

Document downloaded from:

<http://hdl.handle.net/10251/76344>

This paper must be cited as:

Galvez, ED.; Capuz-Rizo, SF. (2016). Assessment of global sensitivity analysis methods for project scheduling. *Computers and Industrial Engineering*. 93:110-120.
doi:10.1016/j.cie.2015.12.010.



The final publication is available at

<http://dx.doi.org/10.1016/j.cie.2015.12.010>

Copyright Elsevier

Additional Information

Assessment of global sensitivity analysis methods for project scheduling

Edelmira D. Gálvez

Department of Mining and Metallurgy Engineering, Universidad Católica del Norte

Av. Angamos 0610, Antofagasta- Chile. E-mail: egalvez@ucn.cl

Tel: +56 552355648

and

Salvador F. Capuz-Rizo

Department of Engineering Projects, Polytechnic University of Valencia,

Cno. de Vera s/n, E-46020 Valencia, Spain. E-mail: scapuz@dpi.upv.es

Tel: 963877000 (Ext.:75653); Fax: 963879869 (Ext.:79869)

In loving memory of Mario E. Mellado, 1971-2015

Abstract

Product development requires scheduling that considers the interdependence between activities. The definition of the interdependencies and duration of activities, communication times and the level of overlap between activities is needed for project scheduling. However, these parameters have epistemic uncertainties that can affect project scheduling. In this work, different global sensitivity analysis techniques were applied to identify the parameters that had the greatest effect on project scheduling. It was concluded that standardized regression coefficients as well as the Morris and Sobol' methods were the most appropriate. It was also found that global sensitivity analysis can help to focus resources based on the definitions and control the uncertainty of key activities. Furthermore, it was concluded that control of the uncertainty of key activities reduces the uncertainty and duration of projects.

Keywords: Dependency structure matrix; Project duration; Sensitivity analysis; Overlapping; Schedule.

Assessment of global sensitivity analysis methods for project scheduling

Manuscript submitted to Computers & Industrial Engineering, June 2015

Abstract

Product development requires scheduling that considers the interdependence between activities. The definition of the interdependencies and duration of activities, communication times and the level of overlap between activities is needed for project scheduling. However, these parameters have epistemic uncertainties that can affect project scheduling. In this work, different global sensitivity analysis techniques were applied to identify the parameters that had the greatest effect on project scheduling. It was concluded that standardized regression coefficients as well as the Morris and Sobol' methods were the most appropriate. It was also found that global sensitivity analysis can help to focus resources based on the definitions and control the uncertainty of key activities. Furthermore, it was concluded that control of the uncertainty of key activities reduces the uncertainty and duration of projects.

Keywords: Dependency structure matrix; Project duration; Sensitivity analysis; Overlapping; Schedule.

1. Introduction

A profitable and effective product is the key to success in today's ever-changing and competitive market. As a result, strong competition in several industries has forced manufacturing firms to develop innovative and higher quality profitable products at an increasingly rapid pace (Kirshnan et al., 1997). The scheduling and management of large and complex projects is a difficult commission that requires effective tools (Herroelen, 2005). The dependency structure matrix, or design structure matrix (DSM), has been shown to be a powerful tool for the management of complex projects because a) it can accurately represent the interdependence and / or relationships between different components of a

system; b) overcomes the size and complexity limitations of digraphs; c) it is easy to understand and able to handle the processes in their entirety; and d) the matrix format is suitable to program and calculate using computers (Chen and Lin, 2003). Chen et al. (2003) presented a project scheduling framework based on DSM to handle sequencing, monitoring, and control of a collaborative product development. Sosa et al. (2004) investigated how the organizational and system boundaries, design interface strength, indirect interactions, and system modularity impact the alignment of design interfaces and team interactions. They used DSM to study complex product architectures in terms of component interfaces and to build statistical models for proper hypothesis testing using DSM data. The use of parameter-based DSM as a process modeling and system analysis tool for building design in the architecture/engineering/construction industry was proposed by Pektaş and Pultar (2006). Tang et al. (2010) studied how to capture and trace the design knowledge through a single-domain and multi-domain DSM. They proposed a DSM-based design knowledge management system that allows for efficient knowledge capturing, searching, and tracing in product design.

Project scheduling is an important element of project management. The procedures range from the traditional models of CPM and PERT to sophisticated optimization models (Węglarz et al., 2011), algorithms and heuristics based methods (Liang, 2009). Project scheduling research concentrates on the generation of a procedure that optimizes the scheduling objective, usually the project duration, and that should serve as a baseline schedule for executing the project (Herroelen and Leus, 2005). Research has been conducted for project scheduling using DSM where the interdependence of activities has been considered in the schedule as well, as DSM has been shown to be a powerful tool for the management of complex projects.

Project scheduling has considerable uncertainty because project activity parameters are also subject to uncertainties (Dixit et al., 2014, Chtourou and Haouari, 2008). These uncertainties are usually epistemic (due to lack of knowledge) and not aleatory (inherent randomness of the system). Examples of epistemic uncertainties are activities that can take more or less time than originally estimated, such as material arriving behind schedule, unavailable resources, and incorrect estimation of activity overlap. There are some aleatory uncertainties, such as weather conditions or natural events that cause

delays. From the viewpoint of project management, little can be done to control aleatory uncertainties, but actions can be taken regarding epistemic uncertainties. In addition activity durations and overlap factors can have different values by allocating different monetary resources to its execution (Zamani, 2013). However, what are the key uncertainties that require more control or study?

Herroelen and Leus (2005) define five approaches for dealing with uncertainty in the scheduling environment: reactive scheduling, stochastic scheduling, scheduling under fuzziness, proactive scheduling, and sensitivity analysis. The first four approaches are related to uncertainty analysis, and therefore use different methods to represent and address uncertainty. Sensitivity analysis has recently emerged in the project scheduling environment, but only "what if ...?" type of questions have been addressed. There are several sensitivity analysis methods that need to be explored in the scheduling environment.

The objective of this work is to assess global sensitivity analysis (GSA) methods for project management. The DSM-based project duration is used as an example. The focus of this study is the use of GSA to identify key input uncertainties for the reduction of uncertainty in the project duration.

2. DSM-based project duration

DSM has been used for project scheduling in the past. Browning (1998) used DSM to enable critical path calculations by defining the amount of effort or work as the duration of the activities. Wang and Lin (2009) developed an overlapping process model to analyze the impact of the process structure on the lead-time of a development project with multiple activities. A DSM was used to represent the complex interaction patterns between the development activities. A triangular distribution was used to represent the uncertainties in the activity duration and reworks. Srour et al. (2013) provided a method to automatically generate a fast-track design schedule without violating the dependency information. They also extended the basic DSM method to construction projects. Maheswari and Varghese (2005) developed methods for the estimation of project durations including the communication time and natural overlaps between activities. The dependency between the activity duration, communication time, and overlap time factors were used to estimate the project duration. Uncertainty was not

considered in the work of Maheswari and Varghese, which motivated the development of different studies to represent uncertainty in the input parameters. Gálvez et al. (2012) studied the effect of the uncertainty associated with task programming using DSM and grey theory, or interval arithmetic. Shi and Blomquist (2012) extended the DSM method proposed by Maheswari and Varghese using fuzzy numbers.

