A Comparison Of Approaches To Stepwise Regression For The Indication of Variables Sensitivities Used With A Multi-Objective Optimization Problem

Mengchao Wang, PhD Candidate **Jonathan Wright, PhD CEng** *Member ASHRAE* **Alexander Brownlee, PhD**

Richard Buswell, BEng PhD

ABSTRACT

Global sensitivity analysis can be used to identify and rank variables importance (sensitivities) for design objectives and constraints, where the solution space is sampled and a linear regression model is normally adopted in the stepwise manner. The relative importance of variables can be examined by three indicators: the order of variables entry into the linear regression model; the absolute values of the standardized regression coefficients or their rank transformation coefficients; and the size of the R^2 *changes (coefficient of determination) attributable to additional variables at each step. However, the robustness of the linear regression model constructed from a stepwise regression is related to the choice of procedure options, e.g. the set of samples and data formulation. Different procedure options could lead to different linear regression models, and therefore influence the indication of variables global sensitivities. Thus, this paper investigates the extent to which the procedure options of a stepwise regression can influence the indication of variables global sensitivities, measured by three different sensitivity indicators, for energy demand, capital costs and solution infeasibility, when using both the randomly generated samples and the biased solutions obtained at the start of a multi-objective optimization process (based on NSGA-II). It concludes that the most important variables are always ranked on the top no matter the choice of procedure options, but it is better to adopt both the entry-orders of variables and their standardized (rank) regression coefficients or the contributions to* R^2 *changes, to provide robust orderings of variables importance, for design objectives and constraints. Moreover, when the sample size is smaller, re-generated separate set of samples for sensitivity analysis can avoid misleading variables importance, especially for the variables ranked in the middle. Finally, to improve computational efficiency, this paper concludes that the first 100 solutions obtained from a multi-objective optimization can be used to perform global sensitivity analysis, to identify the important variables for design objectives.*

INTRODUCTION

Sensitivity analysis (SA) is designed to apportion the variation in the output to the variations from different input variables (Satelli et al., 2000), which is a concern in all areas of building simulation (Evins, 2013). The techniques used for SA are at least grouped into global and local forms (Saltelli et al., 2000). A local SA is conducted similarly to numerical differencing, where each variable is incremented by a pre-defined amount one-at-once, to evaluate its impact on a given problem uncertainty. A global SA is often based on a linear regression model in the stepwise manner, to compare the

Mengchao Wang is a PhD candidate in the Department of Civil & Building Engineering, Loughborough University, UK. Jonathan Wright is a professor in the Department of Civil & Building Engineering, Loughborough University, UK. Alexander Brownlee is a senior research assistant in the Computing Science and Mathematics, University of Stirling, UK. Richard Buswell is a senior lecturer in the Department of Civil & Building Engineering, Loughborough University, UK.

relative importance (sensitivities) of variables. The regression method is most popularly applied for building design, due to the fast computation and easy understanding (Breesch & Janssens, 2005; Tian, 2013).

Three indicators can be used in the stepwise regression to measure the relative importance of variables: the order of variables entry into the linear regression model (Helton et al., 2006); the absolute values of SRCs (standardized regression coefficients) or their rank transformation SRRCs (standardized rank regression coefficients) (de Wilde & Tian, 2009; Ballarini & Corrado, 2012); and the size of R^2 changes (coefficient of determination) attributable to additional variables at each step (Hopfe & Hensen, 2011). According to Iman and Conover (1982), the correlations between variables within a Latin hypercube or random sample are indeed close to zero, where the ordering of variables importance based on each of indicators are expected to be same (Satelli et al., 2000). Thus, in the previous researches, the determination of variables importance is usually based on a particular sensitivity indicator from the stepwise regression.

The linear regression model constructed from stepwise regression depends on the choice of procedure options. Each procedure option has its strengths and weaknesses. Kletting and Glatting (2009) state F-test is often used as default to stop the stepwise regression process, but it has been shown to perform poorly relative to other criteria, e.g. the corrected AIC (Akaike information criterion). Wright et al. (2012) state the identified numbers of variables is related to the choice of sample size. Wang et al. (2013) state a similar linear regression model can be concluded through both raw data and their ranks, but this is not the case when the samples are strongly biased, e.g. at the end of a single-objective optimization.

