Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Thermal Engineering 25 (2005) 283–293
APPLIED THERMAL ENGINEERING
www.elsevier.com/locate/apthermeng

Experimental design methods and flowsheet synthesis of energy systems

Tor-Martin Tveit *

Energy Engineering and Environmental Protection, Helsinki University of Technology, P.O. Box 4400, FIN-02015 HUT, Finland

Received 25 January 2004; accepted 23 May 2004
Available online 6 July 2004

Abstract

This work presents a discussion of how to utilise well known methods from the field of experimental design in the flowsheet synthesis of energy systems. The work is based on an earlier work, where a methodology for improving large scale energy systems using a combination of simulation, experimental design and mathematical programming was presented. The methodology is suitable for synthesis of large scale and complex problems with few degrees of freedom. A conceptual bio-fuel indirectly fired micro-turbine is used to illustrate how a simulation model and experimental design can be used to build an optimisation model of an energy system, and how this model performs compared to a model, where all the units are modelled in detail. The new methodology has a good potential for reducing the optimisation problem and subsequently for solving more complex problems than the traditional methods are able to solve. For the example problem presented in this work, the developed model performs at least as well as the MINLP model with all units modelled in detail. However, as the degrees of freedom of the system increase, the value of the new methodology decreases, since more computational effort is needed to obtain a representative regression model.

Keywords: Experimental design; Energy systems; Flowsheet synthesis

* Tel.: +358-9-451-3635; fax: +358-9-451-3418.
E-mail address: tor-martin.tveit@hut.fi (T.-M. Tveit).

1359-4311/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.applthermaleng.2004.05.015

Nomenclature

A, B, C, K	factors
a, b, c	high level for factors A, B and C respectively
n	number of replications of an experiment
p	pressure (bar)
R^2	the fraction of the variance in the data that is explained by a regression
T	temperature (°C)
y	response variable or binary variable

Greek symbols

β_n	the nth regression model coefficient
η_{el}	electrical efficiency

Superscript

k	number of factors

1. Introduction

The problem of flowsheet synthesis of energy systems is to find the optimal configuration of a flowsheet and its operating conditions. This is a difficult and challenging task. The work presented in this article is based on the earlier work [1], where a methodology for improving large scale energy systems using a combination of simulation, experimental design and mathematical programming was introduced. The current article gives a more detailed description of some classical symmetrical experimental designs and a discussion about how these can be utilised in the earlier developed methodology is also presented. This is illustrated using a conceptual bio-fuel indirectly fired microturbine process. The developed optimisation model is compared to a more traditional model where all the units are modelled in detail, as in the work by Bruno et al. [2] and by Manninen [3]. The comparison shows that the model developed using the new methodology performs very well compared to the detailed model.

Mathematical programming has proved to be well suited for solving flowsheet synthesis problems, and mixed-integer non-linear programming (MINLP) models have been developed extensively for many years. A good example related to the flowsheet synthesis of power plants is the work by Bruno et al. [2], where a rigorous MINLP model for structural and parametric optimisation of utility plant was developed. By using their formulations it is possible to solve some power plant synthesis problems of practical size. An overview of the optimisation techniques for process system engineering can be found in the work by Biegler and Grossmann [4]. However, in order to get a realistic model of the system, often non-linearities and combinatorial complexities will be included in the models. For many problems, this will often result in models for which no efficient optimisation algorithm exist. One alternative is to improve the algorithms. Regarding the non-linearities, most efficient algorithms require that the functions are convex in order to guarantee a global optimum solution. Much work has been done to improve these algorithms, for instance Westerlund et al. [5] developed a MINLP algorithm, where global optimum can be guaranteed for models including pseudo-convex functions. Examples of global optimisation algorithms are the algorithms developed by Ryoo and Sahinidis [6] and by Adjiman et al. [7], which are both based on branch-and-bound strategies. Unfortunately these algorithms are still not efficient enough to solve many flowsheet synthesis problem of sizes relevant to cases in the industry. Another approach is to avoid the problems by trying to reduce the problem. In their work Manninen and Zhu [8] tried to reduce the MINLP problem size by using thermodynamic analysis. The problem was reduced by giving tighter bounds on the variables as well as reducing the size of the superstructure. Similar ideas were presented by Hostrup et al. [9] where the focus was on the synthesis of flowsheets for chemical processes. This paper is based on the work by Tveit [1], where the mathematical problem size was reduced by using a regression model of the behaviour of the system. The regression model of the system's behaviour is made based on the simulation model and experimental design. This simplified model is then used in a mathematical programming model including a superstructure and cost functions. The methodology is suitable for synthesis of large scale and complex problems with few degrees of freedom. The purpose of the work presented in this article is to develop this methodology further and to illustrate using an example how the methodology compares with a more traditional model, where all units included in the system are modelled in detail.

