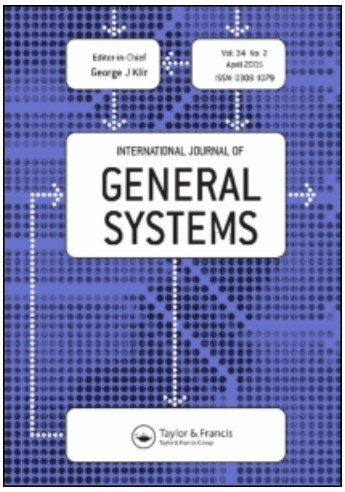

International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713642931

Determination of the correct criteria for K-systems reconstruction

D. R. McCarthy [a]; G. P. Shaffer [a]; C. Jeansonne [a]
[a] College of Science and Technology, Southeastern Louisiana University, Hammond, LA, USA

First Published on: 22 July 2009

To cite this Article McCarthy, D. R., Shaffer, G. P. and Jeansonne, C.(2009)'Determination of the correct criteria for K-systems reconstruction',International Journal of General Systems,99999:1,

To link to this Article: DOI: 10.1080/03081070903029261

URL: http://dx.doi.org/10.1080/03081070903029261

International Journal of General Systems
iFirst article, 2009, 1–10

Taylor & Francis
Taylor & Francis Group

Determination of the correct criteria for K-systems reconstruction

D.R. McCarthy*, G.P. Shaffer and C. Jeansonne

College of Science and Technology, Southeastern Louisiana University, Hammond, LA 70402, USA

(*Received 15 December 2008; final version received 8 May 2009*)

One of the essential components of the K-system reconstruction process is the determination of which state or substate contributes the greatest to the reconstruction. In the literature, there have been two competing criteria used, both of which are based upon a quantity known as the 'information-theoretic transmission', which is defined as $k \log k/k'$, where k is the information in the actual substate, and k' is the information in the reconstructed substate. One approach uses this quantity, while the other uses the quantity $k \log k/k' + (1 - k) \log (1 - k)/(1 - k')$. It had been implicitly assumed that both approaches gave the same result (i.e. the substate that would make the greatest contribution to the reconstruction would have a maximum value for both of these quantities). In this paper, we show empirically that in general they do not give the same results, and that the latter criterion predicts the substate that yields the optimal reconstruction. We then derive the correct criterion and demonstrate that this result is consistent with fundamental principles of information theory.

Keywords: K-systems analysis; reconstructability analysis; information theoretic transmission; substates; system function

1. Introduction

Systems with one or more dependent quantities that are governed by several independent variables are ubiquitous in nature. Indeed, it is hard to find any discipline, be it in the natural sciences, the social sciences, industrial situations or even financial systems, that cannot provide scores of examples of such systems. Not only are they found everywhere, but the understanding of these systems is the subject of numerous active research projects: what factors or combinations of factors lead to the greatest material strength? What factors determine the greatest change in the population or growth of a living organism? Of age, race, educational level, income level, which have the greatest impact on voter turn-out? As varied as are these examples, just as varied is the complexity of the governing systems, many of which are characterised by strongly nonlinear behaviour. The goal of most studies of this nature is to explain the behaviour of the dependent system based upon the independent variables and their interactions, which can be extremely difficult for highly nonlinear systems with many interacting variables (Shaffer 1997).

The most common approach to understand the complex systems is to collect multi-variate data and subject them to statistical analyses, often using multiple regression or analysis of variance (Tabachnick and Fidell 2001, Johnson and Wichern 2002, Zar 2008). The basic idea is to map the data onto some predetermined functional form (e.g. general linear model) and then find the function that produces the best fit. The 'best fit' is typically determined by a

