0.12
exponential	0.13
Ratio to interval scale	
Random zero point	0.04
+2 zero point	0.07
+4 zero point	0.14
+6 zero point	0.20
Random variability (Means = 0-10)	
= 0.5	0.04
= 1.0	0.08
= 2.0	0.15
= 3.0	0.27

reduce the error attributable to categorization to less than that attributable to a "medium" level of uncertainty ( $\sigma = 1.0$ ). The problem is exacerbated by nonlinear perception of categories, a common problem with this method,<sup>3</sup> and by a categorization scheme that is based at one when it should be based at zero. Given this large potential for introducing error attributable to methodology, it is probably best to avoid the Categorization method of weight estimation.

Results not shown here demonstrate that these different classes of error are not strictly additive in their combined effects.

#### 4 DISCUSSION

Because of the heterogeneous composition of the various panels used to test site evaluation methodologies, and the low level of information available for evaluation in each, one would expect that results should be highly variable -- more so than for siting specialists given real-world information. In contrast, opportunities for learning and carryover among methods reduced variability. Readers should, therefore, exercise some skepticism about the magnitude of the variability of results. Type of variability, however, in most cases should not be affected by magnitude of variability, so that many of the results reported here represent extreme cases of sensitivity of the methods to the various sources of error under consideration.

There is no "correct" result against which to compare outcomes, which presents a particularly difficult problem of interpretation. Since we cannot evaluate the relative success of each method in providing a "correct" answer, we must instead use consistency of results among methods, under an assumption that, at a minimum, consistency implies that different methods address the same preferences. Whether or not the numbers correctly represent the internal preference system of the individuals involved can only be determined by the extent to which they agree with methods for which theoretical considerations suggest more valid results.

Among the weight estimation and amalgamation techniques chosen for study, there is a range of theoretical validity and difficulty of application. In general, the more theoretically valid methods are also more difficult to apply. There is, therefore, a tradeoff between potential for error due to theoretical problems and potential for error due to inability of inexperienced persons to provide the kinds of responses required. This is well illustrated by results from the two most extreme methods tested, Global Evaluation and Decision Analysis lotteries. Global Evaluation is the simplest method tested. It is multiobjective decision making in the absence of formalized techniques. With this method it is generally not possible to keep all of the information needed for site evaluation in mind at one time; therefore, each individual isolates a smaller subset of information about those attributes which he considers most important and evaluates sites with respect to that smaller subset. Correlations with methods which include all attributes are necessarily low and, since each person evaluates a different subset of attributes, agree-

ment among individuals is also low. Although this method may produce responses most like informal decision-making processes, the objective of formal multiobjective decision making-methods is to disaggregate the decision making process so as to improve quality of results by forcing decision makers to consider all aspects of a problem. One would expect, therefore, that other formal decision making should produce results more consistent with decision makers preferences than Global Evaluation. Whether or not other methods tested in this study produce "better" results cannot be determined directly. Nevertheless, it is clear that they produce different results which we assume are better because of Global Evaluation's theoretical problems.

Goal Programming and Goal Attainment techniques have similar theoretical problems in that they tend to downgrade a site which is better than the goals entered into the evaluation. Results from these methods are also different from other methods and, by extension, we assume that they are less "good."

Decision Analysis is the most theoretically valid and most difficult method tested. It was applied once by the BNL Panel and once by the Long Island Panel, both composed of persons inexperienced in use of this method. The Long Island Panel's results are so inconsistent among themselves that no feasible Decision Analysis weights can be calculated at all. The BNL Panel was able to produce some consistent results, but only for lottery questions phrased in a manner opposite from normal. Panel members appear to have considered the worst alternatives to be so unacceptable that they were unable to make valid judgments about good levels of the attributes in question. Several Panel members made comments to this effect. Because of the obvious problems Panel members had with this method, the quality of even the feasible weights is suspect. This method is clearly not appropriate for application by inexperienced site analysts without extensive training and consistency checking.

Long Island Panel results suggest that categorization may not work well for scaling attribute values. The reason for this is obvious. Categories in this study were too poorly defined and interpretations too individual-specific to lend themselves well to the level of quantification required by the Weighting Summation decision rule. Use of categories need not always be as variable as is shown here; categories can be better defined and a decision making group can agree on interpretations of each category. Nevertheless, the method in general has less capability to deal with rigorous quantification than other methods tested.

Several different levels of knowledge about the siting problem were tested for influence on results. The Long Island Panel was more familiar with the general environment of the test area and they were given two different levels of specificity of site descriptions and attribute quantification. In each case, improved information reduced variability of weight estimates among the different methods used. This suggests that a large portion of the between-methods variability may be individual uncertainty not attributable to specific method.

Lack of information on specific combinations of impacts at each site (global knowledge) particularly bothered Long Island Panel members and, although they did not pursue the matter in the larger group, it also appeared to bother many members of the Advisory Panel. One participant expressed his concern as follows, "How can I evaluate the relative importance of cost when I don't know what I'm getting for my money?" The same can be said about other tradeoffs. The implication is that participants did not feel they could evaluate attributes in isolation because the relative importance of each attribute is specifically related to magnitudes of other attributes. The relative importance of each attribute is viewed with respect to the total package of sites and impacts and not in isolation or as pairwise tradeoffs among individual attributes. The concept that seems to be involved here is that of goal attainment or satisficing. So long as there exists among the alternatives a site that is generally good in all attributes, although perhaps not best in any particular attribute, participants are willing to be forgiving about the relative importances of the attributes. If no generally satisfactory site exists, however, and all sites are bad in at least one attribute, then participants feel a need to "fight" for their own preferred attributes. In the jargon of multiattribute utility theory, this is lack of value independence and possibly lack of preference independence as well.<sup>3</sup> Independence is a fundamental assumption of most decision rules.

