González-Cruz, M.; Ballesteros-Pérez, P.; Lucko, G.; Zhang, J. (2022). Critical Duration Index: Anticipating Project Delays From Deterministic Schedule Information. Journal of Construction Engineering and Management. 148(11):1-12. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002387

Critical Duration Index: Anticipating Project Delays From Deterministic Schedule Information

Maria Carmen González-Cruz, Ph.D. [1]; Pablo Ballesteros-Pérez, Ph.D. [2];
Gunnar Lucko, Ph.D., M.ASCE [3]; Jing-Xiao Zhang, Ph.D. [4]

Abstract

Classical scheduling techniques are well known to underestimate the average project duration, yet they remain widely used in practice due to their simplicity. In this paper the new Critical Duration Index (*CDI*) is proposed. This index indirectly allows anticipating the probability of a project ending late, as well as the average project duration extension compared to a deterministic project duration estimate. The accuracy of two simple regression expressions that use the *CDI* are tested on two representative datasets of 4,100 artificial and 108 empirical (real) projects. Results show that these regression expressions outperform the only alternative index found in the literature. Besides allowing enhanced forecasting possibilities, calculating the *CDI* only requires basic scheduling information that is available at the planning stage. It can thus be easily adopted by project managers to improve their project duration estimates over prior deterministic techniques.

Keywords

Project; Scheduling; Delays; Forecasting; Regression

---

[1] Project Management, Innovation and Sustainability Research Centre (PRINS). Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia (Spain) ORCID: https://orcid.org/0000-0002-6987-5732 Email: mcgonzal@dpi.upv.es

[2] Project Management, Innovation and Sustainability Research Centre (PRINS). Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia (Spain) ORCID: https://orcid.org/0000-0002-4629-9664 Email: pabbalpe@dpi.upv.es

[3] Catholic University of America, 620 Michigan Avenue NE, Washington, DC 20064 (USA) https://orcid.org/0000-0002-7355-3365 Email: lucko@cua.edu

[4] School of Economics and Management, Chang'an University, 710064 Xian (P.R .China) Email: zhangjingxiao@chd.edu.cn

Practical applications

(for non-academic or practitioner audiences)

Classical scheduling techniques like Gantt charts or the critical path method are known to underestimate the project duration. However, they remain widely used in practice due to their simplicity. In this paper we have proposed the Critical Duration Index (*CDI*). This index allows anticipating the probability of a project ending later than the estimate produced with the Gantt chart or the critical path method. It also allows estimating how much longer the project might take to be completed, that is, the extension of the delay.

The accuracy of two simple regression expressions that use the *CDI* are tested on two representative artificial and real project datasets. Our results show that these regression expressions outperform the only alternative index found in the literature. Besides allowing to forecast the probability and extension of a project delay, calculating the *CDI* only requires basic scheduling information that is available at the planning stage. Hence, it can be easily adopted by project managers to improve their project duration estimates over other deterministic techniques.

Introduction

Projects ending late and costing more than planned is unfortunately a pervasive problem in the construction industry (Ansar *et al.* 2016; Flyvbjerg 2011). This happens in most countries and in all types of projects (Ballesteros-Pérez, 2017; Hamzah et al., 2011; Keane & Caletka, 2008).

There is a myriad of factors that cause a project to be delayed or suffer from a cost overrun (Mahamid *et al.* 2012). Among them are changes in the project scope, cash flow or payment problems, harsh weather conditions, resource scarcity, legal disputes, and others (Ballesteros-Pérez et al., 2012; Chudley & Greeno, 2016). Yet one that is frequently cited among the most common is poor project planning and scheduling practices (Zidane and Andersen 2018).

In many industries – construction included – classical scheduling techniques still are the default option used by most project managers (Ballesteros-Pérez et al. 2018; Wilson 2003). They include Gantt bar charts (Taylor 1903), the critical path method (CPM) (Kelley and Walker 1989, 1959), and the program evaluation and review technique (PERT) (Malcolm *et al.* 1959). Such techniques are easy to learn and their results are easy to convey to others, including untrained project stakeholders. For practice, a vast array of commercial software implementation is available as well (e.g. Agantty, GanttPro, GanttProject, Ganttic, Liquid Planner, Microsoft Project, nTask, Oracle Primavera).

Classical scheduling techniques are intrinsically deterministic. Deterministic means that they assume that activity durations are fixed, i.e., not suffer from duration variability (Ballesteros-Pérez et al. 2020c). It is noted that this is true even in the case of PERT. PERT is a pseudo-probabilistic project monitoring and control technique, because the way in which it estimates the average project duration is deterministic. Namely, PERT simplistically assumes that the average project duration equals exactly the sum of the critical activities' average durations (Khamooshi and Cioffi 2013; Nelson *et al.* 2016).

Yet the simple appeal of these techniques comes at a cost. Among other limitations, it has long been known that deterministic techniques tend to underestimate the average project duration (MacCrimmon and Ryavec 1964; Clark 1961). This means that they are too optimistic for real projects (Ballesteros-Pérez *et al.* 2018; Ballesteros-Pérez 2017a). This renders them unrealistic for complex projects whose execution is characterized by a diverse set of resources in simultaneous activities (Barrientos-Orellana *et al.* 2021).

Unfortunately, this project duration underestimation cannot be overcome by merely tweaking the calculation approach of such scheduling techniques. Instead, to produce more accurate and representative project duration estimates one must resort to other probabilistic alternatives. Examples are stochastic network analysis (SNA) (Pontrandolfo 2000) or schedule risk analysis (SRA) (Ballesteros-Pérez *et al.* 2019a). Both SNA and SRA explicitly use probabilistic calculations, hence are subject to statistical distribution assumptions (Valadares Tavares *et al.* 1999). These advanced scheduling techniques (and others that will be reviewed later) are more complex and thus more computationally expensive (Vanhoucke *et al.* 2016; Vanhoucke 2011). Also, most practitioners are not trained in them, and only few software tools facilitate their implementation (Sanz-Ablanedo *et al.*, 2020; Trietsch and Baker 2012).

However, the need of obtaining reliable project duration estimates in practical contexts remains. That is why a new index named *Critical Duration Index* (*CDI*) is proposed. The *CDI* indirectly allows anticipating the probability of a project ending later than the project duration estimate produced by a deterministic schedule. It also allows approximating the extension of the project delay, i.e., the time difference between the actual and planned project finish dates. Calculating the *CDI* only requires knowing the activity planned durations and their total floats (slacks). This information is readily available in any deterministic schedule. The performance of some regression expressions that use the *CDI* as the only independent variable will be tested on a representative set of artificial projects with diverse topologies, as well as in a set of empirical projects. Finally, it will also be compared how the *CDI* fares against the only similar alternative index found in the literature.

This research will solely focus on the project time (duration) dimension, not in the cost (money) dimension. Producing accurate project *duration* estimates has long been found harder than generating accurate project *cost* estimates (Herrerías-Velasco *et al.* 2011; Clark 1962). This is because costs are merely additive. Hence, by the Central Limit Theorem (CLT), the actual cost of a project statistically approaches the sum of the activities' average cost estimates (Ballesteros-Pérez *et al.* 2020c). However, the project duration depends not just on the activities' average durations, but also on their duration variability, their order of execution, interruptions, overlaps, time lags, etc. (Ballesteros-Pérez *et al.* 2020b). This is why producing reliable project duration estimates from deterministic schedules has proven to be more challenging.

Literature Review

The number of scheduling techniques that can anticipate the average project duration is vast and can only be outlined in broad strokes here. Each technique is suited to a particular context, type of project (e.g. linear and repetitive projects, etc.), project planning or tracking information requirements, statistical knowledge of the scheduler, available computer resources, and other factors (Vanhoucke 2013). This literature review establishes four desirable criteria by which it evaluates existing scheduling approaches. Those should be (a) deterministic, (b) as simple to calculate as the scheduling technique itself, (c) not need information beyond what is known at the planning stage, and (d) overcome, at least partially, the problem of project duration underestimation.

One group of literature encompasses PERT and EVM (earned value management) extensions. Both PERT and EVM are project planning and monitoring techniques. This means that besides establishing a project baseline, they allow controlling whether the project is progressing as expected (Hajdu and Bokor 2014; Hajdu 2013). The way in which they estimate the average project duration is deterministic (also in PERT, as we have noted earlier). But this is not true in most PERT extensions where a truly probabilistic approach is adopted [*cf.* Ballesteros-Pérez (2017a) for a recent review]. This substantially increases their complexity. Hence, they are neither (a) deterministic, nor (b) simple.

Regarding EVM extensions, both probabilistic and deterministic extensions can be found in the literature [*cf.* Barrientos-Orellana *et al.* (2021) for a review of deterministic EVM extensions aimed at forecasting the project duration]. Their limitation is that access to updated project tracking information is required. Tracking information includes the percentage of progress, duration and actual costs of completed or ongoing activities (Khamooshi and Golafshani 2014; Wauters and Vanhoucke 2014). Without such information, EVM cannot estimate the project duration better than any deterministic technique (Kerkhove and Vanhoucke 2017; de Koning and Vanhoucke 2016; Batselier and Vanhoucke 2015a; Colin and Vanhoucke 2015). Hence, condition (c) is not fulfilled.

Other literature addresses network topological indicators that fulfill the first three conditions [(a) deterministic, (b) simple, (c) only planning information]. A topological indicator is a metric that describes a particular trait of a project network (the structure of the activities' precedence relationships). A wide variety of indicators exists, including the coefficient of network complexity (*CNC*), order strength (*OS*), serial-parallel (*SP*), activity distribution (*AD*), length of arcs (*LA*), and topological float (*TF*) [*cf.* comprehensive reviews by Vanhoucke (2010, 2008)].

