Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

AN ANALYTICAL COMPARISON OF OPTIMIZATION PROBLEM GENERATION METHODOLOGIES

Raymond R. Hill

Department of Operational Sciences
Air Force Institute of Technology
2950 P Street, Bldg 640
Dayton, Ohio, 45433, U.S.A.

ABSTRACT

Heuristics are an increasingly popular solution method for combinatorial optimization problems. Heuristic use often frees the modeler from some of the restrictions placed on classical optimization methods required to constrain problem complexity. As a result, modelers are using heuristics to tackle problems previously considered unsolvable, improve performance over classical optimization methods, and open new avenues of empirical study. Researchers should fully understand key test problem attributes and sources of variation to produce efficient and effective optimization studies. These problem attributes and sources of variation are reviewed. Problem correlation structure significantly effects algorithm performance but is often overlooked or ignored in empirical studies. This paper analyzes the correlation structure among a set of standard multidimensional knapsack problems and recommends an improved approach to synthetic, or randomly generated optimization problems for the empirical study of solution algorithms for combinatorial optimization problems.

1 INTRODUCTION

Increasingly researchers are using heuristics to solve combinatorial optimization problems across a diverse range of applications. A fundamental aspect of research into and application of heuristics is empirically testing heuristic performance across a representative range of problems. This range of problems is a focus of this paper. In particular, we address the primary issues to consider when adopting or generating test problems.

Why the concern about test problem generation? An inadequate set of test problems does not provide the full range of heuristic performance information available from adequate test sets. A properly designed industrial experiment, simulation study, or survey would never ignore influential factors. However, we contend that many empirical tests of algorithms and heuristics are unknowingly guilty of such an omission due to a reliance on standard problem sets and "legacy" problem generation methods. This reliance on test sets and legacy generation methods overlooks the influence of problem correlation structure on algorithmic performance.

To defend this assertion, this paper is organized as follows. §2 addresses why heuristics are important to modelers. §3 discusses testing of algorithms and heuristics while §4 discusses the characteristics of adequate test problems. §5 develops the experimental design and sources of variation. §6 presents the results of recent analysis into the correlation structure of accepted standard problem sets while §7 discusses and compares alternative approaches to generating optimization problems. §8 presents our conclusions. Throughout this paper, algorithm is used as a general term to include heuristics.

2 WHY HEURISTICS?

There has always been a symbiotic relationship between computer science and operations research. This is especially true in the increasing use of modern heuristics to solve combinatorial optimization problems. In fact, combined simulation and heuristic codes are helping to solve incredibly complex problems previously considered beyond the realm of classical optimization approaches.

Heuristics do not guarantee optimal solutions. However, a wealth of empirical evidence suggests that, in general, properly implemented heuristics provide reasonable answers quickly. More importantly, these heuristics allow the formulation and solution of more complex, reality-based problems. Therein lies the true contribution of heuristic solution procedures; they provide modelers the ability to address the harder, yet more interesting problems.

This paper is about heuristics and the test problems employed to evaluate heuristics. In particular, this paper lays out what one should consider when generating test problems for empirical testing. Our fundamental thesis is that current heuristic testing is too narrow in scope and thus lacks the empirical basis for theoretically comprehending the general applicability of heuristics.

3 TESTING ALGORITHMS AND HEURISTICS

We test and implement algorithms for at least three reasons. One is to find a feasible solution to a previously unsolvable problem. Past reliance on classical optimization techniques prevented consideration of certain problem attributes to reduce problem complexity. Essentially, we made simplifying assumptions for computational tractability. Thus, we found exact solutions to approximate problems. The aircraft loading problem is an example. Prior cargo loading systems employed classical techniques to obtain feasible cargo loads (and thus airlift requirements) based primarily on cargo weight and volume restrictions. These systems avoided 3-dimensional aspects, aircraft center-of-gravity, or even aircraft floor-loading considerations because the problems became too complex, even nonlinear. However, with new algorithms, such considerations are now practical as demonstrated by Chocolaad's application of tabu search to the Air Force's Airlift Loading Model (ALM) (1998).

