CHEMICAL ENGINEERING TRANSACTIONS
VOL. 35, 2013
A publication of AIDIC
The Italian Association of Chemical Engineering
www.aidic.it/cet
Guest Editors: Petar Varbanov, Jiří Klemeš, Panos Seferlis, Athanasios I. Papadopoulos, Spyros Voutetakis

ISBN 978-88-95608-26-6; **ISSN** 1974-9791
DOI: 10.3303/CET1335205

A Methodology for Creating Sequential Multi-Period Base-Case Scenarios for Large Data Sets

Stéphane L. Bungener[a,*], Greet Van Eetvelde[b], François Maréchal[a]

[a]Industrial Process and Energy System Engineering (IPESE), Institute of Mechanical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland
[b]Environmental and Spatial Management, Faculty of Engineering and Architecture, Ghent University, Vrijdagmarkt 10-301, B-9000, Gent, Belgium
stephane.bungener@epfl.ch

Key performance indicators in engineering problems include but are not limited to financial, operational, management and environmental factors, which are significantly affected by aspects such as seasonality, fouling, economic climate, production rates, supply and demand. The search for an optimal solution to a problem must take into consideration this variability, otherwise running the risk of critical dimensioning or cost estimation errors.
Testing solutions using full data sets covering large periods of time can be a computational challenge, and the analysis of results complicated. For the feasibility of such a study, it is therefore necessary to reduce the large data sets to a number of base case scenarios, which simultaneously reduce the number of data points to be handled while still representing the variability of the system. A novel method is therefore developed to address this problem.
This method offers a way of designing an index of sequential periods common to each production level, which when averaged accurately represent periods of nominal values for each level. The method exploits a multi-objective evolutionary algorithm, minimising the standard deviation of the base cases compared to the real data as well as respecting crucial null value periods. Null value periods are typically found in turnarounds or supply and demand problems and are usually incorrectly represented in other methods. Lastly, the resulting base cases are sequential periods, which is important when dealing with scheduling, shutdown or storage problems. The method is tested using anonymised data and is compared to previously existing methods, with results showing improvement in the performance of the base cases with respect to the objective functions.

1. Introduction

Large scale industrial systems are most often strongly influenced by time. This influence can be seen on the production rates of end products, feedstock consumptions or availabilities, energy market seasonality and others. The variability of operations over time must be taken into consideration when studying optimisation, retrofit, production planning or maintenance problems as demonstrated by Pan et al (2012).
When desiring to reduce the number of studied time periods by studying base case scenarios, a problem arises. The profiles follow a logic easy to understand by operators or engineers of specific processes. However, when looking at multiple profiles, be they correlated or not, it can be difficult to obtain a clear picture of the temporal relations and similarities of the considered profiles. Furthermore, the stochastic and unpredictable nature of such profiles adds an additional dimension when it comes to defining a typical operation profile for project evaluation.
When looking for solutions concerning such time influenced problems, engineers must take into consideration the variation of profiles over long periods, such as years, rather than looking at limited punctual data. Otherwise the engineer risks proposing suboptimal or under/oversized solutions. Studying the 365 individual days of a year for several years is too complicated from a results generation and

Please cite this article as: Bungener S., Van Eetvelde G., Marechal F., 2013, A methodology for creating sequential multi-period base-case scenarios for large data sets, Chemical Engineering Transactions, 35, 1231-1236 DOI:10.3303/CET1335205

interpretation point of view; therefore a way has to be found to reduce the data while maintaining a high level of detail.
A simple solution is to average large periods of time, for example one or several years, as can be seen in Figure 1. This solution is however not satisfactory as it removes too much information, under/overestimating production, and completely ignoring days of zero flow rates.
In fact, in design or retrofit problems, in order to decide on an investment, the engineer must define the sizes of equipments that define the investment and estimate the profit of the solution.
The often neglected zero flow days are caused by turnarounds, shutdowns, resource or environmental constraints and many others. They can critically contribute to or disturb operations of systems regardless of size. Planification, logistics and other problems depending on multiple variables has to be sure to take these days into consideration in order to avoid unexpected consequences. These frequent days of zero flow rates exist in all systems, in a small unit's operations, in a business park's energy consumption, or in a petrochemical site's use of resources and production rates. Gross averages entirely neglect these zero flow days. Another solution is to break down the year into several periods, creating base cases instead of single conditions. The question then becomes how to best choose these periods when considering multiple profiles. Monthly averages will keep a certain level of detail, though they are unlikely to protect days of zero flow as can be seen in figure 1 where averaged values are never equal to zero when they should be. The idea of creating an index of segments of nominal values over the year does remain attractive though.