The methods of Gálvez et al. (2012) and Shi and Blomquist (2012) allow for representation of the uncertainty in the input parameters and calculation of the uncertainty in the project duration. One of the drawbacks of these methods is the need to characterize all the input factors, which are typically defined through an expert review process. Definition of the distribution that characterizes the epistemic uncertainty in the duration of activities and the time overlap factors can be one of the most important parts of uncertainty analysis because these distributions can determine the uncertainty in the project duration. These distributions must be defined through an expert review process, and their development can constitute a major analysis cost. The process of extracting expert knowledge about an unknown quantity or quantities and formulating that information as a probability distribution is known as elicitation (O'Hagan et al., 2006; Meyer and Booker, 2001). The scope of elicitation can vary widely depending on the purpose of the analysis, size of the analysis, and resources available to perform the analysis. One possible analysis strategy is to perform GSA with crude definitions of the distribution functions for the input factors (i.e., activity duration and time overlap factors) to identify key input factors and to understand the behavior of the project duration uncertainty. Then, resources can be concentrated where they are needed.

As previously mentioned, the objective of this work is to survey GSA methods for project management. An example using both the traditional method (sequential: an activity starts once its predecessors are completed) and phased method (some amount of overlap occurs between pairs of activities) is used to illustrate and assess the GSA methods. The example is given below.

The example consists of five activities from A to E. The DSM representation of the example is given in Figure 1. The DSM is a square matrix containing a list of activities in the rows and columns in the same order. The order of activities in the rows and columns in the matrix indicates the sequence of execution.

Values on the diagonal are the mean duration of the activities (days). For example, Figure 1 shows that the mean duration of activity A is 2 days. The marks in the off-diagonal cells indicate that these activities are information predecessors, with activity inputs in its row and activity outputs in its column. For instance, activity B needs information from activity A and provides information to activity D. In Figure 1, the traditional, sequential method of project scheduling is shown. In this method, an activity starts once its predecessors have been completed. Based on the mean duration of the activities, the conventional project duration was estimated to be 14 days (Figure 1). Note that activity C has no effect on the project duration, and all other activities are shown in the order of execution with no time between activities. The conventional project duration is estimated using the following equations:

$$(EF)_i = (ES)_i + A_{ii} \quad 0 < i \leq n \quad (1)$$

$$(ES)_j = \text{Max}[(EF)_i] \quad 0 < i \leq n, 0 < j \leq n \quad (2)$$

$$\text{Conventional project duration} = \text{Max}[(EF)_j] \quad 0 < j \leq n \quad (3)$$

Where n is the number of activities, i denotes all the immediate predecessors of activity j , j is the current activity chosen in the order identified by the DSM, ES means early start, EF early finish, and A_{ii} denotes the diagonal values of the DSM (duration of the activity).

The values in Figure 1 correspond to the expected values of the activity durations. Two situations were analyzed. First, a variation of ± 1 day in the activity durations represented by a uniform distribution (all the activity durations were equally likely) was considered. The values for activities A to E were U(1,3), U(3,5), U(4,6), U(4,6), and U(2,4), respectively. U(x,y) represents a uniform distribution between the values x and y. This study will henceforth be called Case 1. For Case 2, a variation of $\pm 50\%$, also represented by a uniform distribution, was considered. This made the activity duration uncertainties proportional to the nominal values. The values for activities A to E in Case 2 were U(1,3), U(2,6), U(2.5,7.5), U(2.5,7.5), and U(1.5,4.5), respectively.

In Figure 2, the fast-track, overlap or phased method for project scheduling is shown. In this method, some overlap occurs between pairs of activities. Figure 2 shows the same example as Figure 1 with

overlap between activities being allowed. The overlap is represented in the DSM in the form of ratios called time factors, which were introduced by Maheswari and Varghese (2005). Two time factors were used, the time factor for receiving the information for the successive activity (represented by matrix B_{ij} , given by the off-diagonal cell in Figure 2b), and the time factor for sending information from the predecessor activity (represented by matrix C_{ij} , given by the off-diagonal cell in Figure 2c). For example, 0.8 in B_{BA} implies that A can send the required information through B after 80% of its duration has passed; 0.15 in C_{BA} implies that to continue, it is essential that B receive information from A after 15% of its duration has passed, instead of at the beginning of the task. Values of 1 and 0 in B_{ij} and C_{ij} indicate that overlap is not possible/allowed between the activities. Based on the mean values of the activity durations and factor time, the natural overlap project duration was estimated to be 10 days (Figure 2a).

The natural overlap project duration is estimated with the following equations,

$$(ES)_j = \text{Max}[(ES)_i + B_{ji} B_{ii} - C_{ji} C_{jj}] \quad 0 < i \leq n, 0 < j \leq n \quad (4)$$

$$(EF)_i = (ES)_i + B_{ii} \quad 0 < i \leq n \quad (5)$$

$$\text{natural overlap project duration} = \text{Max}[(EF)_j] \quad 0 < j \leq n \quad (6)$$

Note that B_{ii} and C_{ii} are the diagonal values of the DSM (duration of activity).

The nominal values are shown in Figure 2. The uncertainty is applied as follows. 1) The same uncertainty as in case 1 was considered for the duration of the activity. The values for the durations of activities A to E were U(1,3), U(3,5), U(4,6), U(4,6), and U(2,4), respectively. 2) A common uniform distribution, U(0.65, 0.95), was used for the B_{BA} , B_{DB} , and B_{ED} time factors. Similarly, a common uniform distribution, U(0.05,0.25), was used for the C_{BA} , C_{DB} , and C_{ED} time factors. 3) The time factors B_{CA} , B_{EC} , C_{CA} , and C_{EC} did not have uncertainties, and therefore had the values given in Figure 2. 4) Values of 1 for factors B_{CA} , B_{EC} , and 0 for C_{CA} , and C_{EC} denoted that activity C did not overlap activities A and E.

Communication times between activities are easily incorporated by modifying equation (2) by $(ES)_j = \text{Max}[(EF)_i + A_{ji}]$ in equations 1 to 3. A_{ji} is the communication time between activities j and i. This was not included here because the communication time can be considered an activity between dependent activities.

3. Assessment of global sensitivity analysis methods

Sensitivity analysis refers to the determination of the contribution of individual uncertainty inputs to the uncertainty in the outputs (Helton et al. 2006). The first approach to sensitivity analysis is known as the local approach, where the effect of small input perturbations on the model is considered. Local sensitivity analysis has several limitations such as linearity and normality assumptions, as well as local variations. In contrast to local sensitivity analysis, global sensitivity analysis (GSA) considers the entire range of input variations. According to Saltelli et al. (2008), the GSA can be defined as “the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input”. These techniques have been widely used in different engineering areas to determine the most significant variables in a model. The general objectives of GSA are the following: a) the identification of significant and insignificant factors and possibly a reduction of the number of design variables in the optimization problem, and b) the improvement of the understanding of the model’s behavior (highlighting interactions among factors, finding combinations of factors that result in high or low values for the model output, etc.). GSA corresponds to the evaluation of an output model when all the model factors are simultaneously evaluated, and is generally resolved by numerical methods. This methodology has the advantage of simultaneously assessing all factors. However, it requires a large amount of data for the model to be evaluated, and the mathematical techniques are more complex.

In this section several GSA methods are applied to the examples described in the previous section. Only a brief description is given for each method; readers who are interested in a more thorough description can see the following references: Saltelli et al. (2008, 2009); Iooss and Lemaître (2015); Helton et al. (2006); Ionescu-Bujor and Cacuci (2004); Cacuci and Ionescu-Bujor (2004); and Frey and Patil (2002).

The software R (R Core Team, 2013) and sensitivity package (Pujol et al., 2014), which is a free software environment for statistical computing and graphics, were used for all the calculations.