Another reason is to improve performance over existing methods. We need to briefly define "improved performance." For a given problem, a heuristic solution approximates the optimal solution, but obtains this approximation much quicker than a pure optimization algorithm. Thus, performance is partially defined as time to solution. However, because heuristics can accommodate more complex problems, performance must consider obtaining solutions to more realistic problems. This is an important consideration when embedding combinatorial sub-problems within operational systems or within a simulation, or when conducting a series of time sensitive analyses within some analytical setting. Consider the Uninhabited Aerial Vehicle (UAV) routing problem (i.e., the traveling salesman problem). UAVs offer long flight times, extended dwell times over a target, multiple targeting and flexible re-targeting. Pre-mission and in-mission route planning requires rapidly obtaining reasonable solutions to the routing problem. Charlton (1995), Sisson (1997) and Ryan (1998) investigate the performance of tabu search techniques on such routing problems and report favorable results. Such results offer promise for both off-line analyses and embedded route-planning software applications.

The third reason is to compare competing algorithm performance and develop an understanding of how algorithms (and heuristics) perform on various classes of problems (see for instance Zanakis (1977) and Hill (1996)). Much research has been done in this area. However, more needs to be done and should be done with adequate sets of test problems. The rationale for this statement is discussed in the remainder of this paper.

4 TEST PROBLEM CHARACTERISTICS

Test problems are the basis for empirically examining algorithms. Test problems can be drawn from practical settings, from libraries of "standard" problems, or randomly generated (Barr, *et al.*, 1995). Greenberg (1990) calls the last two sources *library analysis* and *statistical analysis*, respectively. Each source has benefits and drawbacks.

Practical test problems can provide realism to the study (Golden, *et al.*, 1986). Conjectures regarding algorithm applicability are straightforward but limited to the range covered by these "real" applications. Such problems also provide benchmarks to compare studies and algorithm performance. However, such problems are generally limited in number and availability, and performance conjectures are limited to the set of problems examined (Lin and Rardin, 1980). Coefficient perturbation methods can expand the limited set of problems, but the validity of the perturbation method is difficult to assess. Further, it is often impractical or impossible to control the values of various attributes of the test problems. Controlling problem parameterization is important for empirical experiment design.

Standard problem sets, available via the internet, strive to overcome the availability problem. Some standard problems are based on real applications while others are synthetic (i.e., randomly generated) but widely employed and accepted. Such problem sets provide benchmark capabilities but again yield conjectures limited to the range of problems defined by the set of problems. Barr, *et al.* (1995) suggest *all* computational studies of heuristics employ standard problem sets to promote comparisons across experiments. While this is sound practical advice, the experimenter should consider all problem attributes in these test sets. This is addressed below and concerns are demonstrated using the multi-dimensional knapsack problems (MKP) available from Beasley (1998).

A third approach is to randomly generate test problems, creating a set of "synthetic" optimization problems. Results from synthetic problems are random variables so statistical analysis techniques are appropriate (Golden, *et al.*, 1986). Problem generation procedures can mimic "real world" problem attributes if known. Good problem generation procedures can control a variety of problem attributes. Some common problem attributes are the:

1. number of variables,
2. number of constraints,
3. marginal distributions of objective and constraint coefficients,

4. method of setting right-hand side values,
5. non-zero entries in constraint coefficient matrix,
6. relationship between objective function values and constraint sums, or
7. correlation structure among objective function and each constraint.

Further, clearly defined problem generation procedures provide an efficient means for distributing and reproducing problems (Barr, *et al.*, 1995). Some drawbacks include generating problem instances harder (or easier) than seen in practical application, some possibly unrealistic problem instances generated in a comprehensive study, and the sometimes tenuous definition of real-world problem attributes.

5 EXPERIMENTAL DESIGN AND SOURCES OF VARIATION

Synthetic optimization problems for empirical investigations facilitate experimental design. Lin and Rardin (1980) and Golden, *et al.* (1986) discuss experimental design and statistical analysis of heuristic results. Each problem characteristic listed above is a potential experimental factor. Properly constructed problem generators can create the problem instances called for in specific experimental designs. As noted by Lin and Rardin (1980), proper experimental design and statistical analyses produce inferences valid for all problems produced by the particular problem generator. By extension, defining problem generation parameters corresponding to real-world problem attributes extends those inferences to all real-world instances.