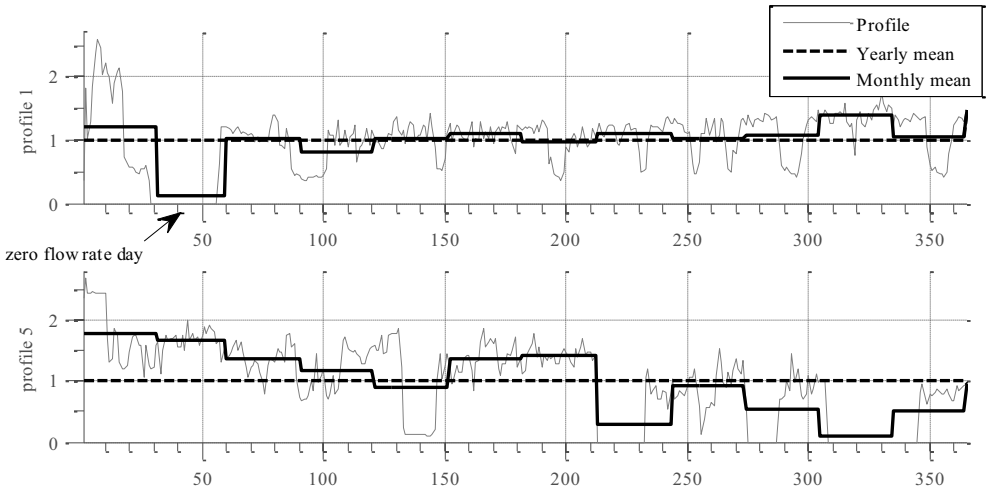

Figure1: Yearly and monthly mean values of two selected production profiles.

Wallace and Fleten (2003) carry out a detailed study of the different stochastic programming models in energy related problems. These address the stochasticity of variables without really addressing the longer term time problem. The discretisation and parameterisation of load duration curves is also covered by Wallace and Fleten, as well as by Poulin et al (2008). These curves are very adapted for studying uni-dimensional problems, but as soon as more than one profile is considered, as is the case in the process industry for example, they loose their use. Furthermore when storage problems are considered, the load duration curves cannot be used as the continuity of sequences in time becomes important. The notion of typical days has been explored by Dominguez et al. (2011), in which a k-medeoids algorithm is used to cluster days into typical representative days. Using this method one is not able to choose the number of segments being studied. Continued work by Fazlollahi et al (2012) defined a general methodology for the creation of sequential typical days, applicable to storage problems, also using the k-medeoids method. Maréchal and Kalitventzeff (2003) studied a non-linear minimisation problem in order to break a year down into periods while minimising the difference between real and averaged values of the profiles, through the choice of an optimal number of periods and data sets to consider. Another solution would be to examine the entire solution space, that is to say all the possible ways of considering number segments in a given time-period. The number of indexes to analyse very quickly becomes overwhelming as the number of segments increases. For example, there are 1.02 x 10^19 ways to choose 10 segments in a year.
For these reasons, this paper proposes an efficient methodology for optimally choosing a way to break down the year into limited numbers of segments over which the profiles can be averaged. Two performance indicators are defined to judge the quality of the results, the first one pertaining to the standard deviation between the averaged and real profiles and the second to the respect of the zero flow

days. This method exploits a genetic Evolutionary Multi-Objective Optimisation (EMOO) (Leyland 2002) to find the best solution based on both objectives. The method is tested in a case study using 5 normalised production profiles and compared to the results of a Monte-Carlo sampling evaluation and a simple monthly averaging.
An added advantage of this method is that it creates sequential elements which can therefore be used in scheduling and storage problems, a particularly important feature to take into consideration in business parks, district management and industrial clusters.