For building design, both sensitivity analysis and model-based optimization can be use to inform design decisions. Combined those two methods would provide designers with a set of optimum solutions and an understanding of which variables have the most impact on those solutions, particularly when some uncertainties exist in the choice of solutions (Wright et al., 2012). Building design is an inherently multi-objective process, there being a trade-off to be made between two or more conflicting design objectives, it leads to wide development of the model-based multi-objective optimization methods, e.g. NSGA-II (non-dominated sorting genetic algorithm II) (Brownlee & Wright, 2012). Therefore, this paper examines the extent to which the procedure options of a stepwise regression can influence the indication of variables global sensitivities, measured by three different sensitivity indicators, for both design objectives and constraints (energy demand, capital costs and solution infeasibility), when using the randomly generated samples and the biased solutions obtained from the start of a multi-objective optimization process (based on NSGA-II).

EXPERIMENTAL APPROACH

Stepwise Regression Analysis

The global SA adopted here is based on a stepwise linear regression model. The more important the variable is (the topper the rank-order of importance), the earlier the entry into the linear regression model; the larger the absolute value of SRC or SRRC; the bigger the contribution to model R^2 change (Saltelli et al., 2000). A measure of variables importance provided by standardized (rank) regression coefficients is based on the effect of moving each variable away from its expected value by a fixed fraction of its standard deviation while retaining all other variables at their expected values, which is free from the effects of units and distribution assumptions. Calculating SRCs or SRRCs equals to performing the regression analysis with the input variables and outputs normalized to mean zero and standard deviation one. The difference in the R^2 values of the regression models constructed at successive steps equals to the fraction of the total uncertainty in the output, that is accounted for by the individual variables being added at each step (if variables are uncorrelated).

The construction of linear regression model is dependent upon the choice of procedure options. This paper focuses on the influences of sampling methods, sample size and variable formulation, where the stepwise regression with the use of bidirectional elimination (it is combined with forward selection and backward elimination, used as default) and BIC (Bayesian information criterion, it is an alternative method to AIC, to avoid overfitting; Wang et al., 2013), is produced by R statistic software (V2.15.0; 2012). Normally, the results of linear regression model are considered valid, when R^2 is larger than 0.7.

For sample size of 100 and above, the difference in the results from different sampling methods becomes slightly, it is feasible to use simple random sampling method and about 100 runs in global SA for typical building simulation applications (Macdonald, 2009; Lomas & Eppel, 1992). Thus, the global SA applied here is based on several different sample sizes:

- The first 100 and all of 1000 randomly generated solutions from Random Sample A;
- The first 100 and all of 1000 randomly generated solutions from Random Sample B;
- First 100 solutions obtained from a multi-objective optimization process (based on NSGA-II).

For most global SA, input variables are real-valued quantities (Helton, 1993). However, the rank transformation based on monotonic relationship can mitigate the problems associated with fitting linear models to nonlinear data. There are two representations of input variables are considered in this research:

- Input variables in their raw forms.
- A rank-transformation of variables (and outputs).

Optimization Algorithm

In this paper, the multi-objective optimization is carried out using the NSGA-II (non-dominated sorting genetic algorithm II). The specific implementation of NSGA-II is listed as below:

- Gray-coded bit-string encoding of the problem variables (163 bits);
- Binary tournament selection;
- Uniform crossover (100% probability of chromosome crossover with 50% probability of gene crossover);
- Single bit mutation (a probability of 1 bit per chromosome);
- Stochastic Ranking fitness assignment (with a 45% probability of infeasible solution bias);
- A population size of 20 with 5000 unique simulations.