In any statistical experiment, error and variation are present. Designed experiments reduce the error by controlling the experimental factors of interest (Lin and Rardin, 1980). Naturally, not all sources of variation are controllable. However, those sources or factors imparting a significant influence should be controlled.

Lin and Rardin (1980) discuss sources of variation in experiments involving integer programming algorithms. They list:

1. variation among algorithms,
2. variation among levels of factors,
3. variation among the problems generated, and
4. measurement error.

The variation among algorithms and levels of factors is *good* variation. This variation drives the inferences emanating from the empirical study; some algorithms outperform others, and some factors are more influential on algorithm performance than others. The variation between the problems generated and measurement error has received less attention and may involve significant "oversight error."

Variation between problems is not unusual. Sampling error is a well-known phenomena in random variate generation. Techniques such as control variates can reduce the effects of sampling error but are rarely, if ever, applied in optimization studies. Lack of random number synchronization is another source of variation and empirical studies do sometimes synchronize random variables in test problem generators.

Measurement error is another significant source of variation. For instance, measuring processing time of an algorithm must account for imprecision in clock sampling, internal representation, and alternate activities of the processor (e.g., multiprocessing systems, automatic backups, etc.)

We define and classify *oversight error* as a component of measurement error for one reason; the researcher may not be aware of a significant factor. For instance, Hill (1996) showed the significant influence of correlation structure on solution procedures for the two-dimensional knapsack problem. His experiment employed synthetic optimization problems, controlling the number of variables, number of constraints, tightness of right-hand side values, problem correlation structure, and type of correlation induced. Correlation structure is often overlooked in empirical studies of optimization algorithms, but it is present in all test problems.

This correlation structure presence leads to the questions, "What type of correlation structure exists in standard optimization problem sets?" and "Is this correlation structure a significant, yet unaccounted for, performance factor?" The MKP sets from Beasley are used to demonstrate our concerns.

6 STRUCTURE OF SOME STANDARD TEST PROBLEMS

Beasley's standard test problems are available via the internet (1998). Data available for each MKP problem includes the number of variables, number of constraints, constraint tightness values, best feasible solution value, and the value of the LP relaxation. Correlation structures were calculated for each problem, with the results summarized in Figures 1 through 4.

Beasley offers 270 synthetic problems distributed in 9 files. Figure 1 displays the range of objective function to constraint correlation values across the 30 test problems in each of the nine files referenced along the Y-axis. The graph depicts correlation value ranges from just below

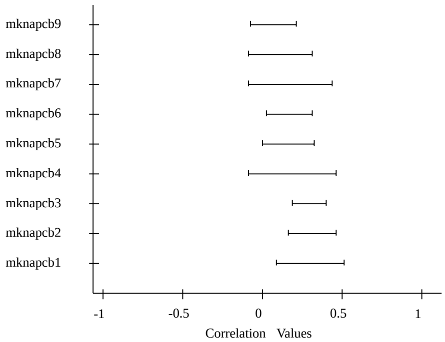

Figure 1: Range of Correlation Values Between Objective Function and Constraint Coefficients in the Nine MKP Problem Files Available From Beasley Website

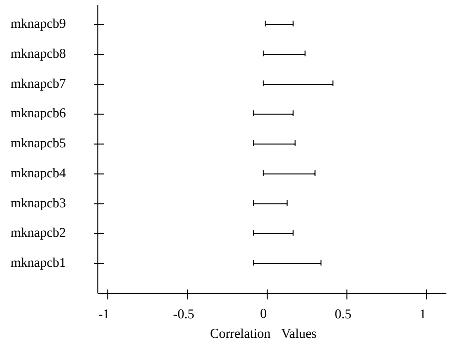

Figure 2: Range of Correlation Values Between Constraint Coefficients in the Nine MKP Problem Files Available From Beasley Website

zero to approximately 0.5. These ranges are narrow with respect to the entire range of feasible correlation values, so are insufficient to provide real insight regarding algorithm performance as a function of problem correlation structure. However, the range is probably wide enough to influence algorithm performance. Figure 2 summarizes the interconstraint correlation values for the same 270 test problems. These ranges are tighter than in Figure 1 and clustered on the positive side of zero correlation, or uncorrelated. Notice the minimum values of the ranges in Figures 1 and 2 rarely attain negative values, and even then, these values are essentially zero (uncorrelated). Hill (1996) demonstrated the significant effect of negative correlation values between objective function and constraint coefficients on algorithm performance.

