CONF-9610207--2

SAND96-2428C

SAND--96-2428C

GREY ANALYSIS

Dr Gary D Cable*
Sandia National Laboratories
Albuquerque, NM, USA

RECEIVED
NOV 06 1996
OSTI

Abstract

A summary of the basic concepts of grey analysis is provided, with descriptions of its application to several classes of problems. Calculational methods are provided for grey relational analysis, and for modeling and prediction using grey methods.

1. Introduction

Grey logic is not another name for fuzzy logic. Grey logic -- also called grey analysis or grey system theory -- is a new technology, a group of techniques for system analysis and modeling. Like fuzzy logic, grey logic is useful in situations with incomplete and uncertain information. Grey analysis is particularly applicable in instances with very limited data and in cases with little system knowledge or understanding.

Grey analysis was invented by Professor Deng Julong of the Huazhong University of Science and Technology in Wuhan, China. Though it is almost unknown in the West, grey analysis has been applied in China to a wide variety of problems. Successes have been claimed in areas as diverse as agriculture, ecology, economic planning and forecasting, traffic planning, industrial planning and analysis, management and decision making, irrigation strategy, crop yield forecasting, image analysis, historical analysis, military affairs, target tracking, propulsion control, communications system design, geology, oil exploration, earthquake prediction, material science, process control, biological protection, epidemiology, environmental impact studies, medical management, and the judicial system.

According to grey analysis' proponents, different grey analysis elements are applicable to all systems.

* This work was supported by the United States Department of Energy under Contract DO-AC04-94AL85000. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

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.

Grey relational analysis is an improved method for identifying and prioritizing key system factors, provides a straightforward mechanism for proposal evaluation, and is useful for variable independence analysis. Grey models provide new tools for forecasting which are particularly helpful in circumstances where system complexity prevents a more complete treatment, where systems are incompletely understood, or where system operational data are limited.

Are the successes claimed real? Is grey analysis a valuable new technology? Or is it something else -- something applicable to classes of problems not yet clearly defined? Final answers to these questions are not available, and this paper will not attempt to answer them. What it will do is provide a summary of the basic concepts of grey analysis, and descriptions of its application to several classes of problems.

2. Grey Concepts

Grey analysis is applied to systems, and most particularly to the information in the system output values. It uses fairly basic definitions for both systems and information. It draws conclusions about the nature of system solutions.

The most basic thing about systems is that they have inputs and outputs. These inputs and outputs are related by the system's structure and characteristics. Their behaviors are tied together by system parameters and relations. All systems have these elements. Analysis of these elements can be used for system evaluation and for system control -- whether the system and its elements are understood or not.

Grey analysis is applied to systems or, more particularly, to the most important system output which is information. Each element of information has a value in which all the system characteristics are embedded. The approach used is to apply grey analysis techniques in a deductive process, evaluating the explicit forms, the output values, to determine the system characteristics or implicit forms.

Grey analysis uses a specific concept of information. It defines situations with no information as black, and those with perfect information as white. Neither of these idealized situations ever occurs in real problems; real (non-ideal) situations vary from dark to bright. Situations between these extremes are described as being grey, hazy, or fuzzy. (Hazy and fuzzy, as used in grey analysis contexts, have this meaning rather than the more specific meanings usually given them in hazy set theory and fuzzy logic.)

With this definition, information quantity and quality form a continuum from a total lack of information to complete information -- from black through grey to white. Since uncertainty always exists, one is always somewhere in the middle, somewhere between the extremes, somewhere in the grey.

Grey analysis then comes to a clear set of statements about system solutions. At one extreme, no solution can be defined for a system with no information. At the other extreme, a system with perfect information has a single, unique solution. In the middle, grey systems have a variety of available solutions. Grey analysis does not attempt to find the right solution, or the best solution, but does provide techniques for determining a good solution, an appropriate solution for the system.