*Corresponding author. Email: dmccarthy@selu.edu

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081070903029261
http://www.informaworld.com

statistical criterion, such as minimising the variance between the actual data and the estimated function. Such techniques are well established and numerous programs exist (e.g. SAS, SYSTAT, SPSS) that allow the researcher to perform such analyses. Regression methods have the advantage that they produce a mathematical relation with parameters that quantify the function and this can help when one is trying to build a first principles model of the system. However, such techniques also have some weaknesses, the most obvious one is that traditional statistical techniques effectively 'force' the data into an assumed model. For simple systems, this is fine; but for complex nonlinear systems, this can be problematic and misleading (Shaffer and Cahoon 1987, Gouw 1995, Gouw and Jones 1995, 1996). Moreover, for data that are expansive, and cover a broad range of variables, the functional forms that truly describe the system can vary from linear, to periodic, to chaotic, to singular and the nature of relationships can and do vary in space and time (Shaffer 1988a) and may even include phase shifts (Shaffer and Sullivan 1988). Although multiple regression analysis may explain the majority of the variability in a dependent variable, it generally does not explain the underlying processes that determine system behaviour (Shaffer and Cahoon 1987, Shaffer 1997). About two decades ago, an alternative technique for analysing complex systems was developed. The technique, which is not based at all upon statistical mapping, but rather is based upon the principles of information theory (Klir 1976, 1981, 1985, Shaffer and Sullivan 1988) and reconstructability analysis (RA; Cavallo and Klir 1981, 1982a, 1982b, Klir 1986, Zwick 1996, 2004a, 2004b, Zwick and Johnson 2004, Zwick and Shervais 2004, Zwick and Shu 1997, 2004), and is known as 'K-systems analysis' (Shaffer and Cahoon 1987, Shaffer 1988a, 1988b, Jones 1989, Trivedi 1993, Gouw 1995, Gouw and Jones 1995, 1996, Trivedi and Jones 1997, Trivedi *et al.* 1998, 2002) or 'entropy data analysis' (Jones 1984, 1985a, 1985b, 1985c, 1985d, 1989). The fundamental technique was originally proposed by Klir (1981) and was developed extensively by Bush Jones in 1984–1985 with a series of seminal papers (Jones 1984, 1985a, 1985b, 1985c, 1985d, 1989) rich in mathematical rigour that developed the theory of K-systems analysis further. To date, K-systems analysis has been used to describe systems including: antibiotic fermentation and extraction (Liu *et al.* 2000, Zhang *et al.* 2000), baldcypress swamp production (Hoeppner and Shaffer 2004), dam impact analysis (Shu 2000), data mining (Chen 1994), estuarine ecosystems (Shaffer and Cahoon 1987, Shaffer 1988a, 1988b, 1997, Shaffer and Sullivan 1988, Trivedi *et al.* 1998), fisheries management (Zhang *et al.* 2000), gene sequencing (Zhihong *et al.* 2004), macroeconomic policy (Liu *et al.* 2000) and mining operations (Hong *et al.* 2000). Nevertheless, the theory of K-systems analysis has had little development (Hoeppner and Shaffer 2004, Zwick and Johnson 2004, Johnson 2005) since the work of Jones (1984, 1985a, 1985b, 1985c, 1985d, 1989). Indeed, at this point, the only commercially available software package for K-systems analysis is a program developed in 1985 (Jones 1985d, 1989) which was written for DOS, and nearly every paper on K-systems analysis is descriptive in nature, and either discusses an application of K-systems analysis or redescribes Jones' work (Trivedi and Jones 1997, Trivedi *et al.* 1998, 2002).

Just because there has been little theoretical development of K-systems analysis, that is not to say that the theory is fully developed and understood. In fact, there are a number of unresolved issues and misunderstandings that currently exist in K-systems analysis, to the point that there are recent publications that use different techniques to perform the analysis (Zwick and Johnson 2004), but provide no explanation as to why this is being done. In this paper, we will provide a brief overview of the techniques and the principles behind K-systems analysis and point out one of the main discrepancies that exists in the theory. We resolve this discrepancy by showing empirically that these two techniques produce different results and that one gives the correct answer. We then derive what the correct criterion should be and relate our findings to the underlying principles of information theory.