Example Building and Performance Model

The example building is based on a mid-floor of a commercial office building with 5 zones located in Birmingham, UK (**as shown in Figure 1**). The size of two end zones and three middle zones are 78.7ft x 26.2ft (24m x 8m) and 98.4ft x 26.2ft (30m x 8m) separately, with floor to ceiling height of 8.9ft (2.7m). Each zone has typical design conditions of, 1 occupant per 107.6ft^2 (10m^2) and equipment loads of 3.7Btu/h*ft^2 (11.5W/m^2). Maximum lighting loads are set at 3.7Btu/h*ft^2 (11.5W/m^2), with the lighting output controlled to provide an illuminance of 500 lux at two reference points located in each of the perimeter zones. Infiltration is set at 0.1 air change per hour, and ventilation rates at 126.8Gpm (8L/s) per person. The heating and cooling is modeled by an idealized system that can provide sufficient energy to offset the zone loads and meet the zone temperature set points during hours of operation (from 9am to 5pm all year around). The internal zone is treated as a passive unconditioned space. It is simulated through EnergyPlus (V7; 2011a), with the weather data based on the CIBSE test reference year (CIBSE, 2002).

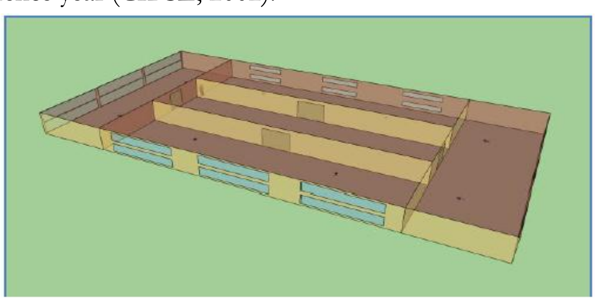

Figure 1 Example building (Wright et al., 2012)

Input Variables, Objective Functions and Design Constraints

Input variables. 16 input variables associated with perimeter zones are considered and optimized (**as shown in**

Table 1). The longest facades of the building face north (and south), when the Orientation is set at 0^o. Dead band is used to avoid an overlap of the heating and cooling set points. The window-to-wall ratio refers to the window area of 6 equal size windows placed in three groups against the wall area in each facade, where the names of variables reflect their positions in perimeter zones. The start and stop times are hours of the day. Three construction types are available for external wall and ceiling-floor: heavy weight, medium weight and light weight. Similarly, there are two internal wall types (heavy weight & light weight), and two double glazed windows types (plain glass & low-E glass). The alternative constructions are taken from the ASHRAE handbook (ASHRAE, 2005). For categorical construction variables, the heavy weight construction is corresponds to a value of 0, with the construction weight decreasing with increasing variable value.

Table 1. Input Variables

Index	Input Variables	Units	Lower Bound	Upper Bound	Increment
1	Heating setpoint	(^{o}F/^{o}C)	64.4^{o}F (18.0^{o}C)	71.6^{o}F (22.0^{o}C)	32.9^{o}F (0.5^{o}C)
2	Heating set-back	(^{o}F/^{o}C)	32.0^{o}F (0.0^{o}C)	46.4^{o}F (8.0^{o}C)	32.9^{o}F (0.5^{o}C)
3	Dead band	(^{o}F/^{o}C)	33.8^{o}F (1.0^{o}C)	41.0^{o}F (5.0^{o}C)	32.9^{o}F (0.5^{o}C)
4	Orientation	(^{o})	-90.0	90.0	5.0
5	North window-wall ratio	(-)	0.2	0.9	0.1
6	South window-wall ratio	(-)	0.2	0.9	0.1
7	East window-wall ratio	(-)	0.2	0.9	0.1
8	West window-wall ratio	(-)	0.2	0.9	0.1
9	Winter start time	(hrs)	1	8	1
10	Winter stop time	(hrs)	17	23	1
11	Summer start time	(hrs)	1	8	1
12	Summer stop time	(hrs)	17	23	1
13	External wall type	(-)	0	2	1
14	Internal wall type	(-)	0	1	1
15	Ceiling-floor type	(-)	0	2	1
16	Window type	(-)	0	1	1

Design objective. The design objectives are to minimize building annual energy demand (for heating, cooling and artificial lighting), and the capital costs (using a model derived from cost estimating data).

