Increasing the accuracy of a prefab building design process simulation using simulated annealing

Martin Haller*, Lars Stehn

*Luleå University of Technology, Division of Structural and Construction Engineering, 971 87 Luleå, Sweden

Abstract

Monte-Carlo simulation analysis has been discussed in project management literature as tool for proactive scheduling and to gain better insights into projects which are characterized by a high level of complexity and uncertainty, such as the design phase of prefab building projects. The application of simulation as proactive scheduling tool in construction projects is hampered by limited accessibility of proper input data, though, because of long project duration, the often temporary organization and multidisciplinary nature of such projects. In this study we use simulated annealing to adjust parameters of a simulation model for which the simulation outcome is sensitive to data perturbation by making use of data from related parameters which is easier to estimate. The applicability of the approach was demonstrated on a real life project, the construction of a 1100 m$^2$ residential building in Sweden. More precisely, we used Design Structure Matrix simulation, i.e. an activity network based Monte-Carlo simulation technique with which stochastic project evolution (deviations from the planned activity sequence due to unexpected iteration of sub-processes) can be simulated, to model the workflow of the design process of the observed project. Then, by means of the simulated annealing approach, we adjusted the rework probabilities (model parameter) such that the frequencies of executed activities in simulated activity sequences fitted the frequencies as observed in the real project. Adjusting input data by using prior knowledge of the dependencies of the project activities and cross analysis with related data that is easy to estimate would help to increase the accuracy of simulations when access to statistical data of the input variable in question is limited. The suggested approach is interesting for practitioners who work with standardized design processes (e.g. as part of standardized building systems) and continuous improvement.

Keywords: building design process; continuous improvement; design structure matrix; model calibration; proactive scheduling.

1. Introduction

Precise design and accurate scheduling of the design phase are crucial success factors in managing prefab building projects because of the high production rate of the standardized building elements (usually manufactured off-site) and their interdisciplinary nature. A high production rate implies that once the production has been started design errors will be multiplied. The interdisciplinary nature implies that inaccurate scheduling will further complicate the cooperation between the project members. As one of the major factors that adversely affect design and scheduling precision, iteration induced by design changes have been identified in numerous studies (e.g. Love et al. 2008; Hwang et al., 2009; Olawale and Sun, 2010). The design process of building projects is often complex, which implies that already a small change during execution can cause significant effects on the project outcome (Austin et al. 2002, Pektas and Pultar 2005). The complexity together with uncertainty about project conditions makes it a challenge for project managers to predict the effects on the project outcome of any design change during the execution of the project.

An established means in engineering design to get insights into and to develop robust design processes of projects characterized by a high level of complexity and uncertainty is Design Structure Matrix (DSM) simulation (e.g. Cho and Eppinger, 2005; Karniel and Reich, 2009). DSM simulation is an activity-network based technique with which variation in project duration is simulated to origin both from variation in activity duration and variation of the activity execution sequence. Usually, the execution sequence is modeled to follow the activity dependency logics as defined by an activity network. But the logical sequence is randomly perturbed in that sub-processes have to be iterated (Cho and Eppinger, 2005). By iterations it is simulated that, as consequence of an unforeseen information change during the execution of a design project, successor activities that have already been carried out have to be reworked (Yassine, 2007; Srour et al. 2013). It is obvious that the quality of process optimization which is based on a numerical model significantly depends on the quality of the numerical model used, and consequently of the accuracy of the model input data. As indicated above, the
probability with which iteration of a sub-process is triggered (i.e. the rework probability) constitute a core input parameter of a DSM simulation model. Usually rework probability is estimated by using historical data or, if not accessible, by expert interviews. Also indirect interview techniques have been applied to facilitate the estimating for practitioners (e.g. Yassine, 2007). However, the use of historical data or expert interviews to assess the rework probability for the design phase of building projects is often hampered by the temporal organizing and long project durations, i.e. a lack of accessible data.

In this paper we propose a new approach to assess the rework probabilities of a DSM simulation model. The idea is to reduce estimation uncertainty by making use of related data that is easier to access than the rework probabilities, i.e. the mean frequencies of activities of observed project execution sequences.

The mean frequency of an activity depends on the underlying process structure (which defines the precedence constraints) and the rework probabilities. The task is to find a configuration of the rework probabilities such that the activity frequencies of the simulated execution sequences fit the observed ones. Because the stochastic nature of DSM simulation it is not possible to find an analytical solution for this problem. However, DSM simulation can be used itself to generate the frequencies (by generating execution sequences) and then simulated annealing (a Monte Carlo optimization technique) can be used to iteratively fit them to the target frequencies.

The applicability of the approach was demonstrated on a real life project; the design process of a 1100 m$^2$ residential building in Sweden.