2. K-systems overview

The theory of RA has existed informally since the early 1960s (Ashby 1964) and was more fully developed in the 1970s (Klir 1976, Cavallo and Klir 1979, Klir and Uttenhove 1979) and early 1980s (Cavallo and Klir 1981, 1982a, 1982b). The fundamental premise behind RA is that a probabilistic system that contains all of the information of the system can be 'reconstructed' from a system that contains no information, i.e. a system in which all dependent variable values are identical or a 'flat' system. The way that this is done is to measure the relative information within each state or substate (this concept will be described in more detail later) and then use this substate to add information to the flat system. Continuing with this process, eventually the initial system will be recovered, and the process of reconstruction helps to describe which substates are the most important. In 1981, Cavallo and Klir recognised that a general system that consists of real physical data could be transformed into a probabilistic form by normalising the dependent variable so that the sum of the function added to unity. Therefore, if the general system is defined by the function $g(x_1, x_2, \ldots, x_n)$ where x_i are the independent variables of which there are n, the transformed function is given by

$$k(x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_n)/\tau, \tag{1}$$

where

$$\tau = \sum_{x_i} g(x_1, x_2, \ldots, x_n), \tag{2}$$

where the sum is over all combinations of the independent variables. If each variable x_i has m_i values, then the full system will have $m_1 \times m_2 \times \cdots \times m_n$ values. The transformation from this general state (known as the g-system) to the normalised state (dubbed the K-system in homage to George J. Klir) is the initial part of the K-systems analysis process. (Note: throughout this article, we are employing a notation that is considerably different from that of Jones and the other authors. It is less compact but more descriptive, and is in a form that we feel helps elucidate the underlying mechanisms of K-system analysis.)

The next step is to evaluate the substates. From a mathematical point of view, the substates are sums of the system function over specified variables with the other variable(s) held fixed. From a physical point of view, the substates contain all of the information and behaviour of function for the variable for which the substate is defined. For example, if one has a function with three independent variables, in addition to the system function itself, there would be six additional substates defined as follows:

$$k_{1,2}(x_1, x_2) = \sum_{x_3} k(x) \quad k_{1,3}(x_1, x_3) = \sum_{x_2} k(x) \quad k_{2,3}(x_2, x_3) = \sum_{x_1} k(x), \tag{3}$$

$$k_1(x_1) = \sum_{x_2, x_3} k(x) \quad k_2(x_2) = \sum_{x_1, x_3} k(x) \quad k_3(x_3) = \sum_{x_1, x_2} k(x), \tag{4}$$

where we have used the notation $x = (x_1, x_2, x_3)$. The substates shown in Equation (3) are those that include the interactions between two variables, and those in Equation (4) show the main effects of each variable on the system. For this system, the substates $k_{i,j}$ will have $m_i \times m_j$ values each, and the substates k_i will have m_i values. Note also that each substate is a separate K-system itself and if one sums over all values within each subsystem, the sum will add to unity.

We are now at the stage where we can begin the actual reconstruction process. As was stated earlier, the basic idea is to start from a state of no relative information (or maximum entropy), and then add substates until we recover a sufficient representation of the actual system. The state with the maximum entropy, or greatest amount of disorder, is the one in which all values of the function have the same value, and that value would be $1/(m_1 \times m_2 \times \cdots \times m_n)$. All of the substates would have uniform initial values that are based upon the dimensionality of those functions. With this flat function, we can calculate the total 'information-theoretic transmission' (Johnson 2005), which can be thought of as a measure of distance between the flat and the actual system. This quantity is defined as follows:

$$T = \sum_{\mathrm{x}} k(x) \log \frac{k(x)}{k'(x)}, \tag{5}$$

where $k'(x)$ is the flat system. This function is closely related to the Shannon entropy $p \log p$, where p is a probability that is used extensively in information theory (Shannon 1948). Typically, for uses related to information theory, the base that is used for the logarithm is 2. For the K-systems application, the proper choice also should be to use base 2; however, for practical purposes, it does not matter which base is used. The absolute value of this quantity does not come into play at all in the reconstruction process, only the relative value, and since different base logarithms vary by the same factor, the relative change will be independent of the base of the logarithm.

