 Open access • Journal Article • DOI:10.1061/(ASCE)CO.1943-7862.0001022

Empirical perspective on activity durations for project management simulation studies

— [Source link](#) 

Jeroen Colin, Mario Vanhoucke

Institutions: Ghent University

Published on: 01 Jan 2016 - Journal of Construction Engineering and Management-asce (American Society of Civil Engineers)

Topics: Project management simulation and Project management

Related papers:

- [Modeling activity times by the Parkinson distribution with a lognormal core: Theory and validation](#)
- [Construction and evaluation framework for a real-life project database](#)
- [On the dynamic use of project performance and schedule risk information during projecttracking](#)
- [Measuring Time: Improving Project Performance Using Earned Value Management](#)
- [Setting tolerance limits for statistical project control using earned value management](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/empirical-perspective-on-activity-durations-for-project-156xq3a6nz>

Empirical Perspective on Activity Durations for Project-Management Simulation Studies

Jeroen Colin¹ and Mario Vanhoucke²

Abstract: Simulation has played an important role in project-management studies of the last decades, but in order for them to produce practical results, a realistic distribution model for activity durations is indispensable. The construction industry often has needed historical records of project executions, to serve as inputs to the distribution models, but a clearly outlined calibration procedure is not always readily available, nor are their results readily interpretable. This study seeks to illustrate how data from the construction industry can be used to derive realistic input distributions. Therefore, the Parkinson simulation model with a lognormal core is applied to a large empirical dataset from the literature and the results are described. From a discussion of these results, an empirical classification of project executions is presented. Three possible uses are presented for the calibration procedure and the classification in project management simulation studies. These were validated using a case study of a construction company. DOI: 10.1061/(ASCE)CO.1943-7862.0001022. © 2015 American Society of Civil Engineers.

Author keywords: Project management; Empirical data; Classification; Simulated data.

Introduction

The value of Monte Carlo simulations (MCS) for project management has long been established (Schonberger 1981; Ragsdale 1989; Kwak and Ingall 2007). Owing to the advances in information technology of the last few decades, simulation studies now stand at the core of newly developed techniques in project management (Trietsch and Baker 2012; Elshaer 2013; Wauters and Vanhoucke 2014), and serve to test them and quantify their performance in large computational experiments (Colin and Vanhoucke 2014; Vanhoucke 2010a). In these simulations, the stochastic activity durations for the activity networks are drawn from a predefined distributional model. The parameters for the distributions that are proposed in the literature, however, are not always as clear and intuitive as one might hope. Moreover, they often lack a clearly outlined calibration method with which, in practice, input distributions can be generated from empirical records of project executions. A recent exception is the Parkinson distribution with a lognormal core (Trietsch et al. 2012), for which a calibration procedure was presented, alongside its theoretical description. This paper expands on the work of Trietsch et al. (2012) by applying it to the database of Batselier and Vanhoucke (2015). This database currently consists of 51 real-life projects, predominantly from the construction industry, which are freely available for researchers (<http://www.or-as.be/research/database>) in the *ProTrack* file format (Vanhoucke 2010a). However, this paper aims not to draw general conclusions,

but rather provide an illustration of how a calibration to a large empirical database can be performed and how its results can be interpreted. In order to generate practically applicable activity duration inputs for project management simulation studies, researchers and practitioners should always consider the use of their own historical data of project executions. But, even when the dataset of Batselier and Vanhoucke (2015) is expanded in the coming years, some project environments will still be too specific to be correctly referenced against the real-life project executions in this dataset.

The contribution of this paper is threefold. First and foremost, the first application is presented of a theoretical distribution to a large database of empirical records in project management. From the thirteen projects that confirmed the theoretical distribution and which were retained for further analysis, eleven were performed in the construction industry and two were performed in the information technology (IT) and education industries, respectively. This study is therefore not restricted to the construction industry, but special attention is paid to the interpretation of the results from construction projects. In order to promote the reproducibility of this study and to encourage its use on practitioner databases, the calibration procedure by Trietsch et al. (2012) has been implemented in the open source statistical programming language *R* (R Core Team 2013). Second, an empirical classification for records of project executions is presented, which can assist in the interpretation of historical data and serve as input for project management simulation studies. Third, it is shown how the calibration procedure and the empirical classification can be usefully applied in the project management practice by validating them in a case study on a Belgian construction company.

The outline of this paper is as follows. The “Literature Overview” section introduces the distributional model and calibration procedure by Trietsch et al. (2012). The “Methodology” section discusses the data cleaning that was performed on the online dataset of Batselier and Vanhoucke (2015) and applies the calibration procedure. Using the outputs of this procedure, the empirical classification of project executions are proposed in the “Results” section. The “Discussion and Conclusions” section offers validation for the use of the calibration procedure and the classification method, using a case study on project management in a Belgian construction company. In addition, both the work that was done for this paper

¹Research Assistant, Dept. of Management Information Science and Operations Management, Faculty of Economics and Business Administration, Ghent Univ., Tweekerkenstraat 2, 9000 Ghent, Belgium. E-mail: jeroen.colin@ugent.be

²Full Professor, Dept. of Management Information Science and Operations Management, Faculty of Economics and Business Administration, Ghent Univ., Tweekerkenstraat 2, 9000 Ghent, Belgium (corresponding author). E-mail: mario.vanhoucke@ugent.be

Note. This manuscript was submitted on November 19, 2014; approved on April 21, 2015; published online on June 25, 2015. Discussion period open until November 25, 2015; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Construction Engineering and Management*, © ASCE, ISSN 0733-9364/04015047(13)/\$25.00.

and its value for future project management research are summarized. In the “Appendix” of this paper, a working R template is presented to reproduce the results from this paper.

Literature Overview

The distributional model proposed by Trietsch et al. (2012) is introduced in this section. First, the level of detail needed in the empirical data is discussed, along with how these data are interpreted. Second, the application in our study of the calibration procedure by Trietsch et al. (2012) is illustrated.

Focus on the Relative Duration of Activities

The desirable level of detail in empirical records for this study lies at the activity level of the work breakdown structure (WBS) of a project. Concretely, it is ideal to have planning and progress data as complete as possible for all activities in a project. The justification for this is that the application of the calibration procedure is targeted to project-management studies that depart from the schedule performance of a project. In these studies, the baseline schedule is considered as a point of reference throughout the execution of the project, and that together with the real-schedule performance and the cost associated with it, are compared relative to the baseline estimates. In the MCS for such project-management studies, the individual activities are given a duration that might deviate from the estimated duration that was used for the construction of the baseline schedule (Vanhoucke 2010b, 2011; Elshaer 2013; Colin and Vanhoucke 2014). The relation to real-life projects is evident, since only very seldom a project will go exactly as planned in terms of the durations of the individual activities and their associated costs. It is not within the scope of this paper to test and to compare all the different theoretical distributions for activity durations that have been proposed for use in project-management studies in the literature. Among these are the generalized beta distribution (Kuhl et al. 2007), the lognormal distribution (Mohan et al. 2007) and a mixture of a beta and a uniform distribution (Hahn 2008). These distributions are valid in their own right for the generation of activity durations in project-management simulations, since they all adequately model the real-world excess variability and overoccurrence of tail area events (McCullagh and Nelder 1989). In order not to risk presenting an oversimplified summary of these distributional models, the original papers are referred to for theoretical background and practical implementations. However, the model describe proposed by Trietsch et al. (2012) will be described in more detail, since it has been chosen to implement it in this study because of the structured manner in which the Parkinson effect and the influence of rounding errors are incorporated in the accompanying calibration method.

Consider a set of projects \mathcal{P} in an empirical database. For a project $j \in \mathcal{P}$, that has a number of activities N_j , d_{ij} represents the real duration of activity $i = 1, \dots, N_j$ in project j . Correspondingly, for that activity, \hat{d}_{ij} denotes the baseline estimate duration that was used in the construction of the baseline schedule. Therefore, this requires the empirical records to have data available such that the relative durations d_{ij}/\hat{d}_{ij} for all activities in all projects can be calculated. If the proportion $d_{ij}/\hat{d}_{ij} > 1$, then activity i of project j has experienced a delay (tardy). The real duration of the activity was larger than its anticipated duration when the baseline schedule was constructed. If $d_{ij}/\hat{d}_{ij} < 1$, the activity was finished in less time than planned (early). Trietsch et al. (2012) assume the proportion d_{ij}/\hat{d}_{ij} to be lognormally distributed, even though this might not always be directly observable. In other words, the natural logarithm of d_{ij}/\hat{d}_{ij} is assumed to be normally distributed.

The lognormal distribution is reasoned to effectively incorporate both the additive and multiplicative effects that are omnipresent in a real-life project execution environment. This distributional model was validated by Trietsch et al. (2012) for activity times on several independent datasets from project scheduling applications by incorporating two important practical considerations. These considerations explain why lognormality is not always apparent at first sight in real-life data. The first consideration is the *Parkinson effect* which is translated into the reported data as hidden earliness. In projects, workers are generally not incentivised to report early activities. Rather, they will effectively use up all time allotted to a given task by starting late by multitasking, shifting resources, or simply reporting the task as on-time (Gutierrez and Kouvelis 1991). Where in reality $d_{ij}/\hat{d}_{ij} < 1$ might have been possible, it is more likely that $d_{ij}/\hat{d}_{ij} = 1$ will be reported. The second consideration handles *rounding effects*. In practice, the reporting of activity performances is mostly done on a coarse time scale and reported times will be influenced by rounding errors. This problem can be recognized easily when a project with large differences in baseline estimate durations is considered. Typically, a single unit for reporting the time performance of activities will be chosen. For example, if a project has a large number of activities that are planned to take a couple of weeks and a small number of activities that take around 3–5 days, the unit for reporting activity performance is likely to be days. The rounding that occurs due to the reporting in days will surely have a larger influence on the analysis from the proportion d_{ij}/\hat{d}_{ij} for the small activities than for the larger activities. Concretely, when the Parkinson effect and the rounding effects are considered, Trietsch et al. (2012) conjectured that the relative empirical distributions d_{ij}/\hat{d}_{ij} follow a Parkinson distribution with a lognormal core.