2. Methodology

2.1 Data preparation

The first step in data preparation is obviously to gather the data which one wishes to study, in the form of k profiles p_j, for a given time duration. This data should then be normalised and/or weighted by a factor ω_j as can be seen in eq(1), creating normalised profile d_j. If each profile is considered equally important, the easiest solution is to set the weighting factor at 1. Certain parameters must also be defined, namely the EMOO initial and total population counts, the tolerance level for the zero day definitions described in section 2.5 and most importantly the number of segments n which will make up the index.

$$d_j(t) = \frac{\omega_j \cdot p_j(t)}{\sum_{t=1}^{T} p_j(t)} \qquad j \in [1,k] \tag{1}$$

$$a \leq x_i \leq b \qquad i = 1,...,n \tag{2}$$

$$\vec{x} = [x_1, x_2, ..., x_i, ..., x_{n-1}, x_n] \tag{3}$$

2.2 Evolutionary Multi-Objective Optimisation

The evolutionary multi-objective algorithm exploited in this method is based on the work of Leyland (2002). The variables of the EMOO are the values of lengths of each segment of the index. Initially, n random values are chosen between a and b Eq(2). These random values make the x array which then serve to create the index of segments Eq(3). These are the values to be optimised according to their influence on the two defined performance indicators which serve as objectives, Eq(4).

$$min[\sigma, \Delta] \quad s.t.\{\vec{x}\} \tag{4}$$

2.3 Index Creation

The n x_i values are cumulatively summed creating an array of n+1 values. This vector is resized by a normalisation and a multiplication by the total time period T under study, Eq(5). Values are rounded down to ensure that they are integers. Graphically this corresponds to putting the x_i values together and elongating them to the desired number of days, as can be seen in figure.

$$\vec{t} = \frac{T}{\sum_{i=1}^{n} x_i} \cdot [1, \sum_{i=1}^{1} x_i, \sum_{i=1}^{2} x_i, ..., \sum_{i=1}^{n} x_i] \tag{5}$$

2.4 Standard deviation performance indicator

The mean values of each profile over each segment are calculated as seen in Eq(6). We call this profile the segmented profile. The standard deviation between these segmented profiles and their original profiles is then calculated for each profile, Eq(7). The mean value of the standard deviations of each profile is calculated and serves as the first performance indicator, Eq(7).

$$\bar{d_j}(t_i) = \sum_{t=t_i}^{t_{i+1}-1} \frac{d_j(t)}{t_{i+1} - t_i - 1} \qquad i \in [0, n-1] \tag{6}$$

$$\sigma_j = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (d_j(t) - \bar{d_j}(t))^2} \qquad \sigma = \frac{\sum_{j=1}^{k} \sigma_j}{k} \tag{7}$$

2.5 Zero flow days performance indicator

This performance indicator calculates the total number of zero flow days which are not respected for all of the production profiles. This basically comes down to summing the days from the segmented profiles which do not respect the zero flow days present in the original profiles. However when using industrial data, , low non zero values may in fact actually correspond to zero flow days. Values will be close to zero, but due to sensor calibration, they are often not, for example values below 5 % of the mean value. This is visible in profile 2 in Figure 1 between days 135 and 145. These days should be taken into consideration when minimising the zero flow day objective. Therefore a tolerance level τ_j is used for each profile when

defining days to be considered as zero flow days. For example, all days whose values are equal to less than 10 % of the mean value can be considered as days of zero flows. This tolerance level should be adapted to each profile.
The data of each real profile are analysed in order to determine which days are zero flow days, as can be seen in Eq(9). A binary value is associated to the z_j array for each time period if the conditions are met. The same is done for the segmented profile, Eq(10). The total number of non respected zero flow days are summed over all the profiles as can be seen in Eq(11).

$$\begin{cases} d_j(t) < \tau_j \cdot \bar{d}_j & \Rightarrow \quad z_j(t) = 1 \\ d_j(t) \geq \tau_j \cdot \bar{d}_j & \Rightarrow \quad z_j(t) = 0 \end{cases} \tag{8}$$

$$\begin{cases} \bar{d}_j(t) < \tau_j \cdot \bar{d}_j & \Rightarrow \quad \bar{z}_j(t) = 1 \\ \bar{d}_j(t) \geq \tau_j \cdot \bar{d}_j & \Rightarrow \quad \bar{z}_j(t) = 0 \end{cases} \tag{9}$$

$$\Delta = \sum_{j=1}^{k} \sum_{t=1}^{T} \mid z_j(t) - \bar{z}_j(t) \mid \tag{10}$$

