O P E R A T I O N S   R E S E A R C H   A N D   D E C I S I O N S

No. 4 2012
DOI: 10.5277/ord120405

Dorota KUCHTA*

APPLICATION OF FUZZY NUMBERS TO THE ESTIMATION OF AN ONGOING PROJECT'S COMPLETION TIME

A dynamic, interactive approach to the control of a project's realization time is proposed. The duration time of the project's activities is assumed to depend on certain factors whose influence may change in time. Based on the project's history up to a certain moment, the change in influence of these factors has been evaluated and estimates of the duration of activities which have not been started yet are updated. The estimates of the duration time of the activities and project are expressed in the form of fuzzy numbers. This allows us to keep a constant track of the risk of the project not keeping the deadline and to be aware of which factors influence delays, thus where to act in order to minimize the final delay while it is not yet too late – in the course of the project's realization, as early as possible.

Keywords: *estimation of a project's completion time, project control, risk of project delay, fuzzy duration of an activity*

1. Introduction

Project delays can be observed everywhere. Although a lot of scientific papers deal with project scheduling and control of a schedule, practice shows (e.g. [12]) that such methods are often not efficient because a large portion of real world projects are delayed. Thus there is a need for new approaches and methods. In the recent literature, many authors concentrate on the identification of factors which cause delays and other problems in realization of projects (cf. [3–5, 7, 9–11, 14, 15]). Most of these approaches focus on specific professions and project types, like construction or IT projects. Moreover, they limit themselves to the identification, before beginning a project, of a simple list of factors which might cause delay. However, there is no approach

*Institute of Organization and Management, Wrocław University of Technology, ul. Smoluchowskiego 25, 50-372 Wrocław, e-mail: dorota.kuchta@pwr.wroc.pl

which would allow us to control the set and influence of these factors during the project's realization. The project's situation, environment, circumstances change while it is being realized and it is not enough to specify potential delay causing factors and their influence at the beginning of the project.

A rudimentary approach allowing us to control the changing influence of the project's environment and features on its completion time during the project's realization was proposed by Cho [3], however it has several drawbacks. First of all, it is based on advanced probability theory, which has only a small chance of being applied in the practice of project management because it requires verifying hypotheses about the probability distributions of duration times of activities. It is not possible to verify such hypotheses for projects which comprise activities only rarely or never performed in the past. Approaches based on fuzzy numbers do not require verification of the fulfillment of such strict conditions as in probability based approaches [7]. Apart from that, the approach of Cho uses complicated formulae for conditional probabilities, which makes it much less suitable in practice than the fuzzy approach proposed here, which, even if it uses some "mathematics", can be presented to the actual users in a rather friendly form of tables containing crisp numbers and described in words. Secondly, Cho [3] takes into account only one single factor per project. Thus the author assumes that there is just one type of problem, i.e. only weather or only technical problems, etc. which may cause a delay in the project. This is unrealistic, even in the smallest project we can usually detect several factors which may cause delays (cf. [8]).

That is why we propose here an approach based on fuzzy numbers and allowing several criteria to influence, in a manner which varies during a project's realization, the actual duration of a project's activities. As no other paper known to the author apart from the one mentioned [3] deals with the estimation of a project's completion time during realization with respect to the changing influence of factors, in our opinion the proposed approach constitutes a substantial novelty in research on project management.

The paper is organized as follows:

In Section 1, we present necessary information about fuzzy numbers. We limit ourselves to triangular fuzzy numbers, although this is not a limiting assumption –generalization would be straightforward.

In Section 2, we explain the way we model projects, their activities and their parameters. Here the limitations of our approach become visible: the way in which we introduce the parameters and the way in which we model their influence on the project's duration. Our approach is only an attempt to generalize the approach of Cho [3] and at the same time to make it more accessible to practitioners, thus it is certainly still far from being perfect.

In Section 3, we present the algorithm for updating information about the estimated duration of individual activities and the risk of the project not keeping the scheduled completion time. The algorithm is based on the model from Section 2.

In Section 4, we illustrate the proposed approach with a very simple example. The paper ends with several conclusions.

2. Basic information about fuzzy numbers

To model uncertainty, we will use triangular fuzzy numbers $\tilde{z} = (z_1, z_2, z_3)$ with the membership function [1]:

$$\mu_z(x) = \begin{cases} 0 \text{ for } x \le z_1 - z_2 \\ \dfrac{x}{z_2} + 1 + \dfrac{z_1}{z_3} \text{ for } z_1 \ge x > z_1 - z_2 \\ 1 \text{ for } x = z_1 \\ -\dfrac{x}{z_3} + 1 + \dfrac{z_1}{z_3} \text{ for } z_1 \le x \le z_1 + z_3 \\ 0 \text{ for } x \ge z_1 + z_3 \end{cases} \tag{1}$$

The membership function $\mu_z(x)$ represents, for each real number x, the possibility that x will be the actual value of an unknown magnitude z. The value z_1 will be called the mean value of $\tilde{z} = (z_1, z_2, z_3)$. The values z_2, z_3 measure the uncertainty linked to the assumption that the unknown magnitude z will be equal to z_1. We consider here a special case, where the membership function has the form of a triangle but a generalization of the proposed approach to other types of fuzzy number (described e.g. in [1] and [6]) would be purely technical. The membership function (1) for $z_1 = 5$, $z_2 = 1$, $z_3 = 2$ has been shown in Fig. 1.