There is little argument that the proper way to perform the reconstruction is to add to the flat system the state (or substate) that contributes to the greatest reduction in T. Johnson directly states: 'At each step, the candidate parameter that results in the greatest reduction in T is selected for inclusion in the model' (Johnson 2005) and such sentiments are stated many times by Jones. Once one has determined the most important substate, the values of the flat system are multiplied by the ratio of the real system to the reconstructed system. All other variables are multiplied by the factor that ensures that the reconstructed system remains probabilistic (i.e. the values all add to unity).

As an example of this procedure, we choose a simple three variable system where each variable has just two values, 0 or 1. Since there are eight total combinations of variables, and eight data points, the flat system will have $k'(x) = 0.125$. Suppose that the substate that makes the greatest contribution to the reconstruction is $k_{1,3}(1, 0) = 0.4$ and that the other terms in this substate have the value 0.2. Therefore, this substate and its associated flat substate will be the ones that will determine the reconstruction. The flat substate will have the values of 0.25 for all terms as there are four substates for $k_{1,3}$. Therefore, all terms in the full (flat) system that have $x_1 = 1$ and $x_3 = 0$ will be multiplied by the factor

$$\frac{k_{1,3}(1,0)}{k'_{1,3}(1,0)} = \frac{0.4}{0.25} = 1.6$$

and all other terms will be multiplied by the factor

$$\frac{k_{1,3}(0,0) + k_{1,3}(0,1) + k_{1,3}(1,1)}{k'_{1,3}(0,0) + k'_{1,3}(0,1) + k'_{1,3}(1,1)} = \frac{1 - k_{1,3}(1,0)}{1 - k'_{1,3}(1,0)} = \frac{0.6}{0.75} = 0.8.$$

With these points in mind, the first reconstruction will now read $k'(1, 0, 0) = 0.2$, $k'(1, 1, 0) = 0.2$ (i.e. 0.125×1.6), and all other terms will equal 0.1 (0.125×0.8). We can then calculate the new value of the transmission, and if it reduces the value of the

initial transmission by 45%, for example, we can state the following result: 'The state with variable 1 equal to 1 and variable 3 equal to zero contributes 45% of the information of the system and has the effect of increasing our dependent variable by 60%'. This process is continued with the new reconstruction, until enough substates have been added to reach a desired level of accuracy.

3. Determining the most important subsystem

The missing item of information that was not addressed in the previous section was, 'how does one determine the substate that will contribute to the greatest decrease in the total transmission?' The answer to this question is the main point of this paper. One way to find this out is to simply try every possible reconstruction from every substate and from there isolate the one that has the greatest impact. Not only is this unappealing from an aesthetic point of view, it is remarkably time intensive. For example, for a system with four variables, each with only three values, there are 81 states, and an additional 174 substates! Jones originally derived a criterion, and he concluded that the following term, which when maximised, would find the state or substate that would lead to the greatest reduction in the total transmission:

$$\Omega_\alpha(x) = k_\alpha(x) \log \frac{k_\alpha(x)}{k'_\alpha(x)} + (1 - k_\alpha(x)) \log \frac{1 - k_\alpha(x)}{1 - k'_\alpha(x)}, \tag{6}$$

where the subscript α refers to any substate including the individual system states. The derivation itself included some non-obvious statements and assumptions that, at least to us, never produced a satisfying answer as to why the second term should exist. Having said that, this was the criterion that Jones and every subsequent publication used to perform the reconstructions. Other authors, in particular those who have published work more recently on K-systems analysis (Johnson 2005), have used the information-theoretic transmission itself, which is Equation (6) without the second term