In addition to the validation of the Parkinson distribution with a lognormal core, Trietsch et al. (2012) validated the assumption of linear association in project management. According to the authors, the lognormal distribution lends itself to be used when statistical dependence is modelled using linear association. This will not be considered further, as this is only relevant for when an actual simulation model is proposed. A recent application of linear association in project management simulation can be found in Colin and Vanhoucke (2014). Here, the primary concern lies with the calibration of the inputs to such simulation studies.

Calibration Procedure Proposed by Trietsch et al. (2012)

In order to guide the reader through the calibration procedure proposed by Trietsch et al. (2012), an illustrative record of an executed project is introduced. This is a record of a construction project for a commercial building, taken from the dataset of Batselier and Vanhoucke (2015) (C2013-09: “Urban Development Project”), that has a total value of around 1.5 million Euro and a planned duration of 291 working days. The scope of the project was development of an integrated building complex, and the project was actually completed within 360 working days and experienced a budget overrun of around 13%. The project schedule consisted of 71 activities for which both baseline estimate durations and actual durations were reported. The relative empirical distribution d_{ij}/\hat{d}_{ij} derived from the $i = 1, \dots, 71$ activities is depicted in the histogram in Fig. 1. This figure depicts the early activities in white ($d_{ij}/\hat{d}_{ij} < 1$), the on-time activities in grey ($d_{ij}/\hat{d}_{ij} = 1$), and the tardy activities in black ($d_{ij}/\hat{d}_{ij} > 1$).

An application of the Parkinson distribution with a lognormal core to the presented project is now presented. The calibration

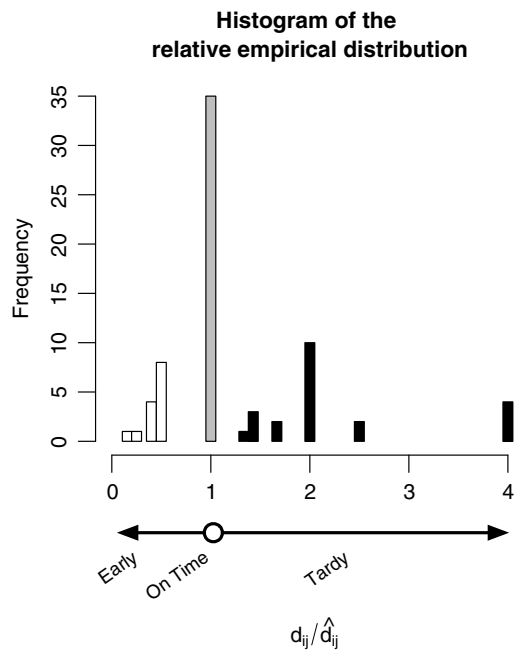


Fig. 1. Relative empirical distribution for the illustrative record of an executed project

procedure proposed by Trietsch et al. (2012) has been structured in the following way. A maximum of four hypothesis tests (P0, P1, P1.1, and P1.2) can be conducted in this order to confirm the Parkinson distribution with a lognormal core. These tests will be discussed in Steps 1–4 of this section. The null hypothesis for these tests will also be explicitly stated.

This procedure follows a sequential testing mechanism, starting with Step 1. If in a certain step, the hypothesis test is not able to confirm the Parkinson distribution with a lognormal core, the procedure proceeds with the next step. Consequently, the test in Step 4 will only be conducted if all prior tests fail to confirm the Parkinson distribution with a lognormal core.

Step 1 (P0)

A visual inspection of Fig. 1 reveals that the assumption of lognormality for the relative empirical distribution d_{ij}/\hat{d}_{ij} seems reasonable. The null hypothesis $H_{0,P0}$ for the test P0 is stated and this hypothesis is tested according to the procedure proposed by Trietsch et al. (2012) as follows:

$$H_{0,P0}: \ln(d_{ij}/\hat{d}_{ij}) \text{ follows a normal distribution}$$

In order to accept or reject $H_{0,P0}$, Trietsch et al. (2012) suggested comparing the Pearson's linear correlation coefficient (R) to the tabulated values by Looney and Gullledge (1985). The linear correlation coefficient can be used for testing the assumption of normality, by the use of normal probability plots. Looney and Gullledge (1985) quantified this use in their paper and shown that a p -value can be obtained for this normality assumption test. The linear correlation coefficient R should be calculated from a linear regression of the empirical values of $\ln(d_{ij}/\hat{d}_{ij})$ on the Blom scores (Blom 1958) for these values (Trietsch et al. 2012). Blom proposed the use of $\Phi(r - 3/8)/(n + 1/4)$ to estimate the normal scores for a sample, where r represents the rank of an observation in a sample, n represents the sample size, and $\Phi(p)$ is used to denote the p th quantile of the standard normal distribution.

Fig. 2(a) shows the Blom scores for $\ln(d_{ij}/\hat{d}_{ij})$ plotted against the empirical values. Fig. 2(a) clearly indicates the large number of activities that have $\ln(d_{ij}/\hat{d}_{ij}) = 0$. This means that in this project, many activities were reported as exactly on-time, since $\ln(d_{ij}/\hat{d}_{ij}) = 0 \Rightarrow d_{ij}/\hat{d}_{ij} = 1$. The linear regression line that is shown in Fig. 2(a) tries to explain the values from $\ln(d_{ij}/\hat{d}_{ij})$ using a linear fit on their Blom scores. According to Trietsch et al. (2012), the intercept and the linear coefficient of this regression model respectively estimate the mean (\hat{m}) and the standard deviation (\hat{s}) of the natural logarithm of the relative empirical distribution d_{ij}/\hat{d}_{ij} . The R value found in this regression results in a p -value of 0.004 when it is compared to the tabulated values of Looney and Gullledge (1985). Given a level of significance $\alpha = 5\%$, the null hypothesis must be rejected (p -value $< \alpha$). The variable $\ln(d_{ij}/\hat{d}_{ij})$ can not be assumed to be normally distributed at this point in the analysis. Therefore, Step 2 of the procedure is now implemented.

Step 2 (P1)

Given the large proportion p_p of activities reported as exactly on-time ($p_p = 35/71 = 49\%$, depicted in grey in Fig. 1), it is very likely that the Parkinson effect is present in these data. In order to prevent this effect from obscuring the interpretation of the data, the first step is to remove all activities that were reported as completed on time. Trietsch et al. (2012) described this step as accounting for the (pure) Parkinson distribution. Fig. 2(b) shows the plot of the empirical data and the Blom scores, with these on-time activities removed. Again, R , \hat{m} and \hat{s} are also shown. A new hypothesis ($H_{0,P1}$) can now be tested using R and the tabulated values in Looney and Gullledge (1985)

$$H_{0,P1}: \ln(d_{ij}/\hat{d}_{ij}) \text{ follows a normal distribution with a Parkinson effect after the removal of the on-time points}$$

The resulting p -value of 0.006 again leads to the rejection of this new null hypothesis, using the significance value α of 5%.

In addition to trimming the sample of on-time points, Trietsch et al. (2012) propose two additional treatments to test whether d_{ij}/\hat{d}_{ij} indeed follows the Parkinson distribution with a lognormal core. For an extensive coverage and justification for these treatments, refer to the original article. Here, only how they affect the present discussion of the illustrative empirical record will be shown. The first treatment is contained in the test that is described in Step 3. If this test is still unable to accept the Parkinson distribution with a lognormal core for the project data, proceed to the test in Step 4, which implements the second treatment.

Step 3 (P1.1)

The first treatment that is implemented is the removal of a proportion p_p (the proportion of on-time points, depicted in grey in Fig. 1) of the strictly tardy points (black in Fig. 1). The justification is that if the Parkinson distribution applies, the resulting data would constitute a trimmed but complete sample in which early and tardy points are approximately in the right proportion (Trietsch et al. 2012)

$$H_{0,P1.1}: \ln(d_{ij}/\hat{d}_{ij}) \text{ follows a normal distribution with a Parkinson effect after the removal of on-time points and after removing } p_p \text{ strictly tardy points}$$

Following the example of Trietsch et al. (2012), and two trimmed samples are used. From the 22 tardy activities (in black), 11 points [49% (p_p) of 22] were removed. This results in two samples, each having 11 different strictly tardy points. Fig. 2(c) shows the plots for these two samples where the empirical values are