The result of the above would be a spreading of weights toward extreme values when there is no generally satisfactory alternative. Site decisions under this condition would be much more sensitive to the particular set of preferences held by participants. Perhaps this implies that decision makers may be more comfortable with results from Exclusionary Screening, Goal Programming, or Goal Attainment Techniques in spite of the fact that these techniques have some theoretical problems with respect to representation of specific preferences.

In spite of participants' complaints about lack of global knowledge, and in spite of some good reasons why global knowledge might change values, there is little difference between judgments made with and without this knowledge. Correlations among the two weight sets are uniformly high. It appears, therefore, not to have affected the quality of the results; however, participants were much more comfortable with weight estimations made with full global knowledge and so felt more confident in the results.

Partly for the same reason, Long Island Panel members felt more comfortable with the Indifference Tradeoff method of weight estimation. This method requires explicit tradeoffs of the relative importance of a unit of one attribute against a unit of another attribute which assures valid weights. This is closer to the complete global knowledge that they would prefer, and also permits more specific tradeoffs than simpler methods which address more general, implicit tradeoffs among all attributes at once. Unlike the Long Island Panel, the BNL Panel specifically did not like the Indifference Tradeoff method. This may in part have been a function of the manner of presentation of the method. The BNL Panel used a graphic approach in which participants were required to locate an indifference point visually. The Long Island Panel used a "list" approach in which participants were required to scan a list of tradeoffs and choose one or the other member of each pair until they found a pair for which they had no preference.

Results of the Indifference Tradeoff method for both Panels were poorly correlated with other methods. The reason for the poor correlation, however, may be different in the two cases. BNL Panel results may be "bad" because the Panel found the method difficult; Long Island Panel results may be "good" because the Panel liked the method. This is not necessarily the case, however, since results from tests with and without global knowledge show that relative happiness with the methodology does not necessarily indicate that weights will be different. The poor correlation may also be related to the fact that this method is one of only two methods tested which assure valid weights if applied correctly.

Results from applications of the various weighting methods tested show that methods do, in fact, make a difference in estimates of weights. The question most central to this analysis, however, is "Does it matter?" Do differences attributable to method affect final site selection? Tables 11 and 13 and Appendix 10 show final site ranks for each method tested by each Panel.

With the exception of Global Evaluation, Goal Programming, and Goal Attainment, which are, for reasons discussed above, obviously different, the results are nearly identical for all methods. Even results for individuals are surprisingly similar considering the lack of correlation of weights among them. BNL Panel individual Decision Analysis weights are probably meaningless, yet group average weights based on mean probability estimates are again nearly identical. Part of the reason for the robustness of these results is that central tendency of means draws weights closer together than they would generally befor individuals. The reduction in range reduces differences among methods and reduces opportunities for extreme tradeoffs and extreme site evaluations.

Group mean weights clearly do not represent group preference. Consider, for example, two attributes to be assessed by members of a group; group members are strongly divided in their assessment of relative importance high and low for one attribute; for the other there is agreement on moderate importance. The group mean weight might be the same for the two attributes, but the relative unhappiness generated by a solution based on the weight on which they disagree is much greater than that on which they agree. There are ways of resolving these kinds of conflicts<sup>4</sup>; most of them involve consideration of individual values or site ranks, not combining individual estimates to form group means.

Another reason for the robustness of the results may be a low level of complexity, as seen in Appendix 3. The top-ranked site is best in only one of seven attributes, but it is not particularly bad for any attribute. It wins, therefore, not because it is good, but because it is not bad, so there are no tradeoffs to be made between very good levels and very bad levels of different attributes.

The influence of numbers of tradeoffs among attributes on the relative importance of weights to a decision is captured quantitatively by the concept of decision complexity. In multiobjective problems of complexity of about 0.5, for example, probability of error is only on the order of 10% even for the grossest of conceptual errors or uncertainty. These problems are correspondingly insensitive to differences in weights attributable to specific circumstances of the problem or to the particular individuals represented by the weights. There is not much that will change the results. At the other extreme, for problems of complexity on the order of 1.3, the expected probability of error can be in excess of 30%, with a 4% chance that the probability

exceeds 50%. These problems are sensitive to any source of variability in weights.

The relative importance of individual uncertainty, method, and complexity to a final decision can be inferred from the ranges of values shown in Table 19. Standard deviations of group weights for the various panels are on the order of half the mean; two units for a weight of five units is representative. For this standard deviation, at any particular problem complexity, individual uncertainty is roughly equal in impact on decisions as are methodological errors. For the data tested, differences in complexity from 0.5 to 1.3, well within the expected range of this index, will increase probability of a decision error by a factor of 2.6. One can infer, therefore, that problem complexity may be on the order of twice as important as other variables in affecting quality of final results.

Complexity of a nuclear power plant site evaluation is controlled by the stage in the siting process immediately preceding final site evaluation in which sites are identified from candidate areas and reduced to a manageable number. There are several aspects of this stage that can introduce systematic bias and thereby cause the complexity of the resulting final site selection problem to be greater than or less than it would be otherwise. The first and most obvious is strong preconceived ideas on the part of the individuals who select the initial slate of sites about what makes a "good" power plant site. Strong biases tend to change complexity by producing sites that are all alike, (reduced diversity) and, therefore, the probability that any one site will be obviously superior or dominant. Reduced diversity increases sensitivity of results to magnitudes of weights, but reduces the significance of weights because all sites are of similar quality under this condition. It is the skill with which these individuals select sites that determines the relative quality of the final decision. The better the sites from which a choice is made, the better the final product is likely to be independent of selection methodology.

Correlations among attribute levels in candidate areas can also affect complexity, but not necessarily by introducing bias in one direction. Complexity will be reduced by positive correlations among attributes and increased by negative correlations. To the extent that individuals choosing sites are aware of correlations and seek or avoid areas likely to have highly correlated attributes, the complexity of the final problem can be biased at this stage.