$$T_\alpha(x) = k_\alpha(x) \log \frac{k_\alpha(x)}{k'_\alpha(x)} \tag{7}$$

to evaluate which substate made the greatest contribution, the implicit reason being that both approaches produce the same result. This is not the case; using data directly from one of Jones' papers (Jones 1985a), we find that in just the first reconstruction, there is a discrepancy. Details of our analysis are shown in Table 1. Using Equation (7) to choose the most important state, we find that the state $k_{2,3}(x_2 = 1, x_3 = 2)$ has the largest value of T_α. However, when we use the criterion as proposed by Jones and defined in Equation (6), we find that the substate $k_2(x_2 = 1)$ has the greatest value of Ω_α. The question remains as to which state is the 'best' one in terms of performing the reconstruction. When the substate with the maximum value of T_α is used in the reconstruction, the transmission drops from 0.529 to 0.216. Therefore, this substate contributes 59.2% of the total information in the system. When the substate with the maximum value of Ω_α is used to reconstruct the flat system, the transmission is reduced to 0.203, capturing 61.6% of the information. Certainly no concrete conclusions can be drawn from one example, however, from this one demonstration clearly the two criteria do not produce the same results, and based upon this one example, it would indicate that the criterion from Equation (6) is the one that produces the best reconstruction.

Table 1. Example using data from Jones (1985a) to isolate the substate which makes the greatest contribution to the reconstruction of the system function.

x_1	x_2	x_3	K	k'	T	Ω	x_1	x_2	k_{12}	k'_{12}	T_{12}	Ω_{12}
0	0	0	0.016	0.083	0.038	0.062	0	0	0.089	0.250	0.133	0.124
0	0	1	0.033	0.083	0.044	0.031	0	1	0.362	0.250	0.194	0.045
0	0	2	0.039	0.083	0.043	0.022	1	0	0.089	0.250	0.133	0.124
0	1	0	0.033	0.083	0.044	0.031	1	1	0.461	0.250	0.406	0.150
0	1	1	0.121	0.083	0.066	0.012						
0	1	2	0.208	0.083	0.275	0.108	x_1	x_3	k_{13}	k'_{13}	T_{13}	Ω_{13}
0	1	2	0.208	0.083	0.275	0.108	0	0	0.049	0.167	0.087	0.094
1	0	0	0.016	0.083	0.038	0.062	0	1	0.154	0.167	0.017	0.001
1	0	1	0.033	0.083	0.044	0.031	0	2	0.248	0.167	0.141	0.030
1	0	2	0.039	0.083	0.043	0.022	1	0	0.082	0.167	0.084	0.044
1	1	0	0.066	0.083	0.023	0.003	1	1	0.187	0.167	0.031	0.002
1	1	1	0.154	0.083	0.137	0.039	1	2	0.280	0.167	0.210	0.058
1	1	2	0.241	0.083	0.369	0.163						
			1.000	1.000	0.529		x_2	x_3	k_{23}	k'_{23}	T_{23}	Ω_{23}
							0	0	0.033	0.167	0.077	0.131
		x_2	k_2	k'_2	T_2	Ω_2	0	1	0.066	0.167	0.088	0.066
		0	0.177	0.500	0.265	0.327	0	2	0.079	0.167	0.085	0.048
		1	0.823	0.500	0.592	**0.327**	1	0	0.098	0.167	0.075	0.028
							1	1	0.275	0.167	0.200	0.053
		x_3	k_3	k'_3	T_3	Ω_3	1	2	0.449	0.167	**0.643**	0.314
		0	0.131	0.333	0.176	0.155						
		1	0.341	0.333	0.011	0.000						
		2	0.528	0.333	0.350	0.115		x_1	k'_1	k'_1	T_1	Ω_1
								0	0.451	0.500	0.067	0.007
								1	0.549	0.500	0.074	0.007

Such an observation merits a more thorough theoretical investigation. Because the change in the total information-theoretic transmission is considered to be the most important quantity, we examine the change in this quantity after one reconstruction. Defining T_0 as the initial transmission, and T_1 as the transmission of the system after the first reconstruction, we have

$$T_0 - T_1 = \sum k \log k/k'_0 - \sum k \log k/k'_1 = \sum k \log k'_1/k'_0, \tag{8}$$

where k'_0 is the initial (flat) state, and k'_1 is the first reconstructed state. We now examine the quantity k'_1/k'_0. Referring to the example from Section 2, this quantity can have either one of two values. If the state is an element of the substate that we are using for the reconstruction, we use the relation