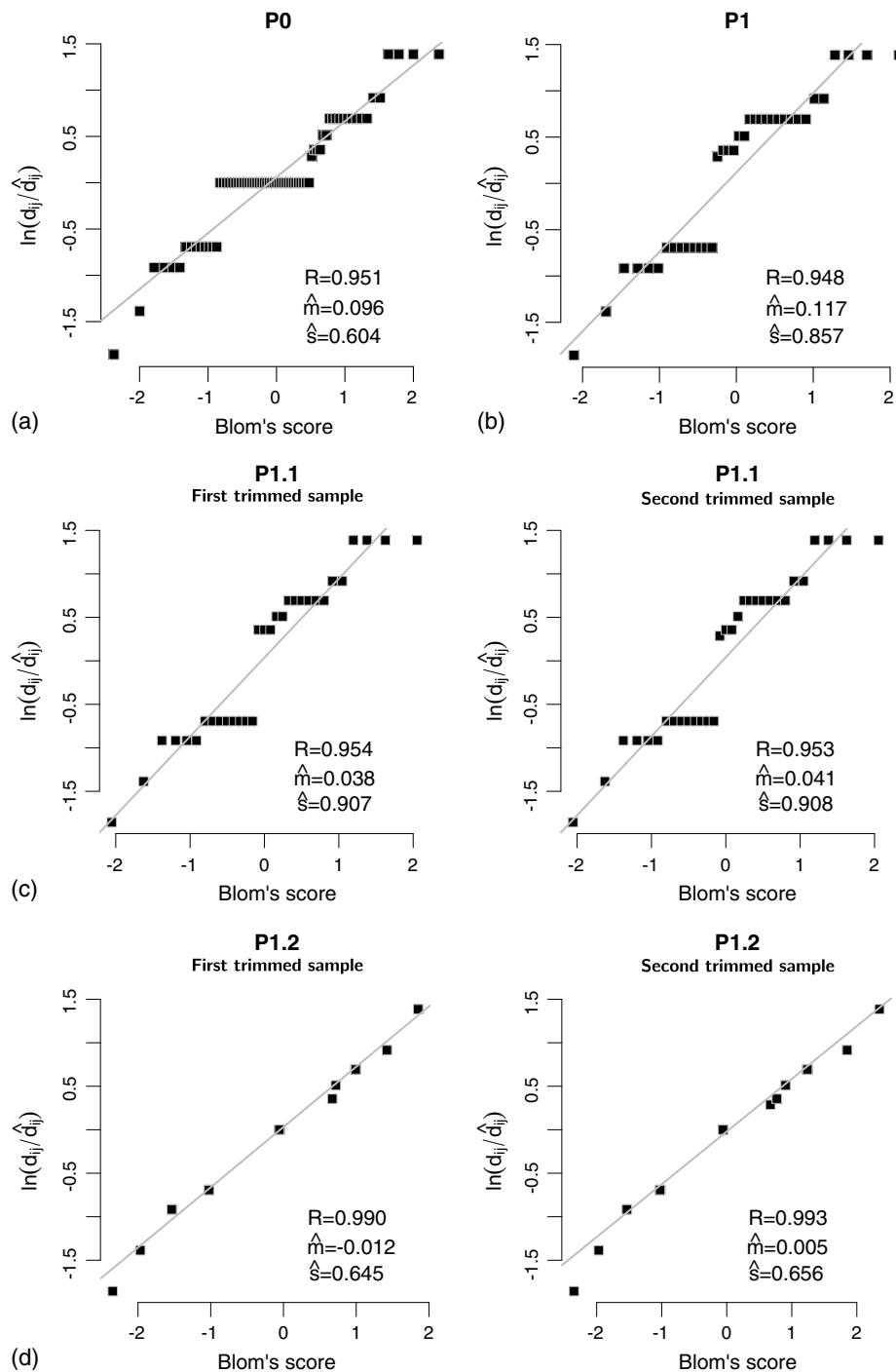


Fig. 2. Application of the four-step calibration on C2013-09 “Urban Development Project”: (a) hypothesis test for P0 on the relative empirical distribution; (b) hypothesis test for P1 on the relative empirical distribution; (c) hypothesis test for P1.1 on the relative empirical distribution; (d) hypothesis test for P1.2 on the relative empirical distribution

depicted against their Blom scores. This first treatment results in two p -values (0.004 and 0.004 for the first and second trimmed sample respectively), which both lead to the rejection of the null hypothesis $H_{0,P1.1}$ at the 5% significance level. Ultimately, Step 4 will need to be implemented.

Step 4 (P1.2)

The second treatment involves the removal of ties. In Fig. 2(b), strong grouping (tied Blom scores) can be observed even when a proportion of the strictly tardy points and the on-time points have

been removed. These ties are due to the rounding that is caused by the coarse time scale in which activity times are reported. A possible remedy for this is the calculation of the average Blom score for all tied points in a group. Fig. 2(d) shows the resulting plots of the empirical data and the averaged Blom scores with the corresponding regression model for the two samples with 11 tardy points removed and the ties resolved. A better fit for the regression models emerges, resulting in higher R values and consequently, higher p -values (0.959 and 0.986 for the first and second trimmed sample respectively). These p -values mean that the new null hypothesis

$H_{0,P1.2}$ can no longer be rejected at the 5% significance level. Consequently, the variable d_{ij}/\hat{d}_{ij} can now be assumed to be following a Parkinson distribution with a lognormal core, according to the theory of Trietsch et al. (2012)

$H_{0,P1.2}$: $\ln(d_{ij}/\hat{d}_{ij})$ follows a normal distribution with a Parkinson effect after the removal of on-time points, after removing p_p strictly tardy points and resolving ties

This section gave an illustrative overview of the mechanism to test the Parkinson distribution with a lognormal core on real activity durations using the four-step procedure of Trietsch et al. (2012). For more detailed information on the mechanism and justification of the methodology, the reader is referred to the original paper. In the next section, the procedure of Trietsch et al. (2012) is used on the project database of Batselier and Vanhoucke (2015).

Methodology

This section discusses the analysis of the online database presented by Batselier and Vanhoucke (2015). First, the issue of which records from this database are suited for this study will be discussed. Second, the Parkinson distribution with a lognormal core procedure is implemented on these data. Moreover, the p -values and the estimated means and standard deviations for the natural logarithm of the relative empirical distributions that result from this procedure are presented.

Discussion of the Empirical Records in the Database of Batselier and Vanhoucke (2015)

The empirical records discussed in this paper are selected from the online database of Batselier and Vanhoucke (2015), which currently consists of 51 real-life projects with actual progress data. Out of the total 51 projects, 39 are construction projects. This database is one of the largest and most diverse yet with respect to planned duration and budget at completion, in the project management literature and is updated and expanded continuously. Furthermore, the quality and authenticity of the data are guaranteed through the application of a database construction tool called project cards (Batselier and Vanhoucke 2015).

From this empirical database, those projects that contain complete project progress information at the activity level were selected, such that for all projects j in this set \mathcal{P} the relative empirical distribution $d_{ij}/\hat{d}_{ij}, \forall i \in 1, \dots, N_j$ could be calculated. Thus, from the total of 51 projects, 24 projects were retained for further study.

Before proceeding with the application of the four-step procedure of Trietsch et al. (2012), some more data cleaning had to be performed. First, the progress data at the work package level of the work breakdown structure (WBS) of the projects had to be removed. The work package level often provides adequate detail to monitor the progress of a project during its execution, but is of no use here given this study's interest merely in the progress data at the activity level. Second, some points from the calculated values in d_{ij}/\hat{d}_{ij} had to be manually removed, because they had no real connection to activity progress. Some very large and some very small values for d_{ij}/\hat{d}_{ij} occurred because of errors that were made in the scheduling phase prior to the project executions. If an activity in the project was planned but not executed, the ratio between the actual activity duration and its baseline estimate was very small

and the data point needed to be removed. Alternatively, if during the project execution, an additional activity was defined that was not previously incorporated in the baseline schedule, the ratio for that activity could be very large. An example of such a situation is when additional work is created, when the scope of the project is redefined during the execution, or when cleaning-up activities are performed in the end of the project that were not accounted for in the scheduling phase. If the project manager did not bother to accurately adapt the baseline schedule to these situations, erroneous points in d_{ij}/\hat{d}_{ij} occur, which consequently need to be removed from the analysis. Altogether, the manual removal of data points was quite futile. Only 7 out of a total of 1,881 activities for the 24 projects had to be removed.

Application of the Parkinson Distribution with a Lognormal Core

The calibration procedure proposed by Trietsch et al. (2012) was applied to the 24 projects of the online dataset. For all 24 projects, it was tested whether the relative empirical distribution d_{ij}/\hat{d}_{ij} could be shown to be Parkinson distributed with a lognormal core. To this end, the four-step procedure (P0, P1, P1.1, and P1.2) was applied to the empirical records.

Table 1 presents the results for the hypothesis tests that provided confirmation of the Parkinson distribution with a lognormal core for the projects in the empirical database. In between brackets, the p -values are shown. In addition, the estimated mean (\hat{m}) and standard deviation (\hat{s}) of the natural logarithm of the relative empirical distribution d_{ij}/\hat{d}_{ij} are presented for each project. Table 1 only shows 13 of the 24 projects. The remaining 11 projects yielded p -values lower than $\alpha = 5\%$ for all hypothesis tests ($H_{0,P0}$, $H_{0,P1}$, $H_{0,P1.1}$ and $H_{0,P1.2}$). For these projects, each of the null hypotheses were rejected at the 5% significance level and the assumption that the data are distributed according to the Parkinson distribution with a lognormal core could no be withheld. For the 13 projects that are shown in Table 1, only the result of the hypothesis test is shown that confirmed the Parkinson distribution with a lognormal core for that project (p -value $> \alpha$). Unlike in the previous section, only a single p value is shown for tests $H_{0,P1.1}$ and $H_{0,P1.2}$ in Table 1, although still the example of Trietsch et al. (2012) was followed and two trimmed samples were used. However, if both p -values were larger than 5%, only the results for the sample with the highest p -value are shown. If one of the two p -values was smaller than 5%, the corresponding hypothesis was rejected. From the 13 projects that were retained for further analysis, only two were not performed in the construction industry. The "Patient Transport System" and "Organizational Development" projects were respectively performed in the IT and Education sectors.