Prespecified physical or geographical conditions can change complexity by eliminating large portions of candidate areas from consideration. In the

past, prespecifying once-through cooling, for example, effectively eliminated all potential sites not close to a large source of water. In most cases, cost considerations would have precluded other cooling methods, so sites without water would have been rejected anyway, but there may have been many borderline areas where prespecifying in this manner introduced bias. Even more than cooling type, system requirements such as transmission distances and reliability considerations, if prespecified, can restrict alternatives and therefore, change complexity. Similarly, political (or NRC) requirements which mandate one or more candidate sites from specific areas (political regions, candidate areas, etc.) or of specific physical types (seashore, river basin, mountains, etc.) force into consideration sites which might otherwise not have been suitable. This would increase diversity in the final slate of sites, but the probable impact on complexity is not clear. If most of the sites forced into the final site would have been left out because they are obviously inferior, then their inclusion will reduce complexity; if they are left out because they involve large tradeoffs, then their inclusion will increase complexity. Again, correlation among attributes will be an important variable in determining complexity.

Because of its importance in determining not only the complexity of the final decision but also quality of the results, the candidate site selection stage, is critical to the success of the entire siting process, perhaps the most critical of all. It is also the least formalized stage and the one most likely to be conducted "by the seat of the pants." Usually, experienced sitters sit down with a large number of maps and hunt up places that "look good." There is no reason why this method should necessarily be less successful than more formal methods. If one assumes that there are many suitable sites available in any candidate area and that only a few need to be found to ensure an acceptable choice, then skillful sitters will have no problem in identifying a slate of candidate sites from which at least a few are "good." A problem will arise, however, if the site evaluators lack skill and experience. The slate of candidates may not contain any high quality sites, and there is no easy way to determine from the manner in which the selection is conducted whether or not this is the case. Final site selection can be made from among inferior sites, and the only way to determine that this has occurred is to search for an obviously superior site -- one which dominates all alternatives. This places a burden of proof on interneners that should



rightfully fall on utilities. In view of the importance of this stage and the relative lack of accountability due to the manner in which it is conducted, more detailed study should be conducted to determine what is necessary to ensure that the results are acceptable.

## 5. CONCLUSIONS

1. Different scaling and weight estimation methods can produce different attribute values and weights.
  - Categorization lacks capability to quantify with the level of rigor required of the Weighting Summation decision rule and can suffer from unquantified nonlinear perception of relative importance.
  - Global Evaluation of more than three or four attributes is unsuccessful because decision makers can not process that much information at once. Instead, they evaluate an unspecified subset of the information.
  - Indifference Tradeoff is the only method tested which both can assure valid weights and can be applied by inexperienced persons. Its results may be different for this reason. The assumptions underlying the methods may not hold, however, so the significance of these "more valid" weights is unclear.
  - Decision Analysis lotteries produced few feasible weights in this study. This method should not be used by inexperienced persons without extensive training and consistency checking.
  - Central tendency of means reduces differences among methods for group mean weights.
2. Different weights can produce different results.
  - Results are strongly influenced by the structure of the data to which weights are applied. Some data structures are less sensitive than others to differences in weights attributable to weight estimation methods.
  - A "decision complexity" index is derived which quantifies sensitivity of results to differences in weights.
3. Different decision rules can produce different results.
  - Goal Programming and Goal Attainment techniques produce obviously different results from other decision rules. This may be because they are insensitive to dominated solutions, or it may be because given lack of value and preference independence, they better represent the satisficing behavior attempted by many of the decision makers in this study.
4. Selection of Candidate Sites from Candidate Areas may be the most important stage of the siting process.
  - Sensitivity of results to differences attributable to weight estimation method and decision rule is strongly influenced by the diversity and complexity of the sites chosen for evaluation.
  - Candidate Site selection is the least quantified and least formalized stage of the site selection process. Because of its importance, it should be conducted with special care.

## REFERENCES

1. Considerations for Rulemaking on Alternative Sites. Staff study document for consideration at the NRC-sponsored Workshop conducted at the MITRE facilities, McLean, Virginia, March 14-16, 1979. Prepared by the Office of Nuclear Reactor Regulation, Office of Standards Development, U.S. Nuclear Regulatory Commission.
2. M.C. Cordaro and W.T. Malloy, "A Methodology for Power Plant Site Selection at the Reconnaissance Level." Nucl. Technol. 23, 223-239 (1974).
3. B.F. Hobbs, "Analytical Multiobjective Decision Methods for Power Plant Siting: A Review of Theory and Applications," BNL-51204 (1979, in press), Brookhaven National Laboratory, Upton, N.Y.
4. R.L. Keeney and H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs, Wiley, New York, 1976.
5. Henri Theil, Principles of Econometrics, Wiley, New York, 1971.

## GLOSSARY

Amalgamation Distance. The straight-line distance in n-dimensional space connecting the centers of two clusters of n characteristics. Used to measure relative similarity of multiattribute phenomena. See Cluster Analysis.

Amalgamation Method. Specific method of combining site attribute information and estimates of attribute weights to form a composite evaluation of relative suitability. See Decision Rule.

Attribute. A physical impact or characteristic used to describe alternatives.

Attribute Value. A subjective estimate of the relative value of a measured attribute level to the suitability of an alternative.

Carryover. Transfer of attitudes and subjective judgments from one assessment to a similar, following assessment.

Categorization. A scaling and weighting method in which relative portion of things with respect to some characteristic is expressed as defined categories (high, medium, low) which are then arbitrarily assigned integer magnitudes.

Cluster Analysis. A technique for identifying groups (clusters) of multiattribute phenomena whose elements are more like each other than like elements of other groups.

Complexity. Here used to indicate numbers and magnitudes of tradeoffs required to make a choice from among alternatives. A special index is derived to express magnitude of decision complexity.

Decision Analysis. A general approach to problem solving which uses estimates of utility functions and attitudes towards risk in making decisions. See References 3 and 4 for details.