2. Illustrative example

A conceptual bio-fuel indirectly fired microturbine is used to illustrate how a simulation model and experimental design can be used to build an optimisation model of an energy system and how this model performs compared to a model where all the units are modelled in detail. The example was chosen so that it contains both linear and non-linear relations.

The basic principle of the bio-fuel indirectly fired microturbine is that air is compressed and then heated by the flue gas from a bio-fuel powered burner. The air is then expanded to atmospheric pressure while producing work that is converted to electricity. The air is still hotter than the environment and part of the air is used as the required combustion air in the burner. The flue gas is cooled further down by heating water for heating purposes. Similar ideas have been developed by for instance by Eidensten et al. [10]. A simulation model of the process can be seen in Fig. 1 and more details about the simulation model can be found in the report by Tveit [11]. In this case the electricity output should be about 200–250 kW. This is achieved by keeping the air flow through the turbine fixed at 1.0 kg/s. [1] The fuel is wood with a heating value of 9.1 MJ/kg. The water content of the fuel is 55 wt.%. The purpose of this work is to demonstrate the use of experimental design combined with simulations and mathematical programming, and not to give a through description of an indirectly fired microturbine. Several important factors regarding the indirectly fired gas turbine process, for instance the design of the heat exchanger and pressure losses, are thus not discussed here.

The simulation model was developed for a project which goal is to design a better heat exchanger which allows higher temperatures. It should be noted at this point that the operating

[1] W_{el} ($p = 5$ bar, $T = 900$ °C) = 180 kW and W_{el} ($p = 10$ bar, $T = 1050$ °C) = 253 kW.