The intrinsic limitation of these topological indicators is that they serve a different purpose: They are aimed at numerically capturing the characteristics of a network structure, and in doing so, they neglect the activity durations which, of course, also impact the eventual project duration. Yet, a project schedule is ultimately a network of activities and these indicators can still provide some useful information of the project duration. Of particular interest is the serial-parallel (*SP*) indicator proposed by Vanhoucke 2010, 2008) as per Equation 1:

$$SP = \frac{m-1}{n-1} \qquad \textbf{Eq. 1}$$

where *m* is the number of activities in the longest chain (not necessarily longest in terms of duration, only in its activity count), and *n* is the total number of activities in the entire network (schedule). The *SP* approaches 0 if all activities are arrayed in parallel and 1 if all activities are in sequence. Most construction projects lie in between *SP* ≈ 0.2-0.8, though (Ballesteros-Pérez et al. 2020c).

The *SP* alone is useless to estimate the project duration. Yet, it is known that projects whose schedules contain more parallel paths (*SP* closer to 0) are more prone to experience delays (Ballesteros-Pérez *et al.* 2019a, 2019b; Ballesteros-Pérez 2017b; Vanhoucke 2010). It is thus possible to improve the formulation of the *SP* indicator and generate a new index that can consider activity durations, not just their relationships. This refinement will become the *CDI*.

On the other hand, Ballesteros-Pérez *et al.* (2020a) developed a related index to approximate the project duration average and standard deviation. The information such index required was mostly (but not exclusively) available at the planning stage. However, that index also had important limitations. While it could be calculated manually, its calculation is cumbersome and its implementation is only feasible for small networks, not real-world sized ones. It also required subjectively setting parameter values to discriminate sub-critical activities (those with total floats close to zero). Additionally, its calculation involved an estimate of the activity durations' variability, a piece of information that is not available in most deterministic schedules. However, since that index is the only alternative to the *CDI*, both will be compared later.

For the sake of completeness, a plethora of other non-deterministic scheduling techniques have also been aimed at estimating a project's duration. They include fuzzy logic (Chen 2007), artificial neural networks (ANN) (Lu 2002), stochastic network analysis (SNA) (Dodin and Sirvanci 1990), Monte Carlo simulation (Liu and Wang 2013), machine learning (Acebes *et al.* 2015), and schedule risk analysis (SRA) (Vanhoucke 2015), to cite just a few. Their problem is that they do not fulfill being (a) deterministic, nor (b) simple, and sometimes not even condition (c) which is relying only on planning information. Hence, these techniques will no longer be considered.

Research methods

Objectives

As mentioned earlier, deterministic techniques are popular in practice, but suffer from the problem of underestimating the average project duration. Hence, the goal of this research is to complement them by developing a new index that:

a) remains deterministic;
b) must be as simple to calculate as the scheduling technique itself;
c) needs no information other than what is known at the planning stage;
d) overcomes, at least partially, the project duration underestimation problem.

CDI Rationale

We have emphasized that most construction projects usually last longer than planned and that classical scheduling techniques tend to underestimate the project duration. This problem is widely known in the construction industry (Altuwaim and El-Rayes 2021; Fan *et al.* 2021; Votto *et al.* 2021). However, it is often neglected that activities' duration *variability* is also directly related to a higher probability of project delays.

Let us consider a project schedule made up of a single activity i whose duration is d_i. The activity duration d_i is approximated by a constant value in a deterministic schedule – normally its expected average duration, let's say 10 days. Yet, project schedulers are aware that d_i is subject to some degree of variability because various risk factors may impact the execution of activity i. This means that the actual (real) activity duration will likely differ from its *average* planned duration (e.g. activity i will eventually last 9 days or 12 days, instead of 10 days). Hence, the larger the variability of d_i, the larger the dispersion of that actual activity duration compared to its average.

In this regard, Ballesteros-Pérez *et al.* (2020c) proved that the average of the actual durations of most construction activities coincides with their planned duration. This means that construction activities approximately end up late 50% of the time (and earlier 50% of the time). In our one-activity project example, this would mean that the probability of the whole project ending late is about 50% and that the average delay would be about zero. But it was also found that the higher the count of parallel paths in a project, the higher the chances of said project ending late.

Let us now consider a project with *two parallel* activities whose average durations are the same. For this project to end on time, both activities must end either early or just on time. Probabilistically speaking, this can occur only in one out of four scenarios. This is because each parallel activity has a 50% chance of finishing on time, that is ½. The combined probability of both activities ending on time is the intersection of both events, that is, ½·½=¼. In other words, it suffices that one activity ends late for the whole project to be delayed.

Consequently, if the project has three parallel activities (again all with same average durations), the chances of the project ending on time will be 2^{-3} = ⅛ = 0.125. From this it is easy to generalize that in projects with j paths with the same average duration, the chances of ending on time will equal 2^{-j}. This probability approaches zero rapidly as the count of paths (j) increases.

Of course, real projects do not have paths that last exactly the same, but most of them have multiple parallel paths. Each path can be made up of one or several activities that are in sequence, in parallel or, more commonly, in a combination of both. Moreover, good project managers keep reallocating resources so that no path gets too delayed. This means that in practice, when a project is properly managed, the actual durations of most paths should not eventually be that different. Hence, we conclude that the probability of a construction project ending late might not be that different from the 2^{-j} theoretical model.

Hence, adopting the 2^{-j} model may prove useful for estimating the chances of a project ending on time. The challenge, though, is how to approximate the number of parallel paths in a project, that is, estimating the value of j. In this regard, few projects have purely independent paths that span from beginning to end. Instead, most paths in real projects diverge, intertwine, overlap, and merge at different points of the project makespan. Also, construction activity durations tend to be quite varied in practice.

Explicitly counting the paths j by directly analyzing the project schedule is a non-trivial task, even more so for medium and large projects. This because in those projects the number of distinct paths skyrockets and many paths separate and merge from each other multiple times. The only alternative is to come up with a (good) approximation of *j*.

We propose that the approximation of the number of paths *j* with a relatively similar duration of a project can be reasonably approximated by the *Critical Duration Index* (*CDI*):

$$CDI = \frac{\sum_{i=1}^{n} \max\{0,\, d_i - tf_i\}}{\max\{PD,\, \sum_{k \in critical} d_k\}} \qquad \textbf{Eq. 2}$$

Equation 2 is straightforward and similar to the inverse of the aforementioned *SP* topological indicator but taking activity durations into account. Basically, the numerator adds up the activity durations (d_i) of all *n* activities in the schedule once their respective total floats (tf_i) have been subtracted. Yet, to avoid adding up negative quantities, the maximum between 0 and $d_i - tf_i$ is taken (this ignores all activities whose float is larger than its duration, i.e., $tf_i > d_i$). The rationale behind subtracting the total float is that we want to consider only durations of subcritical activities (those whose float is relatively small). This because it is expected that the path to which they belong will also have a relatively small total float and their overall duration will not differ much from the critical path.

The denominator of Equation 2 adds up the durations of the critical activities (d_k). Yet, in those projects where the sum of the critical activity durations (Σd_k) exceeds the project planned duration (*PD*), only the former is considered using the maximum operator. A situation where $\Sigma d_k > PD$ can happen if the critical activities are partially overlapped. Conversely, in situations where $\Sigma d_k \leq PD$, the denominator will equal *PD*. This can happen when critical activities' precedence relationships include positive time lags, i.e., if critical activities are not scheduled one immediately after the other. This might yield CDI values in between 0 and 1. However, since good project managers strive for work continuity, in most real projects, the time lags between critical activities tends to be relatively small ($\Sigma d_k \approx PD$). This because critical activities are usually given the highest priority during project execution.

Thus, it is expected that the *CDI* approximates the count of critical and subcritical parallel paths in most construction project schedules. The related probability of a project ending on time (named *PD percentile* henceforth) can be approximated by Equation 3. This is a simple regression expression that implements the previous 2^{-j} theoretical model where j has been directly replaced by the *CDI*:

$$PD\ percentile = \frac{1}{2^{CDI}} = 2^{-CDI} \qquad \textbf{Eq. 3}$$

In equation 3, the *CDI* can range between 0 and + infinity and its calculation must only include meaningful activities, i.e., those that use resources (e.g. hammock-type and/or summary activities are excluded). If the $CDI = 1$, then the project has a dominant critical path and its average delay should be close to zero (as in the one-activity project example). However, the higher the value of *CDI*, the larger the delay the project may suffer. It must be noted, though, that the magnitude of this delay does not grow linearly with the value of *CDI*. After some experimentation with linear, polynomial, exponential, logarithmic and power regression expressions, the authors confirmed that the average project delay (RD / PD, real duration divided by planned duration) can be well approximated by the following logarithmic expression:

$$Average \frac{RD}{PD} \approx 1 + a \cdot \ln(CDI) \qquad \textbf{Eq. 4}$$

where ln is the natural (Euler's) logarithm and a is a constant that depends on the amount and type of activities' duration variability. Generally, a varies between 0 and 0.5.

For real projects, which are the aim of this study, the authors have determined that the value of a is generally close to 0.333. This means that the average project delay compared to the deterministic project duration estimate is approximately ln(*CDI*)/3. The next subsections will focus on demonstrating how the regression estimates of Equations 3 and 4 perform on datasets of artificial and real projects.