3.1. Scatterplot

The scatterplots provide a visualization of the output factor (project duration) versus the input factors (activity durations and overlap time factors) and can reveal nonlinear or other unexpected relationships between the outputs and inputs. Scatterplots can be used as a starting point for complex analyses and can help with the selection of the sensitivity analysis strategy. They can also be used to identify linear relationships, monotonic relations and the existence of thresholds among other potential trends. This is shown in Figure 3 for two input factors in cases 1 and 3. In Figure 3, the polynomial smoother is plotted for each cloud of points to clearly identify the mean trend of the project duration versus each input factor. Figures 3a and 3b show the scatterplot for the B and C activity durations, respectively, for case 1. It is clear that the project duration increases as the duration of activity B increases, whereas there is no dependency on the project duration with duration of activity C. This is to be expected because the duration of activity C has no effect on the project duration (see Figure 1). Figures 3c and 3d show the scatterplot for the B_{DB} and C_{DB} time factors, respectively, for case 3. Both time factors affect the project duration but in the opposite manner. The effect of both time factors is small because the slopes of the smoothed curves are small. Even if scatterplots can be useful, they do not capture some of the interaction effects between inputs and are difficult to analyze for complex systems with many variables, which can be the case in project scheduling.

3.2. Partial Correlation Coefficients

In this section, correlation coefficients will be introduced. They are based on the Pearson correlation coefficient, which is defined as the covariance of the variables divided by the product of their standard deviations.

$$r_{x_j,y} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

Where $\bar{x}_j = \sum_{i=1}^n x_{ij}/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$ are the sample means. Here x_j represents any input factor (activity duration, overlap time factor) and y the output factor (project duration). The Pearson correlation is a measure of the strength and direction of the linear relationship between two variables. The values of this coefficient are between -1 and 1. A positive value means that both y and x_j are decreasing or increasing together, and a negative value means that the input and output factors tend to move in opposite directions. If y and x_j are independent, the coefficient equals 0.

The Partial Correlation Coefficient (PCC) gives the strength and direction of the linear correlation between y and a given input x_j in a case where the influence of the remaining variables is eliminated. In other words, PCC provides a measure of variable importance that tends to exclude the effects of other variables. More precisely, PCC is calculated using the Pearson correlation coefficient between $y - \hat{y}$ and $x_j - \hat{x}_j$, where \hat{y} and \hat{x}_j represent linear regression models ($\hat{y} = b_0 + \sum_{k \neq j} b_k x_k$ and $\hat{x}_j = c_0 + \sum_{k \neq j} c_k x_k$).

As a rule of thumb, the following guidelines regarding the strength of relationship are often useful (though many experts would somewhat disagree on the choice of boundaries). PCC values can be classified by their magnitude as strong (1.0 to 0.6, or -1.0 to -0.6), moderate (0.6 to 0.3, or -0.6 to -0.3), weak (0.3 to 0.1, or -0.3 to -0.1), or very weak (0.1 to 0, or -0.1 to 0) correlations.

Table 1 shows the PCC for the example with a sample size of 200 runs. For case 1, there is a perfect linear correlation between the duration of activities A, B, D and E and the project duration (PCC equal to 1). On the other hand, activity C and the project duration are uncorrelated (PCC near 0). This was expected because all activities have the same level of uncertainty and activity C does not affect the duration of the project. Values of PCC equal to 1 indicate that the model (equations 1-3) behaves linearly, i.e., the early start of activity E is always limited by the early finish of activity D rather than the early finish of C. As such, the maximization function in equation 2 always selects the early finish of activity D. The behavior of case 2, where the uncertainty are proportional to the activity duration, is similar, but there is no perfect linear correlation between the duration of activities A, B, D and E and the project duration. This is because the uncertainties in this case are larger than in case 1. For case 3,

there is not a strong linear correlation between the input factors (duration of activities, overlap time factor) and the project duration. This is also expected because the model (equations 4 to 6) is nonlinear. The duration of activities A, C, E and time factor B_{DB} have PCC values over 0.6. Time factors C_{BA} , C_{DB} , and C_{ED} have negative PCC values which means that these time factors and project duration are inversely related.

The use of transformations may also provide information about input and output factor relations. The most frequently used transformations are ranks. In this transformation the smallest sample value is transformed into 1, the second smallest into 2 and so on up to the largest value. The strongest reason to use the rank transformation in sensitivity analysis is that this transformation converts any monotonic relation between two variables into a linear one. After the sample (x_j, y) is transformed into the rank sample (Rx_j, Ry) , the same procedure is applied to this new sample. This yields the Partial Rank Correlation Coefficient (PRCC) (Ioos and Lemaître, 2015). Table 1 also gives the PRCC values for the examples. The results are similar to the PCC results.

3.3. Standardized Regression Coefficients

In this section, standardized regression coefficients (SRC) will be introduced. SRC are calculated by replacing the variables with standardized variables, which is done by subtracting the sample mean from the variable and dividing the result by its sample standard deviation. They can be calculated using the linear regression coefficient (β_j) associated to the variable x_j :

$$SRC_j = \beta_j \frac{S_j}{S} \quad (8)$$

where S_j and S are the sample standard deviations for the input variable x_j and output variable y . The SRC provides a measure of variable importance. A variable is more important or its uncertainty has a greater effect on the uncertainty in the project duration if its absolute value is greater than the absolute value of another variable. Because the variables are standardized, in our examples, SRC can be more useful for comparing the effects of overlap time factors because these variables are measured in

different units of measurement. Similar to the PCC and PRCC, the SRC can be calculated for the rank variables, and in this case the index is called the standardized rank regression coefficients (SRRC).

Table 2 shows the SRC and the SRRC for our examples and for a sample data of size 200. For case 1, duration of activities A, B, D and E have similar values of SRC and SRRC because they all affect in the same extent to the uncertainty of project duration. This is because the uncertainties of these activities are all equal. The activity duration uncertainty of activity C does not affect the uncertainty of project duration. For case 2, the magnitude of the activities effect on the project duration from highest to lowest was $D > B > E > A$ because the uncertainties in the duration of these activities followed the same order. The uncertainty in the duration of activity C was marginal. For case 3, the behavior was different because activities B and D overlapped the other activities. Furthermore, activity C did not overlap with other activities. As such, the activities affected the project duration in the following order, from highest to lowest, $A > E > C > B > D$. This interesting result occurred because the overlap can affect the scheduling and importance of each activity duration. The overlap time factors that more affect the project duration are B_{ED} , B_{DB} and C_{DB} . These results can be very useful, as will be shown later.

3.4. Morris method

The Morris (1991) method is associated with experimental design methods. The motivation of Morris method comes from models that are complicated enough for classical experimental design and/or they have a moderate-to-large number of input factors. This can be the case in project scheduling, where uncertainties in activity durations and time factors can lead to models with a large number of input factors. This method designs computational experiments to determine which input factors have important effects on an output factor. The experimental plans are composed of individually randomized one-factor-at-a-time designs, and data analysis is based on the resulting sample of observed elementary effects. The Morris method calculates the elementary effect of the j -th input factor at the i -th repetition as:

$$R_j^i = \frac{y(x_1^i, \dots, x_{j-1}^i, x_j^i + \Delta, x_{j+1}^i, \dots, x_n^i) - y(x_1^i, \dots, x_n^i)}{\Delta} \quad (9)$$

where $y(X)$ is the output factor and x_1^i, \dots, x_n^i are the input factors being studied. The method calculates the mean of the value of the elementary effects ($\mu_j = \text{mean of } R_j$) by assessing the overall influence of the factor x_j on $y(X)$ (the larger μ_j is, the more the j-th input contributes to the dispersion of the output), and standard deviation of the elementary effects (σ_j), which is a measure of non-linear and/or interaction effects of the j-th input. In this work, the absolute value of μ_j was used, as was proposed by Campolongo et al. (2007), μ_j^* .