Design constraint. The design constraints are that the thermal comfort in each of the perimeter zones should not exceed 20% of predicted percentage dissatisfied (PPD), for no more than 150 working hours per annum. The constraint functions are configured to return the number of hours above 150, or zero if the constraint is feasible. The infeasibility of a solution is the sum of the squares of each constraint violation (an entirely feasible solution would have an infeasibility of 0).

RESULTS AND ANALYSIS

Tables 2 to 4 state the global sensitivities of variables, based on different sensitivity indicators, for energy demand (Table 2), capital costs (Table 3) and solution infeasibility (Table 4). In each case, the variables are represented in two forms: rank-transformed data and raw data with categorical variables. Each data form applied to global SA has five sets of solutions: the first 100 and all of 1000 randomly generated solutions from Random Sample A & B, and the first 100 solutions obtained from the multi-objective optimization process (based on NSGA-II). In those tables, the global sensitivity of a particular variable is measured in three ways: the order of entry into the linear regression model, the absolute value of SRC (or SRRC), and the size of R^2 changes attributable to individual variables; where the global sensitivity of the variable is indicated by the length of the bars: the more sensitive (important) the variable, the longer the bars of the SRC (or SRRC) and the size of R^2 change, the shorter the bar of entry-order.

From Tables 2 to 4, it can be seen that, in relation to the choice of sample size from a set of randomly generated samples, enlarging the sample size can bring more variables into the linear regression model, and resulting the changes in the relative magnitudes of variables global sensitivities, measured by different sensitivity indicators, for each of design objectives and constraints. For example, when the size of a set of rank-transformed data from Random Sample A is enlarged from 100 to 1000, the entry-order of Summer stop time into the energy demand model is delayed from No.3 to No.8; correspondingly, its SRRC (0.205) and contribution to R^2 change (0.032) are reduced to 0.122 and 0.0148 separately; and therefore, the rank-order of its importance is reduced as well (**as shown in Table 2 columns 2&3**). Furthermore, when the search space is obtained from different sets of random samples, but with same sample size and data form, the linear regression model constructed from the stepwise regression can also be different, identifying different numbers of important variables, especially when the sample size is smaller. For example, 3 more variables are identified from Random Sample B with sample size of 100 and the use of rank transformation, compared to those from Random Sample A with the same sample size and data form, for energy demand, where the SRRCs of the new-identified variables are approximately between 0.1 and 0.2 (ranked in the middle of importance) (**as shown in Table 2 columns 2&4**). In addition, those conclusions are difficult to be judged by the R^2 of the model that is always far better than 0.7 (the reliability level), particularly in the case of using rank transformation.

For a given set of samples, the data form of variables (with or without the use of rank transformation) has limited influences on the indication of variables global sensitivities, especially for the most important variables, even the problems associated with fitting linear models to nonlinear data tends to be mitigated by using rank-transformed data. For instance, in the analysis for solution infeasibility from random samples, the linearity of the regression model can be enhanced, through replacing the raw data by their ranks, indicated by the significantly increased R^2, from about 0.6 to 0.8 (**as shown in Table 4**). As the regression model with rank-transformed data is based on a monotonic relationship rather than a linear relationship. However, for the top two most important variables, Heating setpoint and Dead band, the indicators of their global sensitivities to solution infeasibility are slightly affected by the use of rank transformation, represented by the similar entry-orders, the relative magnitudes of variables standardized regression coefficients and the contributions to R^2 changes.

Thus, in a particular sensitivity analysis of design objectives or constraints to the changes of variables values, the impact on the uncertainty in each variable is varied in a small range, which is related to the choice of samples, sample size, and data form (particularly for the objective having weak linearity). But, the most important variables are always ranked on the top: just as the Heating setpoint and Dead band for the uncertainties of energy demand and solution infeasibility, and the Ceiling-floor and Window types for the uncertainty of capital costs. Furthermore, although there are little correlations (below 0.1) between variables from randomly generated samples, it is hard to conclude identical ordering of variables importance through three of sensitivity indicators. Particularly when variables are equally important, or have very close magnitudes of importance to outputs, their entry-orders are meaningless. Therefore, it is better to adopt both the entry-orders of variables (as the direct outcome from a stepwise regression) and the standardized (rank) regression coefficients or the contributions to R^2 changes, to provide robust orderings of variables importance for design objectives and constraints. In addition, re-generated different sets of samples could avoid misleading variables importance, particularly when the sample size is smaller.