The yet unknown magnitude represented by the fuzzy number in Fig. 1 will be most likely around 5. Values smaller than 4 and greater than 7 are considered to be impossible.

The fuzzy number $\tilde{z}$ models an uncertain, only partially known quantity z. In our case, these unknown quantities will be the durations of project's activities or of the whole project, estimated before the activity or the project has been finished.

Of course, there is always a problem of how to determine $\mu_z(x)$. The idea is that $\mu_z(x)$ should represent, for each x, the possibility of x being the actual value of z. There exist several methods to generate $\mu_z(x)$ in cooperation with the decision maker [13]. However, the most commonly used method is a subjective, intuitive construction of the fuzzy estimate of z. Fuzzy numbers have the nice property of having to fulfil less rigid conditions than probability distributions, thus they are easier to construct intuitively than probability distributions [6].

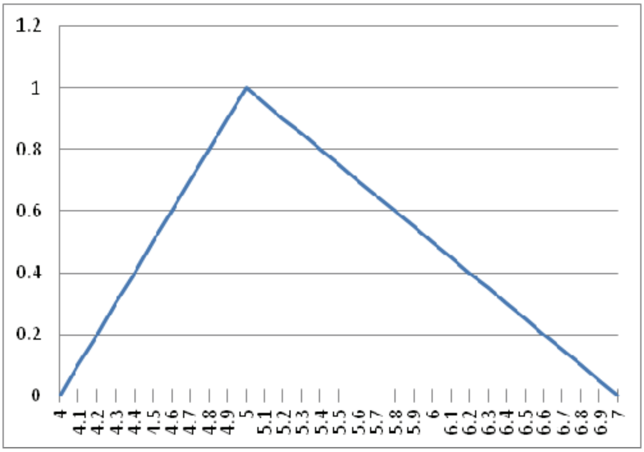

Fig. 1. Membership function of the triangular fuzzy number (5, 1, 2)

It is important to add that [6]:

• Addition of triangular fuzzy numbers can be defined as follows: $(z_1, z_2, z_3) + (w_1, w_2, w_3) = (z_1 + w_1, z_2 + w_2, z_3 + w_3)$.

• A crisp number r can be treated as a special case of a triangular fuzzy number, i.e. $(r, 0, 0)$.

If an uncertain magnitude z occurs repeatedly, it is possible to use the learning process to update its fuzzy estimate, as the estimate may depend on time – circumstances and conditions may change along with the possibility of various x's to become the actual value of z. Thus we can make the estimate $\tilde{z}$ dependent on time and use information from previous moments to better estimate the magnitude z in the future. In such cases, we will consider the fuzzy function $\widetilde{z(t)} = \left(z_1(t), z_2(t), z_3(t)\right)$ with a membership function $\mu(x; t)$, referring to the present moment t which precedes the moment of the actual occurrence of z, when its value is known exactly. $\widetilde{z(t)}$ will stand for the estimate of the actual (still unknown) value of the magnitude z according to our knowledge at moment t. At another moment t the estimate of z may differ, as our knowledge of the situation or circumstances themselves may change.

Let us introduce yet another notion: $\widetilde{z^{\text{post}}(t)}$ for t representing the moment at which the magnitude z actually occurred and its exact value became known. $\widetilde{z^{\text{post}}(t)}$ would stand for a posteriori estimation of z knowing what the situation (the actual realization of z and the circumstances we had at moment t) at a past moment t was like. $\widetilde{z^{\text{post}}(t)} = z_1^{\text{post}}(t),\ z_2^{\text{post}}(t),\ z_3^{\text{post}}(t)$ would be characterized by the membership function $\mu_z^{\text{post}}(x; t)$.

A natural question may be asked at this moment: why use a fuzzy number to represent the possible values of a magnitude at the moment that this magnitude became exactly known? A posteriori fuzzy representation of z makes sense, as even if at a past moment z had already taken a specific value, this does not mean that this specific value at this past moment was certain and completely determined. We assume that at the moment the magnitude z actually occurred, there existed a set of values it might have taken, some of them were more possible, some less. This set of values is represented by $\widetilde{z^{\text{post}}(t)}$. One of these values was actually taken, and it is important to know to which degree this value was possible and whether it constituted a positive or a negative deviation, in order to use this information for the estimation of other, yet unknown magnitudes. Actually, the same idea is used in the previously mentioned paper of Cho, where a posteriori probability distributions are used. The way $\widetilde{z^{\text{post}}(t)}$ can be generated will be now described and illustrated with an example:

Let us consider z representing the actual duration time of a project activity and let us assume that we know that at a past moment t_1 the activity was finished and the actual realization of z became known: it was the crisp number x_1. $\widetilde{z^{\text{post}}(t_1)}$ can be generated by asking the decision maker three questions:

- Given the circumstances at moment t_1 was realization x_1 smaller, greater or equal to the mean value? The answer to this question will tell us on which side of the mean value the actual value x_1 of z was.
- Were we lucky (unlucky) to get value x_1 or was this value unsurprising? To what degree was this value possible? The respective possibility degree will be denoted as $\nu(z, t_1) \in [0, 1]$ and it can be assumed to correspond to $\mu_z^{\text{post}}(x_1; t_1)$.
- Given the circumstances at moment t_1, how big was the variability of z on both sides of the mean value? For instance given the mood in the project team at moment t_1, was it possible to say that only a small range of possible values of the task duration existed or was the range of those potential values rather wide because the situation in the project team was so unusual that almost anything could have happened? This factor will be expressed by two numbers denoted as $h^l(z, t_1)$ and $h^u(z, t_1)$, which should be understood respectively as $z_2^{\text{post}}(t)$ and $z_3^{\text{post}}(t)$.

Having the above information, the membership function $\mu_z^{\text{post}}(x_1; t_1)$. can be unequivocally determined. We have:

$$\widetilde{z^{\text{post}}(t_1)} = \left(z_1^{\text{post}}(t_1), z_2^{\text{post}}(t_1), z_3^{\text{post}}(t_1)\right) \tag{2}$$

where:

- $z_2^{\text{post}}(t) = h^l(z, t_1); \qquad z_3^{\text{post}}(t) = h^u(z, t_1),$

• if x_1 was judged to be above the mean value of $\widetilde{z^{\text{post}}(t_1)}$, $z_1^{\text{post}}(t_1) = x_1 - (1 - \nu(z, t_1))$ $h^u(z, t_1)$,

• if x_1 was judged to be below the mean value of $\widetilde{z^{\text{post}}(t_1)}$, $z_1^{\text{post}}(t_1) = x_1 + (1-\nu(z, t_1))$ $h^l(z, t_1)$,

• if x_1 was judged to be equal to the mean value, $z_1^{\text{post}}(t_1) = x_1$.

Let us consider a situation where the actual duration of an activity became fully known at moment t_1 and turned out to be 8. The decision maker said that this value was a very positive surprise, thus was smaller than the mean value and had the possibility degree 0.2. On top of that, he said that the maximum variability of this value left of the mean value was equal to 5 and right of the mean value equal to 4. We have thus: $\nu(z, t_1) = 0.2$, $h^l(z, t_1) = 5$ and $h^u(z, t_1) = 4$, and we know that $8 < z_1^{\text{post}}(t_1)$. Using Equation (2), we get $\widetilde{z^{\text{post}}(t_1)} = (12, 5, 4)$. This number means that if we were to estimate the activity's duration in similar circumstances to those from moment t_1, this estimate would be represented by the fuzzy number (12, 5, 4), shown in Fig. 2. This number also gives valuable information about possible values of factors influencing the duration time of the activity being analyzed, but also influencing other activities of the project.

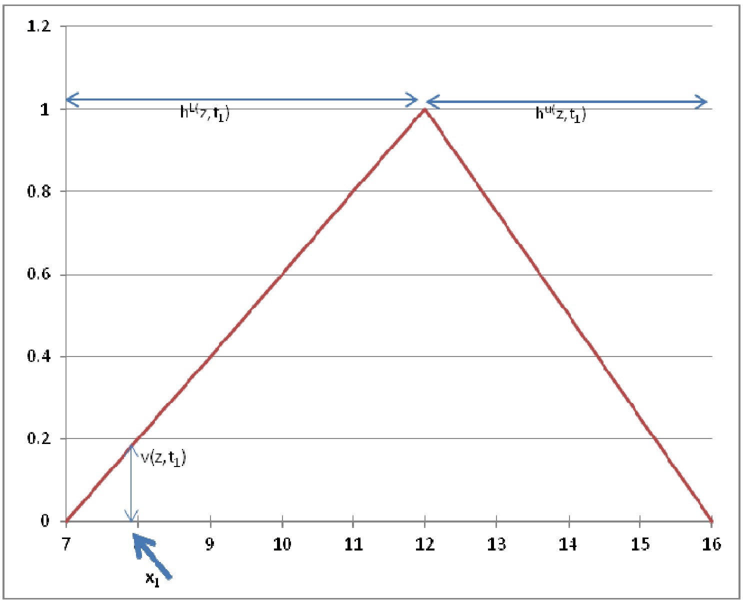

Fig. 2. A posteriori fuzzy membership function of the duration of an activity whose actual duration became known at moment t_1 and was equal to 8, based on the values $\nu(z, t_1) = 0.2$, $h^l(z, t_1) = 5$ and $h^u(z, t_1) = 4$, given by the decision maker

Let us notice that in some cases the fuzzy number $\widetilde{z^{\mathrm{post}}(t)} = \left(z_1^{\mathrm{post}}(t),\ z_2^{\mathrm{post}}(t),\ z_3^{\mathrm{post}}(t)\right)$ can be given directly, without using the value $\nu(z, t_1)$. Everything depends on the decision maker and the concrete situation.