Artificial Projects Dataset

We first test the accuracy of Equations 3 and 4 on an artificial dataset that comprises 4,100 project networks of different topologies (different configurations of activity precedence relationships). Each network has 30 activities plus two extra dummy activities of zero duration that signpost the start and the end of each project. This dataset is curated by the Operations Research & Scheduling Research Group of Ghent University. The complete dataset can be accessed here: https://www.projectmanagement.ugent.be/research/data (MT set).

The networks were generated with RanGen2. RanGen2 is a robust random network generator that was validated in prior studies (Vanhoucke *et al.* 2008; Demeulemeester *et al.* 2003). RanGen2 datasets have been used in multiple scheduling studies due to their representativeness, i.e., wide coverage of network typologies [e.g. Barrientos-Orellana et al. (2021); Ballesteros-Pérez *et al.* (2019b); Batselier and Vanhoucke (2015a); Colin and Vanhoucke (2014)]. Further information about the artificial dataset can be found in Vanhoucke *et al.* (2016).

To resemble real projects, the 30 activities of each artificial network should take on different duration values. As a first step we have randomly generated activity duration averages and their variability by Monte Carlo simulation. The four distributions that we have adopted are summarized in Table 1 (*Dataset ID* column).

< Insert Table 1 here >

As the first one we selected lognormal activity durations to resemble the variety of activity durations and variability that is found in most real construction projects (Colin and Vanhoucke, 2016; Trietsch *et al.,* 2012). Additionally, parameters for the lognormal distribution were set based on those measured in an extensive analysis of over 6,000 construction activities by Ballesteros-Pérez *et al.* (2020c). As the other three, we used Normal, Uniform and Beta distributions to analyze how Equations 2-4 would perform under very different inputs. These distributions were indeed chosen for the sake of generality, as they are very different from each other (symmetrical and skewed, and with different types of support). Also, many researchers have resorted to these distributions in prior studies when modelling construction activity durations (AbouRizk et al., 1994; AbouRizk & Halpin, 1992).

As a first step and for each of the 4,100 projects, we randomly generate a set of average activity durations for the 30 activities of each project with the distributions stated in the second column of Table 1 (*Average Activity Durations* column). This set corresponds to the deterministic activity durations (d_i) that would be used by any deterministic schedule to calculate activities' total floats (tf_i) and the project duration estimate (PD).

Then, we simulate 10,000 runs of each of those 4,100 projects while keeping the *average* duration d_i of each activity constant, but now allowing duration *variability*. This means we simulate 10,000 possible executions of each of the 4,100 projects in which activity durations vary according to the distributions of the third column of Table 1 (*Activity Duration Variability* column). In these 10,000 simulations, activities are scheduled to start as soon as possible (ASAP) and activity preemption (interruption) is not allowed. This simulation yields 10,000 project duration results per project which are labeled real durations (*RD*). They are named *RD* as they represent the possible durations that each of the 4,100 projects, once executed, may have had.

Finally, we count for each project the number of those 10,000 project executions whose real duration is below the deterministic project duration estimate (the one obtained with the initial *average* activity durations). With it, we calculate the proportion of times when $RD \leq PD$ which coincides with the *PD percentile* (probability of the project being early). Also, the average of the 10,000 *RD* values divided by the *PD* coincides with the average delay of each project. The *PD percentile* and *Average RD* / *PD* are the values that Equations 3 and 4 seek to estimate.

Hence, with this simulation exercise we have obtained accurate values for the probability of being late and the average time delay of 4,100 projects. If our equations 3 and 4, which are only expressed as a function of the *CDI*, work well, their outputs should not deviate much from the results obtained by simulation.

Artificial Projects Results

Detailed results for the 4,100 projects and the four activity duration distributions from Table 1 are listed in the supplemental online material (*4100 projects datasets.xlsx* file). For brevity, only the most relevant results are presented here.

Figure 1 shows scatterplots of the actual 4,100 *PD percentile* values as a function of their respective *CDI* with dashed regression curves. Overall, it is found that the coefficients of determination R^2 are high for equation 3 (which corresponds to 2^{-CDI}). Namely, R^2 values vary between 0.85 and 0.88 in Figure 1, which indicates a high goodness-of-fit in the four distributions displayed in Table 1.

< Insert Figure 1 here >

However, besides high R^2 values, other conditions of a representative regression are independence of errors (symmetrically distributed residuals) and constant variance of errors across observations (homoscedasticity). Symmetrically distributed residuals can be checked with PP plots or QQ plots. The latter are included as the supplemental material along with the project results but are not discussed here. The right columns of Table 2, though, show the residuals' first four moments (mean, standard deviation, skewness, and kurtosis).

< Insert Table 2 here >

It can be seen that the *mean* values are very close to zero, and that skewness values do not deviate much from 0 either. However, it is noted that the 4,100 projects do not constitute a homogeneous (balanced) dataset. That is, in the 4,100 projects there are more arrangements with parallel paths than with sequential paths. Therefore, perfectly distributed residuals cannot be expected either.

The residuals homoscedasticity can be checked with standardized residuals plots. These are also found in the supplemental material. Those plots reveal that the point clouds resemble 'gunshots' (absence of residuals directionality). This is also substantiated by the low R^2, intercept, and slope values of their respective regression expressions.

Figure 2, on the other hand, plots Equation 4, where the x-axis represents the natural logarithm of CDI, and the y-axis the average projects' real duration compared to its planned duration, that is, their average delay (in per unit basis, i.e., /1).

< Insert Figure 2 here >

Of particular interest is Figure 2a, which reaches $R^2 = 0.90$ with $a = 0.333$ and represents the variability of real construction projects. Figures 2b to 2d need slightly different slope parameter values (*a* values between 0.02 and 0.14, respectively), yet they also hold a good linear correlation with ln(*CDI*).

Similar conclusions on the independence of errors and homoscedasticity can be made from the regression plots of Figure 2. Table 3 shows the residuals' first four moments after applying Equation 4 to approximate the average project delay extension (*average RD* / *PD*). Again, the mean and skewness are quite close to zero, even closer than in Table 2. For a more detailed analysis the reader is referred to the QQ plots and standardized residuals plots that can be found in the supplemental material.

< Insert Table 3 here >

Finally, as mentioned earlier, other (polynomial, exponential, logarithmic and power) regression expressions were also tried for Equations 3 and 4 which sporadically rendered slightly higher R^2 values compared to those of Table 3. However, the authors decided to retain Equations 3 and 4 due to their extreme simplicity. This should make them more suitable for their adoption in the daily practice of deterministic scheduling.

Empirical Projects Dataset

Next, the accuracy of Equation 4 was also tested on a dataset of 108 real (empirical) projects. Equation 3 cannot be tested with such projects because real projects have only one realization (one *RD* value as the project only happened once). A single *RD* value does not allow measuring the actual *PD percentile*, as this is a probabilistic parameter. Also, most real projects have many paths and their probability of being early is well below 50% (close to 0 as per 2^{-CDI}). That is why so many projects end late and cannot be used for testing Equation 3.

The complete dataset of empirical project involved 133 projects and can be accessed at https://www.projectmanagement.ugent.be/research/data/realdata. Only 108 of them included the as-built schedules (and a *RD* value) and could be used for testing equation 4. The dataset has been described by Vanhoucke *et al.* (2016) and Batselier and Vanhoucke (2015b). This dataset keeps growing and has also been used by many scheduling researchers for validation purposes [e.g. (Ballesteros-Pérez *et al.* 2019b; Martens and Vanhoucke (2018); Colin and Vanhoucke 2016)]. Most projects are construction-related such as building, civil engineering, industrial, and services projects. Most projects were carried out in Belgium, the Netherlands, Italy, the U.S., and Azerbaijan. Further information about the 108 projects is given in the Table of the Appendix I. In that Table the original project code (*Project ID* column) has been retained for replicability. The last columns of the Table list the *CDI* values, plus the actual and estimated *RD* / *PD* values of each project (the latter being the value obtained with Equation 4).

Some of the 108 projects required substantial editing before they could be used in this research. Hammock or summary activities that did not correspond to actual work were identified. Those activities spanned many days but did not actually involve any use of resources. Since they could distort the value of *CDI*, they were removed before analysis.

Empirical Projects Results

Figure 3 compares the actual *RD* / *PD* values of the 108 projects with the logarithm of *CDI*. The regression curve of Equation 4 with slope $a = 0.333$ is shown as a dashed line.

< Insert Figure 3 here >

Overall, the coefficient of determination is moderately high ($R^2 = 0.801$) even though this dataset is also quite heterogeneous, and despite the presence of a few outliers.

More detailed results can be found in the supplemental material (see *108 Empirical projects dataset.xlsx* file).

Discussion

From examining Figures 1-3, as well as Tables 2 and 3, we conclude that Equations 3 and 4 provide good approximations of the probability of a project being early, and of the average project delay. We can also conclude that calculating the *CDI* might be useful for project schedulers who use deterministic scheduling techniques at the expense of a minimal extra calculation effort.

One check remains to perform, though. This is to test whether the *CDI* offers a significant advantage over the only alternative index found in the literature which was proposed by Ballesteros-Pérez *et al.* (2020a). Per Equations 2-4, the calculation of the *PD percentile* and *RD / PD* via *CDI* takes little effort. It also is completely deterministic; it does not involve subjectivity, nor any information that is unknown at the planning stage. On these objectives, the *CDI* already exhibits significant advantages over Ballesteros-Pérez et al.’s (2020a) index. However, it is also necessary to check whether equation 4 is also more accurate (equation 3 cannot be compared as Ballesteros-Pérez et al.’s (2020a) index did not estimate the probability of a project ending late). With that purpose, Table 4 is included.

