SIMULATION OF SOCIO-ECONOMIC CONDITIONS IN A CANADIAN RIVER BASIN USING CROSS IMPACT ANALYSIS

Anand S. Aggarwal and Gerald E. Brown
Water Planning and Management Branch
Environment Canada, Ottawa, Canada

Summary

KSIM, a "cross impact procedure" developed by Kane and his associates at the University of British Columbia, Canada, provides a way of incorporating "soft subjective variables" into a mathematical simulation technique which attempts to by pass the lack of data and complex mathematical structures found in many problems in the socio-economic and environmental fields. The approach gives "knowledgeable" but mathematically unsophisticated people an opportunity to participate actively in model formulation by specifying and defining variables, and gives an insight into the expected performance of each variable and their interrelationships.

This paper explores the possibility of extending this basic "cross impact analysis" to aid in quickly simulating and forecasting problems for which a sophisticated mathematical model may be planned. In this way, an insight into the significance of each variable (both endogenous and exogenous) can be studied before a major modelling effort is undertaken. Alternatively, this extension to the basic KSIM approach can be used in simulating problems for which some data may exist but for which a fully developed model may be impossible or unjustified.

A number of shortcomings were identified in KSIM when the method was applied to problems having a mix of subjective/objective variables. Most of these have been overcome and the modified "cross impact approach" was employed to simulate and forecast socio-economic variables in a Canadian river basin. The results compare favourably with long term forecasts made using an Input/Output model of the basin's economy.

Introduction

The success of any simulation or modelling attempt depends heavily on the experience, knowledge and expertise of the people who formulate the problem, who build the model and finally who ultimately use the model. Traditionally, the variables chosen are such that they can be clearly defined and expressed mathematically to form a system of equations. The coefficients in the system of equations are calibrated or adjusted based on numerical data collected over years. A serious difficulty arises when one moves into the realm of the social, life and environmental sciences and other similar areas. It is often impossible to quantify some parameters or to describe the problem mathematically. There may also be virtually no numerical data available for calibration and past trends may not continue because of technological innovations and other factors. This means that the usual structured models built for simulation and forecasting can not be built. The existence of these soft or subjective variables is not really confined only to these areas because in economics and engineering models, attempts to handle soft variables takes the modelling process into the field of lagged variables and other complicated modelling procedures. The purpose of this paper is to discuss very generally this problem of soft variables, to look at one approach which shows some promise, and to show how it can be modified to handle problems having a mix of hard and soft variables with limited amounts of data.

In attempting to deal with simulation problems having only soft variables and no data, Kane and his associates (1,2,3 and 4) recognized that by requiring an explicit statement of the variables and the relationship between them a basic forecasting simulation is then possible. To develop this possibility and to overcome the problems of mental models of soft variables, Kane developed a cross-impact modelling method which utilizes expert opinion and knowledge to first clearly define the variables which contribute to the model and then, using a matrix format, obtain a statement of the relationship between each pair of variables. These relationships are the impact of one variable on another. In completing the matrix, Kane asks the question "If variable A changes positively, will variable B change positively also or will its change be negative with respect to the first variable?" The answer to this question is then given in terms of an arbitrary scale of intensity say from 1 to 3 so that the relative magnitude and sign of the change between variables can be tabulated in the matrix.

The cross-impact matrix of these simple interrelationships between the variable represents the total information input into the model. Although this information is very simple, it does reflect the expected changes between each pair of variables or, if combined in some logical way, should give a reasonable indication of how the variables will change in time. In a latter section, an example is given which shows the validity of this statement.

The forecasting or projection equation which Kane has developed involves a rather arbitrary procedure for totalling the positive and negative impacts and generating a forecast from one time interval to another. To increase the information content in his simulations, Kane has included a second matrix and a method of modifying the impacts within the matrix.

Potential Applications

The ability to utilize expert opinion in a non-structured format has potential use in at least four clearly definable fields. In the social, life and environmental sciences mentioned above, the dominate variables are soft and subjective. These variables are hard to quantify and usually there is little or no data available. The development of simulation models here has been limited by the unavailability of practical methods of handling such variables.

The second area is where the model objectives are less stringent and the unstructured model can replace the more structured and complex mathematical model. In this situation the hard variables might dominate and numerical data could be available.