Tables 2 to 4 also state that the first 100 solutions obtained from a multi-objective optimization process can be used to perform global SA, to identify the important variables for both energy demand and capital costs (design objectives), where the correlations between variables and outputs could be enhanced, but no variables are dropped out of the stepwise regression due to the correlations. The importance of this conclusion is that it is not necessary to generated separate random sample of solutions to inform variables importance, and then saving computational efforts. In particular for capital costs, the relative magnitudes of variables global sensitivities are close to those found from the random samples. As the capital costs objective is indirectly driven by most variables variations. For energy demand, even the relative magnitudes of variables standardized (rank) regression coefficients from the first 100 solutions of NSGA-II could be affected due to the correlations, the resulting rank-order of variables importance is similar or equal to those based on other sensitivity indicators (e.g. variables entry-orders or contributions to R^2 changes). Moreover, it is interesting to see the importance of Orientation is enhanced significantly at the start of NSGA-II search, which maybe suggest, the Orientation varied in a narrow value range is locally sensitive to the optimum solutions along the energy-cost trade-off; further research is required.

However, this is not the case of applying the same analysis for solution infeasibility. There are limited numbers of infeasible solutions (against thermal comfort) exist even at the start of the search to perform global SA, which is because of the stronger convergence behavior of NSGA-II.

CONCLUSION

This paper examines the extent to which the procedure options of a stepwise regression can influence the indication of variables global sensitivities, measured by three different sensitivity indicators (the entry-order of variables, the relative magnitudes of SRCs or SRRCs, and the size of R^2 changes attributable to individual variables), for both design objectives and constraints, when using the randomly generated samples and the biased solutions obtained from the start of a multi-objective optimization process (based on NSGA-II). It firstly concludes that, in a particular sensitivity analysis of design objectives or constraints, the impact on the uncertainty in each variable is varied in a small range, which is related to the choice of samples, sample size, and data form (particularly for the model having weak linearity). But, the most important variables are always ranked on the top no matter the choice of the procedure options: just as the Heating setpoint and Dead band for the uncertainties of energy demand and solution infeasibility, and Ceiling-floor and Window types for the uncertainty of capital costs. Furthermore, it is better to adopt both the entry-orders of variables and their SRRCs (or SRCs) or the contributions to R^2 changes, to provide robust orderings of variables importance for design objectives and constraints. The entry-order of variables is the direct outcome from a particular stepwise regression, but it is meaning less when variables have equal or very close magnitudes of importance to outputs. Moreover, re-generated different sets of samples could avoid misleading the importance of variables, particularly for the cases having smaller sample size.

This paper also concludes that, the first 100 solutions obtained from a multi-objective optimization can be used to perform global SA, to identify the important variables for design objectives, especially for capital costs. The importance of this conclusion is that it suggests that a separate random sample of solutions is not necessary to provide variables importance. But, this is not the case of applying the same analysis for solution infeasibility. There are limited numbers of infeasible solutions (against thermal comfort) exist even at the start of the optimization, due to the stronger convergence behavior of NSGA-II.

REFERENCES

ASHRAE, 2005. ASHRAE Handbook of Fundamentals, Chap.30, Tab.19 & Tab.22.

Ballarini, I. & Corrado, V., 2012. Analysis of the building energy balance to investigate the effect of thermal insulation in summer conditions. Energy and Buildings, 52, pp.168-180.

Breesch, H. & Janssens, A., 2005. Building simulation to predict the performances of natural night ventilation: uncertainty and sensitivity analysis, Pro. 9th Int. IBPSA Conf. 2005.

Brownlee, A.E.I. & Wright, J.A., 2012. Solution analysis in multi-objective optimization. Loughborough University, IBPSA-England.

CIBSE, 2002. Guide J: weather, solar and luminance data CIBSE London. ISBN: 978-1-903287-12-5.