< Insert Table 4 here >

Ballesteros-Pérez et al. (2020a) had tested the accuracy of their index on the same artificial dataset of 4,100 networks that this study has used, but with 13 parametrization options (I to XIII). The very same dataset, distributions, and parameters are therefore used for Equation 4 to compare its outputs with those of Ballesteros-Pérez et al. (2020a). For brevity, only the coefficients of determination have been included (see the last two columns of Table 4). The column *Previous* R^2 displays the results of Ballesteros-Pérez *et al*'s (2020a) index, while the column *Current* R^2 lists the results of Equation 4 calculated with the *CDI*. Detailed results by project for the 13 dataset configurations can be found in the *supplemental material* (*Ballesteros-Pérez et al 2020a index comparison.xlsx* file).

By comparing the last two columns of Table 4, though, we find that Equation 4 (right column) outperforms Ballesteros-Pérez et al. (2020a) (left column) for most parametrizations. The average R^2 in the last row is also slightly higher for Equation 4.

Hence, we conclude that using the *CDI* to estimate project delays is:

- much simpler (the calculation of Ballesteros-Pérez *et al.*'s (2020a) index is much more complex than that of the *CDI*),
- less information-demanding (Ballesteros-Pérez *et al.*'s (2020a) index requires information about the activities' duration variability, as well as subjectively setting the values of some regression parameters. None of these is available in a deterministic schedule).
- more accurate (as shown in Table 4).

Hence, the *CDI*, as well as equations 3 and 4, should be the default option for project schedulers who want to produce quick project delay estimates from deterministic schedule information.

The *CDI* could also allow a quick categorization of a projects' level of complexity. In this regard *CDI* captures the number of subcritical and critical activities that a project manager must handle simultaneously on average during execution. This interpretation enables the project manager to convey to stakeholders, especially the owner, if their project has more or less chances of ending late. For example, a project with *CDI* = 1 may be considered a project of basic complexity whose probability of ending late would be

50% and whose average expected delay is zero. A project with *CDI* = 3 will end late $1 - 2^{-3} = 0.875 \approx 88\%$ of the time and its average delay will be $\ln(3) / 3 = 0.366 \approx 36\%$ longer than its deterministic planned duration. This is relevant information for practice.

Conclusions

Contributions to the Body of Knowledge

The use of deterministic scheduling techniques is common in construction practice to plan and control projects. However, these techniques are prone to underestimating the actual project duration. To overcome this problem, a new index has been proposed called *Critical Duration Index* (*CDI*). The *CDI* is a simple and deterministic index whose calculation only involves the activity durations and total floats. The *CDI*, via two extremely simple regression expressions, allows approximating the probability that a project will have of ending later than planned with minimal extra calculation effort. The accuracy of the *CDI* and its regression expressions have been tested on two project datasets. Results suggest that the *CDI* provides good estimates of both the probability of a project ending late and the average project delay. It also outperforms a previous index proposed by Ballesteros-Pérez *et al*. (2020a). Finally, a possible interpretation of measuring the project complexity with the *CDI* has also been proposed.

Limitations

Our new approach also has some limitations. As the regression analyses have shown, the actual duration of a project can sometimes differ significantly from the estimates that the *CDI* produces. Of course, it cannot capture in a single magnitude the entire complexity of a project network. Yet, we feel that its accuracy is high enough for quick and manual estimates. The authors tried other mathematical configurations and alternative parameters for the calculation of the *CDI* as well as for equations 3 and 4. For example, in equation

3 we considered replacing the base 2 with Euler's number, that is, using $e^{-\mathrm{CDI}}$ instead of $2^{-\mathrm{CDI}}$. We also tried to use *free floats* instead of *total floats* when calculating Equation 2. However, those alternatives tended to be more prone to overestimating the number of critical and subcritical paths of a deterministic schedule, and also the probability of a delay. Hence, free floats were discarded, but their values have been included in the supplemental material should future researchers wish to pursue that line of inquiry.

Additionally, slightly more accurate, but also more complex regression expressions were found. Those were also rejected as they mostly involved higher order polynomials whose regression coefficients were difficult to anticipate in practice. Hence, as simplicity was a requirement for the *CDI* and its related regression expressions, we eventually decided to retain equations 2 to 4. Still, perhaps more accurate yet simple-to-use parametrizations of equations 2 to 4 might be found in future research. Similarly, a more accurate approach for anticipating the value of a (the slope of equation 4) might also be explored by future researchers.

Next, the practitioners' community should extensively test our expressions in a wider set of projects. It would be useful to receive some feedback about the usefulness of the concept the *CDI* represents, as well as the convenience and simplicity of its derived regression expressions for anticipating project delays. As a second step, the *CDI*-derived classification of project complexity may help raising awareness on a higher probability of projects suffering from resource conflicts. This due to the simultaneous execution of multiple activities sharing the same resources. Eventually, all this will hopefully help the practitioners' community to better understand the limitations of current deterministic scheduling techniques, as well as the need to resort to other tools and techniques that overcoming those limitations.

Appendix I: A summary of the 108 empirical projects characteristics along with the CDI project duration estimates

A summary of the 108 empirical projects characteristics along with the *CDI* project duration estimates is included in the following table.

Table. 108 empirical projects summary.

Project	**Project**	**Project**	**Planned Cost**	**Real Cost**	**Planned Dur.**	**Real Dur.**	**Activity**	***CDI***	***AD* / *PD* (/1)**	
ID	**Name**	**Type**	***PC* [€]**	***RC* [€]**	***PD* [d]**	***RD* [d]**	**Count**		**Actual**	**Estimate**
C2011-05	Telecom System Agnes	Service	180,485.27	180,485.27	43	53	20	1.92	1.23	1.22
C2011-07	Patient Transport System	Service	180,759.44	191,065.06	389	444	49	1.46	1.14	1.13
C2011-10	Building a House	Building	484,398.41	494,947.71	195	203	32	1.09	1.04	1.03
C2011-12	Claeys-Verhelst Premises	Building	3,027,133.19	3,102,395.91	443	453	49	1.10	1.02	1.03
C2011-13	Wind Farm	Civil Eng.	21,369,835.51	26,077,764.74	525	600	107	1.43	1.14	1.12
C2012-13	Pumping Station Jabbeke	Industrial	336,410.15	350,511.31	125	140	74	1.43	1.12	1.12
C2012-15	The Master Project	Service	185,472.45	185,113.10	32	32	121	1.00	1.00	1.00
C2012-17	Building a Dream	Building	241,015.00	314,856.14	145	204	33	2.15	1.41	1.26
C2013-01	Wiedauwkaai Fenders	Civil Eng.	1,069,532.42	1,314,584.58	152	152	39	1.27	1.00	1.08
C2013-02	Sewage Plant Hove	Civil Eng.	1,236,603.66	1,146,444.38	403	408	175	1.04	1.01	1.01
C2013-03	Brussels Finance Tower	Building	15,440,865.89	16,338,027.20	425	426	55	1.00	1.00	1.00
C2013-04	Kitchen Tower Anderlecht	Building	2,113,684.00	2,512,524.00	333	453	244	2.03	1.36	1.24
C2013-05	PET Packaging	Service	874,554.28	874,554.28	521	632	28	1.99	1.21	1.23
C2013-06	Govmnt. Office Building	Building	19,429,810.51	21,546,846.18	352	344	275	1.00	0.98	1.00
C2013-07	Family Residence	Building	180,476.47	175,030.65	170	174	46	1.00	1.02	1.00
C2013-08	Timber House	Building	501,029.51	576,624.05	216	235	41	1.23	1.09	1.07
C2013-09	Urban Develop.Project	Civil Eng.	1,537,398.51	1,696,971.79	291	360	71	1.91	1.24	1.22
C2013-10	Town Square	Civil Eng.	11,421,890.36	15,218,926.38	786	785	186	1.00	1.00	1.00
C2013-11	Recreation Complex	Building	5,480,518.91	5,451,028.00	359	277	159	0.60	0.77	0.83
C2013-12	Young Cattle Barn	Building	818,439.99	879,853.17	115	188	27	5.23	1.63	1.55
C2013-13	Office Finish. Works (1)	Building	1,118,496.59	955,929.22	236	217	11	0.82	0.92	0.93
C2013-14	Office Finish. Works (2)	Building	85,847.89	75,468.30	80	88	9	1.36	1.10	1.10
C2013-15	Office Finish. Works (3)	Building	341,468.11	308,343.78	171	115	17	0.52	0.67	0.78
C2013-16	Office Finish. Works (4)	Building	248,203.92	198,567.00	196	108	7	0.29	0.55	0.59
C2013-17	Office Finish. Works (5)	Building	244,205.40	203,605.97	161	107	23	0.35	0.66	0.65
C2014-01	Mixed-use Building	Building	38,697,822.73	39,777,643.30	474	448	41	1.00	0.95	1.00
C2014-02	Playing Cards	Industrial	191,492.70	190,266.50	124	146	21	1.77	1.18	1.19
C2014-03	Organizational Develop.	Service	43,170.15	83,712.15	229	260	112	1.50	1.14	1.14
C2014-04	Compres. Station Zelzate	Industrial	62,385,597.58	65,526,930.04	522	844	24	7.69	1.62	1.68
C2014-05	Apartment Building (1)	Building	532,410.29	591,410.53	228	274	25	1.48	1.20	1.13
C2014-06	Apartment Building (2)	Building	3,486,375.47	3,599,114.11	547	611	29	1.27	1.12	1.08
C2014-07	Apartment Building (3)	Building	1,102,536.78	1,289,696.78	353	404	25	1.31	1.14	1.09
C2014-08	Apartment Building (4)	Building	1,992,222.09	2,380,299.86	233	275	39	1.49	1.18	1.13
C2015-01	Young Cattle Barn (2)	Building	612,769.44	646,473.65	131	210	27	3.12	1.60	1.38
C2015-02	Railway Station (1)	Civil Eng.	1,121,316.94	967,988.79	417	501	216	1.78	1.20	1.19
C2015-03	Industrial Complex (1)	Building	2,244,090.74	1,868,796.28	257	278	135	1.25	1.08	1.07
C2015-04	Apartment Building (5)	Building	2,750,938.00	2,590,796.73	160	205	56	1.84	1.28	1.20