Figure 4 shows graphs of σ_j versus μ_j^* , which makes it possible to distinguish between three groups of input factors: a) input factors that are influential without non-linear and/or interaction effects (high values of μ_j^* and low values of σ_j); b) input factors that are influential with non-linear and/or interaction effects (high values of μ_j^* and high values of σ_j); and c) input factors that have no effect (low values of μ_j^*).

The Morris results confirm the results obtained by the SRC method. In case 1, the duration of activities A, B, D and E are influential (high value of μ_j^*) and have equals value of μ_j^* because they all affect the uncertainty of project duration to the same extent. The duration of activity C had no effect. In addition, the project duration was linearly dependent on the duration of activities A, B, D and E (because σ_j values are very low in comparison with the values of μ_j^*). For case 2, the activities affected the project duration in the following order, from highest to lowest, $D > B > E > A$, and the effect of the uncertainty in the duration of activity C was marginal. The influence of activities B and D had non-linear and/or interaction effects (because σ_j values are not very low in comparison with the values of μ_j^*). In case 3, the behavior was also similar to the results obtained using the SRC method; the only difference was that for SRC $B_{ED} > B_{DB} > C_{DB}$, and with the Morris method $B_{DB} > B_{ED} > C_{DB}$. It can be observed that almost all the input factors had non-linear and/or interaction effects.

The number of model calls (n) depends on the number of input factors (p) and repetitions (r), $n = r(p + 1)$. In example 15, repetitions were used, and as a result, the numbers of model calls were 90, 90 and 180 for cases 1, 2 and 3, respectively.

3.5. Sobol' Method

In the Sobol' (1993) method, the variance of the model output can be decomposed into terms of increasing dimension, called partial variances, which represent the contribution of the inputs (i.e., single inputs, pairs of inputs, etc.) to the overall uncertainty in the model output. Statistical estimators of partial variances are available to quantify the sensitivities of all the inputs and groups of inputs through multi-dimensional integrals. The calculation of all the partial variances of input groups has a high computation cost, which is why Homma and Saltelli (1996) introduced the concept of a total sensitivity index. The total sensitivity index indicates the overall effect of a given input by considering all the possible interactions of the respective input with all the other inputs. In this paper, the Sobol' method and the improved formulas of Jansen (1999) and Saltelli et al. (2010) for the Sobol' method were applied.

The Sobol'-Jansen method calculates two indices, the first order effects sensitivity index, S_j , corresponding to a single factor (x_j) and the total sensitivity indices S_j^T . The interpretation of the sensitivity indices is straightforward; the larger the sensitivity index the more influential the corresponding input factor or set of input factors. The first order sensitivity index measures only the main effect contribution of each input factor on the output variance. It does not account for the interactions between factors. The first-order sensitivity index (S_j) is important when the objective is to determine the most important input uncertainties. The total sensitivity index (S_j^T) is important when the objective is to reduce the uncertainty in the output model (Adeyinka, 2007). If the first-order sensitivity index (S_j) of the j input factor is very small, then the uncertainty in x_j does not affect the uncertainty in the output model y . Therefore, x_j is non-influential, or unimportant. This does not say anything about input interactions or high-order sensitivity indices such as $S_{j,i}$ or $S_{j,i,k}$. If the total sensitivity index (S_j^T) is also small, then apart from being unimportant, x_j does not interact with other

factors (high-order effects of x_j are negligible). The implication of a small S_j and S_j^T , is that the uncertainty in j has no effect on the uncertainty in y . As such, in subsequent analyses, x_j can be fixed to its nominal value (mean or median) and further research, measurement, analysis and data gathering can be directed towards other factors. Conversely, regardless of the magnitude of S_j^T , a large value of the first-order sensitivity index, S_j , implies that x_j is influential. The arithmetic difference between S_j^T and S_j indicates the magnitude of the interactions between x_j and other factors.

The Sobol'-Jansen method allows for the estimation of both first-order and total indices at the same time, at a total cost of $(n + 2) * s$ model evaluations, where n is the number of input factors and s is the sample size.

The Sobol'-Jansen method was applied to the example for conventional project duration with five input factors (size of 50,000) with Monte Carlo sampling and had a cost of 350,000 model calls. For the example for natural overlap project duration, it had a cost of 6,500,000 model calls.

The results given in Figure 5 are similar to the SRC in the sense that they delivered practically the same ranking of input factors, but additional information regarding interactions is also given. For cases 1 and 2, the interaction between input factors was very low. For case 3 there was little interaction.

3.6 Discussion

Based on the results given for the methods SRC, Morris and Sobol' for case 1, the uncertainties in the duration of activities A, B, D and E affect the uncertainty in project duration. The uncertainties in the duration of activity C does not affect the uncertainty in project duration. To prove this result, three numerical experiments were performed. In the first experiment (E1), it was considered that all durations in the activities had uncertainty, and Monte Carlo simulation (10,000 model calls) was used to estimate the maximum, mean and minimum value of the project duration, 17.7, 14.0 and 10.5, respectively. In the second experiment (E2), the duration of activity C was considered to have no uncertainty, and the durations of the other activities had uncertainty. Monte Carlo simulation was used to estimate the maximum, mean and minimum project duration. The results are practically the same (17.9, 14.0 and 10.3) as those in E1, indicating that the removal of uncertainty in the duration of

activity C did not reduce the uncertainty in the duration of the project. In the third experiment (E3), the duration of activity C had uncertainty, and the durations of the other activities had no uncertainty. Monte Carlo simulation was used again, and the results show that the project duration had no uncertainty (maximum, mean and minimum values are equal to 14). This indicates that the elimination of uncertainties in the duration of the activities A, B, D and E removes uncertainty in the duration of the project. These results are summarized in Figure 6.

Similar numerical experiments were realized for case 2 using Monte Carlo simulation with 10,000 model calls. For case 2, the uncertainties in the duration of activities B and D are the ones that most affected the uncertainty in the project duration. The uncertainties in the duration of activity C did not affect the uncertainty in the project duration. To prove this result, three numerical experiments were performed. In the first experiment (E1), all activity durations had uncertainty. In the second experiment (E2), the uncertainty in the duration of activity C was removed, and in the third experiment (E3), the uncertainties in the duration of activities B and D were removed. The results (maximum, mean and minimum values) for the experiments were E1 (20.4, 14.0, 7.7), E2 (20.4, 14.0, 7.6) and E3 (16.4, 14.0, 11.6), which confirms that the elimination of uncertainties in the duration of the activities B and D reduces the uncertainty in the duration of the project. These results are also summarized in Figure 6.

Finally, numerical experiments were performed for case 3. For case 3, the uncertainties in the duration of activities A and E most affected the uncertainty in the project duration. The uncertainties in the duration of activities B and D are the ones least affected the uncertainty in the project duration. To prove this result, three numerical experiments were performed. In the first experiment (E1), all activity durations had uncertainty. In the second experiment (E2), the uncertainty in the duration of activities B and D were removed, and in the third experiment (E3), the uncertainties in the duration of activities A and E were removed. The results (maximum, mean and minimum values) for the experiments were E1 (14.6, 10.4, 7.4), E2 (13.4, 10.4, 7.2) and E3 (13.3, 10.4, 9.0). These results are not as easily interpreted as in the previous cases. This is because in case 3 there were 11 input factors, and

comparisons should be made considering more input factors (e.g., setting the 5 input factors that most affect the project duration versus the 5 input factors that least affect the project duration).

These results confirm that the SRC, Morris and Sobol' methods can be used to identify input factors that affect the uncertainty in the project duration. The final decision on where to focus efforts to reduce uncertainty or manipulate the input variables (e.g. by hiring extra amounts of the current types of resources or some different types of sophisticated resources) depends on these results and other aspects, such as the associated cost, availability of resources and feasibility of reducing the uncertainty in the activity duration.