3. Grey Relational Analysis

Grey relational analysis is used to assist in the determination of a system's key factors, and in the identification of system factor correlations. It is described as another approach, as "another tool in the toolbox," as an alternative to statistical and fuzzy analyses. The basics of this approach are described below.

The set of system output values -- for the overall system or for each system output -- is a countably infinite set, and may be a finite set; this set can be handled as a time sequence. The values for each input factor, like the output values, can be treated as a sequence. The input and output sequences are compared to identify the system's key factors.

In most cases, system values must be processed -- made compatible -- before they can be compared in any reasonable manner. All input and output factors must be represented in numeric forms. Once this is done, factor compatibility preparation is accomplished by means of three operations: orientation, normalization, and scaling. Orientation involves putting all factors into compatible forms. The sequences must be oriented so that larger values are better, or so that larger values are worse, for all of the sequences. In addition, all factor sequence values must be positive definite. Normalization involves putting all factors into compatible units. The use of dimensionless measures is preferred where these are available. Scaling is getting all factors into compatible ranges. Compatible ranges are not necessarily identical ranges, though most practitioners show a preference for factor values in a zero to one numeric range.

The greatest flexibility is available in the scaling process. Several types of scaling are described in grey analysis materials. Values may be scaled by a sequence's largest or smallest value, or by a "target" (best) value; these three methods are often called "effective measure" methods. Values may also be scaled by a sequence's first or average value, or by an interval value. Scaling by an interval value is reported to have been used often with geological data.

Once the input and output sequences are made compatible, sequence comparisons can begin. The comparisons are made as follows: An "ideal" or "reference" sequence $x_o = \{x_o(k);\ k=1,2,3,\ldots,m\}$ is identified; this sequence is most commonly the sequence of primary output values for system factor analyses, or may be an idealized or noiseless output sequence. Each of the other value sequences x_i is numerically compared to the reference sequence by computing each sequence's grey relational grade as described below. The sequence x_i which best matches the reference sequence x_o will have highest relational grade value Γ_i.

More generally, the highest values of Γ_i belong to the most important system sequences, defined as those sequences whose behavior is most closely correlated with the system's output behavior. Lower Γ_i values are associated with sequences which are less important to the system's output behavior. Correlated sequence numerical behavior is taken to indicate a causal relationship between the sequences, or a causal relationship between both sequences and some common causal factor.

Looked at another way, the set of values Γ_i form a vector in the grey relational space Γ. Variations in system behavior result in variations in the grey relational vector. The solid figure swept out by the grey relational vector for various system value sequences defines the range of the system's behavior in response to the defined system inputs. This solid figure will be close to a single line if a complete set of system inputs has been identified. A substantially more solid figure indicates that one or more system input factors may not have been identified or included in the system evaluation.

An individual grey relational grade showing the correlation between sequences x_i and x_j is $\Gamma_{ij} = \Gamma(x_i, x_j)$. In a similar manner $\Gamma_i = \Gamma(x_o, x_i)$ shows the correlation between the factor sequence x_i and the defined reference sequence x_o. Γ_i, in turn, is composed of grey relational coefficients $\gamma_{oi}(k) = \gamma(x_o(k), x_i(k))$ as

$$\Gamma_i = \Gamma(x_o, x_i) = \Sigma_k \gamma_{oi}(k) = \Sigma_k \gamma(x_o(k), x_i(k))$$

where

$$\gamma(x_o(k), x_i(k)) = \frac{\Delta_{min} + \zeta \Delta_{max}}{\Delta_{oj}(k) + \zeta \Delta_{max}}$$

$$\Delta_{oi}(k) = |x_o(k) - x_i(k)|$$

$$\Delta_{min} = \min_i\{\min_k\{\Delta_{oi}(k)\}\}$$

$$\Delta_{max} = \max_i\{\max_k\{\Delta_{oi}(k)\}\}$$

The symbol ζ represents the equation's "contrast control," sometimes also be referred to as the environmental coefficient or the distinguishing coefficient. This coefficient is a free parameter. Its value, over a broad appropriate range of values, does not affect the ordering of the grey relational grade values, but a good value of the contrast control is needed for clear identification of key system factor(s).