EnergyPlus (V7), 2011. Available online: www.apps1.eere.energy.gov/buildings/energyplus/

Evins, R., 2013. A review of computational optimization methods applied to sustainable building design. Renewable and Sustainable Energy Reviews, 22, pp.230-245.

De Wilde, P. & Tian, W., 2009, Identification of key factors for uncertainty in the prediction of the thermal performance of an office building under climate change. Building Simulation, 2, pp.154-157.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Sys Safe, 42(2), pp.327-367.

Helton, J.C., Johnson, J.D., Sallaberry, C. & Strolie, C.B., 2006, Survey of sampling-based methods for uncertainty ans sensitivity analysis. Reliability Engineering & System Safety, 91, pp.1175-1209.

Hopfe, C.J. & Hensen, J.L.M., 2011. Uncertainty analysis in building performance simulation for design support. Energy and Building, 43, pp.2798-2805.

Iman, R.T. & Conover, W.J., 1982. A distribution-free approach to inducing rank correlation among input variables. Commun. Statist. Simul. Comput. B, 11, pp.311-334.

Kletting, P. & Glatting, G., 2009. Model selection for time-activity curves: The corrected Akaike information criterion and the F-test. Z. Med. Phys., 19 (3), pp.200-206.

Lomas, K.J. & Eppel, H., 1992. Sensitivity analysis techniques for building thermal simulation programs. Energy Buildings, 19(1), pp.21-44.

Macdonald, I.A., 2009. Comparison of sampling techniques on the performance of Monte Carlo based sensitivity analysis. Proc. Building Simulation 2009, pp.992-999.

R statistical software (V2.15.0), 2012. Available online: www.r-project.org/

Saltelli, A., Chan, K. & Scott, E.M., 2000. Sensitivity analysis. UK: Wiley & Sons Ltd.

Tian, W., 2013. A review of sensitivity analysis methods in building energy analysis. Renewable and Sustainable Energy Reviews, 20, pp.411-419.

Wang, M., Wright, J.A., Brownlee, A.E.I. & Buswell, R.A., 2013. A comparison of approaches to stepwise regression for global sensitivity analysis used with evolutionary optimization. Pro. 13th Int. IBPSA Conf. 2013.

Wright, J.A., Wang, M., Brownlee, A.E.I. & Buswell, R.A., 2012. Variable convergence in evolutionary optimization and its relationship to sensitivity analysis, Loughborough University. 2012 IBPSA-England.