2. DSM-based simulation

Different models using DSM-based simulation have been presented in literature. The following description relates to the model created by Cho and Eppinger (2005), which is basically a discrete event model that makes use of DSM to define the rework probabilities.

The basic concept in DSM-based simulation considers variations in project costs and time largely a function of the iteration required in the project’s execution (Yassine, 2004). Figure 2 shows the activity sequence of a sampled project and the underlying process structure. The process structure is represented by a digraph; the activities of the process are depicted by the vertices A1 to A6 and the precedence constraints by the straight arcs. A process structure defines the order in which activities should be carried out and whether activities can be carried out concurrently. The project sequence is represented by the bold arcs. It represents the order in which a project was carried out.

Theoretically, the project in figure 2 requires only 5 transitions from start to finish. But because a change in activity 2 after the 5$^{th}$ transition occurred (first order rework) 8 transitions were required instead. Since A3 depends on input from A2 it had to be reworked too (second order rework). The progression of the project sequence is defined by the dependency logics, the rework probabilities, and a priority role. The rework probabilities are represented by a matrix, where the $i,j^{th}$ element gives the probability that if activity $i$ is the current one activity $j$ will be carried out next. The duration of a project is simulated by interlinking the chronological arrangement of the activities (i.e. concurrent or sequential execution), and the sampled duration of each activity. A more detailed description of DSM-based simulation can be found in e.g. Cho and Eppinger (2005).
3. Simulated annealing

Simulated annealing is a probabilistic numeric method utilized to find a global minimum over a discrete search space \( S = \{ s_1, \ldots, s_n \} \) by iteratively improve a candidate solution \( s_i \in S \) with regard to a quality measure \( f: S \rightarrow \mathbb{R} \), called the energy function. That is, to find a \( s_i \in S \) which minimizes \( f(s_i) \). The method was inspired by the annealing process in metallurgy and first presented by Kirkpatrick et al. (1983). The basic procedure is that \( s_i \) is altered until the global minimum is reached, or until \( f(s_i) \) is sufficient small. The alteration of the candidate solution is often called ‘picking a neighbour’ and is done systematically. The altered candidate solution will be accepted with the probability

\[
\min\{\exp\left(\frac{f(s_i)-f(s_{i+1})}{T}\right), 1\}, \quad \text{where} \quad T \in \mathbb{R}^*.
\]

(1)

That means, the new solution is accepted: in any case if \( f(s_{i+1}) \leq f(s_i) \), and only with a certain probability if \( f(s_{i+1}) > f(s_i) \). The acceptance rate depends on the temperature \( T \) which is decreased towards 0 with the number of iterations; the lower the temperature the lower the acceptance rate. A deeper theoretical background is given in e.g. Häggström (2002) and Kincaid and Cheney (2012).

4. Rework probability estimation

As conclusion from the preceding discussion we propose the following approach to estimate the rework probabilities (figure 3):

- **Objective**: The objective is to find a configuration of the rework probability matrix such that the average frequency for each activity of the simulated sequences will fit the frequencies from the observed sequences. In other words, we want to find a configuration of the rework probability matrix \( \xi \) such that \( f(\xi) = \|h_{\text{sim}} - h_{\text{obs}}\|_2^2 \) gets minimal, where \( h_{\text{sim}} \) is a vector with the average frequency for each activity of the simulated projects as its elements and \( h_{\text{obs}} \) the corresponding frequencies of the observed projects.

- **Required input**: When starting the approach, observed activity sequences of several projects as well as their underlying process structure (common, standardized process) must be available. Methods to derive a DSM-model of the underlying process structure are described in e.g. Steward (1981) and Yassine (2004).

- **Initialize**: The first steps are to calculate \( h_{\text{obs}} \) from the observed projects and to do an initial guess \( \xi^{(0)} \) for the rework probability matrix. \( \xi^{(0)} \) is derived from the observed activity sequences by simply recording the number of transitions from a certain activity to the next in an adjacency matrix (figure 4.b) and then dividing each matrix row by its own sum (we require the sums of each matrix row to be 1). \( \xi^{(0)} \) corresponds then to the lower triangular entries of the adjacency matrix.

- **Loop**: Now the simulated annealing approach can start. The first step is to choose a new neighbour (see section 3). This is done by changing some of the elements of the current rework probability matrix: \( \xi \rightarrow \xi' \). In order to restrict the search to the neighbourhood we allow only elements bigger 0 to change. The next step is to simulate sufficient many project activity sequences (sufficient many in terms of Monte-Carlo simulation) and calculate thereof \( h_{\text{sim}} \), and subsequently \( f(\xi') \). Finally, it has to be decided whether \( \xi' \) is accepted as the new current configuration (equation 1). Repeat these steps until \( f(\xi) \) is sufficient small.
5. Prefab building design process application

In order to test the applicability of the suggested approach, it was applied on a real-life project. The data for the underlying process model originate from a design-build construction project i.e. the construction of a two-storey apartment building with an area of 1100 m². The structural frame of the apartment building consists of prefab timber frame system elements (Masonite’s Flexible Building system MFB®). The design was drawn up by a project team of SME technical consultants.