When appropriate, weighting factors β can be used with the sequence values. With these factors, the calculation of the grey relational grades takes on the more general form

$$\Gamma_i = \Gamma(x_o, x_i) = \Sigma_k \beta_k \gamma_{oi}(k) = \Sigma_k \beta_k \gamma(x_o(k), x_i(k))$$

where

$$\Sigma_k \beta_k = 1$$

The different sequence value weights β_k can be specified from experience, or appropriate weights can be computed by processes such as singular value decomposition using preliminary grey relational grade values. Weighting factors can be extremely valuable in the right circumstances. In the absence of information to the contrary, however, equal weights should be used. One should note that the use of weighting factors is not equivalent to changes in the sequence value units used or the choice made for sequence normalization.

It should be noted that no information about the system itself is used or required in this process; one works solely with the input and output value sequences. This suggests that grey relational analysis is most useful when little is known about the operational details of the system being evaluated. In the absence of other information, the grey relational grades can be used to identify system factor relationships. These relationships can then be used to suggest appropriate modeling approaches, or used in other system analyses.

4. Variable Independence Analysis

A closely related application is used to determine whether or not various system factors are actually independent of each other. Instead of a single grey relational vector, a full matrix of grey relational values is computed. Each value Γ_{ij} describes the correlation between sequences x_i and x_j, and thus between system factors i and j. Examination of the values in each vector of this matrix, treating each factor's sequence in turn as the "ideal" sequence, allows one to identify correlations between factors in the same manner as correlations with the ideal sequence were used above to identify key system factors. Again, the highest Γ_{ij} values identify the most closely related factors. High relational grades between input factors may indicate factor interdependence, suggesting the factors are not truly independent. Indications of the presence of dependent factors should be separately and carefully evaluated.

5. Proposal Evaluation

Proposal evaluation using grey analysis is simply a special case of grey relational analysis. In this instance, the value sequences used are not system input and output value time sequences as they were in the system evaluations described above. Instead, the value sequence associated with each proposal is composed of the values corresponding with the proposal evaluation criteria. For example, the criteria for a road building project might include cost, material strength, amount of steel reinforcement to be used, performance schedule, pollution control, labor force qualifications, workmanship quality, and so forth.

As is indicated by this partial list of criteria, some preliminary effort may be needed before the processes of orientation, normalization, and scaling can be accomplished. Elements such as workmanship quality may not normally be expressed in numeric terms, as they must be for use in mathematical processes. If quality is rated on a qualitative scale (*e.g.*, poor, fair, good, better, best), these ratings can be converted into numeric forms on a five point scale -- either one to five or five to one, on a scale of zero to one, or any other scale that might be appropriate. The scales used need not be linear. As noted above, the values used must be positive definite.

Once all factors are expressed in numeric terms, they must be properly oriented -- oriented in a common direction so that larger values are worse (or better) for all factors. If cost is to be minimized, for example, pollution control might be better expressed in terms of the amount of pollution released rather than in terms of the quality of pollution control, workmanship quality might be expressed in terms of the number of quality problems found, and so forth. Ideal sequence values for proposal evaluation are the appropriate target values, where they can be defined, or selected values from among the proposal data. If cost is to be minimized, for example, the ideal cost value might be the smallest cost value from among the set of proposals.

Depending on the circumstances, it may be desirable to verify the independence of the factors used. In the example given, material strength may or may not be independent variables, depending on whether the amount of steel reinforcement added to the road material is the determining factor in material strength. It must be stressed that variable independence as used here is not a physical property, even though it may seem so from this description. These factors may be independent within the context of the proposal evaluation, even if they are physically dependent factors.

Proposal evaluation is one of the places where the use of weighting factors is appropriate. Project cost may be the most important factor, for example, which would be reflected in its being given the largest weight. A factor like material strength or usable life may also be quite important and have a large factor weight. Other factors may be desirable but less critical, and have smaller weights.