3. Case Study

The developed method is coded in Matlab as is the EMOO algorithm. A case study is carried out on this method using 5 normalised production profiles originating from an industrial complex. All profiles are considered to have the same weight. Each profile presents zero flow days corresponding to turnaround or inactivity periods. The method is applied for a varying number of segments. Results are compared to best results from a Monte-Carlo sampling of the solution space. The Monte-Carlo method uses the same method as described above (section 2.3) to create the index using a random choice of n values. 100,000 indexes are tested using the Monte-Carlo method so as to obtain a fair idea of the solution space. An initial population of 800 individuals and a total number of individuals is fixed at 50,000 for the EMOO algorithm. The tolerance for zero days is set at 10 % of mean profile values. In order to judge the performance of the EMOO and Monte-Carlo results, the performance indicator values for the best performing indexes are plotted in Figure 2.

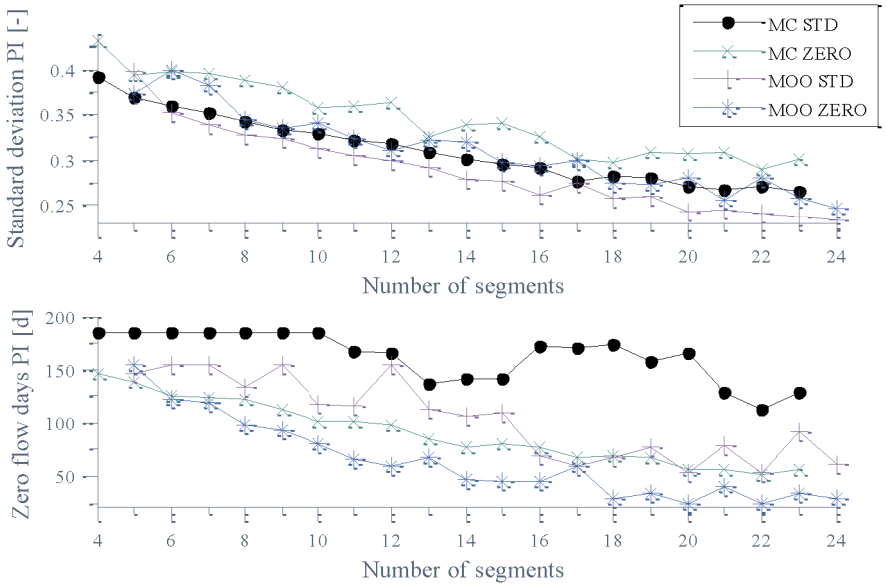

Figure 2: Performance indicator values for best performing indexes for both tested methods

Results from Figure 2 show that the EMOO is able to find better results than the Monte-Carlo sampling method in general. Furthermore, results also show that the best EMOO zero flow rate indexes generally perform as well or better than the best Monte-Carlo results for both objectives. Figure 3 indicates that the EMOO method does take a substantially longer amount of time to find the best results, though once again results are clearly better for the EMOO method. Furthermore while it appears that the Monte-Carlo method reaches a tangent from which little progress is made, the EMOO method continues to find better solutions as time progresses.

Table 1: Comparison of results for different methods and 12 segments

Index	Standard deviation PI [-]	Zero flow days PI [d]
Yearly average	0.456	155
Monthly average	0.332	155
Best EMOO Standard deviation PI	0.284	106
Best EMOO Zero flow days PI	0.295	57
Best Monte-Carlo Standard deviation PI	0.301	146
Best Monte-Carlo Zero flow days PI	0.332	80