Decision Rule. Specific method of combining site attribute information to form a composite evaluation of relative suitability from which a decision can be made. See Amalgamation Method.

Diversity. Range of attribute values represented among alternatives.

Dominant. Superior in all respects.

Global Evaluation. A weighting method and a decision method in which alternatives are ranked or rated holistically, considering all attributes at once. If used as a decision method, the "best looking" alternative is selected. If used as a weighting method, multiple regression is used to calculate weights implied by holistic ratings.

Global Knowledge. Full knowledge of all attribute levels for all alternatives under evaluation.

Goal Attainment. A decision rule which seeks the solution having the smallest maximum weighted deviation of attribute levels from a set of prespecified criteria (goals).

Goal Programming. A decision rule which seeks the solution having the smallest sum of weighted deviations (or squared deviations) from a set of prespecified criteria (goals).

Impacts. Here used to mean potential negative effects were a nuclear power plant located at a candidate site.

Indifference Tradeoff. Weighting method which assigns weights by specifying the change in one attribute that is exactly equal in value to a specified change in another.

Interval-scaled. Measures (attribute values or weights) for which intervals are meaningful, but which have an arbitrary zero point, usually assigned to the smallest magnitude. Ratios among interval-scaled measures are not meaningful.

Lottery. A Decision Analysis weighting method in which evaluations are made of preferences for a "sure thing" against a lottery of known characteristics in which a win is better than the "sure thing" and a loss is worse than the "sure thing" by specified amounts.

Maximax. A decision rule which seeks the alternative that maximizes the highest level of desirable characteristics.

Metfessel Allocation. A weighting method which assigns attribute weights by allocating 100 points in proportion to their relative importance.

Minimax. A decision rule which seeks the alternative that minimizes the highest level of undesirable characteristics.

Power Law. An amalgamation method in which total suitability is estimated as the product of the relative suitability of each attribute level raised to the power of the relative importance of that attribute with respect to the other attributes under consideration.

Ranking. A scaling and weighting method which assigns integers from 1 to n indicating the positions of n alternatives with respect to relative suitability or importance.

Rating. A scaling and weighting method which assigns scores from 0 to a specified maximum (usually 1 or 10) indicating relative positions of alternatives with respect to suitability or importance.

Ratio Estimation. Weighting method which derives weights from estimates or ratios of relative importance of attributes.

Ratio-scaled Weights. Weights in which a 0 level means zero importance and for which ratios among different weights are meaningful.

Robust. Insensitive to changes of inputs.

Scaling. Converting measured levels of attributes to subjective judgments of relative significance to suitability of alternatives. See Attribute Value.

Tradeoff. A choice in which some amount of one attribute must be foregone in order to obtain a larger amount of another attribute.

Weight. Importance of one attribute relative to all other attributes included in a decision. Importance should be expressed in terms of relative willingness to trade off a unit of one attribute for a unit of another. This is not always the case.

Weighting Summation. An amalgamation method in which total suitability is estimated as the sum of the relative suitability of each attribute level multiplied by the relative importance of that attribute compared to the other attributes under consideration.

APPENDIX 1. REPRESENTATIVE SITE DESCRIPTION EVALUATED BY THE  
LONG ISLAND PANEL

Site A

Land Use. The 285-acre site is entirely owned by LILCO. Within the site boundaries are 190 acres of woodland, 15 acres of abandoned fields, 12 acres of marsh, 20 acres of beach and sand bluffs, and a 60-acre recreation camp. One hundred fifty acres of woodland (abandoned apple orchard and poor quality oak and black locust) will be leveled (1% of Suffolk County's woodland), but 115 acres will eventually be restored and planted with native Long Island flora plus a few apple trees. All 15 acres of abandoned fields will be used for construction roads and parking lots of which 5 acres will be replanted with grass after construction. The marshland will be fenced off from encroachment and left untouched. Five acres of the beach and gently sloping bluff area will be removed from its present limited recreational use during plant construction and operation, but ten acres will be donated to the town along with a public access road. The recreational area (high operational costs forced the closing) will be "suitably" relandscaped into wildlife habitat.

Total site area will occupy 50 acres. With the exception of the 10-acre beach area the remaining 245 acres will be limited access, by LILCO permission only.

The site is presently zoned commercial and one acre residential. Present land use in the vicinity of the site is agricultural.

Transmission. Seven miles of two new parallel double-circuit 345-kV lines will be needed from the site to a system-connecting substation. Seventy-five percent of the corridor will utilize existing right-of-way. Twenty-five percent of the corridor will need a new right-of-way. All of the corridor crosses through farmland. Single-pole transmission line supports will be used to minimize the number of poles or legs which interfere with land use. Impact should be minimal.

Water Quality and Marine Ecosystems. There are adequate offshore currents which will disperse thermal effluents. Biological productivity is high in the area because of a sandbar which extends almost 9,000 feet offshore. The highly productive 12-acre marsh area and its small stream also serve to increase local species diversity and productivity. Careful plant siting and marsh protection measures will prevent any harm to the marsh area.

Dredging for long outfall and intake pipes to prevent thermal impacts to the sandbar area and a barge offloading area will mean a one time loss of some crabs and mollusks. Almost 105 acres will be disturbed, but this is rather insignificant in comparison to their total area habitat. Clams and mollusks will reinhabit the dredged area over a period of years. Any beach erosion caused by the dredging or other activities will be corrected.

Terrestrial and Fresh Water Ecosystems. The 12 acres of marshland on the site will be fenced off and protected throughout the plant's lifetime. There are no uncommon, rare, or endangered species on the site property. During construction most flora and fauna, with the exception of birds and large mammals will be lost when the 150 acres of forest habitat is cleared. Given a habitat similar to this on either side of the area to be cleared and the fact that almost 23% of the County's land is in woodland, the impact will be insignificant. Wildlife habitat will be enhanced during plant operation through reforestation of 115 acres of cleared woodland, continued fallowing of 5 acres of abandoned fields, and relandscaping of the 60-acre recreation area. The entire site, with the exception of 10 acres of beach area, will have limited access.