Project	Project	Project	Planned Cost	Real Cost	Planned Dur.	Real Dur.	Activity	*CDI*	*AD* / *PD* (/1)	
ID	Name	Type	*PC* [€]	*RC* [€]	*PD* [d]	*RD* [d]	Count		Actual	Estimate
C2015-06	Family Residence (2)	Building	143,673.20	186,107.00	260	290	184	1.51	1.12	1.14
C2015-07	Industrial Complex (2)	Building	5,999,600.00	5,414,544.00	297	313	138	1.31	1.05	1.09
C2015-08	Garden Center	Building	467,297.21	461,900.17	191	186	186	1.00	0.97	1.00
C2015-09	Railway Station (2)	Civil Eng.	1,457,424.00	2,145,682.26	354	569	340	6.53	1.61	1.63
C2015-10	Tax Return System (1)	Service	18,990.00	8,010.00	85	85	15	1.00	1.00	1.00
C2015-11	Staff Authoriz. System	Service	14,400.00	9,105.00	55	55	7	1.00	1.00	1.00
C2015-12	Premium Payment System	Service	132,570.00	58,410.00	184	184	35	1.00	1.00	1.00
C2015-13	Broker Acc.Conv. System	Service	12,735.00	9,990.00	117	117	16	1.00	1.00	1.00
C2015-14	Sup. Pensions Database	Service	34,260.00	18,285.00	124	124	17	1.00	1.00	1.00
C2015-15	FACTA System	Service	11,700.00	7,035.00	57	57	13	1.00	1.00	1.00
C2015-16	Generic Doc. Output Syst.	Service	64,620.00	64,125.00	270	270	22	1.00	1.00	1.00
C2015-17	Insurance Bundling Syst.	Service	281,430.00	281,070.00	208	236	86	1.45	1.13	1.12
C2015-18	Tax Return System (2)	Service	39,450.00	25,380.00	128	128	15	1.00	1.00	1.00
C2015-19	Receipt Numb. System	Service	43,800.00	37,530.00	182	182	20	0.96	1.00	0.99
C2015-20	Policy Numbering System	Service	12,645.00	11,100.00	171	161	6	0.95	0.94	0.98
C2015-21	Investment Product (1)	Service	4,020.00	3,240.00	37	37	12	1.00	1.00	1.00
C2015-22	Risk Profile Questionnaire	Service	29,880.00	17,400.00	151	151	22	1.00	1.00	1.00
C2015-23	Investment Product (2)	Industrial	46,920.00	32,805.00	122	120	33	0.99	0.98	1.00
C2015-24	CRM System	Service	44,130.00	36,870.00	233	233	21	1.00	1.00	1.00
C2015-25	Beer Tasting	Service	1,210.00	1,780.00	14	14	18	1.00	1.00	1.00
C2015-26	Debt Collection System	Service	458,112.37	512,546.15	148	154	214	1.22	1.04	1.07
C2015-27	Railway Station Antwerp	Building	22,703.52	25,313.12	68	81	18	1.28	1.19	1.08
C2015-28	Web. Tennis Vlaanderen	Service	219,275.00	382,475.00	201	212	20	1.22	1.05	1.07
C2015-29	Fire Station	Building	1,874,496.82	1,887,087.25	284	298	204	1.25	1.05	1.07
C2015-30	Social Apts. Ypres (1)	Building	440,940.89	440,940.89	244	254	40	1.10	1.04	1.03
C2015-31	Social Apts Ypres (2)	Building	1,310,723.46	1,282,185.98	271	364	29	2.80	1.34	1.34
C2015-32	Social Apts Ypres (3)	Building	2,509,031.42	2,509,031.42	358	265	48	0.56	0.74	0.81
C2015-33	IJzertoren Memor. Square	Civil Eng.	214,417.71	224,789.67	50	94	12	6.98	1.88	1.65
C2015-34	Roadworks Poperinge	Civil Eng.	511,325.86	440,394.16	120	193	13	5.75	1.61	1.58
C2015-35	Retirement Apartments	Building	14,956,314.25	16,068,878.30	850	951	11	1.42	1.12	1.12
C2016-01	Railway Bridge (1)	Civil Eng.	671,383.50	703,703.50	225	274	26	1.97	1.22	1.23
C2016-02	Railway Bridge (2)	Civil Eng.	962,181.56	972,341.56	229	239	23	1.15	1.04	1.05
C2016-03	Railway Bridge (3)	Civil Eng.	926,888.01	910,728.01	203	220	25	1.19	1.08	1.06
C2016-04	Railway Bridge (4)	Civil Eng.	906,253.87	906,253.87	248	242	26	0.89	0.98	0.96
C2016-05	Railway Bridge (5)	Civil Eng.	832,497.46	832,497.46	195	197	32	1.05	1.01	1.02
C2016-06	Defense Building	Service	4,331,260.49	4,331,260.49	252	232	96	1.00	0.92	1.00
C2016-07	Shop. Village Walkways	Civil Eng.	930,179.09	932,757.25	224	316	110	3.42	1.41	1.41
C2016-08	SCM System	Service	375,253.34	438,741.66	725	725	99	1.00	1.00	1.00
C2016-09	Data Loss Prevent. System	Service	584,951.77	1,425,155.96	195	189	113	1.00	0.97	1.00
C2016-10	Biofuel Refinery	Industrial	14,362,625.00	14,466,100.00	360	375	23	1.25	1.04	1.07
C2016-11	Residential House (1)	Building	162,472.00	163,189.00	241	254	55	1.15	1.05	1.05
C2016-12	Residential House (2)	Building	222,858.00	226,285.00	291	291	59	1.00	1.00	1.00
C2016-13	Residential House (3)	Building	367,952.00	379,300.00	306	330	51	1.24	1.08	1.07
C2016-14	Residential House (4)	Building	218,366.00	222,021.78	321	320	48	1.00	1.00	1.00
C2016-15	Resid. House Struct. Work	Building	95,694.00	100,763.00	126	130	13	1.12	1.03	1.04
C2016-16	Resid. Finish. Works (1)	Building	54,577.76	64,526.76	90	90	24	1.00	1.00	1.00
C2016-17	Resid. Finish. Works (2)	Building	54,703.17	64,580.17	86	86	24	1.00	1.00	1.00
C2016-18	Resid. Finish. Works (3)	Building	51,115.52	60,829.52	91	91	25	1.00	1.00	1.00

Project	**Project**	**Project**	**Planned Cost**	**Real Cost**	**Planned Dur.**	**Real Dur.**	**Activity**	***CDI***	***AD* / *PD* (/1)**	
ID	**Name**	**Type**	***PC* [€]**	***RC* [€]**	***PD* [d]**	***RD* [d]**	**Count**		**Actual**	**Estimate**
C2016-19	Resid. Finish. Works (4)	Building	51,303.38	53,351.38	91	91	25	1.00	1.00	1.00
C2016-20	Resid. Finish. Works (5)	Building	52,021.28	53,783.28	91	91	25	1.00	1.00	1.00
C2016-21	Resid. Finish. Works (6)	Building	54,324.22	54,996.22	101	101	24	1.00	1.00	1.00
C2016-22	Resid. Finish. Works (7)	Building	56,969.40	57,822.40	101	101	24	1.00	1.00	1.00
C2016-23	Resid. Finish. Works (8)	Building	56,182.71	56,645.71	101	101	24	1.00	1.00	1.00
C2016-24	Resid. Finish. Works (9)	Building	52,262.83	53,176.83	101	101	24	1.00	1.00	1.00
C2016-25	Resid. Finish. Works (10)	Building	54,580.33	56,748.33	91	91	24	1.00	1.00	1.00
C2016-26	Resid. Finish. Works (11)	Building	51,286.24	53,319.24	91	91	24	1.00	1.00	1.00
C2016-27	Apt. Build. Found. (1)	Building	813,663.06	879,701.06	78	88	16	1.45	1.13	1.12
C2016-28	Apt. Struct. Work (1)	Building	569,177.85	586,086.85	71	79	19	1.40	1.11	1.11
C2016-29	Apt. Struct. Work (2)	Building	1,797,873.62	1,860,330.62	129	148	19	1.43	1.15	1.12
C2016-30	Apt. Struct. Work (3)	Building	1,319,736.29	1,353,361.29	85	96	23	1.52	1.13	1.14
C2016-31	Apt. Struct. Work (1)	Building	488,936.00	498,473.00	105	117	23	1.40	1.11	1.11
C2016-32	Apt. Struct. Work (2)	Building	477,381.00	496,991.00	89	97	22	1.23	1.09	1.07
C2016-33	Apt. Struct. Work (3)	Building	377,282.00	394,829.00	116	129	23	1.55	1.11	1.15
C2016-34	Apt. Struct. Work (4)	Building	362,476.00	383,871.00	83	92	23	1.40	1.11	1.11
C2019-01	Project Lepelstraat	Building	1,292,979.00	1,315,819.86	533	673	87	0.85	1.26	0.95
C2019-02	Social Housing	Building	734,602.11	748,555.80	352	355	18	1.00	1.01	1.00
C2019-04	Nuclear Healthcare	Building	4,318,950.00	4,232,553.41	373	520	33	2.84	1.39	1.35
C2019-05	Fuel Tank Filter	Industrial	1,456,000.00	1,476,290.00	510	515	15	1.00	1.01	1.00
C2019-06	Production Line Change	Industrial	1,512,000.00	1,534,060.00	480	501	23	1.15	1.04	1.05
C2019-07	Gluing Machine	Industrial	107,500.00	116,800.00	150	189	8	1.58	1.26	1.15
C2019-08	Labeling Machine	Industrial	114,700.00	128,200.00	115	182	7	3.34	1.58	1.40

Data Availability Statement

All data, models, and code generated or used during the study appear in the submitted article. Namely, all artificial and empirical project datasets, as well as the two indices' comparison have all been included as *Supplementary Materials*.