With this preparation, a grey relational analysis is performed just as described above. The highest grey relational grade values are associated with those proposal value sequences with the best match to the ideal sequence of desired factor values.

Depending on circumstances, especially if several of the top-rated proposals have very similar grey relational grades, performance of a sensitivity analysis may be desirable. Calculation of grey relational grades with reasonable variations (as determined from project characteristics and criteria) in factor normalizations, scaling, and weighting factors will demonstrate whether these variations produce any reordering of the proposal rankings. No further considerations are necessary if no reorderings are produced among the highest ranked proposals.

The relatively mechanical nature of this process may be a particularly desirable characteristic when large numbers of proposals must be evaluated and when separate subteams are performing portions of the evaluations. Predefined weighting factors and value ranges can support, for example, the seamless combination of proposal technical and cost evaluations. This process also provides a clear evaluation audit trail, which may prevent challenges to subsequent contract awards.

6. Grey Modeling and Prediction

We now turn to a completely different kind of grey analysis, related to the types of analysis discussed above only through the basic grey concepts. Grey relational analysis and its applications make use of system input and output information, and of the analyst's knowledge of how the system works. The primary process used is the determination of value correlations.

In grey modeling, by contrast, very little system information is used. No information is used about system input value sequences, or about the mechanics of how the system works. The only system information used is the sequence of output values. Input and system information is replace solely by the assumptions that (1) the output value sequence results from the operation of a system, and (2) system parameters are not changing radically but are fixed or slowly varying. The output values are processed in a fixed, standard manner that does not depend on the system being modeled.

The sequence of output values can be thought of as the system's output signal. Grey modeling, applied to the value sequence, is similar to other varieties of signal processing. As in signal processing, the analysis concentrates entirely on the data values without considering the data source. With this comparison, it should be no surprise that the grey modeling was developed in a signal processing environment at an engineering college..

The numerical modeling process requires the use of one or more Accumulated Generating Operations (AGOs) to transform the system output data into a generated state, modeling the numeric sequence in a standard form, using the derived model to project needed values and values corresponding to the output data generated state, and the using Inverse Accumulated Generating Operations (IAGOs) to produce model output values in the original units. Model output values provide predictive capabilities based solely on the behavior of the system as reflected in the system output values. System and model output data values can also be used together in error analyses, either for evaluation of the accuracy and quality of the grey models or for other applications as described in the following section.

This modeling process is logically like the use of Fourier analysis. The use of one or more AGOs corresponds to the use of a Fourier transform. The AGO operation itself is very much like integration. The modeling and analysis performed with the generated state data values is conceptually just like performing the comparable operations with the transformed data. The use of the inverse AGOs is exactly like the use of inverse Fourier transforms to move the data back into their original coordinate space.

The application of an Accumulated Generating Operation (AGO) to original system data to produce the first generated data state, and the similar transformation from the r^{th} to the $(r+1)^{th}$ generated state, are represented as

$$x^{(0)} \rightarrow AGO \rightarrow x^{(1)} \qquad x^{(r)} \rightarrow AGO \rightarrow x^{(r+1)}$$

The corresponding inverse transforms using model output data may be represented as

$$x'^{(1)} \rightarrow IAGO \rightarrow x'^{(0)} \qquad x'^{(r+1)} \rightarrow IAGO \rightarrow x'^{(r)}$$

Each state is a sequence of system data values, which may be represented for the initial state as

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) = (x^{(0)}(k); k=1,n)$$

As is evident from these representations, the k^{th} sequence value in the original system data state is represented notationally as $x^{(0)}(k)$; the corresponding model output value is $x'^{(0)}(k)$.