4. Case study

This section presents the application of the SRC and Sobol'-Jansen methods of GSA to a case study presented by Maheswari and Varghese (2005) with the uncertainties given by Gálvez (2015). This case study consists of 10 tasks (A through J), as shown in Table 3. Information is also provided about the previous tasks and uniform distributions. Table 4 presents the uniform distributions for the elements of the matrices **B** and **C**.

There are a total of 40 input factors, and identifying the input factors that influence the uncertainty in the project duration most is not an easy task. The results of applying the SRC and Sobol'-Jansen methods are shown in Table 5. Only the total Sobol'-Jansen indices are shown because the first order indices are very similar to the total indices (i.e., the interaction between input factors is very small). The results delivered by SRC and Sobol'-Jansen are virtually identical. The last column of table 5 includes a ranking from highest to lowest according to the results of the SRC. It can be seen that this ranking is also valid for the total Sobol'-Jansen index.

The first 10 input factors that affect the project duration are the following: $B_{HI} > A_{EB} > A_{BB} > A_{HH} > A_{II} > A_{CC} > B_{EB} > C_{HI} > C_{EB} > B_{IE}$, where A_{ii} represents the duration of activity i . The last 10 input factors that affect the project duration are the following: $C_{IG} > B_{GJ} > B_{ID} > C_{JB} > A_{JJ} > B_{FA} > C_{JF} > B_{GF} > C_{GJ} > C_{GF}$. To verify the results, Monte Carlo simulations were performed as follows. In numerical E 1, all input factors were considered to have the uncertainties given in tables 3 and 4. In numerical E 2,

the first 10 input factors affecting the project duration were set to their mean values, and the other input factors still had uncertainty. In numerical E 3 the last 10 input factors affecting the project duration were set to their mean values, and the other input factors remained uncertain. The results obtained in E 1 indicate that the duration of the project takes a mean of 32.6 days with maximum and minimum values of 35.2 and 30.2, respectively (i.e., an uncertainty of 5). By fixing the input factors that least affect the duration of the project, as in E 3, the mean project duration remained 32.6 and the maximum and minimum durations were 35.1 and 30.2, respectively. However, by fixing the input factors that most affect the project duration, as in E 2, the uncertainty was reduced by 62%. The mean, maximum, and minimum values obtained in experiment 2 were 32.6, 33.1 and 32.2, respectively.

The identification of input factors that most affect the uncertainty in the project duration can also be used to reduce the project duration by setting the input factors to convenient values (e.g., reducing the duration of activities). If the first 10 input factors that affect the project duration are set to the most convenient value (the lowest value for a positive SRC or highest value for a negative SRC), it was found that the mean project duration was 29.2, a reduction of 10%. However, if the last 10 input factors were fixed at convenient values, the mean project duration is 32.6 (i.e., no change).

5. Conclusion and final comments

Based on the studies conducted on the effect of input factors (duration of activities and overlap time factors), the following conclusions can be made:

- GSA can be used to identify input factors that affect the uncertainty of the duration of projects.
- The reduction in the uncertainty of input factors that affect the uncertainty of the duration of projects reduces the uncertainty in project duration.
- The identification of input factors that most affect the uncertainty of the project duration can also be used to reduce the project duration by setting the input factors to convenient values.
- PCC allows for the identification of the strength and direction of the linear correlations between the project duration and given input factors.

- SRC, Morris, and Sobol' methods allow for the sorting of the input factors that most affect the uncertainty of the project duration. However, the Sobol' method has a higher computational cost and is more difficult to implement.
- SRC also allows for identification of the direction of the correlation between the project duration and given input factors. Only the C_{ij} time factors had opposite directions to the project duration.
- Based on the results of the Sobol' method, it can be concluded that the interaction between input factors was insignificant.
- The Morris method allows for identification of non-linear and interaction effects, but based in the results of the Sobol' method, the σ_j index of Morris gives more information regarding the non-linear effects. Additionally, if the μ Morris index is used (not the absolute value, μ^*), the Morris method identifies the direction of the effect of the input factor to the project duration.
- The introduction of overlap time factors can change the effect of the uncertainty of the duration of an activity.

If GSA is used a more robust project scheduling is expected because not only the mean values of the input variables are used, but also the variance of these variables.

References

- Adeyinka A.L. (2007). Applications of Sensitivity Analysis in Petroleum Engineering, Thesis University of Texas at Austin.
- Browning TR. (1998). Use of dependency structure matrices for product development cycle time reduction. In: Proceedings of the fifth ISPE international conference on concurrent engineering: research and applications, Tokyo, Japan.
- Campolongo, F., Cariboni, J., Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. Environ. Model. Softw. 22, 1509–1518.
- Cacuci D.G., M. Ionescu-Bujor. (2004). A comparative review of Sensitivity and uncertainty analysis of large-scale systems. II: Statistical methods, Nuclear science and engineering 147 (3), 204-217.

- Chen C.H., S.F. Ling, W. Chen. (2003). Project scheduling for collaborative product development using DSM, *International Journal of Project Management*, 21, 291-299.
- Chen S., Lin L. (2003). Decomposition of interdependent task group for concurrent engineering, *Computers & Industrial Engineering* 44, 435-459.
- Chtourou H., Haouari M. (2008). A two-stage-priority-rule-based algorithm for robust resource-constrained project scheduling, *Computers & Industrial Engineering* 55, 183-194.
- Dixit V., Srivastava, R.K. , Chaudhuri, A. (2014) Procurement scheduling for complex projects with fuzzy activity durations and lead times, *Computers & Industrial Engineering*, 76, 401-414.
- Frey H.C., S.R. Patil. (2002). Identification and review of sensitivity analysis methods. *Risk Anal* 22(3), 553-578.
- Gálvez E.D., S.F. Capuz-Rizo, J.B. Ordieres. (2012). Study of the uncertainty of task programming using the dependency structure matrix, *Información Tecnológica* 23(1) 19-34.
- Gálvez E.D., J.B. Ordieres, S.F. Capuz-Rizo. (2015). Evaluation of Project Duration Uncertainty using the Dependency Structure Matrix and Monte Carlo Simulations, *Revista de la Construcción*, in press.
- Herroelen W., R. Leus. (2005). Project scheduling under uncertainty: survey and research potentials, *European J. of operational research* 165, 289-306.
- Herroelen W. (2005). Project scheduling - Theory and practice, *Production and Operations Management*, 14 (4), 413-432.
- Helton J.C., J.D. Johnson, C.J. Sallaberry, C.B. Storlie, C.B. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis, *Reliability Engineering & System Safety* 91, 1175-1209.
- Homma T., A. Saltelli. (1996). Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety* 52, 1-17.
- Iooss B., Lemaître. (2015). A review on global sensitivity analysis methods. In *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, C. Meloni and G. Dellino (eds), Springer.