3. Model of a project

We understand a project as a set of activities $\{A_i\}_{i=1}^{I}$ with certain dependences between them (precedence relationships, the fact of using the same resource unit, etc.). Of course, for each activity in the planning phase of the project we need to estimate its planned duration time. As mentioned before, in most projects there will be a considerable group of activities whose duration times at the planning stage will be uncertain and will depend on many factors. In the literature, the fact of the existence of a multitude of such factors, as well as their dynamic nature (the influence of various factors on the project may change in time), is rarely taken into account. In this paper we make an attempt to remove these lacunae.

We assume that the mean value of an activity's duration may depend on a number of factors. These factors may be the weather, mood in the activity team, skills of the activity team, attitude of certain stakeholders, etc. Most of these factors are hardly measurable and yet they may strongly influence the activity times. Let us define the list of these factors for the whole project as $\{\alpha_s\}_{s=1}^{S}$. The impact of these factors on the estimate of the mean values of the durations of project activities at each moment of the project $t \in [0, T]$ (where T is the time horizon, beyond which the execution of the project will certainly not go, 0 stands for the planning phase of the project), will be denoted as $\alpha_s(t)$, $s = 1, ..., S$. The factors α_s, $s = 1, ..., S$, or rather their impact on the mean values of activities' durations, are thus assumed to be quantitatively characterizable, although in an imprecise way. In [3] only one factor is introduced for all the project activities. Here we allow an arbitrary number of factors per project, chosen by the decision maker.

Apart from the factors influencing the mean duration time of activities, we consider factors influencing the uncertainty (variability) in the estimation of an activity's duration. These factors may partially correspond to the factors $\{\alpha_s\}_{s=1}^{S}$ but may also be quite different. Let us denote them as $\{\beta_r\}_{r=1}^{R}$. The crisp numbers $\beta_r(t)$, $r = 1, ..., R$ will represent the impact of the corresponding factors on the uncertainty of the estimates of the duration of the activities of the project at a given moment t.

The distinction between the factors α_s, $s = 1, ..., S$, influencing the mean value of the activities' duration, and factors β_r, $r = 1, ..., R$, influencing the variability of estimation around the mean value, is important. The first group are decisive for activity duration and determine the most possible value of the estimate, the second group decide about the variability of the estimate around the mean value. For example, in construction projects the weather is decisive for the duration of some activities and may change the mean value of the estimate completely (forecasts of long rainy periods or long sunny periods often give quite different mean values of the estimates) but other factors, like technical problems or the experience of the team members, are not so decisive, they simply mean that we cannot be completely sure about the mean value and have to take into account its variability in both directions.

Evaluations of the impact of individual factors on the estimation of the duration of project activities at each moment of time have to be dependent on time. Many factors change with time or we learn about their influence on the project only with time, e.g. weather may be good in some periods of a project's realization and worse in others, but also such not easily measurable values as the mood in the project team, the experience of the project team, the quality of relations with the customer change in time, sometimes drastically. It is important to emphasize that many of these factors, thus many of the values $\alpha_s(t)$, $s = 1, ..., S$, and $\beta_r(t)$, $r = 1, ..., R$, will not be easily measurable, thus the aim of our method will be to help the decision maker to get to know them – as soon and as accurately as possible – during the project's realization.

We have assumed that the values $\alpha_s(t)$, $s = 1, ..., S$, and $\beta_r(t)$, $r = 1, ..., R$ refer globally to the whole project. But it is obvious that their influence on the estimate of the duration of each project activity may be different, e.g. for some activities the mean value of the estimate will be strongly influenced by the weather and the variability of this estimate by the reliability of machines and for other activities the mean value will depend on the quality of customer cooperation and the variability on relations within the project team. That is why in our model we introduce a triangular fuzzy number $\tilde{d}_i(t) = \left(d_i^1(t), d_i^2(t), d_i^3(t)\right)$ standing for the estimate of the duration of activity A_i at a moment t before this activity has been finished, where d_i^1 is a function of one of α_s, $s = 1, ..., S$, and d_i^2 and d_i^3 functions of one of $\beta_r(t)$, $r = 1, ..., R$. Thus we have:

$$\tilde{d}_i(t) = \left(f_i^1\left(\alpha_{si}(t)\right), f_i^2\left(\beta_{pi}(t)\right), f_i^3\left(\beta_{ri}(t)\right)\right), \qquad i = 1, ..., I \tag{3}$$

where f_i^1, f_i^2, f_i^3 are invertible functions from and into the set of nonnegative real numbers, s_i is one of the indices $s = 1, ..., S$ and p_i and r_i are selected indices (they may be identical) from the set $r = 1, ..., R$.

We assume that the mean value of the estimate, as well as both its variability measures depend each on exactly one parameter. This is a limitation which in theory can be overcome very easily but in practice the extended solution may lead to a too complicated model. We think that in most cases one major factor can be selected – one for the mean value and one for each of the two variability measures. We underline here that with respect to Cho [3] our model constitutes a significant step towards modeling reality, as we assume that the duration of each activity depends altogether on three factors which may be specific to this activity, and not on one single factor common to all the project activities.