Beasley also provides a file containing 48 test problems drawn from the literature (file MKNAP1.TXT). The correlation structures for these problems are summarized in Figures 3 and 4. Figure 3 summarizes the range of objective function to constraint correlation values for each of the 48 test problems. The average correlation value is also indicated. There are more instances of negative correlation and the ranges appear centered around the zero correlation value versus skewing to the positive side as seen in Figures 1 and 2. Again these ranges are sufficiently wide to likely influence algorithm performance, but not variable enough to draw general conclusions regarding correlation effects on algorithm performance.

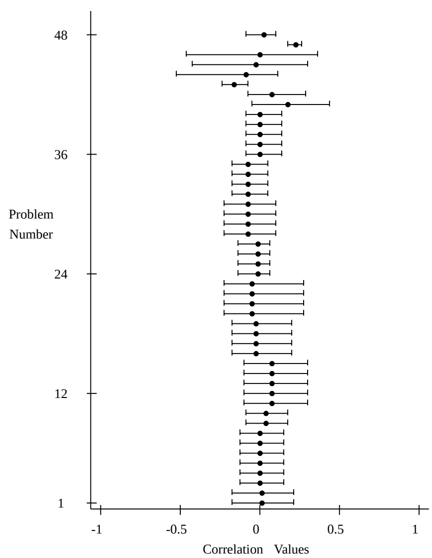

Figure 3: Range of Objective Function to Constraint Coefficient Correlation for 48 Problem From Literature Available From Beasley Website

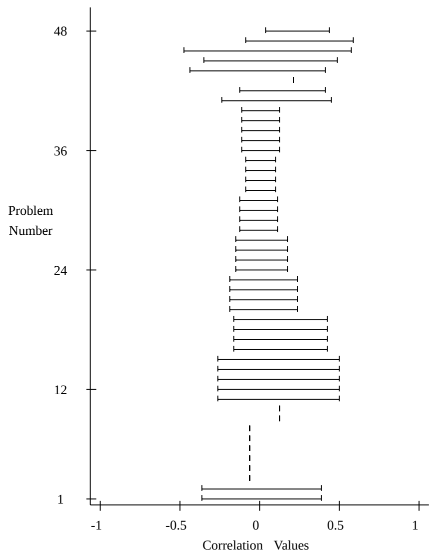

Figure 4: Range of Inter-Constraint Coefficient Correlation for 48 Problem From Literature Available From Beasley Website

Figure 4 summarizes the interconstraint correlation values for the same 48 test problems. Overall, the range of correlation values is centered at zero and are narrow with respect to the entire range of feasible correlation values. As with the previous data displayed, these ranges are insufficient to provide real insight regarding algorithm performance as a function of problem correlation structure. However, once again the range of values provided by these problems is probably wide enough to influence algorithm performance.

Figures 3 and 4 demonstrate two artifacts of this problem set. First, since many problems vary only by the right-hand side values of the constraints, we see many repeated correlation structures. Second, nine problems had two constraints (i.e., a single interconstraint correlation value) that were essentially uncorrelated.

The message of Figures 1 through 4 is test problems have a correlation structure which can affect algorithm performance. Failure to consider correlation effects can lead to oversight error. Empirical studies using these problems should consider the correlation structure attributes and supplement these problems with synthetic problems involving a wider range of correlation structures. Thus, an empirical experiment should consider correlation structure as a experimental factor in the analysis of results.