$$k'_1 = k'_0 \frac{k_\alpha}{k'_{\alpha 0}}, \tag{9}$$

where k_α is the value of the real substate, and $k'_{\alpha 0}$ is the value of the substate from the initial reconstruction (i.e. the flat state). If the state is not an element of the substate that we are using for the reconstruction, we use the relation

$$k'_1 = k'_0 \frac{1 - k_\alpha}{1 - k'_{\alpha 0}}. \tag{10}$$

Inserting Equations (9) and (10) into Equation (8), we have the relation

$$T_0 - T_1 = \sum_{k \in k_\alpha} k \log \frac{k_\alpha}{k'_{\alpha 0}} + \sum_{k \notin k_\alpha} k \log \frac{1 - k_\alpha}{1 - k'_{\alpha 0}}. \tag{11}$$

Recognising that the quantities $k_\alpha/k'_{\alpha 0}$ and $(1 - k_\alpha)/(1 - k'_{\alpha 0})$ are constants, we rewrite Equation (11) as

$$T_0 - T_1 = \log \frac{k_\alpha}{k'_{\alpha 0}} \sum_{k \in k_\alpha} k + \log \frac{1 - k_\alpha}{1 - k'_{\alpha 0}} \sum_{k \notin k_\alpha} k. \tag{12}$$

The first summation is the sum of all states that are elements of the substate that we have chosen, so the value of this sum is k_α itself. Using a similar reasoning, the second sum equals $1 - k_\alpha$. Therefore, the change in the total information-theoretic transmission due to a substate k_α will equal

$$T_0 - T_1 = k_\alpha \log \frac{k_\alpha}{k'_{\alpha 0}} + (1 - k_\alpha) \log \frac{1 - k_\alpha}{1 - k'_{\alpha 0}}, \tag{13}$$

which is the same relation stated in Equation (6). It is important to note that this derivation is general and holds true for any substate k_α. Clearly the substate that produces the greatest value of Ω_α will produce the greatest reduction in the transmission, as the quantities are indeed equivalent. Equation (13) can be readily generalised for the nth reconstruction to read

$$T_n - T_{n+1} = k_\alpha \log \frac{k_\alpha}{k'_{\alpha n}} + (1 - k_\alpha) \log \frac{1 - k_\alpha}{1 - k'_{\alpha n}}. \tag{14}$$

4. Discussion and conclusions

We have demonstrated empirically and derived a condition that shows that the best subsystem to be used in a K-systems reconstruction is one that has the largest value of the quantity $\Omega_\alpha = k_\alpha \log k_\alpha / k'_{\alpha 0} + (1 - k_\alpha) \log (1 - k_\alpha)/(1 - k'_{\alpha 0})$. Further insight as to why this is an expected result can be gained by referring to the original work by Shannon on information theory. In this work, Shannon derived the quantity $S = -p \log p$ where S denoted the 'entropy' or information within a system, and p is the probability of occupying a state that exists within that system. The important point is that this quantity only has meaning when the term $p \log p$ is summed over all states. Therefore, relating this back to the K-system reconstruction process, when one evaluates the quantity $k_\alpha \log k_\alpha / k'_{\alpha 0}$ on its own, it does not have any meaning as related to overall system behaviour. To know what the impact of a substate is on the entire system, one has to add the information (or transmission in this case) that the substate is occupied, and the information that the substate is unoccupied. Therefore, it follows that the term $k_\alpha \log k_\alpha / k'_{\alpha 0}$ (the transmission association with occupying the substate k_α) must be added to $(1 - k_\alpha) \log (1 - k_\alpha)/(1 - k'_{\alpha 0})$ (the transmission of not occupying k_α).