The use of non-structured simulation approaches to modelling can be useful in building a serious structured model. By using an unstructured model, the importance of the various variables and data can be tested before the final model is built. In this way, many of the problems encountered in building the structured model might be eliminated. The quick response of this method to produce projections allows the model builder to explore more fully which variables are important and how they interact.

The fourth and most important application is the combining of the structured model with the unstructured one. The advantage here is that many of the factors added to mathematically sophisticated models for calibration and adjustment are often added to handle the soft variables associated with the problem. The inclusion of soft variables at the beginning of any serious modelling undertaking should produce a model having few calibration and adjustment problems.

To test the feasibility of using an unstructured format, a model of a Canadian river basin having hard variables and numerical data was built using a modified version of Kane's cross-impact model, KSIM. The particular river basin chosen had recently been subjected to a formal economic study so that the 50 year forecast generated by the modified KSIM could be compared to these results. The economist who produced the original study also provided the expert opinion for the cross impact matrix.

Kane's Cross-Impact Simulation: KSIM

Kane based the development of his simulation method on the idea that the growth and decay patterns of many ecological, biological and economic processes are sigmoidal in nature. In other words, growth of these variables begins slowly, passes through a period of relatively rapid growth and finally slows again as the upper bound on growth is approached. To achieve this characteristic in his simulation model Kane adopted the following projection equation.

$$x_i(t+\Delta t) = x_i(t)^{P_i(t)}$$

where $P_i(t)$ in its simplest form is

$$P_i(t) = \frac{1 + \frac{1}{2}\Delta t \sum_{j=1}^{N} [|a_{ij}x_j| - a_{ij}x_j]}{1 + \frac{1}{2}\Delta t \sum_{j=1}^{N} [|a_{ij}x_j| - a_{ij}x_j]}$$

where a_{ij} are matrix elements giving the impact of x_j on x_i or in other words

$$P_i(t) = \frac{1 + \Delta t \;|\text{Sum of negative impacts on } x_i|}{1 + \Delta t \;|\text{Sum of positive impacts on } x_i|}$$

When the negative impacts are greater than the positive ones, $P_i>1$ and x_i decreases. Similarly when the reverse is true, $P_i<1$ and x_i increase. If $P_i=1$, x_i remains constant.

To achieve a uniform sigmoidal form for all variables, Kane uses a 0 to 1 scaling where 1 represents the maximum value of variable x_i. The initial conditions of each variable is proportional to its starting value at its maximum or a fraction between 0 and 1.

Reference is made to the bibliography for a more detailed description of Kane's KSIM. For the purpose of this paper, the cross-impact matrix and the above equations are adequate.

A Modified Approach to KSIM

A number of modifications to Kane's approach were made so that the method could be applied to the river basin problem described earlier. The first modification was the introduction of a weighting function to improve the overall homogeneity of the cross-impact matrix. Normally, the matrix is completed column by column; however, a difficulty arises in that although the impact of variable x_i on each x_j are based on the same scale of relative values, there is a tendancy for this scale to change as each column is completed. This means that if each x_j variable has a different level of importance within the model, this relative importance is lost to a large degree in completing the cross-impact matrix. The matrix loses a great deal of the homogeneity required for good projections.

To help improve this situation, a vector of variable weights was introduced so that when each column of the matrix are multiplied by the corresponding element in the row vector the overall homogeneity of the matrix is improved. Ideally this should not be necessary but such a correction was found necessary.

A further difficulty which was found when applying KSIM to the river basin problem was the 0 to 1 scaling. This scaling means that all variables are indistinguishable from each other. Minor variables appear equal to important ones. To overcome this problem and allow the introduction of exogenous variables the 0 to 1 scaling was converted to numerical values, outside the model. In this way the projections made for the river basin are in terms of actual growth dollars. Exogenous variables are handled within the model by converting them to the 0 to 1 scaling.

This modification identifies a further difficulty in that the maximum value of each variable must be known so that the initial value between 0 and 1 can be determined. In most cases this maximum long term value is unknown and may in fact be the desired output from the simulation model. This problem represents a definite limitation to this application of unstructured modelling. In the river basin example this was overcome in a very arbitrary way.