- Ionescu-Bujor M., D.G. Cacuci. (2004). A comparative review of Sensitivity and uncertainty analysis of large-scale systems. I: Deterministic methods, *Nuclear science and engineering* 147 (3), 189-203
- Jansen M.J.W. 1999. Analysis of variance designs for model output, *Computer Physics Communication* 117, 35–43.
- Kirshnan V., S.D. Eppinger, D.E. Whitney. (1997). A model-based framework to overlap product development activities, *Management Science*, 43(4) 437-451.
- Liang, T. (2009), Fuzzy multi-objective project management decisions using two-phase fuzzy goal programming approach, *Computers & Industrial Engineering*, 57, 1407-1416.
- Morris M.D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments, *Technometrics*, 33(2), 161-174.
- Maheswari J.U., K. Varghese. (2005). Project scheduling using dependency structure matrix, *International Journal of Project Management* 23 (3) 223-230.
- Meyer M.A., J.M. Booker. (2001). *Eliciting and Analyzing Expert Judgment: A Practical Guide* (ASA-SIAM Series on Statistics and Applied Probability).
- O'Hagan A., C.E. Buck, A. Daneshkhah, J. R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakley, T. Rakow. (2006). *Uncertain Judgments: Eliciting Experts' Probabilities*, John Wiley & Sons Ltd., England.
- Pektaş S.T., M. Pultar. (2006). Modelling detailed information flows in building design with the parametric-based design structure matrix, *Design Studies* 27, 99-122.
- Pujol G., B. Iooss, A. Janon. (2014). *Sensitivity: Sensitivity Analysis*. R package version 1.10.1. <http://CRAN.R-project.org/package=sensitivity>.
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Saltelli A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola. (2008). *Global sensitivity Analysis: The primer*, John Wiley & Sons Ltd.
- Saltelli A., K. Chan, E. M. Scott. (2009). *Sensitivity Analysis*, Wiley.

- Saltelli A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto S. Tarantola. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications* 181, 259–270.
- Shi Q., T. Blomquist, (2012). A new approach for project scheduling using fuzzy dependency structure matrix, *International Journal of Project Management*, 30 503-510.
- Sobol' I.M. (1993). Sensitivity analysis for non-linear mathematical model, *Math. Modelling Comput. Exp.* 1, 407–414.
- Sosa M.E., S. D. Eppinger, C.M. Rowles. (2004). The Misalignment of Product Architecture and Organizational Structure in Complex Product Development, *Management Science*, 50(12) 1674 – 1689.
- Srouf I.M., M.U. Abdul-Malak, A.A. Yessine, M. Ramadan. (2013). A methodology for scheduling overlapped design activities based on dependency information, *Automation in Construction*, 29, 1-11.
- Tang D., R. Zhu, J. Tang, R. Xu, R. He. (2010). Product design knowledge management based on design structure matrix, *Advanced Engineering Informatics*, 24, 159-166.
- Wang J., Y.I. Lin. (2009). An overlapping process model to assess scheduling risk for new product development, *Computers & Industrial Engineering*, 57, 460-474.
- Węglarz J., J. Józefowska, M. Mika, G. Waligóra. (2011). Project scheduling with finite or infinite number of activity processing modes-A survey, *European journal of Operational Research*, 208(3) 177-205.
- Zamani R., (2013). An evolutionary search procedure for optimizing time-cost performance of projects under multiple renewable resource constraints, *Computers & Industrial Engineering* 66, 451-460.

Table 1. Partial correlation coefficients and partial rank correlation coefficients for the example.

	PCC			PRCC		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
A	1.000	0.974	0.869	0.959	0.852	0.849
B	1.000	0.993	0.562	0.957	0.960	0.502
C	0.035	0.186	0.641	0.066	0.146	0.574
D	1.000	0.996	0.577	0.955	0.976	0.469
E	1.000	0.990	0.873	0.961	0.942	0.828
B_{BA}			0.278			0.225
B_{DB}			0.661			0.589
B_{ED}			0.566			0.476
C_{BA}			-0.507			-0.448
C_{DB}			-0.483			-0.402
C_{ED}			-0.324			-0.251

Table 2. Standardized regression coefficients and standardized rank regression coefficients for the example.

	SRC			SRRC		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
A	0.531	0.291	0.507	0.516	0.278	0.503
B	0.499	0.540	0.213	0.491	0.530	0.213
C	0.000	0.024	0.266	-0.015	0.010	0.273
D	0.542	0.709	0.172	0.516	0.709	0.157
E	0.568	0.437	0.530	0.569	0.398	0.514
B_{BA}			0.077			0.072
B_{DB}			0.177			0.168
B_{ED}			0.272			0.260
C_{BA}			-0.111			-0.106
C_{DB}			-0.124			-0.123
C_{ED}			-0.045			-0.064

Table 3: List of tasks with the uniform distribution in the duration (A_{ii}), as given by Gálvez et al. (2015).

Task Identification	Previous Information	Duration, U(min,max)
<i>A</i>	-	U(5.7,6.1)
<i>B</i>	<i>D</i>	U(7.3,9.1)
<i>C</i>	<i>A</i>	U(6.2,7.5)
<i>D</i>	<i>A, F</i>	U(3.8,4.0)
<i>E</i>	<i>B</i>	U(8.3,9.4)
<i>F</i>	<i>A, C</i>	U(0.9,1.0)
<i>G</i>	<i>F, J</i>	U(2.0,2.1)
<i>H</i>	<i>I</i>	U(9.4,10.7)
<i>I</i>	<i>D, G, E</i>	U(4.5,5.4)
<i>J</i>	<i>F, B</i>	U(2.9,3.0)

Table 4: Uniform distribution of time factors B_{ji} and time factors C_{ji} . (Gálvez et al., 2015)

j	i	Time	Time
		Factors B_{ji}	Factors C_{ji}
		U(min,max)	U(min,max)
C	A	U(0.79,0.80)	U(0.10,0.11)
F	A	U(0.99,1.00)	U(0.00,0.05)
F	C	U(0.57,0.60)	U(0.30,0.32)
D	A	U(0.79,0.80)	U(0.09,0.10)
D	F	U(0.67,0.70)	U(0.19,0.21)
B	D	U(0.87,0.90)	U(0.35,0.40)
J	F	U(0.89,0.90)	U(0.18,0.21)
J	B	U(0.92,1.00)	U(0.00,0.05)
G	F	U(0.46,0.50)	U(0.46,0.50)
G	J	U(0.96,1.00)	U(0.30,0.32)
E	B	U(0.83,0.90)	U(0.10,0.11)
I	D	U(0.56,0.60)	U(0.18,0.21)
I	G	U(0.79,0.80)	U(0.49,0.63)
I	E	U(0.96,1.00)	U(0.00,0.05)
H	I	U(0.63,1.00)	U(0.38,0.43)

Table 5. SRC and Sobol'-Jansen results for the case study.

Input factor	SRC	Sobol'-Jansen	Ranking	
Duration of Activities	A	0.1115	0.0131	11
	C	0.2135	0.0517	6
	F	0.0113	0.0002	21
	D	0.0396	0.0024	14
	B	0.3210	0.1015	3
	J	0.0013	0.0000	35
	G	-0.0046	0.0000	25
	E	0.3371	0.1213	2
	I	0.2489	0.0658	5
	H	0.2816	0.0783	4
Time factor B_{ji}	CA	0.0368	0.0018	15
	FA	-0.0011	0.0000	36
	FC	0.0782	0.0056	13
	DA	0.0030	0.0000	30
	DF	0.0118	0.0001	20
	BD	0.0366	0.0018	16
	JF	0.0046	0.0000	24
	JB	0.0031	0.0000	28
	GF	-0.0006	0.0000	38
	GJ	-0.0028	0.0000	32
	EB	0.1810	0.0432	7
	ID	0.0028	0.0000	33
	IG	0.0030	0.0000	29
	IE	0.1281	0.0166	10
HI	0.6606	0.4398	1	
Time factor C_{ji}	CA	-0.0300	0.0006	18
	FA	0.0036	0.0000	27
	FC	-0.0086	0.0000	22
	DA	0.0058	0.0000	23
	DF	-0.0349	0.0008	17
	BD	-0.1389	0.0224	9
	JF	-0.0008	0.0000	37
	JB	0.0015	0.0000	34
	GF	0.0001	0.0000	40
	GJ	0.0005	0.0000	39
	EB	-0.0296	0.0010	19
	ID	-0.0036	0.0000	26
	IG	-0.0030	0.0000	31
	IE	-0.0870	0.0081	12
HI	-0.1746	0.0332	8	