Equation (3) shows that we assume that before activities are started, their duration times are considered as uncertain, the mean (most possible) duration time for each activity depends on some factor from the set $\{\alpha\}_{s=1}^{S}$ (potentially different for each activity) and the uncertainty of the estimate depends on one or two elements (potentially different for each activity) from the set of factors $\{\beta_r\}_{r=1}^{R}$. Moreover, the estimated influence of both groups of factors may change in time.

One important point is that the above notation introduces a structure into the set of activities. First of all, we can (and should) consider subsets of all the activities consisting entirely of activities whose estimates of the most possible (mean) duration depend on the same parameter α_s, $s = 1, ..., S$. Let us denote the set of activities whose estimate of the most possible duration depends on α_s as $\mathbb{Q}_s$. We thus have $\bigcup_{s=1}^{S}\mathbb{Q}_s = \{A_i\}_{i=1}^{I}$ and $\mathbb{Q}_{s_1} \cap \mathbb{Q}_{s_2} = \varnothing$ for each s_1, $s_2 = 1, ..., S$, $s_1 \neq s_2$. Once one of the activities from $\mathbb{Q}_s$ for a selected s has been finished, we will try to make conclusions about the impact of the parameter α_s on the estimate of the most possible duration of those activities from the set $\mathbb{Q}_s$ which have not been started yet.

Secondly, for each parameter $\{\beta_r\}_{r=1}^{R}$. we can define the subset $\wp_r$ of the set of all the activities $\{A_i\}_{i=1}^{I}$ whose uncertainty depends on the parameter β_r. It is true that $\bigcup_{r=1}^{R}\wp_r = \{A_i\}_{i=1}^{I}$. However, we may have $\wp_{r_1} \cap \wp_{r_2} \neq \varnothing$ for any r_1, $r_2 = 1, ..., R$, $r_1 \neq r_2$, as the uncertainty of an estimate may depend on two different parameters – one for the left hand side and one for the right hand side. Again, if an activity from the set $\wp_r$ for a selected r has been finished, we can use the information about the uncertainty of the activity's duration for those members of $\wp_r$ which have not been started yet.

4. Updating the estimates of the duration times of activities and the project completion time

In this section, we will present the proposed algorithm for updating the estimates of the duration of activities which have not been started yet and as a consequence, updating the estimate of the total duration time of the project at selected control moments $\{t_j\}_{j=0}^{J}$ such that $t_j \geq 0$, $t_J \leq T$ and t_J is smaller than the actual moment of project completion, $t_{j-1} < t_j$ for $j = 2, ..., J$. We assume that at each moment t_j all the activities are either finished or not started. The latter assumption can be weakened, we introduce it for the sake of simplicity. We consider moments t_j to be kinds of milestones, as done by Cho. The control moments $\{t_j\}_{j=0}^{J}$ are chosen accordingly to the nature of the project – the more risky the project and the more serious the consequences of its delay, the smaller the intervals $[t_{j-1}, t_j]$ for $j = 1, ..., J$ should be.

For a selected t_j, $j \geq 0$ let us denote by $\tilde{C}(t_j)$ the estimated total completion time of the project at moment t_j. $\tilde{C}(t_j)$ will be defined as the maximum of the estimated lengths of all the paths in the project network, using the actual completion time of the activities which have been finished at moment t_j and using the estimated duration time of the unfinished activities in the form of fuzzy numbers derived from Eq. (3):

$$\tilde{d}_i(t_j) = \left(f_i^1\left(\alpha_{si}\left(t_j\right)\right), f_i^2\left(\beta_{pi}\left(t_j\right)\right), f_i^3\left(\beta_{ri}\left(t_j\right)\right)\right), \qquad i = 1, ..., I, \ j = 1, ..., J \tag{4}$$

The sums of fuzzy numbers considered (the estimated lengths of various paths, which are triangular fuzzy numbers), are not directly comparable, thus the maximum of them is not uniquely defined. However, this is a problem from another domain richly discussed in the literature: that of comparing fuzzy numbers (e.g. [4]). This problem of determining the maximum of several fuzzy numbers will not be discussed here in detail. We simply assume that the decision maker has decided how to choose the longest path in the project network and knows the fuzzy estimate $\tilde{C}(t_j)$ of the project duration. Analyzing it, he may judge its mean value and its variability, deciding whether he can accept it or not. In the latter case, concrete steps have to be taken in order to try to decrease the estimate of the project duration – such steps will usually be possible, as t_j is a moment at which not all the project activities have been finished yet.