Acknowledgments

This research is supported by the National Social Science Fund projects (No.20BJY010); National Social Science Fund Post-financing projects (No.19FJYB017); Sichuan-Tibet Railway Major Fundamental Science Problems Special Fund (No.71942006); Qinghai Natural Science Foundation (No.2020-JY-736); List of Key Science and Technology Projects in China's Transportation Industry in 2018-International Science and Technology Cooperation Project (No.2018-GH-006 and No.2019-MS5-100); Emerging Engineering Education Research and Practice Project of Ministry of Education of China

(No.E-GKRWJC20202914); Shaanxi Province Higher Education Teaching Reform Project (No.19BZ016); and Humanities and Social Sciences Research Project of the Ministry of Education(21XJA752003) ; and Humanities and Social Sciences Research Project of the Ministry of Education(21XJA752003).

Supplemental Materials

The complete 4,100-project and 101-project activity databases and the extended calculation results are available online in the ASCE Library (www.ascelibrary.org).

References

AbouRizk, S. M., & Halpin, D. W. (1992). “Statistical Properties of Construction Duration Data.” *Journal of Construction Engineering and Management,* 118(3): 525–544.

AbouRizk, S. M., Halpin, D. W., & Wilson, J. R. (1994). “Fitting Beta Distributions Based on Sample Data.” *Journal of Construction Engineering and Management,* 120(2): 288–305.

Acebes, F., Pereda, M., Poza, D., Pajares, J., Galán, J. M. (2015). “Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques.” *International Journal of Project Management,* 33(7): 1597-1609.

Altuwaim, A., El-Rayes, K. A. (2021). “Multiobjective Optimization Model for Planning Repetitive Construction Projects.” *Journal of Construction Engineering and Management,* 147(7), 04021072(12).

Ansar, A., Flyvbjerg, B., Budzier, A., Lunn, D. (2016). “Does infrastructure investment lead to economic growth or economic fragility? Evidence from China.” *Oxford Review of Economic Policy,* 32(3): 360-390.

Ballesteros-Pérez, P. (2017a). “M-PERT: Manual Project-Duration Estimation Technique for Teaching Scheduling Basics.” *Journal of Construction Engineering and Management,* 143(9): 04017063(13).

Ballesteros-Pérez, P. (2017b). “Modelling the boundaries of project fast-tracking.” *Automation in Construction,* 84(December): 231-241.

Ballesteros-Pérez, P., Cerezo-Narváez, A., Otero-Mateo, M., Pastor-Fernández, A., Vanhoucke, M. (2019a). “Performance comparison of activity sensitivity metrics in schedule risk analysis.” *Automation in Construction*, 106(February): 102906.

Ballesteros-Pérez, P., Cerezo-Narváez, A., Otero-Mateo, M., Pastor-Fernández, A., and Vanhoucke, M. (2020a). “Forecasting the Project Duration Average and Standard Deviation from Deterministic Schedule Information.” *Applied Sciences*, 10(654): 22 pp.

Ballesteros-Pérez, P., González-Cruz, Ma. C., Pastor-Ferrando, J. P., & Fernández-Diego, M. (2012). “The iso-Score Curve Graph. A new tool for competitive bidding.” *Automation in Construction*, 22 (March): 481-490.

Ballesteros-Pérez, P., Larsen, G. D., González-Cruz, M. C. (2018). “Do projects really end late? On the shortcomings of the classical scheduling techniques.” *Journal of Technology and Science Education*, 8(1): 86-102.

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Cerezo-Narváez, A., Lucko, G., Pastor-Fernández, A., Otero-Mateo, M., Contreras-Samper, J. P. (2020b). “Forecasting accuracy of in-progress activity duration and cost estimates.” *Journal of Construction Engineering and Management*, 146(9): 1-13.

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Mora-Melià, D., González-Cruz, M. C., Fuentes-Bargues, J. L., Pellicer, E. (2019b). “Earned Schedule min-max: Two new EVM metrics for monitoring and controlling projects.” *Automation in Construction*, Elsevier, 103(April): 279-290.

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Soetanto, R., González-Cruz, M. C., Larsen, G. D., Cerezo-Narváez, A. (2020c). “Duration and cost variability of construction activities: an empirical study.” *Journal of Construction Engineering and Management*, 146(1): 04019093(13).

Barrientos-Orellana, A., Ballesteros-Pérez, P., Mora, D., González-Cruz, M. C., Vanhoucke, M. (2021). “Stability and accuracy of deterministic project duration forecasting method in Earned Value Management.” *Engineering, Construction and Architectural Management*, 29(3): 1449-1469.

Batselier, J., Vanhoucke, M. (2015a). “Empirical Evaluation of Earned Value Management Forecasting Accuracy for Time and Cost.” *Journal of Construction Engineering and Management*, 141(11): 05015010.

Batselier, J., Vanhoucke, M. (2015b). “Construction and evaluation framework for a real-life project database.” *International Journal of Project Management*, Pergamon, 33(3): 697-710.

Chen, S.-P. (2007). “Analysis of critical paths in a project network with fuzzy activity times.” *European Journal of Operational Research*, 183(1): 442-459.

Chudley, R., Greeno, R. (2016). *Building Construction Handbook*. Routledge. Abingdon, United Kingdom.

Clark, C. E. (1961). “The Greatest of a Finite Set of Random Variables.” *Operations Research*, 9(2): 145-162.

Clark, C. E. (1962). “Letter to the Editor – The PERT Model for the Distribution of an Activity Time.” *Operations Research*, 10(3): 405-406.

Colin, J., Vanhoucke, M. (2014). “Setting tolerance limits for statistical project control using earned value management.” *Omega*, 49(December): 107-122.

Colin, J., Vanhoucke, M. (2015). “A comparison of the performance of various project control methods using earned value management systems.” *Expert Systems with Applications*, 42(6): 3159-3175.

Colin, J., Vanhoucke, M. (2016). “Empirical perspective on activity durations for project-management simulation studies.” *Journal of Construction Engineering and Management*, 142(1): 04015047(13).

Demeulemeester, E., Vanhoucke, M., Herroelen, W. (2003). “RanGen: A random network generator for activity-on-the-node networks.” *Journal of Scheduling*, 6(1): 17-38.

Dodin, B., Sirvanci, M. (1990). “Stochastic networks and the extreme value distribution.” *Computers & Operations Research*, 17(4): 397-409.

Fan, S.-L., Yeh, I.-C., Chi, W.-S. (2021). “Improvement in Estimating Durations for Building Projects Using Artificial Neural Network and Sensitivity Analysis.” *Journal of Construction Engineering and Management,* American Society of Civil Engineers, 147(7), 04021050(9).

Flyvbjerg, B. (2011). *Over Budget, Over Time, Over and Over Again*. Oxford Handbook of Project Management, eds. Morris, P. W. G., Pinto, J., Söderlund, J., Oxford University Press,

Hajdu, M. (2013). “Effects of the application of activity calendars on the distribution of project duration in PERT networks.” *Automation in Construction*, 35(November): 397-404.

Hajdu, M., Bokor, O. (2014). “The Effects of Different Activity Distributions on Project Duration in PERT Networks.” *Procedia – Social and Behavioral Sciences,* 119: 766-775.

Hamzah, N., Khoiry, M. A., Arshad, I., Tawil, N. M., Che Ani, A. I. (2011). “Cause of Construction Delay - Theoretical Framework.” *Procedia Engineering,* 20: 490-495.

Herrerías-Velasco, J. M., Herrerías-Pleguezuelo, R., van Dorp, J. R. (2011). “Revisiting the PERT mean and variance.” *European Journal of Operational Research*, 210(2): 448-451.

Keane, P. J., Caletka, A. F. (2008). *Delay analysis in construction contracts*. Wiley-Blackwell, Oxford, United Kingdom.

Kelley, J. E., Walker, M. R. (1959). “Critical-Path Planning and Scheduling.” *Proceedings of the Eastern Joint Computer Conference,* Boston, MA, December 1-3, 1959, McMillan, New York City, NY: 160-173

Kelley, J. E., Walker, M. R. (1989). “The Origins of CPM: A Personal History.” *PMnetwork,* 3(2): 7-22.

Kerkhove, L. P., Vanhoucke, M. (2017). “Extensions of earned value management: Using the earned incentive metric to improve signal quality.” *International Journal of Project Management*, 35(2): 148-168.

Khamooshi, H., Cioffi, D. F. (2013). “Uncertainty in Task Duration and Cost Estimates: Fusion of Probabilistic Forecasts and Deterministic Scheduling.” *Journal of Construction Engineering and Management,* 139(5): 488-497.