This identifies how each data state is represented. The next element needed is the meaning of the AGO and of each of these states. The k^{th} value of a generated state is the sum of the first k values of the prior state. The computation of the k^{th} value of the first generated state from original data values is thus

$$x^{(1)}(k) = \Sigma^{k}_{i=1}\, x^{(0)}(i)$$

For example, $x^{(1)}(1) = x^{(0)}(1)$, $x^{(1)}(2) = x^{(0)}(1) + x^{(0)}(2)$, $x^{(1)}(3) = x^{(0)}(1) + x^{(0)}(2) + x^{(0)}(3)$, and so forth. The IAGO is then represented as

$$x^{(0)}(k+1) = x^{(1)}(k+1) - x^{(0)}(k)$$

Most immediate applications of grey modeling make use of a grey model of first order with one parameter, represented notationally as GM(1,1). In the generated data space, this model satisfies the differential equation

$$dx^{(1)}/dt + ax^{(1)} = b$$

The corresponding difference equation is

$$d(k) + az^{(1)}(k) = b$$

where

$$z^{(1)}(k) = (x^{(1)}(k) + x^{(1)}(k-1))/2$$

$$d(k) = x^{(1)}(k+1) - x^{(1)}(k) = x^{(0)}(k)$$

Thus the difference equation, for each value of the index k, becomes

$$x^{(0)}(k) + az^{(1)}(k) = b$$

The resulting (n-1) simultaneous equations are solved for the coefficients a and b. A least squares method is most commonly used by grey analysis practitioners. In terms of these coefficients, the solution for the model data points is

$$x'^{(1)}(k) = ce^{-ak} + b/a$$

where

$$c = [x'^{(1)}(1) - b/a]\, e^{a}$$

This produces a solution in the generated model data space as

$$x'^{(1)}(k+1) = [x^{(0)}(1) - b/a]e^{-ak} + b/a$$

Use of the IAGO and simplifying produces a solution in the model data space as

$$x'^{(0)}(k+1) = (1-e^{-a})[x^{(0)}(1) - b/a]e^{-ak}$$

Some systems cannot be modeled to use a single parameter, no matter what degree of situation and system oversimplification is used. For such cases, there are available mathematical extensions of the grey models. These are the GM(M,N) models -- grey models of order M with N parameters. The mathematical support for these models is under development. Thus far, some applications have been found for the GM(1,N) first-order N parameter models (usually with N being 2 or 3). No applications have yet been found that require the use of second- or higher-order models.

As normally occurs with any new methodology, grey modeling was applied first to the most basic kinds of systems and processes. Modeling uses and projects simple data sequences from oversimplified versions of extremely complex systems. Projections of agricultural yields are used to manage operation of the supporting irrigation systems, and projections system utilization and demand inform management decisions for medical and judicial service systems.

The grey modeling process, as currently practiced, does have some operational limitations. Like the rest of grey analysis, grey modeling works best where other approaches fail -- in circumstances with little information. Operationally, the best results are obtained from models generated using four to six system data points. The use of too many Accumulated Generating Operations also worsens the accuracy of grey model results, probably by transforming the generated state data values into a super-exponential condition.

7. Grey Model Applications

The simplest applications for grey models are generalized systems with single output values. The best such applications are those single output systems for which all the appropriate inputs and operations are not known and/or for which the ways in which the systems operate are not understood. These are the systems which cannot be modeled by other means, either because insufficient information is available or because the system mechanics are too ill-known or too complex to be modeled.

Prime examples of this kind of systems and data are crop yields, annual rainfall amounts, medical system demand, and legal system demand. In each case, the data sequences are regular -- aggregate rainfall per year, crop yield for a given region per year or per growing season, and number of civil or criminal cases filed per year or per quarter. System and input variations, generally ill-defined in their behavior and impacts, cause variations in the data values. Modeling of the rainfall, yields, and demands allows for reasonable projection of future values.

Other applications model and project time variations of events. Examples of particular interest to governmental organizations include projection of the times of the first frost, or the last freeze, or the first snow or severe rainstorm of the season. As in crop yield projection, the data values are sequenced by year. In this case, however, the data values are the date of the event in question. An easy representation for each date is the date's day number within the year (1-365). The grey model then projects when these events may be expected to occur in coming years.