In order to get a reliable and informative estimate $\tilde{C}(t_j)$ at each control moment t_j, it is necessary to have the best possible estimates $\tilde{d}_i(t_j)$ of the durations of those activities A_i which have not been finished up to moment t_j. The proposed algorithm for updating the estimates $\tilde{d}_i(t_{j+1})$ is as follows (we assume that moment t_j has already occurred and we know the actual durations of all the activities finished up to this moment) :

Step 1: For all the activities A_i that are not finished at moment t_{j+1} set $\tilde{d}_i(t_{j+1})$ = $\tilde{d}_i(t_j)$. This is the first step, whose aim is to make sure that all the unfinished activities will be assigned a value $\tilde{d}_i(t_{j+1})$. However, some of the values will be updated further in the algorithm (Step 3 and Step 4);

Step 2: Let us denote by S_{j+1} the set of those activities A_i which are finished at moment t_{j+1} but were not finished at moment t_j. Find out the actual duration of each activity A_i belonging to S_{j+1} and denote it as d_i^a. Then ask the executors of the respective activities the following questions:

- To what extent was the value d_i^a possible at moment t_{j+1}, given the state of the factors α_{si}, α_{pi}, α_{ri} (given the weather, the mood in the team, the state of the machines etc. that prevailed while executing activity A_i)? In other words, given the conditions the executors of the activity had, was the activity realized in a time that was normal or rather less normal, rather, very or not at all surprising? The answer – the possibility degree of the duration d_i^a at moment t_{j+1} – must be a number from the interval [0, 1]. Let us denote it as $\nu(d_i, t_{j+1})$.
- If d_i^a was not the most possible (mean) value of the duration of A_i at moment t_{j+1}, was it smaller or greater than the most possible value, given the state of the respective factors at moment t_{j+1}?
- What was the variability (to the left and to the right hand side of the mean value) of the estimate of the duration of activity A_i at moment t_{j+1}? Denote the respective values as $h^l(d_i, t_{j+1})$ and $h^u(d_i, t_{j+1})$. The interpretation of these values can be expressed as follows: how much was it possible at moment t_{j+1} to over- or underestimate the duration of activity A_i? These values can be given in first approximation as percentage deviations from the (yet unknown) mean value. Then using an iterative process, the final absolute values $h^l(d_i, t_{j+1})$ and $h^u(d_i, t_{j+1})$ can be found.
- Based on the values d_i^a, $\nu(d_i, t_{j+1})$, $h^l(d_i, t_{j+1})$ and $h^u(d_i, t_{j+1})$ find (as in Eq. (2)) the post factum fuzzy number

$$\tilde{d}_i^{\text{post}}\left(t_{j+1}\right) = d_i^{\text{post, 1}}\left(t_{j+1}\right), d_i^{\text{post, 2}}\left(t_{j+1}\right), d_i^{\text{post, 3}}\left(t_{j+1}\right)$$

From Eq. (3), we can then determine (for $A_i \in S_{j+1}$):

$$\begin{aligned}
\alpha_{si}\left(t_{j+1}\right) &= \left(f_i^1\right)^{-1}\left(d_i^{\text{post, 1}}\left(t_{j+1}\right)\right) \\
\beta_{pi}\left(t_{j+1}\right) &= \left(f_i^2\right)^{-1}\left(d_i^{\text{post, 2}}\left(t_{j+1}\right)\right) \\
\beta_{ri}\left(t_{j+1}\right) &= \left(f_i^3\right)^{-1}\left(d_i^{\text{post, 3}}\left(t_{j+1}\right)\right)
\end{aligned} \tag{5}$$

Step 3: For each A_i belonging to S_{j+1}, and for all unfinished A_k such that $A_k \in \mathbb{Q}_{si}$, thus for such activities whose mean value of the estimate depends on the same factor as the mean value of the finished activity A_i, set $d_j^k\left(t_{j+1}\right) = f_k^1(\alpha_{si}(t_{j+1}))$ (Eq. (5)). Here we update the information about the factor influencing the mean values of the estimates on the basis of the information about finished activities, e.g. if an activity was finished at moment t_{j+1} and its mean value depended on the mood in the project team and this mood turned out to be good at moment t_{j+1}, this should affect the estimates of the durations of all the unfinished activities whose mean value depends on the mood in the project team.

Step 4: For each A_i belonging to S_{j+1}, and for all unfinished A_k such that $A_k \in \wp_{pi} \cup \wp_{ri}$, update $d_k^2\left(t_{j+1}\right)$ and/or $d_k^3\left(t_{j+1}\right)$ based on Eq. (5) in an analogous way as in Step 3 for $d_k^1\left(t_{j+1}\right)$. Here, we update the variability of the estimates of the durations of yet unfinished activities on the basis of information from the already finished activities – if a (left hand side or right hand side) variability factor was at moment t_{j+1} greater/less than before, this should be taken into account in the estimation of the duration of yet unfinished activities whose variability depends on the same factors as the variability of activities from the set S_{j+1}.

In the following section we will present a short illustration of the proposed approach.

5. Illustrative example

Let us consider the following example: We have a construction project whose goal it is to repair a damaged section of an expressway. The project consists of 5 activities, linked by precedence relations in such a way that the first activity A_1 has to be finished, then the activity A_2 and so on – no pair of activities can be processed simultaneously.