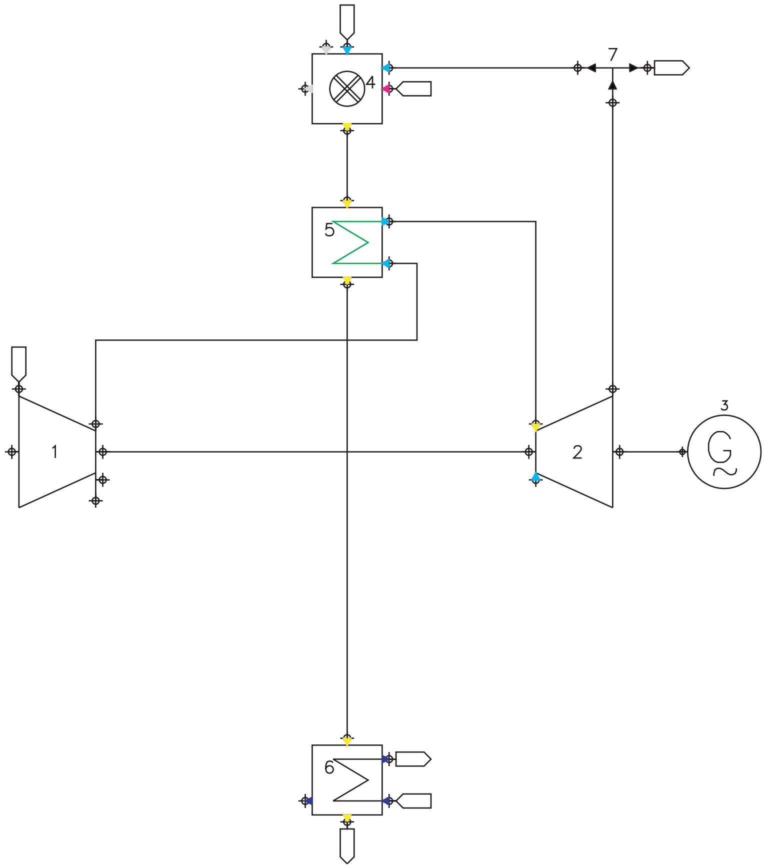

Fig. 1. Bio-fuel indirectly fired microturbine; compressor (1), expander (2), generator (3), burner (4), gas–gas heat exchanger (5), gas–water heat exchanger (6) and splitter (7).

temperatures for the microturbine used in this paper are therefore higher than what is economically feasible today, and that the prices are restricted to only the relevant components, excluding for instance installation costs. However, the purpose of the study is to show how to help decision makers choose the correct pair of compressor and expander for the process to achieve a desired electrical efficiency of the system. This results in a model, where both parametrical (temperature and pressure) and structural (choice of compressor/expander pair) must be considered. The different compressors and expanders can operate at different pressures and temperatures as shown below:

- Compressor/expander alternative 1;
  - Operating conditions: $p = 5 \ldots 7$ bar and $T = 700 \ldots 970$ °C
  - Cost: 1000 €
- Compressor/expander alternative 2;
  - Operating conditions: $p = 5 \ldots 8$ bar and $T = 700 \ldots 1000$ °C
  - Cost: 1200 €

- Compressor/expander alternative 3;
  - Operating conditions: $p = 7 \ldots 10$ bar and $T = 700 \ldots 1050$ °C
  - Cost: 2000 €

where p is the pressure after the compressor and T is the temperature before the expander. The task of the optimisation model is to minimise the costs of the system given a specific electrical efficiency. In this case the electrical efficiency depends on the temperature and pressure before the expander.

The first step is to build a traditional MINLP-model, where all units in the system is modelled in detail. However, the MINLP-model will not be as detailed as the simulation model. The model consists of sets of units and streams, their connections and the mass and energy balances for each unit. The compressor and expander are modelled using an ideal gas model with relative pressure and isentropic efficiencies. The composition of the flue gas is fixed and the adiabatic combustion temperature is calculated using the lower heating value of the fuel. The discontinuities due to the upper and lower bounds' dependence on the choice of compressor/expander pair are modelled using *big-M*-formulations. The model consists of 89 equations and 119 variables, of which three are discrete. The full mathematical formulation of the model can be found in the report by Tveit [11].

The next step is to build an MINLP-model using the new methodology. The discontinuities related to the choice of compressor/expander pair will be modelled in the same way as in the traditional MINLP-model. The main task will be to develop a regression model where the electrical efficiency is related to the pressure after the compressor, p, and the temperature, T, before the expander.

$$\eta_{\mathrm{el}} = f(T, p) \tag{1}$$

where p is the pressure after the compressor and T is the temperature before the expander. To be able to compare and evaluate the regression models developed using the factorial design and the hybrid composite design, 100 simulation runs were made. For more complex problems this kind of data would not be available, and it is presented here only as an illustration.

2.1. Experimental designs

A crucial step in the methodology is to generate the regression model based on a series of simulation runs. The simulation runs should be chosen in such a way that the time it takes to complete the runs is as short as possible, but so that the regression model will be a reliable model of the system. In the case where it is possible to use a symmetrical design, the 2^k factorial design is a well known and mature technique in experimental design. A good introduction can be found in the book by Montgomery [12] or in the book by Dougherty [13]. In the 2^k factorial design, the effect of k factors are investigated. All factors are at two different levels (*high* and *low*). The factors are varied together, instead of one at the time. This makes it possible to investigate the individual effects of the factors as well as the effect of the interaction between the factors. The 2^3 design has three factors, labelled A, B and C. A geometrical interpretation of the design is shown in Fig. 2(a).