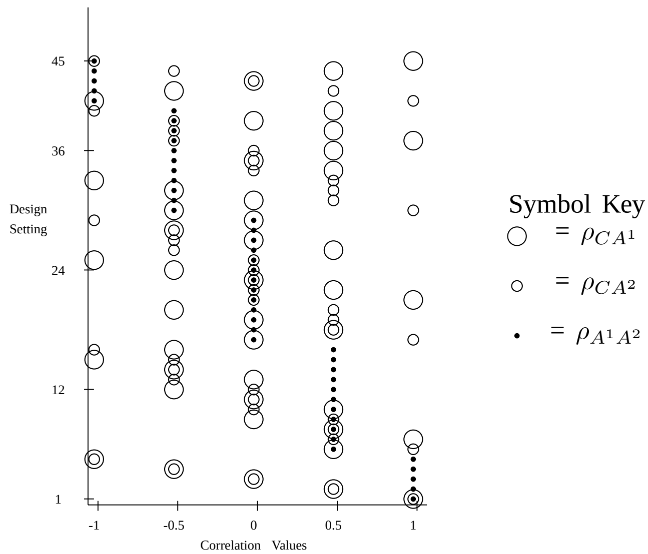

Figure 5: Correlation Values Employed in Experimental Design in Hill (1996)

Fortunately, experimental design and problem generation techniques make this a straightforward task.

Figure 5 plots the correlation values employed in the experimental design used by Hill (1996) to examine algorithm performance of two-dimensional knapsack problems. In Figure 5, ρ_{CA^1} represents the objective function to first constraint correlation value, ρ_{CA^2} represents the objective function to second constraint correlation value, and $\rho_{A^1A^2}$ represents the interconstraint correlation value. The Y-axis of Figure 5 represents each of the 45 correlation structure design settings used in the experimental design. Note ρ_{CA^1}, ρ_{CA^2}, and $\rho_{A^1A^2}$ vary across their entire range of feasible correlation values. Such a design facilitates insight regarding the effect of correlation structure on algorithm performance. A key to obtaining such an experimental design is choosing a proper problem generation methodology.

7 SYNTHETIC OPTIMIZATION PROBLEM GENERATION APPROACHES

We close by discussing three general approaches to test problem generation, the last being our recommended approach. Since test problem generation is essentially a multivariate sampling problem, we focus on generating random samples. This presentation is terse as details are available in Cario, *et al.* (1995), Hill (1996), and Reilly (1997).

One approach is simply generating all coefficients independently. This approach is easy to implement and the marginal distributions selected are easy to maintain. However, the problem correlation structure is not controlled and as the number of variables in the problem increase, the correlation values will converge to zero, i.e., uncorrelated.

A second, and quite popular approach, is to use implicit correlation induction. Moore and Reilly (1995) define an implicit correlation induction method as a multivariate sampling approach in which some population correlation level is implied by the specification of the parameters for the problem generation method. There are two benefits of such an approach. First, it is easy to implement and second it has been used extensively in past research. However, this method has crucial shortcomings. The marginal distributions and the correlation levels induced are dependent, thus confounding analysis of the effects of problem parameters on algorithm performance. Furthermore, due to the typically linear nature of the problem generation scheme, the range of coefficient values, for a given correlation value, are quite limited.

A preferred approach is to use explicit correlation induction, wherein one defines the univariate marginal distributions, selects the target population correlation structure, and then samples from the resulting multivariate distribution. This approach has real benefits. For one, selecting correlation structures directly facilitates experimental design. Then, independence of marginal distributions with selected correlation structure facilitates statistical analysis. Finally, a wider range of coefficient values are realized since sampling is based on a joint distribution versus some functional form of a marginal distribution. A drawback is determining the technique for sampling from the appropriate joint distribution, though multivariate sampling techniques such as Iman and Conover's (1982) or Hill and Reilly (1994) are two examples applicable to test problem generation.

8 CONCLUSIONS

Heuristics and algorithms for combinatorial problems are an active area of research and application. This paper focused on the issues associated with the testing of such algorithms and heuristics, namely what problem attributes effect algorithm performance and what are the sources of variation within a set of test problems. Specifically, this paper calls attention to the sometimes overlooked effect of problem correlation structure on algorithm performance. Demonstration of our concern rested on the unknown correlation structure resident within "standard" problem sets routinely employed within empirical optimization studies. This paper presented the results of computing those correlation structures. The correlation structures computed have a range of values sufficient enough to likely affect algorithm performance, but really insufficient to draw meaningful conclusions regarding the effect of correlation structure on algorithm performance.