Table 1. Activities of the illustrative project

Activity	Description of activity
A_1	delivery and installation of machines and devices
A_2	cleaning of the crack
A_3	filling the crack up with substance x
A_4	control of the repair quality
A_5	cleaning up

We assume that the most possible duration times of all the activities depend on just one factor α_1, communication within the project team, a characteristic which is hardly measurable. We assume that we will try to measure it on a scale from 0 to 100, where 100 stands for non-existent communication and 0 for excellent communication.

In the project planning phase, we also have to identify factors $\{\beta_r\}_{r=1}^{R}$, influencing the variability of the activities' duration times. In our case, it is assumed that there are five of them. Three (β_1, β_2, β_3) give us the right hand side variability of the estimate of duration (thus the possibility that the actual duration will be longer than the mean value). They are measured on a scale from 0 to 100, 100 standing for the worst possible situation and 0 for the best one. Two of them (β_4, β_5) give us the left hand side variability (the possibility that an activity will be in fact done earlier that indicated by the mean value). These factors are also measured on a scale from 0 to 100, where 0 stands for the worst situation and 100 for the best one.

Table 2. Factors $\{\beta_r\}_{r=1}^{5}$ for the illustrative project

Factor	Description of factor	Variabality type
β_1	machine and device quality	negative
β_2	soil condition	
β_3	quality of cooperation with subcontractors	
β_4	weather	positive
β_5	experience of the project team	

In Table 3, we give the necessary information for individual activities in order to apply Eqs. (3), (4). We have included there the dependences of the estimates of the activity durations on the individual factors. The columns 2–4 include the dependence functions, the last three columns contain the indices of the factors on which the estimates depend. The values in the fifth column are identical for all the activities but this is only a feature of the example discussed – potentially, each activity may have a different entry in each of the last three columns.

Table 3. Parameters for the estimation of durations of activities (Eqs. (3), (4))

Activity	$f_i^1(x)$	$f_i^2(x)$	$f_i^3(x)$	s_i	p_i	r_i
A_1	$10 + x$	$2x$	x	1	4	1
A_2	$3 + 4x$	$x/2$	$2x$	1	4	2
A_3	$4 + x^2$	$x/3$	$x/2$	1	5	3
A_4	$5 + 3x$	x	x	1	5	1
A_5	$8 + x$	$x/2$	$x/2$	1	5	2

In the planning phase of the project, thus at moment 0, we assume that we have the following values of the factors influencing the project:

Table 4. Values of the factors at $t_0 = 0$

Factor	Value
α_1	10
β_1	3
β_2	4
β_3	2
β_4	7
β_5	5

Equations (3), (4) and Tables 3, 4 give us the following initial estimates of the durations of all the activities:

Table 5. Activities of the illustrative project and the estimates of their durations at $t = 0$

Activity	$\tilde{d}_i(0)$
A_1	(20, 14, 3)
A_2	(43, 4, 8)
A_3	(104, 2, 1)
A_4	(35, 5, 3)
A_5	(18, 3, 2)

Individual parameters are rounded to the closest integer number, the unit is one day.

For example, the estimated duration of activity A_1 was calculated in the following way:

- Table 3 shows that the left hand side variability of A_1 is influenced by the factor β_4 and the right hand side variability by the factor β_1; the mean value is influenced by the factor α_1, as is the case for all the other activities:
- We thus have (Table 3 and Eq. (3)):

$$f_1^1\left(\alpha_{s_1}(0)\right) = f_1^1\left(\alpha_1(0)\right) = 20$$

$$f_1^2\left(\beta_{p_1}(0)\right) = f_1^2\left(\beta_4(0)\right) = 14$$

$$f_1^3\left(\beta_{r_1}(0)\right) = f_1^3\left(\beta_1(0)\right) = 3$$

We can now estimate the realization time for the whole project according to our knowledge at moment 0: $\tilde{C}(0)$ will be equal to the sum of the fuzzy numbers from Table 5, thus, rounded, to (220, 27, 17). The variability of the estimate of the project's duration does not seem to be very big. Hence the risk of exceeding the mean value, 220, seems to be rather low.

Let us now assume that activity A_1 was finished at $t_1 = 30$. Its actual duration was 30 days and the decision maker said that given the state of communication in the project team and the weather, this duration was really a complete, very positive surprise, almost a miracle. The decision maker assumed that the possibility degree of the duration of 30 days was actually 0, the left hand side variability (with respect to the mean value) was 30 and the right hand side variability 20. Thus we have:

$$d_1^{\text{post}}(30) = \left(d_1^{\text{post},1}(30),\ d_1^{\text{post},2}(30), d_1^{\text{post},3}(30)\right) = (60, 30, 20) \tag{6}$$

Using Equation (5), we get the updated (for $t_1 = 30$) values of some of the factors – those which influenced the finished activity A_1:

$$\begin{aligned} \alpha_1(30) &= \left(f_1^1\right)^{-1}(60) = 50 \\ \beta_4(30) &= \left(f_1^2\right)^{-1}(30) = 15 \\ \beta_1(30) &= \left(f_1^3\right)^{-1}(20) = 20 \end{aligned} \tag{7}$$