The lower-case letters in the figure correspond to the combinations, where at least one of the factors is at high level. (1) corresponds to the case where all factors are at the low level. The point

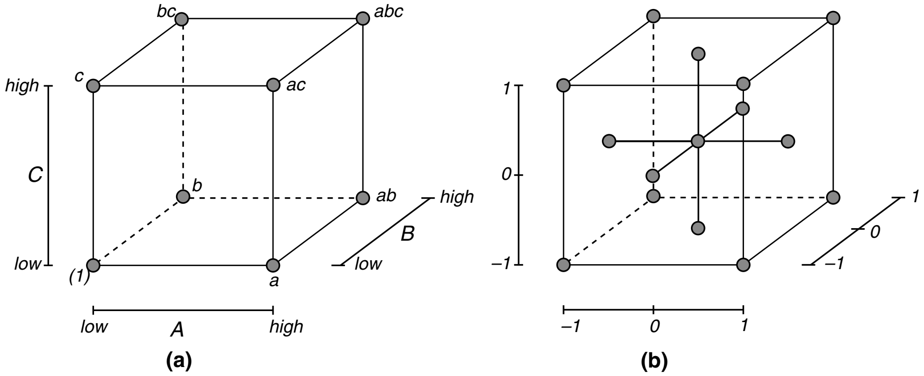

Fig. 2. Examples of two designs (a) The 2^3 factorial design-three factors A, B and C all at two levels (*high* and *low*). (b) The face-centred composite design-three factors.

ac would for example mean that the factors A and C are both at high level and the factor B is at low level. The actual high and low values are arbitrary. For instance in the example below the pressure, p, has 10 bar as the high value and 5 bar as the low. Similarly the temperature, T, has low and high values of 900 and 1050 °C respectively. Letting $A = p$ and $B = T$, the point *ab* would then be $(p, T) = (10\ bar, 1050\ °\mathrm{C})$.

It is usual in relation to the 2^k design to use *effect contrasts*. The effect of a factor is defined to be the change in response produced by a change in the level of the factor, i.e. for the effect A in a 2^2 factorial design, the effects that the factors B and C have on the response are not studied, and subsequently B and C are said not to be included. Any contrast, $[K]$, can be found by expanding the expression given in Equation (2).

$$[K] = (a \pm 1)(b \pm 1) \cdots (l \pm 1) \tag{2}$$

using a minus sign if the factor is included in the effect and a plus sign otherwise. Having found the contrast any average effect, K, can be found using the expression

$$K = \frac{[K]}{2^{(k-1)}n} \tag{3}$$

where n is the number of replications of the experiment. Returning to the 2^3 factorial design, the contrast for $K = AC$ is given by

$$[AC] = (a - 1)(b + 1)(c - 1) = abc - ab + ac - a - bc + b - c + (1) \tag{4}$$

The average effect, AC, is then given by

$$AC = \frac{[AC]}{2^2 n} \tag{5}$$

When the effects are calculated, a regression model can be built. For the 2^2-factorial experiment the regression model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \tag{6}$$

where y is the response, x_1 and x_2 are coded variables that can take values between −1 and 1 and ϵ is the error. For example x_1 is defined as

$$x_1 = \frac{A - \left(A_{\mathrm{low}} + \frac{A_{\mathrm{high}} - A_{\mathrm{low}}}{2}\right)}{\left(A_{\mathrm{low}} + \frac{A_{\mathrm{high}} - A_{\mathrm{low}}}{2}\right)} \tag{7}$$

The first coefficient, β_0, is the average of the response for the observations. The other coefficients are the half of the relevant effect, e.g. $\beta_{12} = \frac{AB}{2}$.