Khamooshi, H., Golafshani, H. (2014). “EDM: Earned Duration Management, a new approach to schedule performance management and measurement.” *International Journal of Project Management,* 32(6): 1019-1041.

de Koning, P., Vanhoucke, M. (2016). “Stability of earned value management do project characteristics influence the stability moment of the cost and schedule performance index.” *Journal of Modern Project Management*, 4(1): 9-25.

Liu, Y., Wang, Z. F. (2013). “Analysis of Project Shedule Risk Indexes in PERT Network Using Monte Carlo Simulation.” *Advanced Materials Research,* 760-762: 2205-2211.

Lu, M. (2002). “Enhancing Project Evaluation and Review Technique Simulation through Artificial Neural Network-based Input Modeling.” *Journal of Construction Engineering and Management*, 128(5): 438-445.

MacCrimmon, K. R., Ryavec, C. A. (1964). “An Analytical Study of the PERT Assumptions.” *Operations Research,* 12(1); 16-37.

Mahamid, I., Bruland, A., Dmaidi, N. (2012). “Causes of Delay in Road Construction Projects.” *Journal of Management in Engineering,* 28(3): 300-310.

Malcolm, D. G., Roseboom, J. H., Clark, C. E., Fazar, W. (1959). “Application of a Technique for Research and Development Program Evaluation.” *Operations Research,* 7(5): 646-669.

Martens, A., Vanhoucke, M. (2018). “An empirical validation of the performance of project control tolerance limits.” *Automation in Construction,* 89(May): 71-85.

Nelson, R. G., Azaron, A., Aref, S. (2016). “The use of a GERT based method to model concurrent product development processes.” *European Journal of Operational Research,* 250(2): 566-578.

Pontrandolfo, P. (2000). “Project duration in stochastic networks by the PERT-path technique.” *International Journal of Project Management,* 18(3): 215-222.

Sanz-Ablanedo, E., Chandler, J. H., Ballesteros-Pérez, P., & Rodríguez-Pérez, J. R. (2020). “Reducing systematic dome errors in digital elevation models through better UAV flight design.” *Earth Surface Processes and Landforms,* 45(9): 2134-2147.

Taylor, F. W. (1903). “Shop Management.” *ASME Transactions,* 24: 1337-1480.

Trietsch, D., Baker, K. R. (2012). “PERT 21: Fitting PERT/CPM for use in the 21st century.” *International Journal of Project Management,* 30(4): 490_502.

Trietsch, D., Mazmanyan, L., Gevorgyan, L., & Baker, K. R. (2012). “Modeling activity times by the Parkinson distribution with a lognormal core: Theory and validation.” *European Journal of Operational Research,* 216(2): 386–396.

Valadares Tavares, L., Antunes Ferreira, J., Silva Coelho, J. (1999). “The risk of delay of a project in terms of the morphology of its network.” *European Journal of Operational Research,* North-Holland, 119(2): 510-537.

Vanhoucke, M. (2015). “On the use of schedule risk analysis for project management.” *Journal of Modern Project Management,* 2(3): 108-117.

Vanhoucke, M. (2008). “Measuring time using novel earned value management metrics.” *Proceedings of the 22nd IPMA World Congress,* International Project Management Association, Zurich, Switzerland, November 9-11, 2008, Rome, Italy, 1: 99-103.

Vanhoucke, M. (2010). *Measuring Time – Improving Project Performance Using Earned Value Management*. International Series in Operations Research and Management Science, 136, Springer, Dordrecht, Netherlands.

Vanhoucke, M. (2011). “On the dynamic use of project performance and schedule risk information during project tracking.” *Omega,* 39(4): 416-426.

Vanhoucke, M. (2013). *Project Management with Dynamic Scheduling*. Springer, Berlin, Germany.

Vanhoucke, M., Coelho, J., Batselier, J. (2016). “An Overview of Project Data for Integrated Project Management and Control.” *Journal of Modern Project Management,* 3(3): 7-21.

Vanhoucke, M., Coelho, J., Debels, D., Maenhout, B., Tavares, L. V. (2008). “An evaluation of the adequacy of project network generators with systematically sampled networks.” *European Journal of Operational Research,* North-Holland, 187(2): 511-524.

Votto, R., Ho, L. L., Berssaneti, F. (2021). “Earned Duration Management Control Charts: Role of Control Limit Width Definition for Construction Project Duration Monitoring.” *Journal of Construction Engineering and Management,* 147(9): 04021108(14).

Wauters, M., Vanhoucke, M. (2014). “Study of the stability of earned value management forecasting.” *Journal of Construction Engineering and Management,* 141(4): 04014086(10).

Wilson, J. M. (2003). “Gantt charts: A centenary appreciation.” *European Journal of Operational Research,* 149(2): 430-437.

Zidane, Y. J. T., Andersen, B. (2018). “The top 10 universal delay factors in construction projects.” *International Journal of Managing Projects in Business,* 11(3): 650-672.

Table 1: 4100 Artificial Project Datasets with activity duration distributions

Distribution	Average Activity Durations		Activity Duration Variability	
	Mean μ	Std. deviation σ	Mean μ	Std. deviation σ
Lognormal	Uniform (0, 3)	0.50	0	Uniform (0.23, 0.70)
Normal	Uniform (0, 30)	$\mu/5$	0	Uniform (0, $0.35\cdot\mu$)
Uniform	Uniform (1, 30)		Uniform (0, $2\cdot\mu$)	
Beta	$(O+4\cdot L+P)/6$ with: $O \sim$ Uniform (1, 30) $P \sim O +$ Uniform (1, 30) $L \sim O +$ Uniform $(0, 1)\cdot(P-O)$		$[(P-O)/6]+$ SQRT[(5/6) + $+ (16/7)\cdot((L-O)(P-L)/(P-O)^2)]$ Bound between O and P	

Note: Parameters are mean and standard deviation for Lognormal, Normal and Beta, and lower and upper bound for Uniform and Beta. Negative values of the Normal distribution were truncated. The standard deviation of the Beta distribution uses the calculation proposed by Herrerías-Velasco et al. (2011) for the PERT technique.

Table 2: *PD* Percentile Regression Results with Residuals

Distribution	***x***	***y***	R^2	**Mean**	**Std. dev.**	**Skewness**	**Kurtosis**
Lognormal	$1/2^{\text{CDI}}$	PD percentile	0.88	-0.003	0.048	-0.698	-0.235
Normal	$1/2^{\text{CDI}}$	PD percentile	0.85	0.005	0.046	1.202	6.578
Uniform	$1/2^{\text{CDI}}$	PD percentile	0.85	-0.022	0.045	-0.379	2.438
Beta	$1/2^{\text{CDI}}$	PD percentile	0.86	0.026	0.044	1.721	2.870

Table 3: *RD*/*PD* Regression Results with Residuals

Distribution	***x***	***y***	R^2	**Mean**	**Std. dev.**	**Skewness**	**Kurtosis**
Lognormal	$1+0.33\cdot\ln(CDI)$	*RD* / *PD*	0.90	0.002	0.027	0.613	6.479
Normal	$1+0.06\cdot\ln(CDI)$	*RD* / *PD*	0.89	0.000	0.008	0.537	11.112
Uniform	$1+0.14\cdot\ln(CDI)$	*RD* / *PD*	0.85	0.001	0.014	0.054	7.642
Beta	$1+0.02\cdot\ln(CDI)$	*RD* / *PD*	0.84	0.000	0.003	1.167	16.094

Table 4: 4,100 Projects for Comparison of equation 4 with the index proposed by Ballesteros-Pérez *et al.* (2020a)

Dataset	Activities/ Project n_i	Activity Durations (d_i)	Activity Duration Variability	Previous R^2	Current R^2
I	30	Lognormal (2, 1)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.849	0.861
II	30	Lognormal (2, 1)	Lognormal with CV_i = 0.1 (constant)	0.852	0.886
III	30	Lognormal (2, 1)	Lognormal with CV_i = 0.3 (constant)	0.919	0.865
IV	15	Lognormal (2, 1)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.739	0.849
V	30	Lognormal (0.25, 0.005)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.928	0.940
VI	30	Lognormal (0.25, 0.75)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.899	0.873
VII	30	Lognormal (6, 0.12)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.939	0.942
VIII	30	Lognormal (6, 1.5)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.745	0.852
IX	30	Normal (25, 7.5)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.945	0.940
X	30	Normal (25, 7.5)	Lognormal with CV_i= 0.1 (constant)	0.947	0.961
XI	30	Normal (25, 7.5)	Lognormal with CV_i = 0.3 (constant)	0.947	0.908
XII	30	Uniform (0, 100)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.926	0.930
XIII	30	Constant (d_i = 10)	Lognormal with CV_i = Uniform (0.1, 0.3)	0.929	0.942
			Average	0.890	0.904

Note: Parameters are mean and standard deviation for Lognormal and Normal, and lower and upper bound for Uniform.

List of Figure captions

Fig. 1: 4,100 artificial projects' Planned Duration (*PD*) probability percentiles regression plots assuming lognormal (a), normal (b), uniform (c), and beta (d) activity durations.

Fig. 2: 4,100 artificial projects' average Real Duration (*RD*) / Planned Duration (*PD*) regression plots assuming lognormal (a), normal (b), uniform (c), and beta (d) activity durations.

Fig. 3: 108 empirical projects' Real Duration (*RD*) / Planned Duration (*PD*) regression plot.