Of course, every new insight into a process raises new questions that merit attention. The reader may notice that in the example used to illustrate the difference between the two criteria, that although each approach gave different results, in reality, they really were not that different. When T_α was used, the most important state had a value of $T_{2,3}(1, 2) = 0.643$ while the second most important state was $T_2(1) = 0.592$. In the case of using Ω_α, we found that the most important state was $\Omega_2(1) = 0.327$ while the second most important was $\Omega_{2,3}(1, 2) = 0.314$. In other words, the results from both approaches are quite similar, and in a qualitative sense, they both do a good job of identifying the important states. In both cases, the difference between the most important state and the second most important state was less than 10%, which begs the question, 'How good were the data in the first place?' If the data themselves have an inherent uncertainty of 20% (which is not uncommon at all in ecological systems), then what really is the 'most important state?' One of the acknowledged weaknesses of K-systems analysis is the lack of statistical rigour, and the coarse and occasionally qualitative nature of the analysis. Results are only as good as the data, and without a quantitative description of the limitations of the data and how these limitations impact the results of the analysis, the relevance of any analytic method is limited at best. The issue of propagation of error as applied to K-systems analysis is a topic that will be addressed in a subsequent manuscript.

References

Ashby, W.R., 1964. Constraint analysis of many-dimensional relations. *General systems yearbook*, 9, 99–105.

Cavallo, R.E. and Klir, G.J., 1979. Reconstructability analysis of multi-dimensional relations: a theoretical basis for computer-aided determination of acceptable systems models. *International journal of general systems*, 5, 143–171.

Cavallo, R.E. and Klir, G.J., 1981. Reconstructability analysis: evaluation of reconstruction hypotheses. *International journal of general systems*, 7, 7–32.

Cavallo, R.E. and Klir, G.J., 1982a. Decision making in reconstructability analysis. *International journal of general systems*, 8, 243–255.

Cavallo, R.E. and Klir, G.J., 1982b. Reconstruction of possibilistic behavior systems. *Fuzzy Sets and Systems*, 8, 175–197.

Chen, Z., 1994. Qualitative reasoning for systems reconstruction using Lebesgue discretization. *International journal of general systems*, 25, 2329–2337.

Chen, Z., 1997. K-systems theory for goal-driven data mining. *Advances in systems science and applications,* Special issue 1, 40–43.

Gouw, D., 1995. Data analysis using reconstructability theory. Dissertation (PhD), Louisiana State University, Baton Rouge.

Gouw, D. and Jones, J.B., 1995. Using K-sysyems theory to compute true main and interaction effects in a design of experiments. *Cybernetics and systems*, 26, 481–498.

Gouw, D. and Jones, J.B., 1996. The interaction concept of K-systems theory. *International journal of general systems*, 24, 163–169.

Hoeppner, S.S. and Shaffer, G.P., 2004. The K-systems glitch: granulation of predictor variables. *Kybernetes*, 33, 962–972.

Hong, S., Zhao, Z., Jian, L. and Tang, D., 2000. Analysis of the factors influencing the output of raw coal based on factor reconstruction analysis. *International journal of general systems*, 29, 457–464.

Johnson, M.S., 2005. State-based systems modeling: theory, implementation and applications. Dissertation (PhD), Portland State University, Portland, OR.

Johnson, R.A. and Wichern, D.W., 2002. *Applied multivariate statistical analysis*. 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Jones, J.B., 1984. Determinations of unbiased reconstructions with arbitrary data. *International journal of general systems*, 10, 169–176.

Jones, J.B., 1985a. Determination of unbiased reconstruction. *International journal of general systems*, 10, 169–176.

Jones, J.B., 1985b. A greedy algorithm for a generalization of the reconstruction problem. *International journal of general systems*, 11, 63–68.

Jones, J.B., 1985c. Reconstructability analysis for general functions. *International journal of general systems*, 11, 133–142.

Jones, J.B., 1985d. Reconstructability considerations with arbitrary data. *International journal of general systems*, 11, 143–151.

Jones, J.B., 1989. A program for reconstructability analysis. *International journal of general systems*, 15, 199–205.

Klir, G.J., 1976. Identification of generative structures in empirical data. *International journal of general systems*, 3, 89–104.

Klir, G.J., 1981. On systems methodology and inductive reasoning: the issue of parts and wholes. *General systems yearbook*, 26, 29–30.

Klir, G.J., 1985. *Architecture of general systems problem solving*. Binghamton, NY: Plenum Press.

Klir, G.J., 1986. Reconstructability analysis: an offspring of Ashby's constraint theory. *Systems research*, 3, 267–271.