Model Development

A computer program was first developed to implement Kane's basic simulation approach. The number of variables, impact matrix, initial values, time and stop time are read in; the impacts computed using the equations developed in Kane's papers and the output displayed graphically. This program was subsequently expanded to incorporate the modifications discussed above. The first modification was included by making a provision for supplying a vector of weighting values specifying the relative importance of variables between themselves. The problem of including data for exogenous variable in the model was overcome by revising the program to read in the numerical values for up to five variables. The introduction of the exogenous time series data meant that the time increment in the model equalled the period of the exogenous data. A new factor was included for calibration. Provision was made to constrain limitless growth of the desired parameters and allow the projections to commence at the suitable point on the sigmoidal curve. The actual data rather than values between 0 and 1 were printed using conversion factors. The program was implemented on an IBM/370 computer and a copy may be obtained by contacting either of the two authors.

Simulation Results

The modified approach was employed to simulate and forecast socio-economic variables in a Canadian river basin. The 29 major economic activities were aggregated into the following five groups:

(i) Agriculture: This group includes fruit trees, livestock, dairying, vegetables, poultry and eggs, and nursery industries.

(ii) Mining and logging

(iii) Resources Based Manufacturing: This group includes fruit and vegetable canners, sawmills and planning mills, and other wood products industries.

(iv) Non-Resource Based Manufacturing: This group includes metal fabricating, non-metallic, and other manufacturing industries.

(v) Services: This group includes packing house, wholesale trade, retail trade, hotel, tourist camp, health and education, other services, transport, finance and insurance, construction, municipal government, and utilities industries.

In addition to the above five endogenous variables, five other exogenous variables having an effect on the socio-economic activity were considered. These are:

(vi) Population

(vii) Exports

(viii) Government Spending

(ix) Land

(x) Logging and Mining Resources

The consensus is that the government spending, available land and logging and mining resources are not influenced by the first seven variables. These three were, therefore, not simulated. Simulation runs were made using the program described earlier. A number of the problems encountered are outlined here. Once the model builder or the expert has provided the impact matrix, how does he go about picking an initial value? Assigning of the starting value is quite important and to our knowledge no method has so far been proposed in the literature. The problem of selecting initial values was overcome by running a number of simulations and then choosing very arbitrarily a value found by dividing the initial 1970 data by twice the projected year 2020 data. Since the model was in fact intended to produce this information the data from the economic study mentioned earlier was used. This whole arrangement was very unsatisfactory and really pointed out the major weakness of Kane's KSIM approach when applying it to problems of this sort. To include the feasibility of some variables changing their impact during simulation, a parameter named XBASE was included. The impacts are computed by multiplying the a_{ij} value by the difference between x_j and $XBASE_j$ rather than x_j. This has the effect of reversing the impact when x_j reaches $XBASE_j$.

Table 1 shows the impact matrix. Variables 6, 7 and 10 were considered exogenous to the model. The initial values for the ten variables were 0.2, 0.8, 0.1, 0.05, 0.03, 0.23, 0.09, 0.10, 0.0, and 0.09 respectively. XBASE was set to zero for all variables except variable 10 for which it was 0.8. The assigned weights were 0.3, 0.2, 0.5, 0.5, 1.0, 0.5, 1.0, 0.0, 0.0, and 10.0. As one will note variable 9, LAND, appears to have no effect on anything. This is not so. It was used to constrain the growth of agriculture.

The results of the simulation for the next fifty years are shown in Figure 1. The number in the graph refers to the variable number. Number 0 refers to variable 10. * indicates that two or more variables have the same value.

TABLE 1

Impact Matrix

	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	1	2	3	0	0	0
2	0	0	3	0	1	0	2	0	0	1
3	1	1	2	2	2	2	2	0	0	0
4	2	1	2	2	3	2	3	2	0	0
5	3	2	2	3	1	3	1	2	0	0
6	1	0	1	3	3	1	0	0	0	0
7	3	2	3	3	2	0	1	1	0	0

Comparison of Results

Table 2 gives the projections for the year 2020 obtained from the economic study of the river basin and present model. The difference between two models, for the Mining and Logging, Non-Resources Based Manufacturing, and Services groups is less than ten percent in the year 2020. But the difference for other groups, namely Agriculture and Resources Based Manufacturing, is more significant. This is a situation model builder often faces while comparing projections of two different types of models rather than validating the projections with the real world data. Figure 2 shows the yearly projections for the five variables for both models on semi-logarithmic

FIGURE 1

Projections to year 2020 as given by the impact matrix of Table 1.

scale. The slopes of the two curves are considerably different. The projections made by the economic study show initial rapid growth followed by a slowing trend where as the curves obtained using the modified KSIM method show an initial slow rate of increase followed by rapid growth. This means although the two projections show similar results in the year 2020 the difference during the mid-year are very large. This is the second major shortcoming in applying KSIM to a problem of this type.