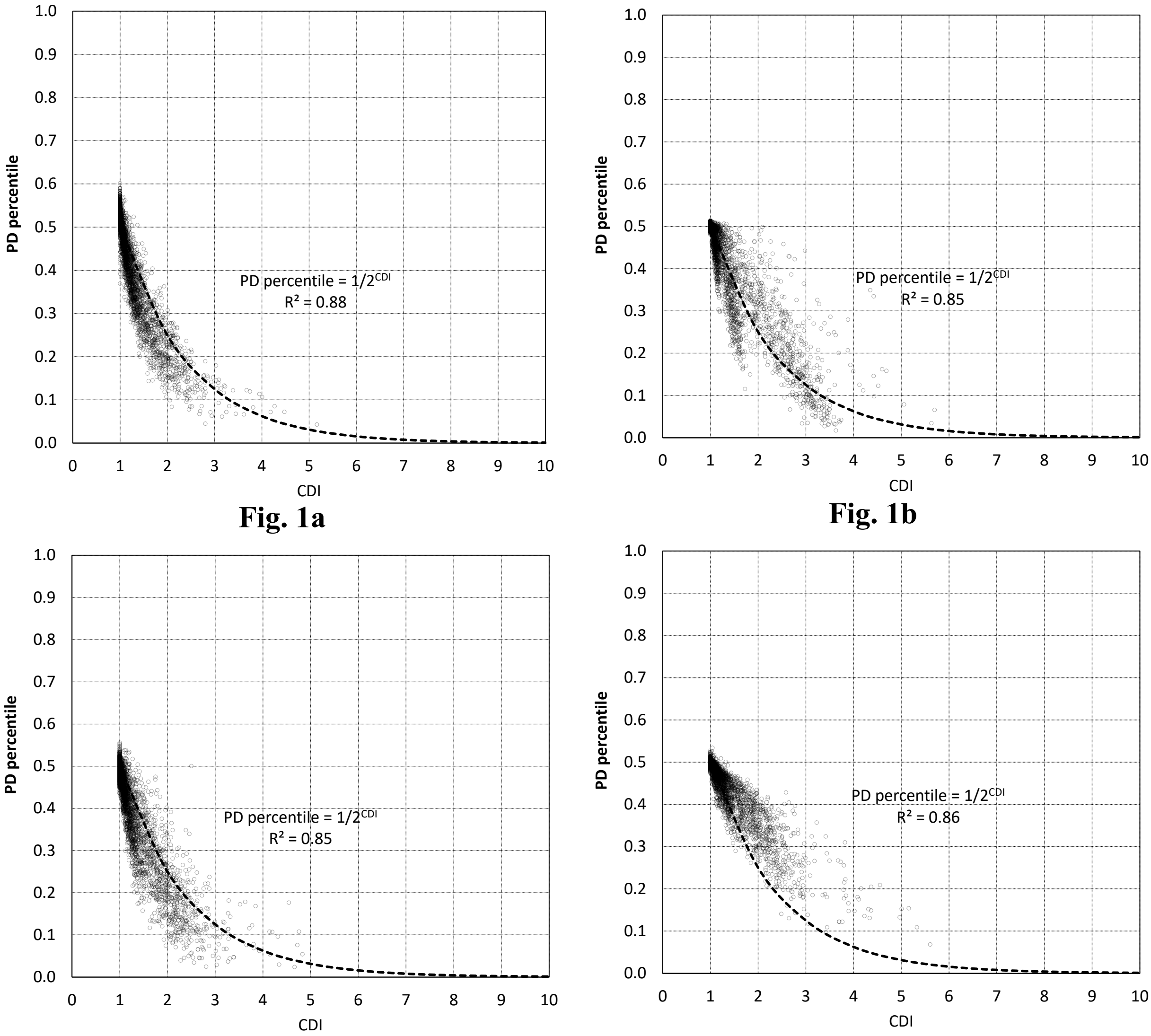

Fig. 1a

Fig. 1b

Fig. 1c

Fig. 1d

Fig. 1: 4,100 artificial projects' Planned Duration (*PD*) probability percentiles regression plots assuming lognormal (a), normal (b), uniform (c), and beta (d) activity durations.

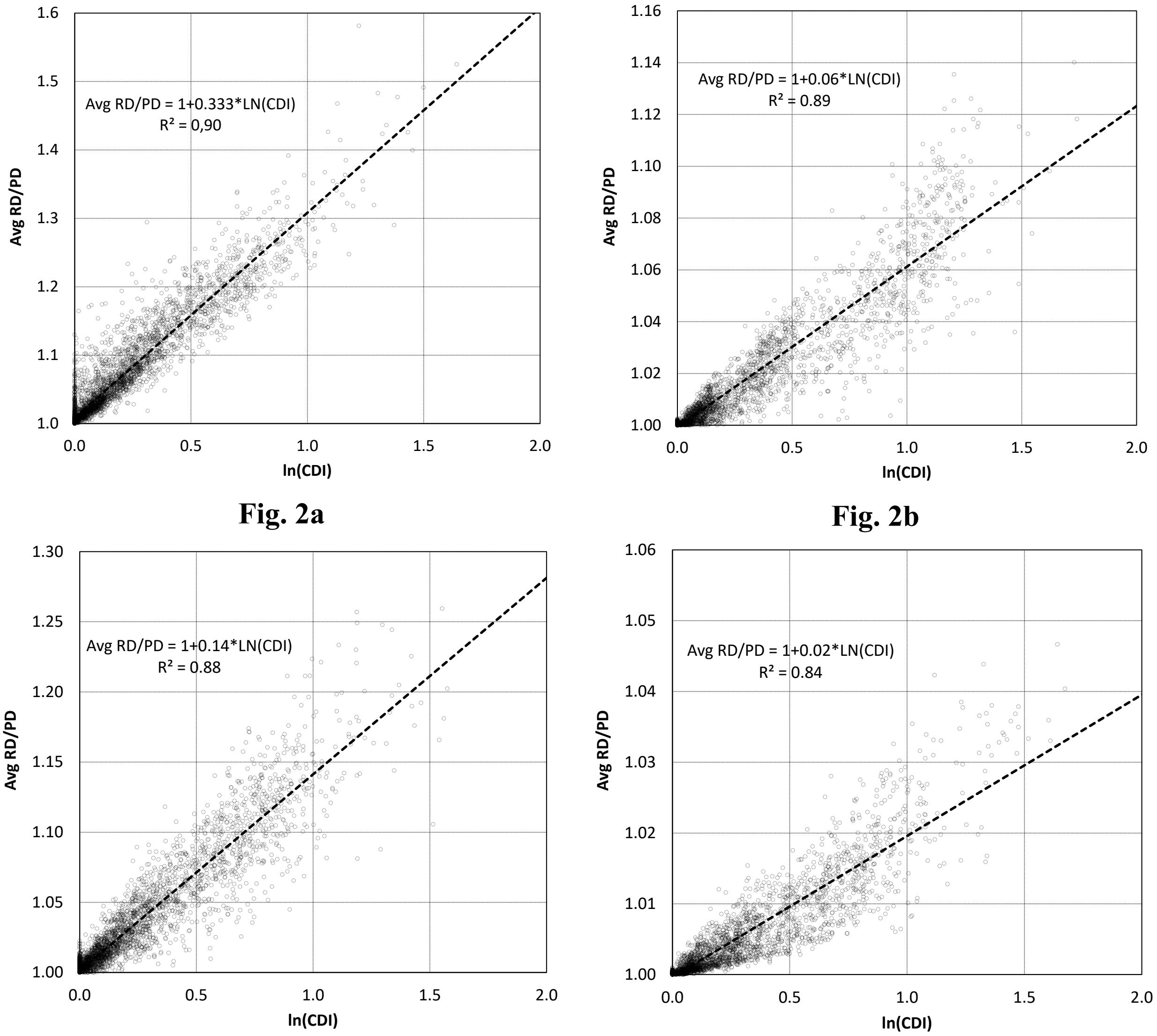

Fig. 2a **Fig. 2b**

Fig. 2c **Fig. 2d**

Fig. 2: 4,100 artificial projects' average Real Duration (*RD*) / Planned Duration (*PD*) regression plots assuming lognormal (a), normal (b), uniform (c), and beta (d) activity durations.

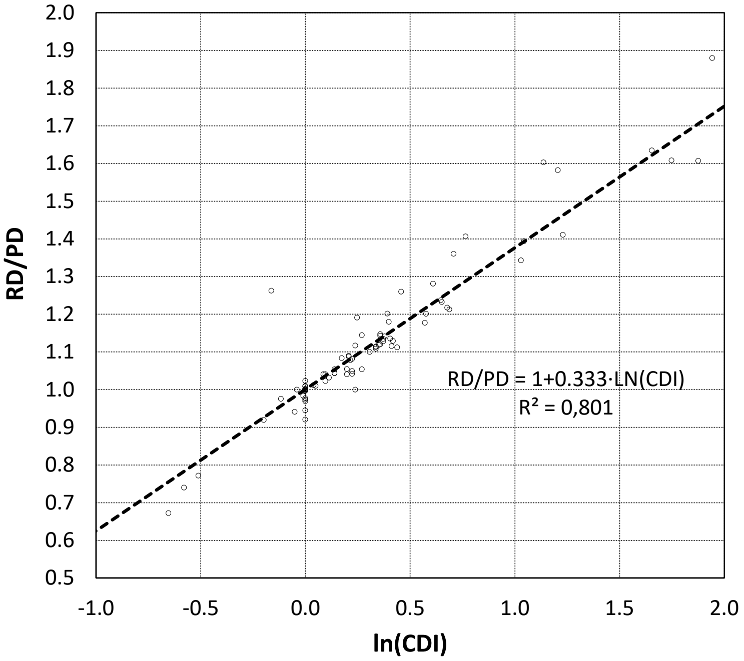

Fig. 3: 108 empirical projects' Real Duration (*RD*) / Planned Duration (*PD*) regression plot.