Klir, G.J. and Uttenhove, H.J.J., 1979. Procedures for generating reconstruction hypotheses in the reconstructability analysis. *International journal of general systems*, 5, 231–246.

Liu, D., Xu, H. and Lu, G., 2000. Factor system reconstruction analysis method and its application in antibiotics fermentation. *International journal of general systems*, 29, 395–410.

Shaffer, G.P., 1988a. K-systems analysis for determining the factors influencing benthic microfloral productivity in a Louisiana estuary, USA. *Marine ecology progress series*, 43, 43–54.

Shaffer, G.P., 1988b. A comparison of benthic microfloral production on the west and gulf coasts of the United States: an introduction to the dynamic K-systems model. *Marine ecology progress series*, 43, 55–62.

Shaffer, G.P., 1997. K-systems analysis: an anti-silver bullet approach to ecosystems modeling. *Advances in systems science and applications, special issue*, 1, 32–39.

Shaffer, G.P. and Cahoon, P., 1987. Extracting information from ecological data containing high spatial and temporal variability: benthic microfloral production. *International journal of general systems*, 13, 107–123.

Shaffer, G.P. and Sullivan, M.J., 1988. Water column productivity attributable to displaced benthic diatoms in a well-mixed shallow estuary. *Journal of phycology*, 24, 132–140.

Shannon, C.E., 1948. A mathematical theory of communication. *Bell system technical journal*, 27, 379–423 623–656.

Shu, G., 2000. Reconstruction analysis of the comprehensive evaluation of social–economic–environmental–ecological impacts of the three Gorges project. *International journal of general systems*, 29, 429–438.

Tabachnick, B.G. and Fidell, L.S., 2001. *Using multivariate statistics*. 4th ed. Needham Heights, MA: Allyn and Bacon.

Trivedi, S.K., 1993. Reconstructability theory for general systems and its applications for automated rule learning. Dissertation (PhD), Louisiana State University, Baton Rouge.

Trivedi, S.K. and Jones, J.B., 1997. Reconstructability of selection systems and U-structures. *Advances in systems science and applications, special issue*, 1, 52–58.

Trivedi, S.K., Jones, J.B. and Iyengar, S.S., 1998. A reconstructability analysis approach to biological systems modeling. *In*: S.S. Iyengar, ed. *Computer modeling and simulations of complex biological systems*. Boca Raton, FL: CRC Press, 126–159.

Trivedi, S.K., Jones, J.B. and Iyengar, S.S., 2002. Why K-systems analysis works. *Systems analysis modelling simulation*, 42, 23–31.

Zar, J.H., 2008. *Biostatistical analysis*. 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Zhang, G., Chang, J. and Shu, G., 2000. Applications of factor-criteria system reconstruction analysis in the reproduction research of grass carp, black carp, silver carp, and bighead in the Yangtze River. *International journal of general systems*, 29, 419–428.

Zhihong, Z., Pengtao, W., Huaqing, L. and Shu, G., 2004. Construction of main sequence of gene based on method of factor reconstruction analysis. *Kybernetes*, 33, 1020–1025.

Zwick, M., 1996. Control uniqueness in reconstructability analysis. *International journal of general systems*, 24, 151–162.

Zwick, M., 2004a. An overview of reconstructability analysis. *Kybernetes*, 33, 877–905.

Zwick, M., 2004b. Reconstructability analysis with Fourier transforms. *Kybernetes*, 33, 1026–1040.

Zwick, M. and Johnson, M., 2004. State-based reconstructability analysis. *Kybernetes*, 33, 1041–1052.

Zwick, M. and Shervais, S., 2004. Reconstructability analysis detection of optimal gene order in genetic algorithms. *Kybernetes*, 33, 1052–1062.

Zwick, M. and Shu, H., 1997. Set-theoretic reconstructability of elementary cellular automata. *Advances in systems science and applications,* Special issue 1, 31–36.

Zwick, M. and Shu, H., 2004. Reconstructability and dynamics of cellular automata. *Kybernetes*, 33, 877–905.