TABLE 2

Projections for Year 2020

(Output is shown in Millions of 1970 Dollars)

Variable Number	Variable Name	Input/Output Model	Present Model	Percentage Difference
1	Agriculture	89.7	123.0	37.0
2	Mining & Logging	13.5	14.0	3.7
3	Resource Based Manufacturing	197.0	254.0	29.1
4	Non-Resource Based Manufacturing	1050.0	1030.0	1.9
5	Services	5500.0	4980.0	9.5

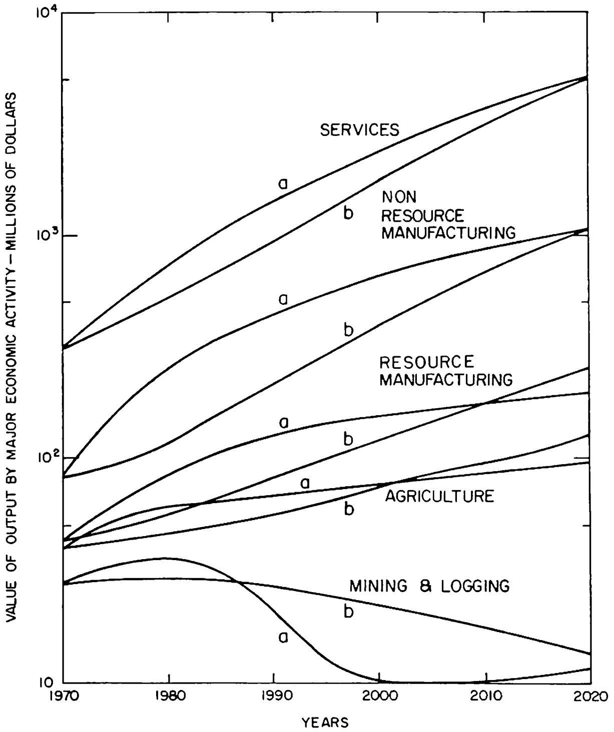

FIGURE 2

Modified KSIM projection for Canadian river basin - "a" projection are those obtained from an economic study of the river basin and the "b" curves are those obtained from the modified KSIM described in the text.

Conclusions

The idea of using an impact matrix to describe the interaction between different socio-economic variable is very appealing. It can provide insights to model builders regarding factors to be selected in a simulation. It helps to define model parameters without going into deep mathematics. The public and other non-researchers can take an active part in the modelling process and because they have participated the model is likely to be more acceptable to them. And with the refinements discussed in this paper, cross impact analysis can be used as a projection tool.

The modified cross impact technique could prove extremely useful in water resources studies because generally these involve a mix of subjective and objective variables. This paper described the first attempt to apply a simplified modelling approach to understand the economy of a river basin, and the results are very encouraging. The authors are continuing further work with the methodology to overcome the limitations imposed by using the 0 to 1 scaling and a projection equation having a basically sigmodial characteristic.

Acknowledgements

The authors wish to express their thanks to Dr. C.M. Woodside for introducing us to the simplified modelling and providing continual stimulus.

References

1. Kane, J., "A Primer for a New Cross-Impact Language - KSIM", Technological Forecasting and Social Change 4, 129-142 (1972).
2. Kane, J. et al, "Environmental Simulation and Policy Formulation - Methodology and Example (Water Policy for British Columbia)", International Symposium on Mathematical Modelling Techniques in Water Resources System (1972)
3. Kane, J. et al, "Policy Simulation", Water Resources Research 9 (1), 65-59 (1973)
4. Kane, J. et al, "Health Care Delivery: A Policy Simulator", Socio-Economic Plann. Sci., 6(3), 283-293 (1972)