The first observation that can be made from Table 1 is that only 1 of the 24 projects in the online dataset is directly lognormally distributed. The construction project described as "Young cattle barn" is the only one that has a p -value larger than $\alpha = 5\%$ for P0. For the remaining 12 projects, presented in Table 1, the Parkinson distribution with a lognormal core is only confirmed for two projects by P1.2. For the other 10 projects, the distributional model was confirmed using P1. No occurrence was found of an empirical record for which P1.1 succeeded in confirming the Parkinson distribution with a lognormal core. As it appears from the online dataset, the first treatment (removing tardy points) is only effective to confirm the distributional model when it is combined with the second treatment (resolving ties).

The next section discusses the \hat{m} and \hat{s} values of Table 1 and presents a classification of the data for further research.

Results

This section discusses the estimated values for the standard deviation \hat{s} and the mean \hat{m} of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} , as shown in Table 1. First, the standard deviation \hat{s} will be discussed. Second, the results for the mean \hat{m} will be examined more closely. These standard deviations and means will be used in an empirical classification for project execution that will be presented at the end of this section.

Standard Deviation of the Natural Logarithm of the Relative Empirical Distributions

The standard deviations \hat{s} of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} are discussed in this section. A visual inspection will be conducted first, followed by a more thorough statistical interpretation.

Visual Inspection of s

Fig. 3(c) shows the estimated standard deviations for the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} that were previously presented in Table 1. In order to incorporate the uncertainty that is associated with estimating the standard deviation, 95% confidence intervals were added to the plot. The confidence interval around \hat{s} is calculated using the assumption that $(n-1)\hat{s}^2/s^2$ follows a χ^2 distribution with $n-1$ degrees of freedom, if the original population of data is normally distributed. s denotes the population standard deviation of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} , whereas \hat{s} was the estimate produced from a sample, and n denotes this sample size. The confidence limits around s can then be calculated from the inequality $\chi_{1-\beta/2}^2 \leq (n-1)\hat{s}^2/s^2 \leq \chi_{\beta/2}^2$, with a confidence level $1-\beta$. The confidence interval is denoted in grey, while the \hat{s} value is plotted as a white circle. Fig. 3(c) shows that the large estimated standard deviation of the natural logarithm of the relative empirical distribution of the “Organizational Development” project (C2014-03). This large standard deviation clearly is unique among the projects examined from the online dataset. For the other projects, a clear classification is not immediately apparent from a visual inspection.

Statistical Interpretation of s

A more robust, statistical interpretation of the values for \hat{s} shown in Fig. 3(c) is now presented. Therefore, Levene’s test for the homogeneity of variances was implemented (Levene 1960). Levene’s test

can be used to verify the assumption that the variances (or standard deviations) are equal across different samples or groups. To that end, a statistic W is calculated that under the null hypothesis (\hat{s} is equal for all projects in a group) is distributed according to an F distribution. For details on the calculation of W , the reader is referred to the original work.

The statistic W is used to test the assumption of homogeneity of variances for different groups of projects from Table 1 and Fig. 3(c). Clearly, the “Organizational Development” project (C2014-03) is unique as no other project can be found in the dataset for which the assumption of equal variance is not rejected by Levene’s test at a 5% significance level. A group (High \hat{s}) was thus created for this project alone. The question then remains whether the other projects can be assumed to have equal variances. Levene’s test confirms this assumption, based on a found p -value of 0.39, at a 5% significance level [$W = 1.06$, $P[W > 1.06] = 0.39$ with ($d.f. = 11; 223$)].

The 13 projects were therefore classified into two groups: a group of projects with a Low \hat{s} and a group of one project with a High \hat{s} . The Low \hat{s} group includes projects C2011-07, C2012-13, C2013-01, C2013-03, C2013-07, C2013-08, C2013-09, C2013-11, C2013-12, C2013-13, C2013-15, and C2013-17, while the High \hat{s} group includes only C2014-03.

Fig. 3(d) reproduces Fig. 3(c) with a color coding for the two groups. The High \hat{s} group is colored in black and the Low \hat{s} group is depicted in a light shade of grey. In addition, the pooled standard deviation, \hat{s}_{pooled} , is added for each group to represent the real standard deviation of that group. This pooled standard deviation is calculated for each group using

$$\hat{s}_{\text{pooled}} = \sqrt{\frac{\sum_{j=1}^k (n_j - 1) \hat{s}_j^2}{\sum_{j=1}^k (n_j - 1) \hat{s}_j}}$$

where \hat{s}_j = estimated standard deviation of the natural logarithm of the relative empirical distribution for a project j in a group; n_j = sample size for that project; and k = number of projects in that group.

Mean of the Natural Logarithm of the Relative Empirical Distributions

The mean \hat{m} of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} is closely examined in this section. Again,

Table 1. Discussion of the Activity Times Data Obtained from Batselier and Vanhoucke (2015) Using the Parkinson Distribution with a Lognormal Core [As Described in Trietsch et al. (2012)]

Project number	Project name	Parkinson distribution with lognormal core \hat{m} \hat{s} (p -value)				
		P0	P1	P1.1	P1.2	
C2011-07	Patient transport system	—	0.47 (0.73)	0.26	—	—
C2012-13	Pumping station	—	0.24 (0.09)	0.62	—	—
C2013-01	Wiedauwkaai fenders	—	0.45 (0.33)	0.70	—	—
C2013-03	Brussels finance tower	—	−0.09 (0.45)	0.62	—	—
C2013-07	Family residence	—	−0.03 (0.59)	0.51	—	—
C2013-08	Timber house	—	0.27 (0.15)	0.64	—	—
C2013-09	Urban development project	—	—	—	0.05 (0.59)	0.73
C2013-11	Recreation complex	—	0.20 (0.05)	0.72	—	—
C2013-12	Young cattle barn	0.01 (0.15)	0.57 (0.15)	—	—	—
C2013-13	Office finishing works (1)	—	−0.89 (0.59)	0.54	—	—
C2013-15	Office finishing works (2)	—	−0.31 (0.57)	0.99	—	—
C2013-17	Office finishing works (3)	—	−0.68 (0.42)	1.03	—	—
C2014-03	Organizational development	—	—	—	1.05 (0.46)	3.06

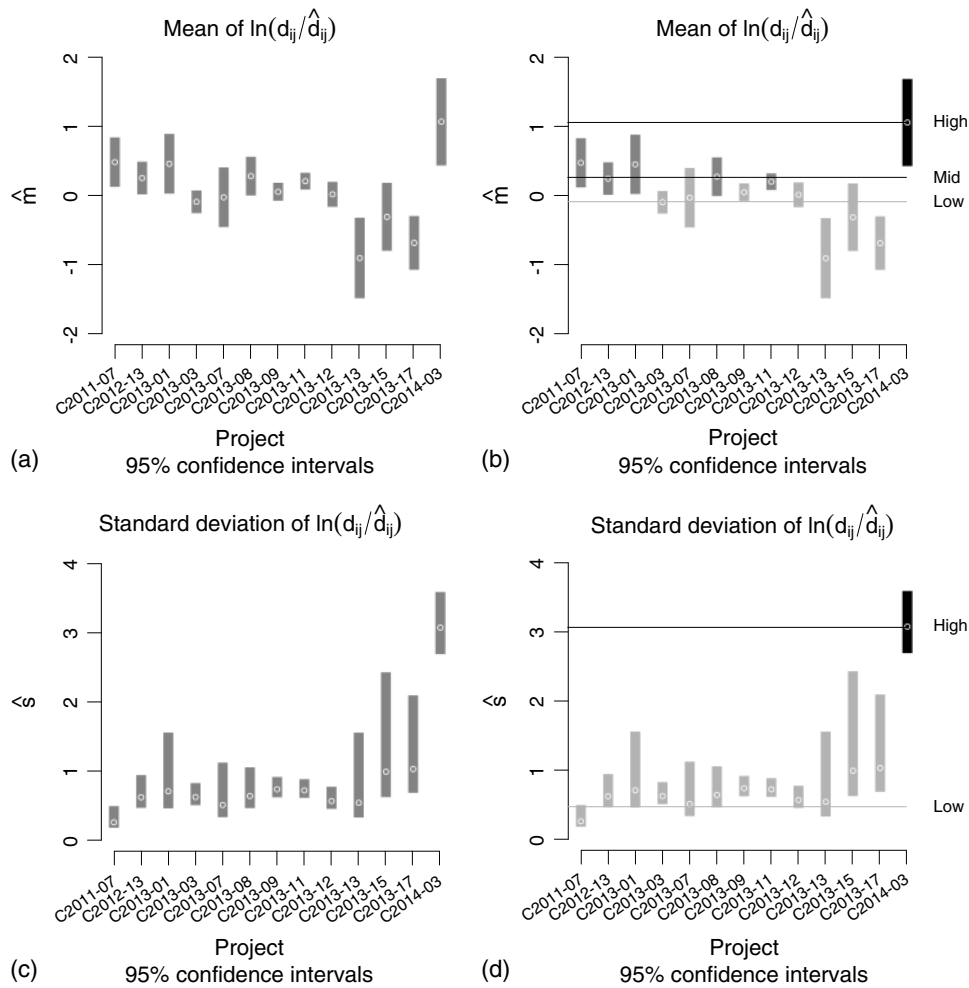


Fig. 3. Discussion of the results obtained from the application of the four-step calibration procedure on the dataset of empirical records of project executions: (a) \hat{m} for the empirical project records according to the Parkinson distribution with a lognormal core; (b) \hat{m} for the empirical project records according to the Parkinson distribution with a lognormal core, with the grouping illustrated in color; (c) \hat{s} for the empirical project records according to the Parkinson distribution with a lognormal core; (d) \hat{s} for the empirical project records according to the Parkinson distribution with a lognormal core, with the grouping illustrated in color

a visual inspection is first undertaken and then a statistical analysis is added.