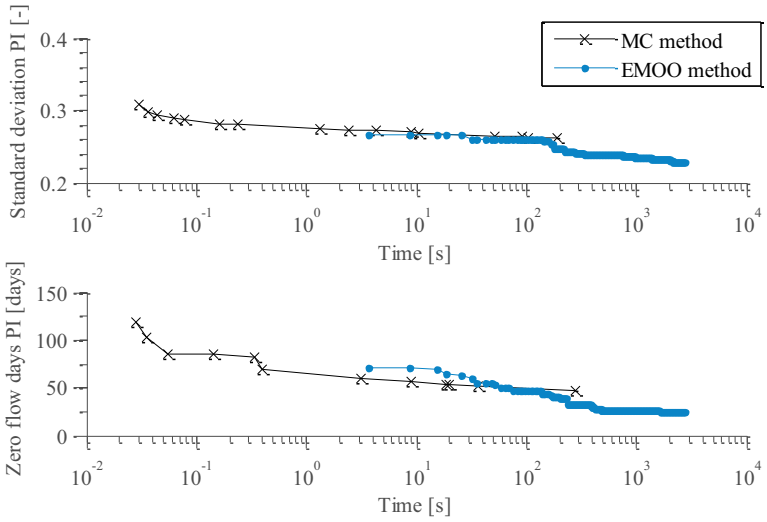

Figure 3: Time taken to reach performance indicator values for 20 segments

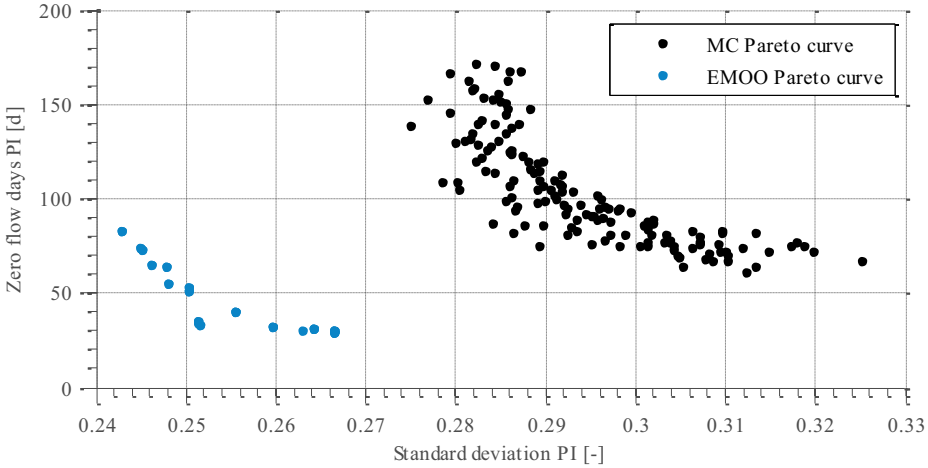

Figure 4: Pareto distribution of best results for EMOO and Monte-Carlo method using 20 segments

Table 1 shows the advantage of using the EMOO segmentation method rather than the more simple monthly averages. The number of non-respected zero flow days is very high when using the monthly averages, for the yearly average method as well. Using the same number of segments it is possible to improve the standard deviation performance indicator by up to 15 % and reduce the number of penalizing flow rate days by up to 98. Figure 5 illustrates the results of using the best index for each indicator using 12 segments. It can be seen how the best index for the second objective fits better into the zero flow days than the first objective or monthly mean method. The indexes are also illustrated. For selecting the index to use for a particular problem, Figure 4 shows a Pareto distribution of the effects of the indexes on the performance indicators, plotted against each other. In this way the engineer is able to choose the compromise between accuracy of profile and respect of the zero flow day constraints using a chosen index

of segments, the engineer will be able to use realistic base cases to simulate several punctual solutions leading to a global solution.