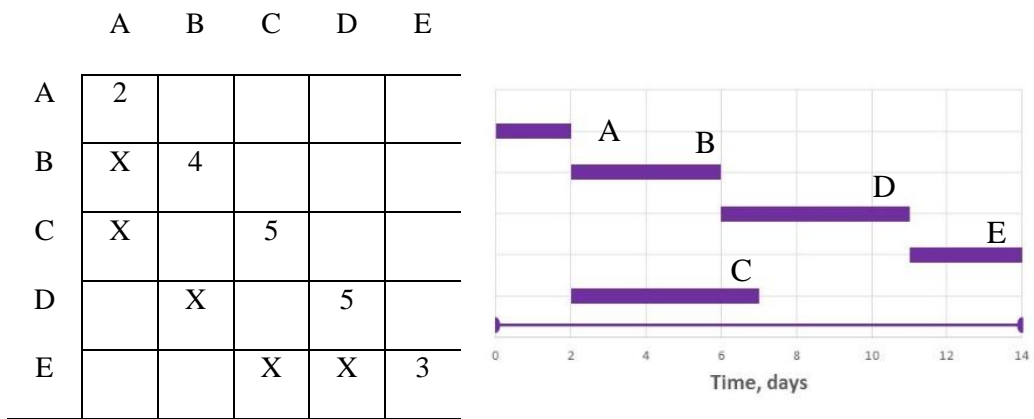
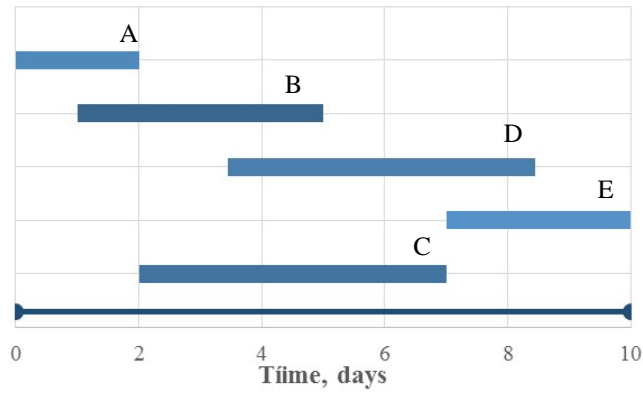


Figure 1. DSM showing the mean values of duration of activities (matrix A_{ij}) and conventional scheduling.



a) Overlap scheduling

	A	B	C	D	E
A	2				
B	0.8	4			
C	1.0		5		
D		0.8		5	
E			1.0	0.8	3

b) Mean time factor of processor activities (B_{ij})

	A	B	C	D	E
A	2				
B	0.15	4			
C	0		5		
D		0.15		5	
E			0	0.15	3

c) Mean time factor of receiving information (C_{ij})

Figure 2. DSM showing the mean values of duration of activities and time factor of transfer of information between activities, and overlap scheduling.

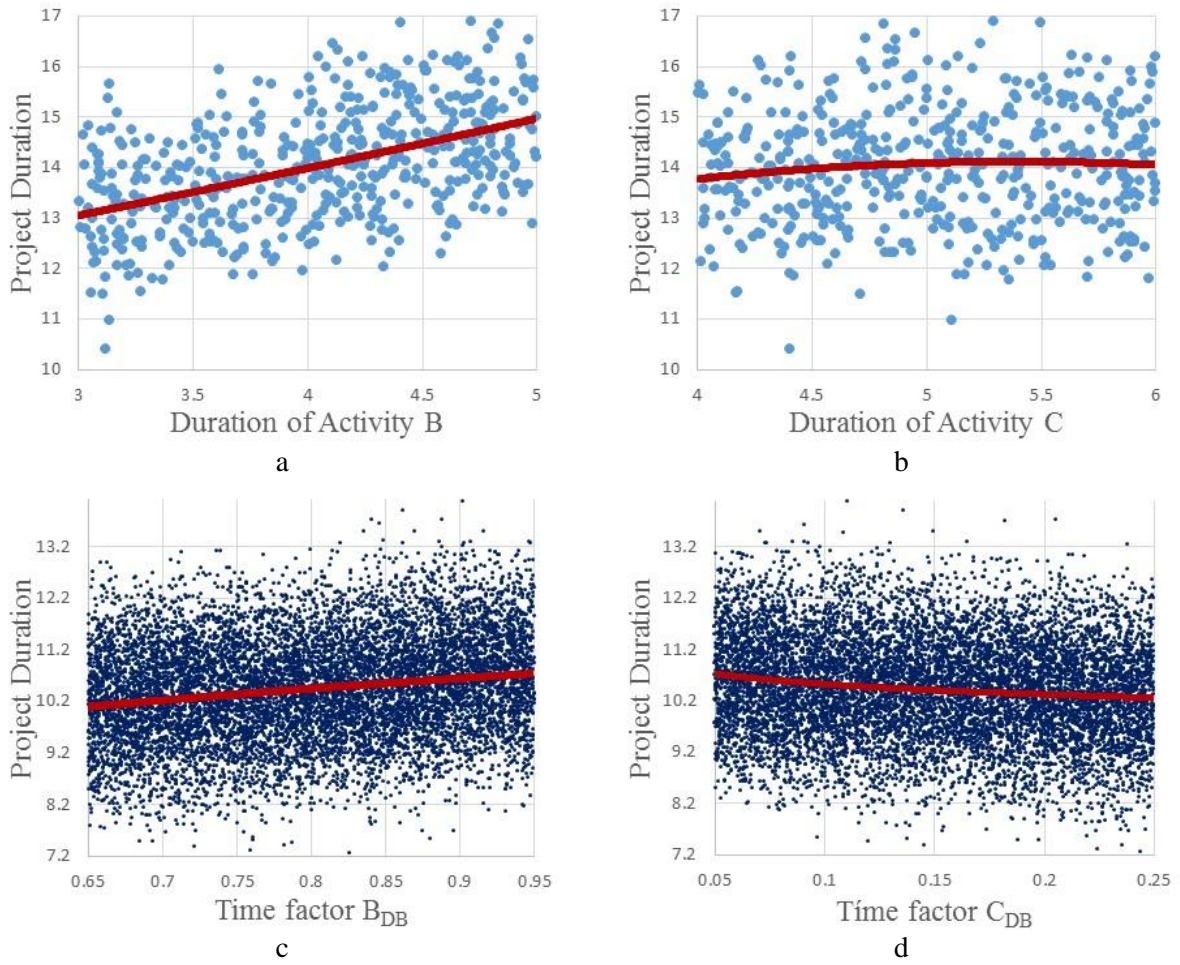
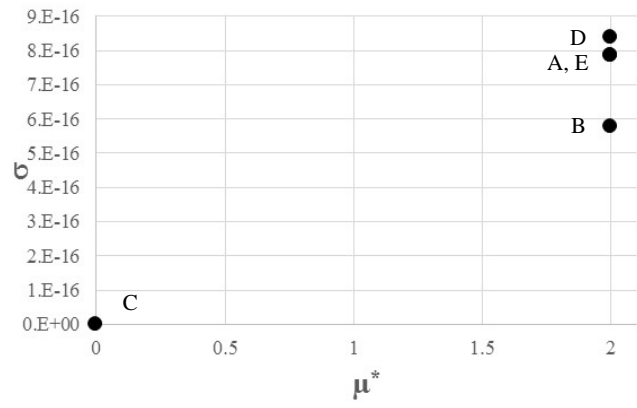
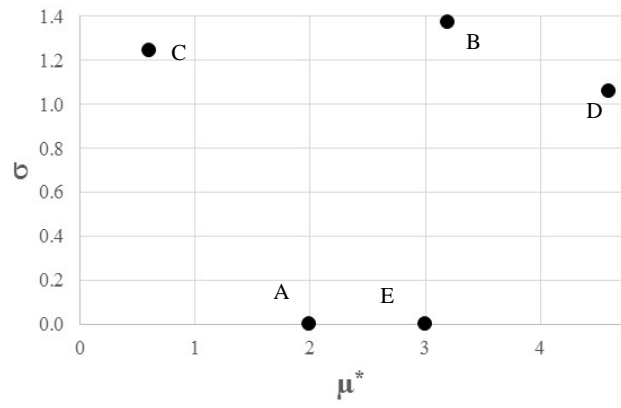