Visual Inspection of m

Fig. 3(a) presents the estimated means \hat{m} of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} that were previously presented in Table 1. The data points are accompanied by 95% confidence limits that were calculated from the inequality $t_{\beta/2, n-1} \leq (\hat{m} - m)s \leq t_{1-\beta/2, n-1}$. Here, m denotes the population mean of the natural logarithm of the relative empirical distributions d_{ij}/\hat{d}_{ij} , whereas \hat{m} was the estimate produced from a sample, and n denotes this sample's size. s is given by the \hat{s}_{pooled} value of the group to which the project belongs, $t_{p, n-1}$ denotes the p th quantile of the t distribution with $n - 1$ degrees of freedom and $1 - \beta$ was chosen as 95%. From a visual inspection of Fig. 3(a), the large mean for the “Organizational Development” project is apparent. This project was already identified as unique in terms of its standard deviation, and its mean does also reflect the unique character of this project. Moreover, a large group of projects can be observed that have a logarithmic mean of around 0. Three distinct projects (C2013-13, C2013-15, and C2013-17) appear to have a somewhat smaller observed logarithmic mean. This observation will now be verified using a statistical analysis.

Statistical Interpretation of m

Welsh's test (Welch 1947) is well-known for its ability to test for a difference in the means across groups or samples, when the variance is not necessarily assumed to be equal. Under the assumption of equal mean (which is the standard null hypothesis for this test) Welsh's F statistic is distributed according to an F distribution. The conclusions for the analysis of the presented \hat{m} values can be summarized along the following lines. First, Welsh's test confirms that not all means for all 13 projects of Table 1 can be assumed to be equal as the null hypothesis is easily rejected at the 5% significance level [$F = 4.16$, $P[F > 4.16] = 1.9 \times 10^{-4}$ with ($d.f. = 12$; 47.14)]. Moreover, the observations from the prior visual inspection could not be confirmed using Welsh's test. It is unlikely that the grouping mentioned earlier for \hat{m} can be withheld. Although the three projects with the smaller \hat{m} can be assumed to have an equal mean [$F = 0.74$, $P[F > 0.74] = 0.49$ with ($d.f. = 2$; 10.30)], the hypothesis of equal means for the remaining projects is rejected [$F = 2.72$, $P[F > 2.72] = 0.015$ with ($d.f. = 8$; 44.62)].

We therefore continue the discussion of \hat{m} in a more systematic manner. The “Organizational Development” project is added to the High \hat{m} group and the remaining 12 projects are organized into a Mid \hat{m} group and Low \hat{m} group using an enumeration algorithm.

Table 2. Summary of Welsch's Tests for \hat{m}

Value	F	p .value ($d.f.1$; $d.f.2$)
Low	2.42	0.06 (6; 21.62)
Mid	1.17	0.34 (4; 24.9)
High	— ^a	(0; -)

^aNo value could be found since the group consists out of a single project.

This algorithm enumerates all possible combinations of the remaining 12 projects and tests the assumption of equal mean for these combinations. Only six possible combinations were found where the assumption of equal mean is confirmed within the groups and rejected between the groups. The combination that had the highest combined p -value for the Welsch's tests for the two groups was selected. The Mid \hat{m} group then consists of projects C2011-07, C2012-13, C2013-01, C2013-08, and C2013-11, while the Low \hat{m} group consists of C2013-03, C2013-07, C2013-09, C2013-12, C2013-13, C2013-15, and C2013-17.

Fig. 3(b) shows the grouping and the corresponding pooled mean estimates \hat{m}_{pooled} . These were calculated for each group as

$$\hat{m}_{\text{pooled}} = \frac{\sum_{j=1}^k n_j \hat{m}_j}{\sum_{j=1}^k n_j}$$

where \hat{m}_j = estimated mean of the natural logarithm of the relative empirical distribution for a project j in a group; n_j = sample size for that project; and k = number of projects in that group. These \hat{m}_{pooled} estimate the real logarithmic means of the groups of project executions, even though this mean is not observed directly from the data. In Fig. 3(b), the High \hat{m} group is colored black, the Mid \hat{m} group has a grey color, and the Low \hat{m} group was given a lighter shade of grey. Table 2 shows the results for the Welsch's test for the constructed groups.

Empirical Classification for Project Executions

It has been established (Williams 1999) that project-management simulation studies need to incorporate empirical data in order to produce practical results. However, it is not a straightforward task to incorporate empirical data into simulation models of project executions. In order to facilitate this task, a classification for project executions is proposed that results from the grouping of the \hat{m} and \hat{s} values in the previous section. Next, two ways to classify new (real or simulated) project executions are proposed. Finally, the proposed classification is discussed from a construction-industry point of view.

Given the grouping of the empirical records of Batselier and Vanhoucke (2015) for both the \hat{m} and \hat{s} values, the classifications given in Table 3 are proposed. The six classes of project executions in Table 3 all combine a value from the \hat{m}_{pooled} estimates with a value from the \hat{s}_{pooled} estimates. A project execution, whether it is real or simulated, can then be referenced against these six classes. This can be done in two ways.

Table 3. Empirical Classification of Project Executions

\hat{m}_{pooled}	\hat{s}_{pooled}	
	0.47	3.06
-0.09	Low \hat{m} /Low \hat{s}	Low \hat{m} /High \hat{s}
0.26	Mid \hat{m} /Low \hat{s}	Mid \hat{m} /High \hat{s}
1.05	High \hat{m} /Low \hat{s}	High \hat{m} /High \hat{s}

The first way to classify a new project execution would be to apply the procedure of Trietsch et al. (2012). If the Parkinson distribution with a lognormal core is confirmed for that project execution, the estimates \hat{m} and \hat{s} can be used. Levene's and Welsch's tests can then be deployed to check whether the natural logarithm of the relative distribution of the new project execution has respectively an equal standard deviation or an equal mean with any of the groups from the analysis in this paper. This would identify the project execution as belonging to a single class of Table 3. However, this would mean that the empirical records, which were needed to obtain the classification in Table 3, need to be consulted for each new project execution.

Therefore, an alternative procedure is proposed that relies only on the estimates for \hat{s}_{pooled} and \hat{m}_{pooled} of Table 3. Again, for each new project execution, the Parkinson distribution with a lognormal core would need to be confirmed, and the \hat{m} and \hat{s} estimates calculated. Calculation of a confidence interval (e.g., the 95% confidence interval) is suggested for both the \hat{s} and \hat{m} value. Then, for each cell in Table 3, one should compare the \hat{m}_{pooled} and \hat{s}_{pooled} values for the corresponding groups to the appropriate confidence interval. If the \hat{m}_{pooled} for a certain group (Low \hat{m} , Mid \hat{m} , or High \hat{m}) is contained within the confidence interval around \hat{m} and if the \hat{s}_{pooled} value of a certain group (Low \hat{s} or High \hat{s}) is contained within the confidence interval around \hat{s} , then it can be assumed that the project execution belongs to the class of Table 3 that is characterized by these \hat{m}_{pooled} and \hat{s}_{pooled} values. This will be illustrated with examples in the "Discussion and Conclusions" section.

Now the use of the classification of Table 3 for construction projects will be discussed. As noted earlier, the database of Batselier and Vanhoucke (2015) predominantly contains records of construction projects. However, some other sectors (IT and education) were included in this study to produce the classification in this section. Most notably, the "Organizational Development" project (education) has influenced the proposed classification greatly. The empirical evidence shows that, for construction projects, only the two classes in the top left corner of Table 3 should be considered. Since, in the dataset of Batselier and Vanhoucke (2015), the construction projects were found to belong to the Low \hat{m} and Mid \hat{m} groups with respect to the mean of the natural logarithm of the relative empirical distributions, and to the Low \hat{s} group with respect to the standard deviation of the natural logarithm of the relative empirical distributions. For these two classes, the coefficient of variation (c.v.) was calculated with $c.v. = \sqrt{e^{\hat{s}_{\text{pooled}}^2} - 1}$. The formula for the c.v. of a lognormally distributed variable X follows from: $E[X] = e^{m+s^2/2}$ and $\text{sd}[X] = e^{m+s^2/2} \sqrt{e^{s^2} - 1} = E[X] \sqrt{e^{s^2} - 1} \Rightarrow \sqrt{e^{s^2} - 1} = \text{sd}[X]/E[X] = c.v.$ The c.v. is independent of the expected value of a lognormally distributed variable and therefore, \hat{m}_{pooled} is not incorporated in the formula. The value of 0.5 for the Low \hat{s} group confirms the findings of Trietsch et al. (2012) for the construction industry, as it is between their values of 0.3 and 0.8, found for respectively low and mid variance construction projects.

The issue of how the empirical classification of Table 3 can be put to work in future simulations studies for project management will be discussed in the next section.

Discussion and Conclusions

This section first provides a discussion of the empirical classification that is proposed in this paper. The discussion offers three possible uses for the empirical classification in the project management practice. These three possible uses are validated through a case

study performed in a construction company. Second, overall conclusions for the study are presented.