For example; using a 2^2 factorial design for developing a regression model of the effect of pressure and temperature on the electrical efficiency of the process shown in Fig. 1 yields the regression model shown in Eq. (8).

$$\eta_{\mathrm{el}}(p,T) = 0.2655 + 0.0108 \cdot \left(\frac{p-7.5}{2.5}\right) + 0.0352 \cdot \left(\frac{T-975}{75}\right) + 0.0052 \cdot \left(\frac{p-7.5}{2.5}\right) \cdot \left(\frac{T-975}{75}\right) \tag{8}$$

where the expressions $(\frac{p-7.5}{2.5})$ and $(\frac{T-975}{75})$ correspond to the coded variables x_1 and x_2 respectively. The pressure, p, varies between 5 and 10 bar and the temperature, T, between 900 and 1050 °C. In this case the error of the regression model at the simulated values is zero. For a more rigorous estimation of the error-function (ϵ), more data is needed.

The 2^k design assumes that the factor effects are linear, even if the interaction terms can model curvature to a certain degree, i.e. twisting the plane. One test of linearity is possible if a *centre-point* is added to the 2^k-factorial design. A comparison between the 100 simulated data points and the points calculated using the regression model in Eq. (8) gives an R^2-value of 0.983. The difference between the centre-point and the average of the factorial points is 0.0085 (i.e. 3.2%), thus indicating a slight curvature. If a linear model is sufficient then the difference of the response between the centre point and the average of the response at the factorial point is small. In the case where the curvature cannot be modelled accurately enough by Eq. (6), one alternative is the *second-order response model*. A second-order response surface model for the example can be formulated as

$$y(x_1,x_2) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_{12} \cdot x_1x_2 + \beta_{11} \cdot x_1^2 + \beta_{22} \cdot x_2^2 \tag{9}$$

The 2^k-factorial design does not produce enough data to fit the parameters in Eq. (9), so more data is needed. There are many different ways of obtaining more data. In the illustrative example in this work the feasible operating point of the system is a square limited by the variables' upper and lower bounds, this means that the region of interest is a square and the *face-centred central composite design* can be used. Fig. 2(b) shows a graphical interpretation of the design. In this case the domain is 2-dimensional, and the design can be extended without generating an excessive amount of simulation runs. The resulting hybrid composite design has 9 points including a centre point. The resulting regression model is shown in Eq. (10).

$$\begin{aligned}\eta_{\mathrm{el}}(p,T) = {} & -1.153 \times 10^{-1} - 5.596 \times 10^{-3} \cdot p + 3.130 \times 10^{-4} \cdot T + 2.357 \times 10^{-5} \cdot pT \\ & - 8.673 \times 10^{-4} \cdot p^2\end{aligned} \tag{10}$$

where β_{22} is equal to 0.

The purpose of the experimental design stage is to reduce the amount of simulation runs and still get good regression models. However, the bio-fuel indirectly fired microturbine was chosen to be simple enough to allow a comparison between the models developed using experimental design methods with a large amount of data points from the simulation model as well as with a model developed using this generous amount of data. In this case the number of simulated data points is 100. A comparison between the 100 simulated data points and the points calculated using the regression model in Eq. (10) gives an R^2-value of 0.996. The second order response surface with the coefficients fitted using the 100 data points has an R^2-value of 0.999. The closer the R^2-value is to unity, the better the variance in the data is explained by the model. The model is shown in Eq. (11).

$$\begin{aligned}\eta_{\mathrm{el}}(p,T) = {} & -9.544\times 10^{-2} - 2.386\times 10^{-3}\cdot p + 2.684\times 10^{-4}\cdot T + 2.736\times 10^{-5}\cdot pT \\ & - 1.351\times 10^{-3}\cdot p^2\end{aligned} \tag{11}$$

The difference between the values calculated using Eq. (8) and the simulation runs is in the range of −3.2% and 0.0%. The average difference is −1.9% and the standard deviation 0.01. The similar ranges for Eqs. (10) and (11) are [−1.7%, 1.0%] and [−1.1%, −0.1%] respectively. The average difference for Eq. (10) is −0.5% with a standard deviation of 0.006. For Eq. (11) the average difference is −0.6% with a standard deviation of 0.002.