Grey models have also been used to predict irregular -- and extremely irregular -- event occurences. As one example, the same sort of information used to project seasonal rainfall can be used in a different way to project the occurrence of unusually wet or dry years. The years with rainfall above some high level, or below some minimum level, are gathered together. The year numbers (usually relative to some baseline, as years after some starting point) become the data values, indexed by their order number in the data sequence. Data value modeling provides for projection of the next year(s) which will have similarly high (or low) rainfall.

A similar type of case, though more extreme and irregular, appears in earthquake prediction. Again, since significant earthquakes occur quite irregularly, the data values are indexed by their sequence numbers in the data series. The data values are the dates (year and decimal fraction relative to some starting point) of the earthquakes in the selected region. The grey modeling process provides for projection of when the next earthquake(s) might be expected to occur. The earthquake magnitude does not come directly from the model -- only the earthquake dates are included in the data modeled, just as the amounts of rainfall in the extremely wet or dry years were not included in the data modeled above. In a similar manner, however, a data sequence may be prepared for earthquakes above a selected strength (magnitude) and a projection made for the estimated date of the next similarly large earthquake in that region.

The variety in these examples shows that grey modeling can be applied to extremely wide ranges of systems. Catastrophic processes can be modeled as reasonably as regular processes; drought year occurences and earthquakes can be projected as reasonably as crop yields and medical demand. This flexibility and the limited data requirements are grey modeling's greatest strengths.

8. Grey Error Analysis

Error analysis using grey models is a new, developing subfield of grey analysis. Grey error analysis is being applied in two very different ways.

One application is the one most likely to occur to model users -- evaluating the accuracy of the models. Here the model data values are compared with the original system data values from which the model was derived to determine how good a fit the model is to its supporting data. More important, or course, are comparisons of model projections with actual data. Both of these error evaluations are directed toward the determination of how well and how accurately the grey model projections match system behavior.

A completely different type of application comes from looking at the error behavior in the light of the first axiom of information: "A difference in information is also information." Here models project approaching data values. The differences between the projected and measured values are not interpreted as errors, but as indicating some change in the system or process being measured. These differences -- the "errors" -- become the primary information sought and used. This process is currently being used in image analysis edge detection, and in location and identification of likely oil-bearing structures.

9. Decision Support

The application of grey analysis to decision support is straightforward once the appropriate modeling support has been prepared. Standard decision support tools are used in very nearly the usual manner. The primary difference is the use of grey models to handle the uncertain elements. Essentially, then, grey models are used to project the needed future values.

The clearest decision support application is in linear programming. Particularly useful applications include resource allocation for maximizing production levels or profits, and management of service organizations using demand and utilization projections.

10. Summary

Grey analysis provides additional analysis tools to assist in the evaluation of systems and situations characterized by limited, missing, and/or uncertain information. It is not at all the same thing as fuzzy logic, but may be used along with fuzzy logic as well as chaos theory and other techniques as appropriate for particular systems and evaluations.

Grey analysis is especially useful when the complete set of factors involved in the system's behavior is unknown or unclear, when the relationships of system factors to the system's behavior and inter-relationships among factors are uncertain, when system behavior is too complex to determine (or model or compute) completely, or when only limited information on system behavior is available. It also provides a means of evaluating competing possibilities, proposals, theories, or models against various types of criteria.

Grey analysis can be used to determine solutions for systems under evaluation. Generally, however, it will not provide a unique solution. This is both because the lack of a unique solution is axiomatic in grey system theory, and because modeling ambiguities inherently produce multiple possible solutions. As in other areas of mathematics, not all developed solutions will be physically reasonable. For this reason, grey analysis practitioners say all solutions should be put through appropriate "reality checks." Nevertheless, even at their least useful, grey analysis solutions can assist in identifying missing information and additional data requirements.