Discussion

The use of the empirical classification presented in Table 3 is validated using a case study on the project management practice in a construction company. First, three projects that were performed in a construction company are discussed, then three possible uses for the classification in Table 3 are shown. In combination with the empirical data from the construction company, it is shown how Table 3 can lead to the implementation of a SPC project control tool, the planning of management reserve and execution of a schedule risk analysis.

Analysis of Construction Projects

Data were obtained from three residential building projects performed by a Belgian construction company in the beginning of 2014. An associated researcher worked as a project management consultant for this construction company in the first six months of 2014, providing both planning and control services. Due to confidentiality, only the necessary details were abstracted from these projects and made available online in R data format (“Validation Dataset” at the statistical project control page on <http://www.projectmanagement.ugent.be>). These new projects, performed by a single construction company, are currently being processed to be included in the database of Batselier and Vanhoucke (2015), and can be retrieved using the names C2014-05, C2014-06, and C2014-07. These projects constitute a homogeneous set in terms of their scope (structural construction work), duration (30–70 days) and budget (140,000–210,000 Euro). A qualitative analysis indicated that from the three projects, only C2014-06 was perceived as successful by the project management team that was responsible for the execution of the projects. A quantitative analysis supported this finding. For project C2014-06, a budget overrun and schedule delay of around 10% were observed, while the two other projects experienced a budget overrun of around 17–19% and were finished with a 24% schedule delay.

Table 4 shows the result of the four-step procedure of Trietsch et al. (2012) applied to projects C2014-05, C2014-06, and C2014-07. The estimates \hat{m} and \hat{s} for these projects are also depicted in Fig. 4(a), along with their 95% confidence limits. Fig. 4(a) allows the projects to be identified as belonging to either of the empirical classes presented in this study. From the \hat{m} and \hat{s} grid in Fig. 4(a), projects C2014-05 and C2014-07 were classified as Mid \hat{m} /Low \hat{s} and project C2014-06 was classified as Low \hat{m} /Low \hat{s} . This detailed inspection of the activity durations for these projects explains the budget overruns and schedule delays that were observed by the project management team. Moreover, the empirical classification effectively allows the quantification of the subjectively perceived success of project C2014-06, by the project management team. In the construction company that is the subject of this case study, a project belonging to the Low \hat{m} /Low \hat{s} class is perceived as successful. Consequently, this class can be used as a reference for future project executions. It will now be demonstrated

how this class can be used in three project simulation applications that incorporate the projects of the validation set. In addition, the use of the two top left classes of Table 3 for construction projects is validated by these independent data.

Statistical Process Control (SPC) Application for Project Management

The implementation of a SPC procedure for project control was proposed for the construction company, based on the preceding classification of the projects in the case study. For such a SPC procedure for project control, a major prerequisite is the ability of the process (in this case the projects that are executed) to “speak for itself” (ReVelle 2004) and the control approach should be able to accurately reflect the “voice of the process” (Wheeler 1995). Effectively, this means that empirical data from the construction company needed to be incorporated in the SPC procedure. Fig. 4(b) shows the SPC procedure proposed by Leu and Lin (2008). A recent study (Colin and Vanhoucke 2015) shows that this approach outperforms the other SPC procedures that were proposed in the literature for project control. For details on the calculation of the control limits (and consequently, the zones A, B, and C) and the simulation from which these can be calculated, the reader is referred to the original article and the comparative study. Here, it suffices that the simulations incorporate the Low \hat{m} /Low \hat{s} class into the definition of the “in control” reference. Moreover, the observations for project C2014-05 were added to the SPC chart (diamonds). If this had been done dynamically during the execution of the project, the SPC chart would have signalled an “out of control” situation when the project was around 50% completed, on account of more than four consecutive observations falling within the C zone (Leu and Lin 2008; Small 1958). Project C2014-05 would have been identified as unlikely to be belonging to the Low \hat{m} /Low \hat{s} , and the project management team would have been given an early warning signal on which they could have made a decision to steer the project back towards this class. In conclusion, Fig. 4(b) shows how the empirical classification of Table 3 allows the implementation of SPC charts for project control.

Management Reserve Planning Application for Project Management

In the construction company that is the subject of our case study, the practice of running MCS in order to estimate possible budget overruns and schedule delays was not previously adopted. The planning of a management reserve was therefore done consistently based on personal experience and rules-of-thumb. The main reason provided for this was that it requires a detailed input analysis, for which the know-how was not present within the construction company. After the identification of the successful historical project (C2014-06) as belonging to the Low \hat{m} /Low class, they were shown how this class could serve as an input to future simulation studies. Fig. 4(c) shows a grey area of actual cost curves for fictitious executions of project C2014-06 that would belong to the Low \hat{m} /Low class. The simulation that was used to produce these fictitious project executions was described in Colin and Vanhoucke (2015). The probability of these actual cost curves is indicated by

Table 4. Discussion of the Activity Times Data of the Validation Dataset from the Case Study in a Belgian Construction Company

Project number	Project name	Parkinson distribution with lognormal core \hat{m} \hat{s} (p -value)			
		P0	P1	P1.1	P1.2
C2014-05	Structural construction works (1)	—	0.37 (0.09)	0.30	—
C2014-06	Structural construction works (2)	—	0.05 (0.61)	0.33	—
C2014-07	Structural construction works (3)	—	0.34 (0.21)	0.31	—

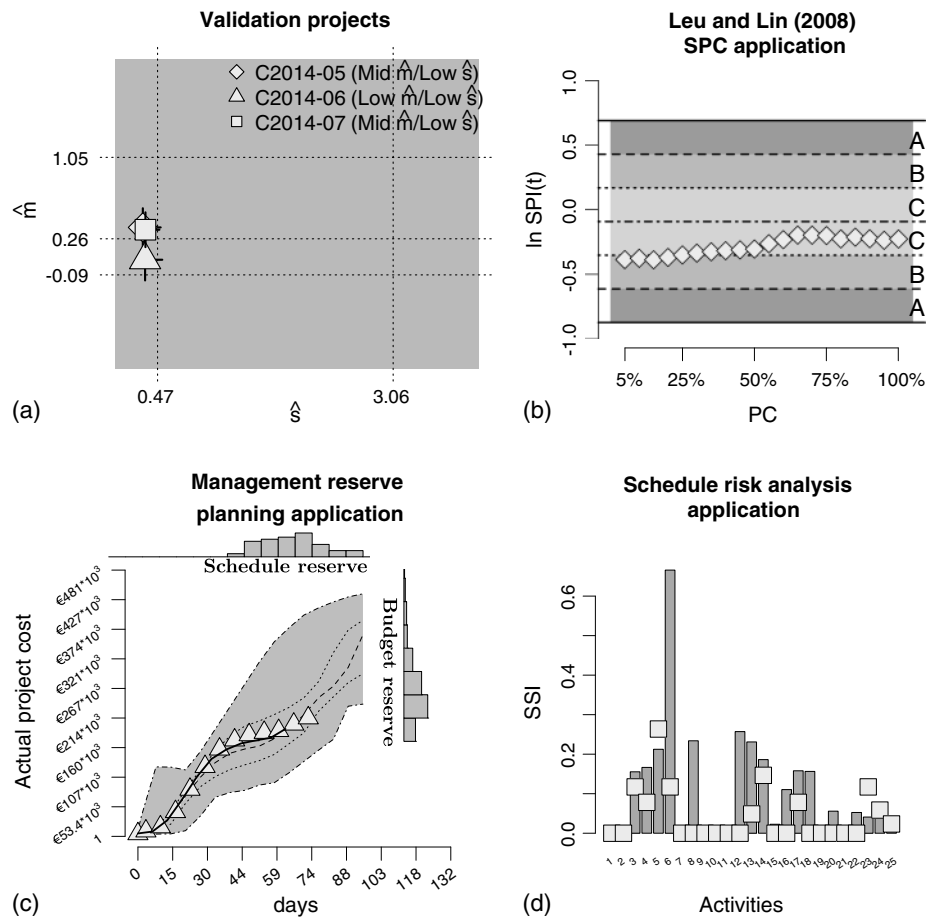


Fig. 4. Validation of the empirical classification for project executions on a case study in a Belgian construction company: (a) \hat{m} and \hat{s} for the validation dataset (C2014-05, C2014-06, C2014-07); (b) a SPC application for project control, depicting project C2014-05; (c) a management reserve planning application for project C2014-06; (d) a schedule risk-analysis application for project C2014-07

the dotted lines that are superimposed onto the grey area. From top to bottom, these respectively represent the 100th, the 75th, the 50th, the 25th, and the 0th quantile. The full black line represents the baseline schedule for project C2014-06, and the real execution is depicted using white triangles. Even though the simulation was run after the completion of the project, it could have been done as easily during its planning and negotiation phase. This would have given the construction company insights into the risks that they are exposed to when performing the project. A schedule reserve could be taken into consideration when engagements are made with respect to the actual duration of the project with the project owner or the other possible stakeholders. In addition, a budget reserve could be taken into account when a price is set for the execution of the project in a contractual agreement. The probabilistic outputs of the simulation for the actual cost and the total project duration were added to Fig. 4(c) as histograms opposing respectively the y-axis and the x-axis. These histograms, and the information that can be retrieved from them, show the value of the presented empirical classification for Monte Carlo simulations in the project planning and negotiation phase.