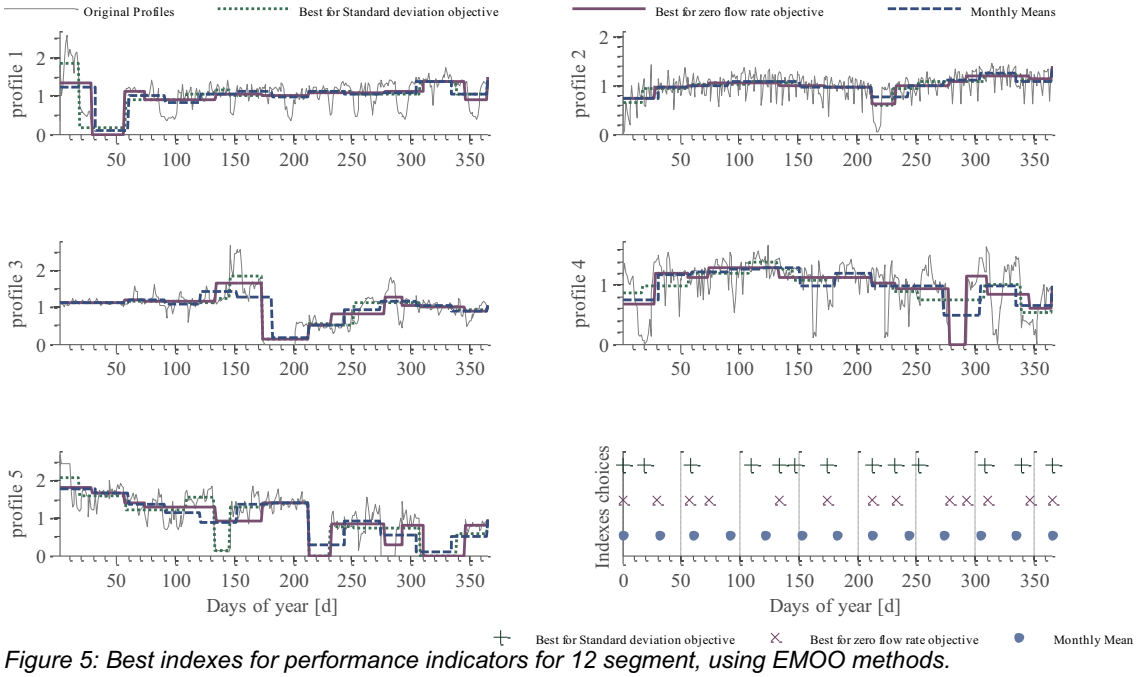

Figure 5: Best indexes for performance indicators for 12 segment, using EMOO methods.

4. Conclusion

Finding a way to deal with time dependence is key to engineering problems, especially those concerning logistics, planification and retrofits, for example on industrial manufacturing sites. When dealing with a complex system, accurately reducing the amount of data one is necessary to studying solutions. This paper shows a method for doing so while maintaining the highest accuracy possible and also respecting days where flow rates are null.

Comparing results to a simple monthly averaging method as well as a Monte-Carlo space sampling method indicates that the use of an evolutionary multi-objective algorithm is suitable for finding the best index of segments to break down the year down into manageable segments. The algorithm does however take a long time to compute. This is not necessarily a problem as in general one needs only execute it once for a given problem. Importantly, the critical zero flow days were much better respected using this algorithm, significantly reducing their neglect. By taking into account these days, it is also expected that the impact of unforeseen shutdowns and logistics problems be reduced.

This method can be applied to multi-period process integration studies in order to find thermoeconomically optimal energy efficiency solutions. This method can also be extended to short and long term storage and scheduling problems as the created base cases are sequential.

References

Domínguez-Muñoz F., Cejudo-López J.M., Carrillo-Andrés A., Gallardo-Salazar M, 2011, Selection of typical demand days for CHP optimization. Energy and Buildings 43, 11 , 3036-3043.

Leyland G., Multi-objective optimisation applied to industrial energy problems, 2002, PhD Thesis. Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland.

Maréchal F., Kalitventzeff B., 2003, Targeting the integration of multi-period utility systems for site scale process integration. Applied Thermal Engineering 23.14, 1763-1784.

Poulin A., Dostie M., Fournier M., Sansregret S., 2008, Load duration curve: A tool for technico-economic analysis of energy solutions. Energy and Buildings 40, 1, 29-35.

Fazlollahi S., Bungener S.L., Becker G., Maréchal F., 2012, Multi-Objectives, Multi-Period Optimization of district heating networks Using Evolutionary Algorithms and Mixed Integer Linear Programming (MILP). Computer Aided Chemical Engineering, 31, 890-894.

Stein W., Fleten S.E., 2003, Stochastic programming models in energy. Handbooks in operations research and management science 10, 637-677.

Pan M., Bulatov I.,,Smith R., 2012, Retrofit Procedure for Intensifying Heat Transfer in Heat Exchanger Networks Prone to Fouling Deposition. Chemical Engineering Transactions, 29, 1423-1428.