Table 2. Variables Sensitivities for Energy Demand

Variables Index	Rank-Transformed Data					Raw Data				
	RandomSampleA 100	RandomSampleA 1000	RandomSampleB 100	RandomSampleB 1000	NSGA-II First100	RandomSampleA 100	RandomSampleA 1000	RandomSampleB 100	RandomSampleB 1000	NSGA-II First100
The Entry-Order of Variables										
1	2	2	2	2	2	2	2	2	2	2
2	7	6	10	7	8	7	6		7	8
3	1	1	1	1	1	1	1	1	1	1
4		13			3		13		13	4
5		11		11			12		11	9
6		9		9			9		9	
7	3	5	4	3	4	4	4	4	3	5
8		12					11		12	
9		10	9	10			10		10	
10										7
11	6	4	3	4	5	5	3	3	4	3
12	4	8	7	8		3	7	7	8	6
13		14								
14			8					8		
15		7	5	6	7		8	5	6	
16	5	3	6	5	6	6	5	6	5	
Variables SRRCs or SRCs										
1	0.40	0.44	0.44	0.47	0.54	0.40	0.45	0.43	0.48	0.57
2	0.13	0.13	0.10	0.11	0.07	0.09	0.13		0.12	0.08
3	0.59	0.67	0.68	0.68	0.88	0.61	0.66	0.65	0.68	0.83
4		0.05			0.31		0.04		0.04	0.22
5		0.05		0.05			0.05		0.04	0.08
6		0.12		0.11			0.12		0.12	
7	0.15	0.18	0.22	0.21	0.16	0.14	0.19	0.20	0.21	0.12
8		0.05					0.06		0.04	
9		0.06	0.13	0.08			0.06		0.07	
10										0.11
11	0.13	0.18	0.22	0.17	0.16	0.15	0.19	0.22	0.18	0.16
12	0.21	0.12	0.14	0.10		0.21	0.13	0.16	0.11	0.18
13		0.04								
14			0.09					0.10		
15		0.13	0.20	0.14	0.06		0.13	0.17	0.15	
16	0.19	0.19	0.19	0.17	0.11	0.14	0.18	0.18	0.17	
Individual R^2 Changes										
1	0.198	0.202	0.202	0.220	0.316	0.196	0.217	0.208	0.223	0.311
2	0.016	0.021	0.008	0.013	0.003	0.007	0.023		0.015	0.007
3	0.534	0.431	0.441	0.457	0.501	0.573	0.430	0.817	0.454	0.553
4		0.002			0.058		0.002		0.001	0.020
5		0.002		0.003			0.002		0.002	0.003
6		0.012		0.009			0.011		0.012	
7	0.035	0.034	0.045	0.039	0.042	0.028	0.038	0.036	0.038	0.013
8		0.002					0.003		0.002	
9		0.004	0.012	0.006			0.004		0.005	
10										0.008
11	0.012	0.035	0.051	0.029	0.018	0.017	0.040	0.044	0.030	0.029
12	0.032	0.015	0.020	0.011		0.039	0.015	0.022	0.013	0.013
13		0.001								
14			0.014					0.010		
15		0.017	0.028	0.020	0.003		0.016	0.025	0.021	
16	0.028	0.042	0.029	0.030	0.012	0.017	0.037	0.025	0.028	
R^2	0.86	0.82	0.85	0.84	0.95	0.88	0.84	0.82	0.84	0.96

Table 3. Variables Sensitivities for Capital Costs

Variables Index	Rank-Transformed Data RandomSampleA 100	Rank-Transformed Data RandomSampleA 1000	Rank-Transformed Data RandomSampleB 100	Rank-Transformed Data RandomSampleB 1000	Rank-Transformed Data NSGA-II First100	Raw Data RandomSampleA 100	Raw Data RandomSampleA 1000	Raw Data RandomSampleB 100	Raw Data RandomSampleB 1000	Raw Data NSGA-II First100
The Entry-Order of Variables										
1	10	10		12	12		11		10	
2				11			13		12	
3	9	9	8	9	4	10	9	9	9	11
4					11					9
5	4	3	2	3	6	4	3	2	3	5
6	3	4	5	5	8	3	4	6	4	6
7	5	5	6	4	3	5	5	5	5	3
8	8	7	4	7	5	6	6	4	6	8
9				13	13		12	10	13	10
10					9		14		14	12
11		11		10		11	10		11	13
12						9			15	
13	6	6	7	6	7	7	7	7	7	7
14	7	8		8	10	8	8	8	8	4
15	1	1	1	1	1	1	1	1	1	1
16	2	2	3	2	2	2	2	3	2	2
Variables SRRCs or SRCs										
1	0.04	0.02		0.02	0.01		0.02		0.02	
2				0.02			0.01		0.02	
3	0.04	0.04	0.06	0.04	0.09	0.03	0.04	0.04	0.04	0.05
4					0.06					0.06
5	0.15	0.16	0.20	0.16	0.13	0.14	0.15	0.16	0.15	0.14
6	0.14	0.15	0.10	0.14	0.14	0.12	0.14	0.12	0.13	0.19
7	0.15	0.14	0.09	0.14	0.18	0.13	0.13	0.13	0.13	0.18
8	0.05	0.09	0.13	0.09	0.16	0.08	0.09	0.12	0.09	0.13
9				0.01	0.02		0.01	0.02	0.01	0.03
10					0.08		0.01		0.01	0.05
11		0.02		0.02		0.02	0.02		0.02	0.02
12						0.03			0.01	
13	0.09	0.09	0.09	0.10	0.15	0.06	0.08	0.08	0.09	0.16
14	0.07	0.06		0.06	0.09	0.05	0.06	0.06	0.06	0.10
15	0.88	0.92	0.93	0.92	1.00	0.92	0.94	0.96	0.94	1.05
16	0.23	0.22	0.19	0.22	0.34	0.20	0.21	0.20	0.20	0.31
Individual R^2 Changes										
1	0.001	0.001		0.0003			0.0005		0.001	
2				0.0004	0.001		0.0001		0.0002	
3	0.002	0.001	0.004	0.002	0.022	0.001	0.001	0.002	0.001	0.001
4					0.002					0.002
5	0.023	0.026	0.051	0.024	0.010	0.019	0.022	0.037	0.020	0.011
6	0.031	0.024	0.010	0.019	0.007	0.028	0.022	0.011	0.016	0.019
7	0.018	0.018	0.007	0.019	0.054	0.015	0.016	0.016	0.017	0.038
8	0.003	0.008	0.022	0.009	0.013	0.004	0.007	0.017	0.009	0.012
9				0.0002	0.001		0.0001	0.0004	0.0001	0.001
10					0.002		0.0001		0.0001	0.001
11		0.0004		0.001		0.0004	0.001		0.0005	0.0002
12						0.001			0.0001	
13	0.007	0.009	0.008	0.010	0.008	0.004	0.009	0.006	0.008	0.008
14	0.003	0.004		0.004	0.003	0.003	0.003	0.003	0.003	0.020
15	0.853	0.837	0.836	0.843	0.725	0.886	0.873	0.869	0.879	0.758
16	0.040	0.046	0.026	0.047	0.139	0.033	0.040	0.032	0.039	0.127
R^2	0.98	0.97	0.96	0.98	0.99	0.99	0.99	0.99	0.99	0.99