Schedule Risk Analysis Application

A third possible use for the proposed empirical classification is in the application of a schedule risk analysis (SRA) for project management. SRA are not conducted regularly for projects in the construction company, due to the difficulties that are associated with deriving a probabilistic input for the MCS. However, given the

identification of the successful historical project as belonging to the Low \hat{m} /Low class, the required input analysis is drastically reduced. Fig. 4(d) shows the SRA that was conducted post-hoc for project C2014-07. Here, the schedule sensitivity index (SSI) is shown as grey bars for all activities. Although the proposed approach is not restricted to the SSI, the illustration is limited to this single indicator, since it was found to deliver the best possible results in a project control application (Vanhoucke 2010b). For more details regarding the characteristics and the calculations required in a SRA, refer to more extensive studies on the subject (Vanhoucke 2010a; Elmaghraby 2000). Moreover, the observed SSI for the real execution of project C2014-07 is shown using white squares in Fig. 4(d). Since the real execution of C2014-07 was found to belong to the Mid \hat{m} /Low class [Fig. 4(a)], substantial differences in the SSI for certain activities can be observed. This indicates a possibly larger variance for activity durations with respect to the Low \hat{m} /Low class, which confirms the findings of the classification in Fig. 4(d). Moreover, Fig. 4(d) illustrates that the proposed classification allows to test the robustness of a SRA, which shows its value for the development and testing of project control approaches which are based on SRA.

Conclusion

This paper analysed the empirical data for activity distributions of an existing dataset using the distributional model of

Trietsch et al. (2012). The purpose of this application was to show the relevance of such an analysis in future project management studies in which data of new projects will be used. This new data, coming from historical observations or simulation studies, can then be compared and validated against the classification, shown in this paper. The potential value of the proposed classification has been demonstrated on a case study in a construction company.

The analysis that was conducted in this paper was made possible by the extensive dataset of project executions provided by Batselier and Vanhoucke (2015). To this dataset, the distributional model and calibration procedure proposed by Trietsch et al. (2012) were applied. This model assumes the relative empirical distribution d_{ij}/\hat{d}_{ij} to be lognormally distributed, when the Parkinson effect and the occurrence of rounding errors are correctly accounted for, known as the Parkinson distribution model with a lognormal core. The large majority of the historical records in the database of Batselier and Vanhoucke (2015), that were suited for this analysis, are from the construction industry. For 54% of these records, confirmed the Parkinson distribution model with a lognormal core was confirmed.

During our analysis, estimates were obtained for the standard deviation \hat{s} and the mean \hat{m} of the natural logarithm of the relative empirical distributions (d_{ij}/\hat{d}_{ij}). These estimates were used to find, respectively, two and three groups of project executions for \hat{s} and \hat{m} . The validity of these groups was confirmed using a statistical analysis with Levene's test for the homogeneity of variances and Welsch's test for equal means. The two groups that were found for \hat{s} have values 0.47 and 3.06 for the pooled standard deviation of the natural logarithm of the relative empirical distributions. The three groups for \hat{m} have values -0.09 , 0.26 , and 1.05 for the pooled mean of the natural logarithm of the relative empirical distributions. For construction projects, the findings of Trietsch et al. (2012) were confirmed with respect to the observed coefficients of variation of 0.5 for the Low \hat{s} group of project executions, and validated the use of the Low \hat{m} /Low and Mid \hat{m} /Low on independent data.

The grouping of the historical records from the dataset of Batselier and Vanhoucke (2015) into two and three groups, for respectively \hat{s} and \hat{m} , supported the proposition of an empirical classification for project executions, using six classes. Subsequently, the ways in which this classification can be useful in the project management practice were demonstrated using a case study in a construction company. The empirical classification can be deployed to develop SPC charts for project control, for planning applications of the management reserve, and for SRA studies. It is also believed that the proposed classification method and the suggested applications can lead to more generalizable results in future research in the project-management domain.

Appendix. R Template for the Proposed Classification of the Project Data

This appendix presents a working R template to reproduce the classification of the empirical records of Batselier and Vanhoucke (2015). The results that are presented in this paper are subject to some numerical instability and therefore, small differences can be expected when the R template is run. In particular, the uncertainty associated with the estimation of the coefficients for the linear regression models might cause slightly different coefficients for each execution of the R template. In this appendix, the files that should be present on the workstation where this R template is executed are listed. Second, the outline of the R template is presented, and finally, the actual R code is shown.

Required Inputs

From the statistical project control research page available at the OR&S website (OR&S 2014) the following files should be downloaded to the working directory of the R process:

- Colin_and_Vanhoucke_2015_template: The R template to run the method proposed in the paper;
- Batselier_Vanhoucke_2015.xlsx: The data for the 24 projects, including baseline and real durations;
- Trietsch_etal_2012.R: The R syntax for the four-step procedure of Trietsch et al. (2012);
- Grouping_tests.R: The R syntax for the classification as presented in the paper; and
- Looney_and_Gulledge_1985.txt: The table that is input for the R files.

Outline

Preamble

The preamble of the R template includes statements to clear the workspace, to load a package to read Excel files, and to define some subsidiary functions.

```

1 #-----
2 # R worksheet to reproduce the results for:
3 #
4 # Colin and Vanhoucke (2015): "Empirical Perspective on Activity Durations
5 # for Project-Management Simulation Studies"
6 # To Appear in "Journal of Construction Engineering and Management"
7 #
8 #-----
9
10
11 #----- PREAMBLE -----
12 rm(list=ls(all=TRUE)) # clear workspace
13 library("xlsx") # to read excel files
14 source("Trietsch_etal_2012.R")
15 source("Grouping_tests.R")
16
17
18 sublist <- function(l,indices){
19   out <- apply(indices,function(i){
20     l[i]}
21   ),simplify=FALSE)
22   names(out) <- names(l)[indices]
23   return(out)
24 }
25
26 chunk <- function(x,n){
27   split(x, factor(seq(1:n))),
28 }
29
30 #----- DATA -----
31
32 # Read the (cleaned) data of
33 # Batselier and Vanhoucke (2015)
34
35 d_over_d_hat <- apply(c("C2011-07","C2011-12","C2011-13","C2012-18","C2013-01","C2013-02",
36   "C2013-03","C2013-04","C2013-05","C2013-06","C2013-07","C2013-08","C2013-09","C2013-
37   -10","C2013-11","C2013-12","C2013-13","C2013-14","C2013-15","C2013-16","C2013-17","
38   C2014-01","C2014-02","C2014-03"),function(project.name){
39
40   print(project.name)
41   d_and_d_hat <- read.xlsx("Batselier_Vanhoucke_2015.xlsx",sheetName=project.name)
42   return(apply(d_and_d_hat[,1],function(x) x[1]/x[2]))
43 })
44
45 #----- FOUR-STEP PROCEDURE -----
46
47 # Apply the four-step procedure of
48 # Trietsch et al. 2012
49
50 d_over_d_hat_model <- apply(1:length(d_over_d_hat),function(i){
51   print(names(d_over_d_hat)[i])
52   #-- Apply "Trietsch et al. 2012"
53   T <- Trietsch_etal_2012(d_over_d_hat[[i]])
54   #-- Add confidence intervals
55   if(!is.logical(T)){
56     beta <- 0.05
57     n <- length(T$a)
58     T[["a_CI"]] <- c(sqrt((n-1)*T$a^2/qchisq(1-beta/2,n-1)),sqrt((n-1)*T$a^2/qchisq((beta/
59     2),n-1)))
60     T[["n_CI"]] <- c(T$a*qt(1-beta/2,n-1)*T$a/sqrt(n),T$a*qt(1-beta/2,n-1)*T$a/sqrt(n))
61   }
62 }