Population and Community Features. The nearest population center has around 1,000 inhabitants year round. Growth in this area has been slow. Economic dependence is on agriculture, fishing and summer tourism. In the summer the average population density, which in winter is 135 persons per square mile almost doubles. There are 25 dwellings located within a half-mile of the proposed reactors.

Local community facilities include several marinas, a golf course, and public beaches within 1-1/2 miles. A hospital is situated two miles away, and a small-plane airstrip is 13 miles away. None of them should be affected by the plant's location.

A large majority of plant workers will commute daily. Though mitigation will be undertaken, traffic will be a problem especially in the summer.

Aesthetics. The site is an attractive setting of woods and meadows near the shore. Slightly higher elevations to the north and west are insufficient to provide screening. The major access road, a two-lane winding state highway, carries both local and tourist traffic by the site.

Retaining woodland at the site borders, landscaping the plant entrance, and using colored panels (blue and white) on the structures will reduce some of the visual intrusion.

Cost. The two-unit cost is 2.657 billion, including transmission costs.



## Site A (Summary)

### LAND USE

- 285 acres owned by LILCO
- 190 acres woodland, 15 acres abandoned fields, 12 acres marsh, 20 acres beach and sandbluff, 60 acres closed recreational camp
- 10 acres beach to be donated to town
- Zoned commercial and 1 acre residential

### TRANSMISSION

- 7 miles new line, 25% new right-of-way
- All farmland

### WATER QUALITY AND MARINE ECOSYSTEMS

- Highly productive sandbar 9000 ft. offshore
- 105 acres dredged, will reseed with clams in time

### TERRESTRIAL AND FRESHWATER ECOSYSTEMS

- 12 acres marsh fenced and protected
- 150 acres forest cleared, similar habitats nearby
- 115 acres reforested, 60 acres relandscaped

### POPULATION AND COMMUNITY FEATURES

- Nearest town 1000 persons, 135 persons per square mile
- Slow growth area
- Population doubles in summer
- 25 dwellings within 1/2 mile
- Marinas, golf course, public beaches, hospital at 2 miles, small airstrip at 13 miles
- Traffic problems in summer

### AESTHETICS

- Woods and meadows
- Some screening
- Visible from 2-lane highway carrying local and tourist traffic

### COST

- \$2,657 billion

Appendix 2  
BNL Panel Mean Weights<sup>a</sup>

METHOD	ATTRIBUTE						
	Site Cost	Land Use	Trans- mission	Water Quality	Terrestrial Ecology	Population	Aesthetics
<u>Ranking</u>							
Mean	0.07	0.19	0.13	0.21	0.16	0.17	0.06
S.D.	0.04	0.03	0.04	0.03	0.05	0.09	0.03
<u>Rating</u>							
Mean	0.05	0.18	0.13	0.21	0.16	0.18	0.09
S.D.	0.05	0.05	0.06	0.04	0.06	0.09	0.07
<u>Ratio Estimation</u>							
Mean	0.04	0.15	0.10	0.24	0.17	0.25	0.05
S.D.	0.05	0.08	0.06	0.09	0.08	0.07	0.05
<u>Indifference Tradeoff</u>							
Mean	0.08	0.18	0.11	0.19	0.20	0.16	0.08
S.D.	0.08	0.11	0.07	0.11	0.12	0.13	0.08
<u>Global Evaluation</u>							
Mean	0.005	0.38	0.15	0.34	0.003	0.39	0.26
S.D.	0.08	0.10	0.10	0.03	0.02	0.11	0.08
<u>Goal Programming (Goals)</u>							
Mean	0.06	0.14	0.13	0.17	0.23	0.16	0.11
S.D.	0.08	0.10	0.10	0.03	0.02	0.11	0.08
<u>Decision Analysis Lottery</u>							
Mean	0.08	0.18	0.11	0.19	0.19	0.16	0.08
S.D. <sup>b</sup>							

<sup>a</sup>All weights normalized to sum to 1.0

<sup>b</sup>These weights were generated with mean probabilities and therefore have no standard deviation.

APPENDIX 3

ADVISORY PANEL MEAN SCALED ATTRIBUTE VALUES

Method		Site cost	Land use	Trans- mission	Water quality	Terrestrial ecology	Popu- lation	Aes- thetics
<u>Categorization</u>								
Site A	Mean	2.63	2.81	3.25	3.38	3.13	2.69	2.81
	S.D.	0.72	0.83	0.68	0.72	1.02	0.79	0.75
Site B	Mean	4.13	1.94	4.19	3.81	1.63	2.31	2.31
	S.D.	0.72	0.85	0.54	0.54	0.89	1.08	0.79
Site C	Mean	1.63	2.44	1.75	3.06	3.19	3.19	1.56
	S.D.	0.50	1.31	0.86	0.68	1.05	0.91	0.51
Site D	Mean	1.19	2.81	2.63	2.00	2.19	3.06	2.25
	S.D.	0.40	0.83	0.72	0.82	0.91	0.93	1.06
Site E	Mean	2.63	3.81	3.00	1.81	1.56	2.56	3.00
	S.D.	0.72	0.91	0.63	0.54	0.73	0.81	0.73
<u>Rating</u>								
Site A	Mean	4.02	4.91	7.41	7.78	7.13	4.31	7.19
	S.D.	1.53	1.90	1.63	1.97	2.19	1.30	2.01
Site B	Mean	10.00	2.31	9.87	9.25	1.69	2.69	5.03
	S.D.	0	3.44	0.36	1.73	3.20	4.27	2.04
Site C	Mean	1.36	2.09	0.19	6.42	9.19	9.19	0.19
	S.D.	1.19	2.76	0.75	2.14	1.76	1.91	0.75
Site D	Mean	0	4.75	3.47	1.54	2.94	5.66	3.38
	S.D.	0	2.40	1.42	2.05	2.05	2.97	2.88
Site E	Mean	3.54	9.25	5.22	0.44	0.81	3.53	9.50
	S.D.	1.57	1.73	1.39	0.89	1.64	1.82	1.15
<u>Metfessel Allocation</u>								
Site A	Mean	16.87	20.64	24.99	27.63	26.56	16.65	23.18
	S.D.	4.93	6.05	4.13	4.75	5.06	2.94	5.27
Site B	Mean	53.55	10.38	39.31	31.38	9.44	15.38	18.41
	S.D.	19.91	9.84	8.55	8.01	5.89	10.75	4.74
Site C	Mean	9.32	15.46	6.75	23.19	36.75	29.84	10.40
	S.D.	6.10	9.45	4.42	6.89	11.02	8.02	5.66
Site D	Mean	4.89	18.30	11.63	9.81	16.50	23.61	17.39
	S.D.	5.82	6.86	4.33	5.61	4.98	5.06	6.03
Site E	Mean	15.37	35.22	17.32	8.00	10.75	14.52	30.62
	S.D.	6.07	12.39	4.19	5.13	7.03	3.19	5.03