Of the two models developed using the experimental designs, Eq. (10) is the best model and will be used in the MINLP model. This is due to the non-linearity behaviour of the system. It can be seen that since the function being modelled is concave, the model (Eq. (8)) is in fact an underestimation of the function. This kind of information is generally not available, and decisions about the best model must be based on information about curvature and similar indications of the system's behaviour. In general linear models are preferred in mathematical programming models, since the resulting optimisation models are easier to solve. However, if, as in this case, the behavior of the system is inherently non-linear, a non-linear model might be preferable.

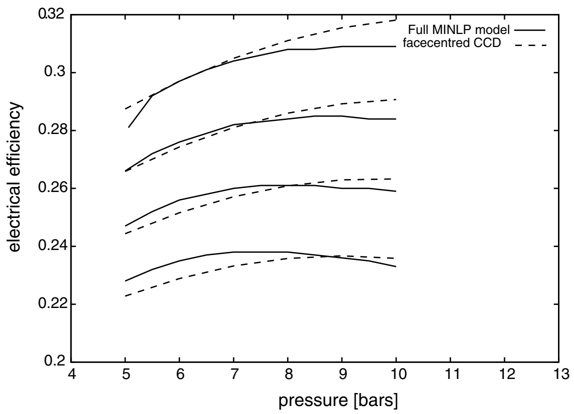

Fig. 3. Comparison of the contour plots of the electrical efficiency as a function of the pressure and temperature for the complete MINLP model and the regression model in Eq. (10) (second order response surface). The temperatures are 900, 950, 1000 and 1050 °C, where the curves with the lowest electrical efficiencies correspond to a temperature of 900 °C.

Now that the regression model is developed, the MINLP model can be formulated. The MINLP model based on the regression model consists of 19 equations and 14 variables, of which 3 are binary. The full mathematical formulation of the model can be found in the report by Tveit [11].

Fig. 3 shows a comparison of the contour plots of the electrical efficiency as a function of pressure and temperature between the regression model (Eq. (10)) and the MINLP model where all units are modelled in detail.

3. Results and discussion

The two MINLP models are implemented in the General Algebraic Modeling System (GAMS) by GAMS Development Corporation, and solved using the MINLP solver DICOPT by the group of Prof. Ignacio Grossmann at Carnegie Mellon University. The LP solver was OSL from IBM and the NLP solver was MINOS5 developed by the Systems Optimization Laboratory at Stanford University. The models were solved on a SGI Origin 2000 with 560 GB of RAM and 128 MIPS R12000 processors. The two models were solved repeatedly minimising the costs while gradually changing the electrical efficiency from 0.23 to 0.31. The results are shown in Fig. 4. The solving time for the full MINLP was reported by GAMS to be 0.010 and 0.000 s for the regression based model. As can be seen from the figure, the difference between the two results is small. There are only two narrow regions in the shift to and from the second alternative where the two models give different results. For $\eta_{el} = 0.266\ldots0.268$ the regression based model gives 1200 € as minimum costs, while the full MINLP model gives 1000 €. For $\eta_{el} = 0.284\ldots0.286$ the regression based model gives 1200 € as minimum costs, while the full model gives 2000 €. Even if the results are quite similar, there is a considerable difference in the size and numerical complexity of the models. In general it is not trivial to compare the computational effort needed to solve two optimisation models. However, a good rule of thumb is that the more variables (in particularly integer/binary variables) and non-linearities the more complicated the problem. The MINLP model, where all

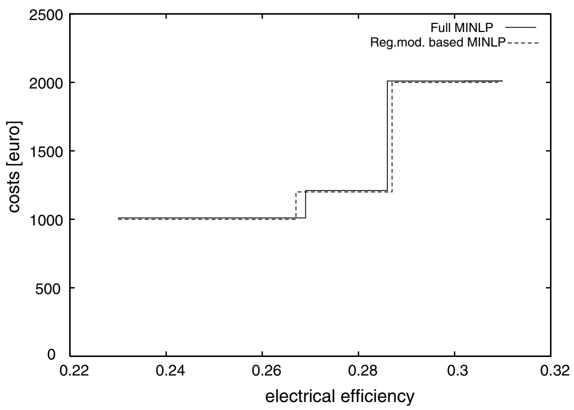

Fig. 4. Minimum costs as a function of the electrical efficiency (η_{el}) for the full MINLP and the regression model based mathematical programming model.