```

Fig. 5. Colin and Vanhoucke (2015) template


```

56 }
57 return(T)
58 })
59 names(d_over_d_hat.model) <- names(d_over_d_hat)
60
61 ## Print the output
62
63 d_over_d_hat.confirmed <- sublist(d_over_d_hat.model,(1:24){apply(d_over_d_hat.model,
64   function(x){is.logical(x)}})
65 d_over_d_hat.confirmed
66
67 ##----- CLASSIFICATION -----
68 ## Statistical interpretation of hat(s)
69 ## Levene's test
70
71 grouping_test(sublist(d_over_d_hat.confirmed,1:12),"sd")
72
73 ## Pooled hat(s)
74
75 pooled_m <- c(sum(apply(sublist(d_over_d_hat.confirmed,1:12),function(T) length(T$z))-1
76   *apply(sublist(d_over_d_hat.confirmed,1:12),function(T) T$m)^2)/sum(apply(sublist(d
77   _over_d_hat.confirmed,1:12),function(T) length(T$z))-1), d_over_d_hat.confirmed[[13]]
78   $s)
79
80 ## Statistical interpretation of hat(m)
81
82 d_over_d_hat.confirmed <- lapply(1:length(d_over_d_hat.confirmed),function(i){ #--
83   recalculate the CI's around hat(m) using the pooled_m
84   T <- d_over_d_hat.confirmed[[i]]
85   beta <- 0.05
86   a <- length(T$z)
87   T[["m_CI"]] <- c((T$m-qt(1-beta/2,a-1)*c(rep(pooled_s[1],12), pooled_s[2])[1]/sqrt(a),T$m
88     +qt(1-beta/2,a-1)*c(rep(pooled_s[1],12), pooled_s[2])[1]/sqrt(a))
89   return(T)
90 })
91 names(d_over_d_hat.confirmed) <- names(sublist(d_over_d_hat.model,(1:24){apply(d_over_d
92   hat.model,function(x){is.logical(x)}})
93
94 ## Welch's test
95
96 grouping_test(sublist(d_over_d_hat.confirmed,1:13),"mean")
97 grouping_test(sublist(d_over_d_hat.confirmed,1:12),"mean")
98 grouping_test(sublist(d_over_d_hat.confirmed,10:12),"mean")
99 grouping_test(sublist(d_over_d_hat.confirmed,1:9),"mean")
100
101 ## Enumeration algorithm
102
103 combination.list <- apply(1:13,function(i) combn(1:13,i)) #-- List all
104   combinations
105
106 F.list <- lapply(combination.list,function(i) apply(1,2,function(x){ #-- Welch's test
107   for all combinations
108   out <- grouping_test(sublist(d_over_d_hat.confirmed,x),"mean",print=FALSE)
109   out["p.value"]
110 })})
111
112 p.list <- apply(1:13,function(i){ #-- Test whether the remaining
113   projects can be organized in another group
114   apply(1:length(F.list[[i]],function(j){
115     if(is.na(F.list[[i]][j])) any(combination.list[[i]][,j]==13)}
116     F.list[[i]][j] <- 0
117   })
118   if(F.list[[i]][j]>0.05){
119     out <- grouping_test(sublist(d_over_d_hat.confirmed,setdiff(1:13,c(combination.list
120       [[i]][,j],13))),"mean",print=FALSE)
121     if(is.na(out["p.value"])){
122       out["p.value"] <- 0
123     }
124     if(out["p.value"]>0.05){
125       out["p.value"] + F.list[[i]][j]
126     }else{
127       return(NA_real_)
128     }
129   }else{
130     return($$real_)
131   }
132 })
133
134 output.list <- apply(1:13){apply(p.list,function(p) max(apply(p,function(p_4) max(p_4,
135   na.rm=TRUE)))>0},function(i){ #-- Print all combinations that confirm in-group
136   equality of mean
137   apply(1:dim(combination.list[[i]][2]){is.na(F.list[[i]][,j])},function(j){
138     comb <- combination.list[[i]][,j]
139     print(paste("Combined p-value: ",p.list[[i]][j],sep=""))
140     print(paste("(",paste(comb," ",collapse=" "),")/(",paste(setdiff(1:13,c(comb,13))," ",
141       collapse=""),")/(",13,""),sep=""))
142   })
143 })
144
145 rbind(grouping_test(sublist(d_over_d_hat.confirmed,c(4,5,7,9:12)),"mean"), grouping_test(
146   sublist(d_over_d_hat.confirmed,c(1,2,3,6,8)),"mean"), grouping_test(sublist(d_over_d
147   hat.confirmed,13),"mean") #-- Print the combination that was chosen (table 2)
148
149 ## Pooled hat(s)
150
151 pooled_m <- c(sum(apply(sublist(d_over_d_hat.confirmed,c(4,5,7,9:12)),function(T) length(T
152   $z))*apply(sublist(d_over_d_hat.confirmed,c(4,5,7,9:12),function(T) T$m))/sum(
153   apply(sublist(d_over_d_hat.confirmed,c(4,5,7,9:12),function(T) length(T$z))),sum(
154   apply(sublist(d_over_d_hat.confirmed,c(1,2,3,6,8)),function(T) length(T$z))*apply(
155   sublist(d_over_d_hat.confirmed,c(1,2,3,6,8)),function(T) T$m))/sum(apply(sublist(d
156   over_d_hat.confirmed,c(1,2,3,6,8),function(T) length(T$z))),sum(apply(sublist(d_o
157   ver_d_hat.confirmed,c(13)),function(T) T$m))/sum(apply(sublist(d_over_d_hat.confirmed,c(13)),function(T)
158   length(T$z))))

```

Fig. 5. (Continued.)

Data

The activity-level data from the projects in the dataset of Batselier and Vanhoucke (2015) are read from the Microsoft Excel file (Batselier_Vanhoucke_2015.xlsx). This Microsoft Excel file contains the planned and actual durations for all activities in the 24 projects that are suited for the analysis.

Four-Step Procedure of Trietsch et al. (2012)

The four-step procedure of Trietsch et al. (2012) is applied to these 24 projects. If the Parkinson distribution with a lognormal core is not confirmed for a certain project, the four-step procedure returns NA as output. If however, the distributional model for a project is confirmed, the function returns a list for that project that includes

- The step (P0, P1, P1.1, P1.2) in which the hypothesis was confirmed;
- The s and m estimates (\hat{s} and \hat{m});
- The p value; and
- A list of the values in the empirical relative distribution.

Classification

The classification of Table 3 was produced from the grouping of the projects with respect to their \hat{s} and \hat{m} values. This grouping can be reproduced through the application of Levene's and Welsch's tests to the output of the four-step procedure. First the statistical interpretation of \hat{s} can be reproduced, and next, the statistical interpretation of \hat{m} can be repeated. The latter includes the implementation of the enumeration algorithm.

Template

The actual R code is shown in Fig. 5.

References

- Batselier, J., and Vanhoucke, M. (2015). "Construction and evaluation framework for a real-life project database." *Int. J. Project Manage.*, 33(3), 697–710.
- Blom, G. (1958). *Statistical estimates and transformed beta-variables*, Wiley.
- Colin, J., and Vanhoucke, M. (2014). "Setting tolerance limits for statistical project control using earned value management." *Omega Int. J. Manage. Sci.*, 49, 107–122.
- Colin, J., and Vanhoucke, M. (2015). "On the performance of statistical process control for project management." *Int. J. Project Manage.*, in press.
- Elmaghraby, S. (2000). "On criticality and sensitivity in activity networks." *Eur. J. Oper. Res.*, 127(2), 220–238.
- Elshaer, R. (2013). "Impact of sensitivity information on the prediction of project's duration using earned schedule method." *Int. J. Project Manage.*, 31(4), 579–588.
- Gutierrez, G., and Kouvelis, P. (1991). "Parkinson's law and its implications for project management." *Manage. Sci.*, 37(8), 990–1001.
- Hahn, E. (2008). "Mixture densities for project management activity times: A robust approach to PERT." *Eur. J. Oper. Res.*, 188(2), 450–459.
- Kuhl, M. E., Lada, E. K., Steiger, N. M., Wagner, M. A., and Wilson, J. R. (2007). "Introduction to modeling and generating probabilistic input processes for simulation." *Proc., 2007 Winter Simulation Conf.*, S. Henderson, B. Biller, M. Hsieh, J. Shortle, J. Tew, and R. Barton, eds., IEEE, NJ, 63–76.
- Kwak, Y. H., and Ingall, L. (2007). "Exploring Monte Carlo simulation applications for project management." *Risk Manage.*, 9(1), 44–57.
- Leu, S. S., and Lin, Y. C. (2008). "Project performance evaluation based on statistical process control techniques." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2008)134:10(813), 813–819.
- Levene, H. (1960). "Robust tests for quality of variances." *Contribution to probability and statistics: Essays in honour of Harold Hotelling*, I. Olkin, ed., Stanford University Press, Redwood City, CA, 278–292.

- Looney, S. W., and Gullledge, T. R., Jr. (1985). "Use of the correlation coefficient with normal probability plots." *Am Stat.*, 39(1), 75–79.
- McCullagh, P., and Nelder, J. (1989). *Generalized linear models, monographs on statistics & applied probability*, CRC Press.
- Mohan, S., Gopalakrishnan, M., Balasubramanian, H., and Chandrashekar, A. (2007). "A lognormal approximation of activity duration in pert using two time estimates." *J. Oper. Res. Soc.*, 58(6), 827–831.
- (OR&S) Operations Research & Scheduling. (2014). (www.projectmanagement.ugent.be).
- Ragsdale, C. (1989). "The current state of network simulation in project management theory and practice." *Omega Int. J. Manage. Sci.*, 17(1), 21–25.
- R Core Team. (2013). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
- ReVelle, J. B. (2004). *Quality essentials: A reference guide from A to Z*, ASQ Quality Press, Milwaukee, WI.
- Schonberger, R. (1981). "Why projects are" always" late: A rationale based on manual simulation of a PERT/CPM network." *Interfaces*, 11(5), 66–70.
- Small, B. (1958). *Statistical quality control handbook*, AT&T, Indianapolis.
- Trietsch, D., and Baker, K. R. (2012). "Pert 21: Fitting PERT/CPM for use in the 21st century." *Int. J. Project Manage.*, 30(4), 490–502.
- Trietsch, D., Mazmanyan, L., Govergyan, L., and Baker, K. R. (2012). "Modeling activity times by the Parkinson distribution with a lognormal core: Theory and validation." *Eur. J. Oper. Res.*, 216, 386–396.
- Vanhoucke, M. (2010a). *Measuring time—Improving project performance using earned value management*, Springer, New York.
- Vanhoucke, M. (2010b). "Using activity sensitivity and network topology information to monitor project time performance." *Omega Int. J. Manage. Sci.*, 38(5), 359–370.
- Vanhoucke, M. (2011). "On the dynamic use of project performance and schedule risk information during project tracking." *Omega Int. J. Manage. Sci.*, 39(4), 416–426.
- Wauters, M., and Vanhoucke, M. (2014). "Support vector machine regression for project control forecasting." *Autom. Constr.*, 47, 92–106.
- Welch, B. L. (1947). "The generalization of student's problem when several different population variances are involved." *Biometrika*, 34(1/2), 28–35.
- Wheeler, D. J. (1995). *Advanced topics in statistical process control*, Vol. 470, SPC Press, Knoxville, TN.
- Williams, T. (1999). "Towards realism in network simulation." *Omega Int. J. Manage. Sci.*, 27(3), 305–314.