Table 4. Variables Sensitivities for Solution Infeasibility

Variables Index	Rank-Transformed Data RandomSampleA 100	Rank-Transformed Data RandomSampleA 1000	Rank-Transformed Data RandomSampleB 100	Rank-Transformed Data RandomSampleB 1000	Rank-Transformed Data NSGA-II First100	Raw Data RandomSampleA 100	Raw Data RandomSampleA 1000	Raw Data RandomSampleB 100	Raw Data RandomSampleB 1000	Raw Data NSGA-II First100
The Entry-Order of Variables										
1	1	1	1	2		1	2	2	2	
2										
3	2	2	2	1		2	1	1	1	
4					5	4				
5		5	4	6	1					1
6		6		5	4					
7	3	3	3	3	3		4		3	
8										3
9		9							6	
10									8	
11		7	5	7		5	3		4	
12	4	8		8			5		7	
13										
14										
15		10		9						
16		4		4	2	3			5	2
Variables SRRCs or SRCs										
1	0.69	0.63	0.66	0.65		0.56	0.55	0.52	0.55	
2										
3	0.52	0.62	0.53	0.62		0.44	0.55	0.54	0.57	
4					0.38	0.16				
5		0.10	0.14	0.10	0.38					0.42
6		0.09		0.10	0.14					
7	0.17	0.15	0.18	0.15	0.31		0.06		0.10	
8										0.32
9		0.04							0.07	
10									0.05	
11		0.07	0.11	0.09		0.15	0.09		0.09	
12	0.08	0.06		0.04			0.06		0.07	
13										
14										
15		0.04		0.04						
16		0.10		0.11	0.25	0.16			0.07	0.34
Individual R^2 Changes										
1	0.61	0.40	0.48	0.38		0.38	0.30	0.27	0.30	
2										
3	0.23	0.38	0.28	0.41		0.21	0.31	0.29	0.31	
4					0.07	0.02				
5		0.01	0.02	0.01	0.16					0.08
6		0.01		0.01	0.04					
7	0.03	0.02	0.04	0.02	0.08		0.004		0.01	
8										0.10
9		0.002							0.005	
10									0.003	
11		0.01	0.01	0.01		0.02	0.01		0.01	
12	0.01	0.003		0.00			0.003		0.005	
13										
14										
15		0.002		0.001						
16		0.01		0.01	0.09	0.03			0.01	0.09
R^2	0.88	0.84	0.83	0.85	0.45	0.67	0.62	0.56	0.64	0.28