the units are modelled in detail, has about 4.5 times more equations and 8.5 times more variables than the MINLP model based on the regression model. In addition to this, the full model has 75 non-zero non-linear entries in the problem matrix, compared to only two for the MINLP model based on the regression model. It should be noted that the number of simulations needed to get a good regression model will increase with the necessary number of independent variables. For example for a 2^k-factorial design the number of simulations increases exponentially with the number of independent variables. Based on the above, it can be concluded that in this case the MINLP model developed using the new methodology considerably reduces the size of the optimisation problem. Thus, the new methodology has a good potential for reducing the optimisation problem and subsequently for solving more complex energy system flowsheet synthesis problems than the traditional methods. However, as the degrees of freedom of the system increase, the value of the new methodology decreases, since more computational effort is needed to obtain a representative regression model.

Acknowledgements

Professor Carl-Johan Fogelholm is gratefully acknowledged for supporting the work. Funding for this work was provided by Nordic Energy Research and TEKES (Finnish National Technology Agency). The Finnish IT center for Science (CSC) is acknowledged for providing access to their SGI Origin 2000 for solving the MINLP problems.

References

[1] T.-M. Tveit, A methodology for improving large scale thermal energy systems, Applied Thermal Engineering 24 (2003) 515–524.
[2] J.C. Bruno, F. Fernandez, F. Castells, I.E. Grossmann, A rigorous MINLP model for the optimal synthesis and operation of utility plants, Institution of Chemical Engineers Trans IChemE 76 (1998) 246–258.
[3] J. Manninen, Flowsheet synthesis and optimisation of power plants, Ph.D. thesis, University of Manchester Institute of Science and Technology, May 1999.
[4] L.T. Biegler, I.E. Grossmann, Retrospective on optimization, Computers and Chemical Engineering 28 (8) (2004) 1169–1192.
[5] T. Westerlund, H. Skrifvars, I. Harjunkoski, R. Porn, An extended cutting plane method for a class of non-convex minlp problems, Computers and Chemical Engineering 22 (1998) 357–565.
[6] H.S. Ryoo, N.V. Sahinidis, Global optimization of nonconvex nips and minlps with applications in process design, Computers and Chemical Engineering 19 (1995) 551–566.
[7] C.S. Adjiman, I.P. Androulakis, C.D. Maranas, C.A. Floudas, A global optimization method, αBB, for process design, Computers and Chemical Engineering 20 (1995) 419–424.
[8] J. Manninen, X.X. Zhu, Thermodynamic analysis and mathematical optimisation of power plants, Computers and Chemical Engineering 22 (1998) S537–S544.
[9] M. Hostrup, R. Gani, Z. Kravanja, A. Sorsak, I.E. Grossmann, Integration of thermodynamic insights and MINLP optimization for the synthesis, design and analysis of process flowsheets, Computers and Chemical Engineering 25 (2001) 73–83.
[10] L. Eidensten, J. Yan, G. Svedberg, Biomass externally fired gas turbine cogeneration, Journal of Engineering for Gas Turbines and Power 118 (1996) 604–609.

[11] T.-M. Tveit, Two optimisation models for a conceptual bio-fuel indirectly fired microturbine, Energy Engineering and Environmental Protection Publications ISBN951-22-7137-0, Helsinki University of Technology, 2004, Available from <http://eny.hut.fi/library/publications/tkk-eny/tkk-eny-16.pdf>.
[12] D.C. Montgomery, Design and Analysis of Experiments, fifth ed., John Wiley & Sons Inc, 2001.
[13] E.R. Dougherty, Probability and Statistics for the Engineering, Computing and Physical Sciences, Prentice Hall International, Inc, 1990.