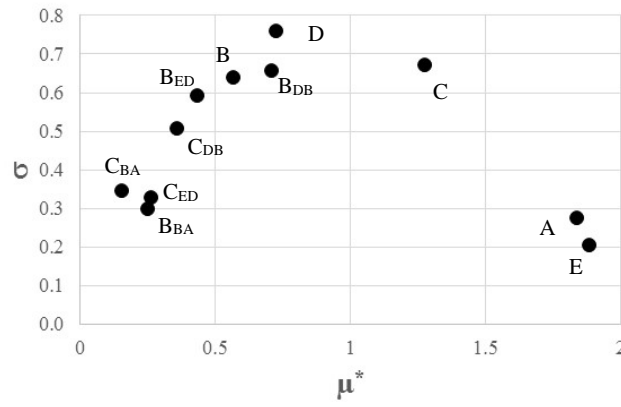
Figure 3. Example of scatterplots and polynomial smoother curves: a) duration of activity B for case 1; b) duration of activity C for case 1; c) time factor B_{DB} for case 3; and d) time factor C_{DB} for case 3.



a



b



c

Figure 4. Results of Morris method for a) case 1, b) case 2, and c) case 3.

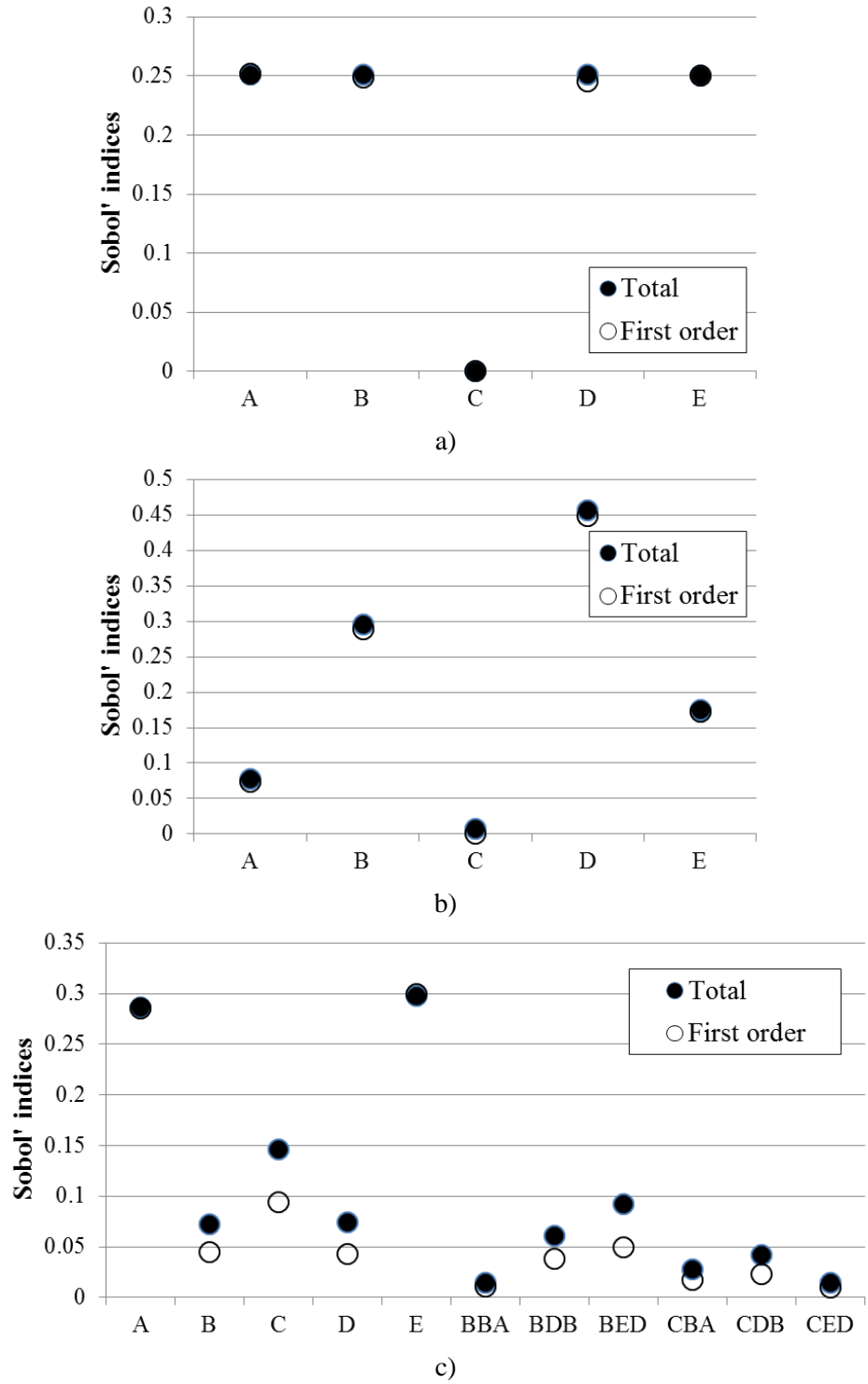


Figure 5. Results of Sobol' method for a) case 1, b) case 2, and c) case 3.

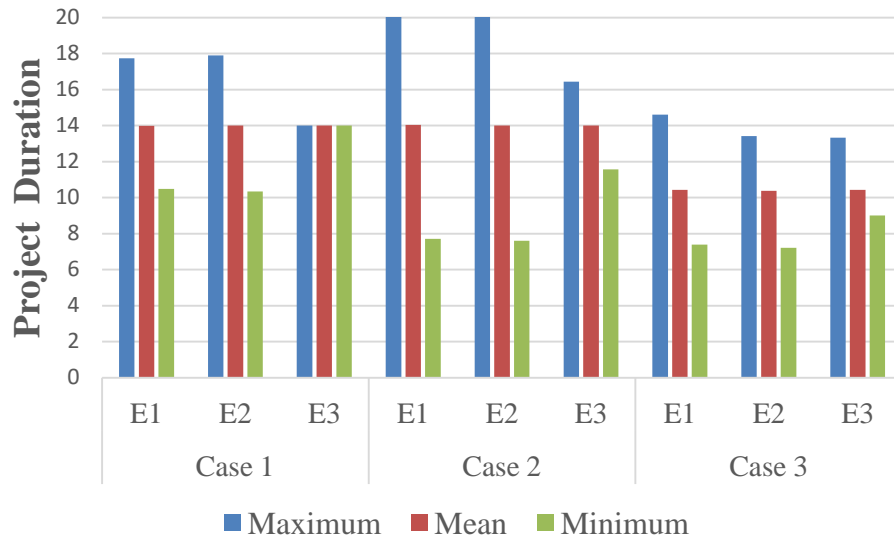


Figure 6. Numerical experiments using Monte Carlo simulation with 10,000 model calls.