Appendix 4  
Advisory Panel Mean Weights<sup>a</sup>

METHOD	ATTRIBUTE						
	Site cost	Land Use	Transmission	Water Quality-Ecosystems	Terrestrial Ecosystems	Population	Aesthetics
<u>Categorization</u>							
Mean	0.14	0.16	0.16	0.17	0.14	0.13	0.10
S.D.	0.05	0.04	0.05	0.03	0.05	0.05	0.04
<u>Rating</u>							
Mean	0.13	0.17	0.16	0.18	0.14	0.15	0.07
S.D.	0.10	0.06	0.07	0.06	0.08	0.08	0.04
<u>Metfessel Allocation</u>							
Mean	0.13	0.19	0.13	0.19	0.14	0.14	0.07
S.D.	0.11	0.09	0.06	0.07	0.07	0.10	0.04

<sup>a</sup>All weights normalized to sum to 1.0.

APPENDIX 5

LONG ISLAND PANEL SCALED ATTRIBUTE VALUES, FIRST MEETING, 31 MARCH 1979

Method		Site cost	Land use	Trans- mission	Water quality	Terrestrial ecology	Popu- lation	Aes- thetics
<u>Categorization</u>								
Site A	Mean	1.71	2.86	2.14	3.57	2.71	3.71	3.43
	S.D.	1.25	0.69	0.90	0.98	1.11	1.11	0.98
Site B	Mean	1.14	2.86	1.86	3.43	2.29	2.86	3.29
	S.D.	0.69	1.21	0.90	1.27	1.11	1.35	0.95
Site C	Mean	2.43	3.14	4.14	2.43	2.00	3.57	2.71
	S.D.	1.90	1.07	0.90	1.27	1.00	0.98	1.11
Site D	Mean	1.29	3.14	3.00	2.86	2.86	3.71	2.43
	S.D.	0.76	1.46	0.58	0.90	0.90	0.95	0.79
Site E	Mean	1.86	4.43	4.29	2.29	3.43	3.71	2.43
	S.D.	1.46	1.13	0.76	0.95	1.27	1.25	0.79
<u>Rating</u>								
Site A	Mean	2.76	3.57	3.43	6.00	5.86	5.30	6.57
	S.D.	1.57	3.87	3.26	3.11	3.63	1.79	2.94
Site B	Mean	0	4.86	0	7.43	1.29	4.13	8.29
	S.D.	0	4.71	0	4.43	2.63	3.02	2.50
Site C	Mean	8.57	6.57	7.86	0.43	0.43	7.14	6.00
	S.D.	3.78	3.46	3.76	0.79	1.13	4.88	4.12
Site D	Mean	1.18	4.14	5.79	7.29	8.57	6.30	4.14
	S.D.	1.00	2.85	2.34	3.50	2.44	2.98	3.53
Site E	Mean	3.37	7.86	8.29	0.86	8.00	4.00	0.86
	S.D.	2.21	3.93	2.06	1.86	2.38	5.03	1.07

Appendix 6  
 Long Island Panel Mean Weights  
 First Meeting, 31 March 1979<sup>a</sup>

METHOD	ATTRIBUTE						
	Site Cost	Land Use	Trans- mission	Water Quality	Terrestrial Ecology	Population	Aesthetics
<u>Categorization Without Global Knowledge</u>							
Mean	0.08	0.19	0.12	0.16	0.16	0.17	0.12
S.D.	0.06	0.05	0.03	0.06	0.05	0.04	0.03
<u>Rating Without Global Knowledge</u>							
Mean	0.06	0.21	0.09	0.17	0.16	0.21	0.09
S.D.	0.08	0.11	0.08	0.11	0.08	0.06	0.07
<u>Metfessel Allocation Without Global Knowledge</u>							
Mean	0.05	0.20	0.09	0.18	0.14	0.27	0.07
S.D.	0.06	0.12	0.08	0.06	0.06	0.16	0.06
<u>Categorization with Global Knowledge</u>							
Mean	0.11	0.16	0.13	0.15	0.15	0.19	0.10
S.D.	0.06	0.04	0.03	0.05	0.05	0.02	0.04
<u>Rating with Global Knowledge</u>							
Mean	0.08	0.20	0.10	0.06	0.17	0.22	0.06
S.D.	0.08	0.09	0.07	0.10	0.06	0.03	0.06
<u>Metfessel Allocation with Global Knowledge</u>							
Mean	0.05	0.18	0.09	0.17	0.16	0.27	0.08
S.D.	0.06	0.09	0.07	0.05	0.05	0.14	0.06

<sup>a</sup>All weights normalized to sum to 1.0.