As a statistical experiment, empirical optimization studies should undergo the same rigorous experimental design employed in other empirical settings. This echoes the work of Lin and Rardin (1980) and Golden, *et al.* (1986). Properly choosing a test problem generation procedure facilitates implementation of the resultant experimental design. Since test problem generation is an application of multivariate sampling, explicit correlation induction schemes are the logical choice for a problem generation mechanism. In other words, the experimenter should select a generation method in which the form of the marginal distributions are unaffected by the correlation structure specified. Coupled with better experimental designs, researchers can learn more about the applicability and utility of algorithms and heuristics for solving increasingly complex problems.

REFERENCES

Barr, R. S., B. L. Golden, J. P. Kelly, M. G. C. Resende, and W. R. Stewart. 1995. Designing and Reporting on Computational Experiments with Heuristic Methods. *Journal of Heuristics*, **1**(1): 9-32.

Beasley, J.E. 1998. OR-Library. Internet access via `http://mscmga.ms.ic.ac.uk/`

Cario, M. C., J. Clifford, R. Hill, J. Yang, K. Yang, C. Reilly. 1995. Alternative Methods for Generating Synthetic Generalized Assignment Problems. *Proceedings of the 4th Industrial Engineering Research Conference*, eds. B. Schmeiser and R. Uzsoy, 1080-1089.

Carlton, William B. 1995. *A Tabu Search Approach to the General Vehicle Routing Problem.* Ph.D. Dissertation. University of Texas, Austin.

Chocolaad, Christopher A. 1998. *Solving Geometric Knapsack Problems Using Tabu Search Heuristics.* M.S. Thesis. Department of Operational Sciences, Air Force Institute of Technology, AFIT/GOR/ENS/98M-05.

Golden, B. L., A. A. Assad. E. A. Wasil, and E. Baker. 1986. Experimentation in Optimization. *European Journal of Operational Research*, **27**,1-16.

Greenberg, Harvey J. Winter 1990. Computational Testing: Why, How and How Much. *ORSA Journal on Computing*, **2**(1): 94-97.

Hill, R. R. and C.H. Reilly. 1994. Composition for Multivariate Random Variables. *Proceedings of the 1994 Winter Simulation Conference*, eds. J.T. Tew, S. Manivannan, D.A. Sadowski, and A.F. Seila. 332-

342. Institute of Electrical and Electronics Engineers, Orlando Florida.

Hill, R. R. 1996. Multivariate Sampling with Explicit Correlation Induction for Simulation and Optimization Studies. Ph.D. Dissertation, Department of Industrial, Welding and Systems Engineering, The Ohio State University, Columbus, OH.

Iman, R.L. and W.J. Conover. 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Communications in Statistics: Simulation and Computation*, **11**(3), 311-334.

Lin, B. W. and R. L. Rardin. 1980. Controlled Experimental Design for Statistical Comparison of Integer Programming Algorithms. *Management Science*, **25**(12): 1258-1271.

Moore, B. A. and C. H. Reilly. 1995. Randomly Generating Optimization Problems with Explicitly Induced Correlation. The Ohio State University. Department of Industrial and Systems Engineering. Working Paper 1993-002 (revised).

Reilly, C. 1997. Generating Coefficients for Optimization Test Problems with Implicit Correlation Induction. *Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics*, Volume 3, 2438-2443.

Ryan, Joel L. 1998. *Embedding a Reactive Tabu Search Heuristic in Unmanned Aerial Vehicle Simulations.* M.S. Thesis. Department of Operational Sciences, Air Force Institute of Technology, AFIT/GOR/ENS/98M-21.

Sisson, Mark R. 1997. *Applying Tabu Heuristic to Wind Influenced, Minimum Risk, and Maximum Expected Coverage Rates.* M.S. Thesis. Department of Operational Sciences, Air Force Institute of Technology, AFIT/GOR/ENS/97M-20.

Zanakis, S. H. 1977. Heuristic 0-1 Linear Programming: An Experimental Comparison of Three Methods. *Management Science*, **24**, 91-104.

AUTHOR BIOGRAPHY

RAYMOND R. HILL is an Assistant Professor of Operations Research in the Department of Operational Sciences at the Air Force Institute of Technology. His research interests include modeling and simulation, applications of optimization, and military applications of modeling. He holds a Ph.D. from The Ohio State University.