Data collection and the compilation of the process model were conducted iteratively between September 2009 and September 2010. In line with the recommendations outlined by van der Aalst (1999; 2011) the project was mapped at the work step level, i.e. all activities in which information (in the form of documents and agreements or material) was transferred between at least two of the main actors were logged. The log file of the project activities provided descriptions of each activity undertaken, their dates of execution, the number of hours worked, and the required inputs. In order to triangulate these data, the contents of the log file were discussed with the project leader over the course of 10 meetings. To further ensure the correctness of the logged activities, further information on the proceedings was gathered from 14 project-meetings. Finally, based on the information about required inputs a topological sort was generated by means of a DSM partitioning algorithm. Note that the process structure model was compiled iteratively. It was presented (in parts) to the main actors during three project meetings to confirm its correctness. The resulting network model (figure 4.a) consists of 98 activities with 376 dependency relations. It spans the process from the compilation of the client’s specifications through the pre-tendering phase to the completion of the building plans. A detailed process description is given in Haller (2012).

Figure 4. The modelled process structure (a) and the observed transitions (b)

Once the required input was compiled, the rework probabilities were estimated by means of the approach described in section 4. Figure 5 shows the residuals i.e. the differences between the activity frequencies from the simulated execution sequences and the observed activity frequencies.

Figure 3. Schematic diagram of the rework probability estimation algorithm
A positive residual means that the simulated frequency was higher than the observed one. The closer the residuals are to 0 the better is the estimation. Note the residuals between activities 16 and 33 (planning of technical details in the pre-tendering phase). A possible explanation for the peaks could be that the main actors did not completely agree about how the process structure of this phase should be modeled. There was some uncertainty about the dependencies of the activities. Thus, perhaps the model partly does not match the real structure. Another explanation could be that it was quite difficult to catch all transitions between the project participants during this phase (figure 4.b), since many technical details were discussed via telephone. These transitions were not recorded but had to be reconstructed afterwards. That means that there is also some uncertainty about the correctness of the observed activity sequence in this phase.

Figure 6 shows the length of the simulated activity sequences (in number of activities). The mean length of 10 000 simulated sequences is 207.5, which is close to the length of the sequences of the observed project (202 activities). In contrast, the mean length of 10 000 simulated sequences generated by using direct observations, i.e. without fitting, was 148.6. The distribution of the simulated project lengths is right skewed, which is in line with Banks (2001) and Robinson (2007). Considering the results of this study, we assume that by using the proposed approach we could increase the plausibility of the project outcome.

6. Conclusions

A core component of DSM simulation models is the rework probabilities. Rework probabilities are usually estimated by direct observation or expert interviews. However, the application of these methods in construction projects is often hampered by limited accessibility of proper input data, because of long project duration, the often temporary organization and their multidisciplinary nature. This study presents a new estimation procedure for rework probabilities that can be used in DSM based simulation models. It is an iterative, activity network based approach where use of related model data is made i.e. the underlying process structure and the activity frequencies of observed project sequences. The suggested approach proceeds in three phases. First, the underlying process structure has to be derived by means of DSM sequencing algorithms and a first guess of the rework probabilities by inference from the observed project sequences. Then, project execution sequences are
sampled by means of DSM based simulation and the activity frequencies of the simulated sequences are compared with the activity frequencies derived from the observed sequences. Finally, the first guess is improved by a Monte-Carlo optimization technique, i.e. simulated annealing, until the simulated activity frequencies are sufficient close to the observed frequencies.

The approach was tested on a real life project, the construction of an 1100 m² residential building in Sweden. The results of the study indicate that by applying the approach it is possible to receive more accurate rework probability estimates than by direct observation. Adjusting input data by using prior knowledge of the dependencies of the project activities and cross analysis with related data that is easy to estimate would help to increase the accuracy of simulations when access to statistical data of the input variable in question is limited. A prerequisite is, though, that the observed project sequences are resulting from a common underlying design process and that this process must be well defined. The suggested approach is interesting for practitioners who work with standardized design processes (e.g. as part of standardized building systems) and continuous process improvement.

In this study it was assumed that adjusting the rework probabilities with reference to the observed activity frequencies will increase the accuracy of the model. However, preliminary additional tests indicate that this is not always the case. Rules have to be developed that will help the practitioner to assess the potential efficiency of the estimation procedure for a given problem in advance.

References