APPENDIX 7

LONG ISLAND PANEL SCALED ATTRIBUTE VALUES, SECOND MEETING, 2 APRIL 1979

Method		Popu- lation	Trans- mission	Social Services	Wet- lands	Farm land	Clam beds	Aes- thetics	Site Cost
<u>Categorization</u>									
Site A	Mean	3.00	1.20	0.40	0.60	1.60	2.20	1.00	1.40
	S.D.	0.71	0.45	0.89	0.55	0.89	1.30	1.73	1.34
Site B	Mean	3.00	2.80	1.40	2.80	2.20	0.80	0.20	0.80
	S.D.	0.71	1.10	1.67	1.30	1.30	1.79	0.45	0.84
Site C	Mean	2.00	2.60	0	2.80	0	2.40	3.20	2.00
	S.D.	0.71	0.89	0	1.30	0	1.14	0.45	1.58
Site D	Mean	1.00	1.80	2.20	0.60	2.20	1.00	2.80	0
	S.D.	1.22	0.45	1.30	0.89	1.30	1.73	1.10	0
Site E	Mean	2.60	0.80	2.80	0.40	0.60	1.20	3.60	2.20
	S.D.	0.55	0.45	1.30	0.55	0.55	1.64	0.55	1.64
<u>Rating</u>									
Site A	Mean	8.60	1.20	1.80	1.00	7.20	9.20	2.00	3.60
	S.D.	3.13	0.76	2.49	1.00	2.59	1.79	4.47	3.51
Site B	Mean	6.40	10.00	4.80	10.00	9.40	2.20	1.40	2.30
	S.D.	2.04	0	3.96	0	0.55	4.38	1.52	2.86
Site C	Mean	2.80	6.40	0	10.00	0	9.80	9.60	4.90
	S.D.	1.44	2.19	0	0	0	0.45	0.89	4.51
Site D	Mean	0.60	3.40	7.20	0.80	10.00	2.20	8.00	0
	S.D.	1.34	1.47	2.28	1.20	0	4.38	1.87	0
Site E	Mean	4.60	0	10.00	0	3.60	4.60	10.00	8.00
	S.D.	1.67	0	0	0	3.85	3.29	0	4.4/

Appendix 8  
Long Island Panel Mean Weights, Second Meeting, 2 April 1979a

METHOD	ATTRIBUTE							
	Popula- tion	Trans- mission Lines	Social Service Cost	Coastal Wetlands	Farm Land	Clam Produc- tivity	Aesthetics	Cost
<u>Categorization Without Global Knowledge</u>								
Mean	0.21	0.09	0.11	0.14	0.11	0.12	0.13	0.11
S.D.	0.00	0.08	0.06	0.06	0.07	0.06	0.05	0.07
<u>Rating Without Global Knowledge</u>								
Mean	0.26	0.13	0.12	0.10	0.09	0.08	0.14	0.08
S.D.	0.00	0.11	0.09	0.09	0.09	0.10	0.12	0.08
<u>Metfessel Allocation Without Global Knowledge</u>								
Mean	0.33	0.09	0.09	0.11	0.09	0.08	0.12	0.10
S.D.	0.16	0.10	0.04	0.05	0.078	0.06	0.08	0.10
<u>Categorization With Global Knowledge</u>								
Mean	0.20	0.10	0.10	0.12	0.10	0.11	0.17	0.12
S.D.	0.05	0.09	0.05	0.07	0.07	0.05	0.10	0.11
<u>Rating With Global Knowledge</u>								
Mean	0.26	0.12	0.10	0.09	0.09	0.70	0.17	0.08
S.D.	0.03	0.10	0.08	0.11	0.10	0.10	0.15	0.10
<u>Metfessel Allocation With Global Knowledge</u>								
Mean	0.29	0.10	0.10	0.09	0.08	0.09	0.14	0.11
S.D.	0.20	0.09	0.07	0.07	0.07	0.05	0.11	0.11

<sup>a</sup>All weights normalized to sum to 1.0



Appendix 9  
 Long Island Panel Mean Weights  
 Third Meeting, 7 June 1979<sup>a</sup>

METHOD

	Popula- tion	Trans- mission Lines	Social Service Cost	Coastal Wetlands	Farm Land	Clam Produc- tivity	Aesthetics	Cost
<u>Categorization Without Global Knowledge</u>								
Mean	0.21	0.13	0.11	0.07	0.10	0.13	0.10	0.17
S.D.	0.03	0.05	0.05	0.03	0.03	0.03	0.05	0.05
<u>Rating Without Global Knowledge</u>								
Mean	0.28	0.14	0.07	0.07	0.05	0.10	0.11	0.18
S.D.	0.03	0.06	0.05	0.10	0.04	0.06	0.08	0.11
<u>Metfessel Allocation Without Global Knowledge</u>								
Mean	0.43	0.11	0.07	0.05	0.05	0.07	0.07	0.15
S.D.	0.16	0.08	0.05	0.03	0.04	0.03	0.05	0.08
<u>Indifference Tradeoff With Global Knowledge</u>								
Mean	0.10	0.10	0.19	0.00	0.00	0.00	0.01	0.59
S.D.	0.11	0.17	0.25	0.00	0.00	0.00	0.02	0.00
<u>Decision Analysis Lottery with Global Knowledge</u>								
Mean	Results too inconsistent with Indifference Tradeoff							
S.D.	to yield feasibility weights							

<sup>a</sup>All weights normalized to sum to 1.0

APPENDIX 10  
 Long Island Panel Individual Site Ranks, First and Second Meetings  
 31 March and 2 April 1979

Person	SITE																										
	A									B									C								
31 March 1979	I	II	III	IV	V	VI	VII	VIII	IX	I	II	III	IV	V	VI	VII	VIII	IX	I	II	III	IV	V	VI	VII	VIII	IX
JD	4	1	1	1	4	4	2	1	1	5	5	5	5	2	5	5	5	5	3	2	2	2	5	3	1	2	2
DH	5	5	3	3	5	3	4	3	4	1	2	2	2	3	1	1	2	2	3	3	5	5	1	3	5	5	3
FL	4	3	3	3	4	3	3	3	3	1	1	2	2	3	1	1	2	2	2	2	1	1	4	2	2	1	1
WM	2	2	2	2	4	2	2	2	2	1	1	1	1	5	1	1	1	1	3	3	3	3	2	3	4	3	3
PM	3	4	4	3	4	2	4	3	3	5	3	2	2	3	5	5	2	2	2	5	5	5	2	3	5	5	5
KR	3	5	5	5	4	3	4	5	5	1	1	1	1	5	1	1	1	1	2	3	3	4	3	2	5	3	4
JT	3	3	3	4	4	3	3	3	3	1	1	1	1	5	1	1	1	1	3	4	4	3	2	3	4	2	4
Group Mean Weights	3	2	2	3	1	3	2	2	2	1	1	1	1	2	1	1	1	1	4	4	4	4	3	4	4	3	4
2 April 1979	I	II	III	IV	V	VI	VII	VIII	IX	I	II	III	IV	V	VI	VII	VIII	IX	I	II	III	IV	V	VI	VII	VIII	IX
JD	2	3	2	2	4	2	2	2	2	5	5	5	5	5	3	4	5	5	3	3	3	3	5	4	4	3	4
DH	2	3	5	5	4	2	4	5	5	4	4	4	4	3	3	3	4	4	5	5	2	2	5	5	5	2	2
FL	2	3	2	2	1	2	2	2	2	3	4	4	5	3	3	3	4	3	4	2	3	3	4	5	5	3	4
WM	2	2	2	2	4	2	2	2	2	4	5	5	5	3	3	4	4	4	3	3	3	3	2	4	3	3	3
PM	4	2	4	4	5	5	4	5	5	5	5	5	5	3	3	4	4	4	3	4	2	2	2	2	2	2	2
Group Mean Weights	2	2	2	4	4	2	2	2	2	3	5	5	5	2	3	4	5	5	5	4	3	2	3	5	5	3	3

APPENDIX 10 (Continued)

Person	D									E								
	I	II	III	IV	V	VI	VII	VIII	IX	I	II	III	IV	V	VI	VII	VIII	IX
31 March 1979	1	4	4	4	1	4	4	4	4	2	3	3	3	3	2	3	3	3
JD	1	4	4	4	1	4	4	4	4	2	3	3	3	3	2	3	3	3
DH	4	4	4	4	1	5	3	4	5	2	1	1	3	3	2	2	1	1
FL	2	4	4	4	2	3	4	4	4	5	5	5	1	3	5	5	5	5
WM	5	5	5	5	3	4	5	5	5	4	4	4	1	5	4	3	4	4
PM	1	1	3	4	5	1	3	4	4	4	1	1	1	3	4	1	1	1
KR	4	4	4	3	2	4	3	4	3	5	2	2	1	5	5	2	2	2
JT	2	5	5	5	3	2	5	5	5	5	2	2	1	5	5	2	4	2
Group Mean Weights	2	5	5	5	4	2	5	5	5	5	3	2	5	2	5	3	4	3
2 April 1979	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
JD	3	3	4	4	3	4	4	4	3	1	1	1	1	2	1	1	1	1
DH	1	1	1	1	1	1	1	1	1	3	2	3	3	2	4	2	3	3
FL	1	1	1	1	2	1	1	1	1	5	5	4	4	5	4	4	5	5
WM	1	1	1	1	1	1	1	1	1	5	4	4	4	4	5	5	5	5
PM	1	1	1	1	1	1	1	1	1	2	3	3	3	3	3	4	3	3
Group Mean Weights	1	1	1	1	1	1	1	1	1	3	3	3	3	4	4	3	4	4

---

APPENDIX 10 (Continued)

---

- I - Categorized attribute values and weights, without global knowledge
- II - Categorized weights applied to rated attribute values without global knowledge
- III - Rated attribute values and weights without global knowledge
- IV - Metfessel Allocation weights applied to rated attribute values, without global knowledge
- V - Global evaluation
- VI - Categorized attribute values and weights, with global knowledge
- VII - Categorized weights applied to rated attribute value, with global knowledge
- VIII - Rated attribute values and weights, with global knowledge
- IX - Metfessel Allocation weights applied to rated attribute values with global knowledge

APPENDIX 11  
 Long Island Panel Individual Site Ranks, Third Meeting  
 7 June 1979.

Person	SITE																								
	A					B					C					D					E				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
DH	4	3	5	3	-	2	4	4	2	-	5	5	2	4	-	1	1	1	1	-	3	2	3	5	-
FL	2	2	2	3	-	3	3	3	2	-	4	4	4	4	-	1	1	1	1	-	5	5	5	5	-
WM	3	3	4	3	-	4	5	5	4	-	2	2	2	1	-	1	1	1	2	-	5	4	3	5	-
PM	4	5	5	2	-	5	4	4	3	-	3	2	2	4	-	1	1	1	1	-	2	3	3	5	-
Group Mean Weights	2	3	5	3	-	4	4	4	2	-	3	2	2	4	-	1	1	1	1	-	5	5	3	5	-

- I - Categorized weights applied to rescaled group mean rated attribute values, with global knowledge
- II - Rated weights applied to rescaled group mean rated attribute values, with global knowledge
- III - Metfessel Allocation weights applied to rescaled group mean rated attribute values, with global knowledge
- IV - Indifference Tradeoff weights applied to rescaled group mean rated attribute values, with global knowledge
- V - Decision Analysis - no feasible weights

RECEIVED BY TIC JAN 12 1981