As the mean values of all the unfinished activities are influenced by the factor α_1, the left hand side variability of A_2 is influenced by β_4 and the right hand side variability of A_4 is influenced by β_1, at $t_1 = 30$, we get the following updates of the estimates of the duration of the unfinished activities (the values in bold have been determined based on Eqs. (6) and (7)):

Table 6. Activities of the illustrative project and the estimates of their durations at $t_1 = 30$

Activity	$\tilde{d}_i(30)$
A_1	30 (actual value)
A_2	(**203**, **8**, 8)
A_3	(**2504**, 2, 1)
A_4	(**155**, 5, **30**)
A_5	(**58**, 3, 2)

Individual parameters are rounded to the closest integer number.

And the updated estimate of the duration of the whole project is $\tilde{C}(30) = (2980, 48, 51)$.

The difference between the mean value of $\tilde{C}(0)$ and $\tilde{C}(30)$ is very big. This is because at the moment 0 we did not know exactly how the communication in the project was. We estimated it with the score 10 but as the project continued, it turned out to be much worse, 50 (or maybe it worsened between moments 0 and 30). Also, two variability factors turned out to be higher than we had supposed, but in this case, the most substantial problem concerns the mean value.

It should be pointed out that $\tilde{C}(30)$ = (2980, 48, 51) is only an estimate of the total duration of the project, carried out on the 30th day of the realization of a project whose duration was initially estimated as (220, 27, 17). Thus, there is still a lot of time till the initially planned final moment of the project and the value $\tilde{C}(30)$ = (2980, 48, 51), representing our knowledge on the 30th day about the possible total duration of the project, should be seen only as an early warning. There is time to make various decisions about the project so that its actual duration will be acceptable. Also, each control moment offers us the possibility to verify the information from Table 3.

The same procedure would then have to be repeated at following control moments.

6. Conclusions

A method of estimating a project's realization time at points during the project's realization has been proposed. The method makes an attempt to take into account all the factors influencing the duration of various activities and learn about their impact on the basis of what happens in the project and it environment during its realization. This approach uses fuzzy estimates and an updating algorithm which has to be applied at important milestones of the project.

The proposed approach has, to our knowledge, no equivalent in the existing literature with respect to its main feature: consideration of the changing influence of various factors on the activities' durations during the realization of the project. We see it as a proposal for further research, as the proposed approach certainly has several drawbacks. First of all, the number of factors influencing an activity is still limited. Secondly, no approach is proposed for the generation of the dependence functions. We assume that they are based on intuition and knowledge however, in practice, they should be supported by a systematic method of gaining the necessary information. Thus practical verification of the approach, together with a method for interactive determination of the required dependence functions, is absolutely necessary. Also, in the proposed approach it is assumed that we can identify all the factors influencing the duration of a project activity at the beginning of the project, which in practice may be difficult or sometimes even impossible. In this respect, further research is also required.

References

[1] ALEFELD G., HERZBERGER J., *Introduction to Interval Computation*, Academic Press, New York 1983.

[2] CHEN S.J., HWANG C.L., DE HWANG F.P., *Fuzzy multiple attribute decision making: Methods and applications*, Springer-Verlag, Berlin 1992.

[3] CHO S., *A linear Bayesian stochastic approximation to update project duration estimates*, European Journal of Operational Research, 2009, 196, 585–593.

[4] GHOSH S., JINTANAPAKANONT J., *Identifying and assessing the critical risk factors in an underground rail project in Thailand: A factor analysis approach*, International Journal of Project Management, 2004, 2 (8), 633–643.

[5] HWANG B.-G., ZHAO X., GAY M.J.S., *Public private partnership projects in Singapore: Factors, critical risks and preferred risk allocation from the perspective of contractors*, International Journal of Project Management, 2013, 4, 424–433

[6] KUCHTA D., *Miękka matematyka w zarządzaniu: Zastosowanie liczb przedziałowych i rozmytych w rachunkowości zarządczej*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2001.

[7] KUCHTA D., SKORUPKA D., *Project risk management taking into consideration the influence of various risk levels based on linguistic approach*, Proceedings of the FLINS 2012 Conference, Istanbul 2012.

[8] KUCHTA D., ŚLUSARCZYK A., *Size of buffers in critical chain method – case study*, [in:] *Project Management Development – Practice and Perspectives*, Second International Scientific Conference on Project Management in the Baltic Countries, University of Latvia, Riga, 11–12.04.2013, 107–119.

[9] SKORUPKA D., *Identification and Initial Risk Assessment of Construction Projects in Poland*, Journal of Management in Engineering, ASCE, 2008, 3, 120–127.

[10] SKORUPKA D., *Method of planning construction projects, taking into account risk factors*, Operations Research and Decisions, 2009, 3, 119–128.

[11] SKORUPKA D., *Method of Construction Project Assessment*, Lambert Academic Publishing Saarbrücken 2012.

[12] Standish Group CHAOS Report, West Yarmouth, Massachusetts, 2012.

[13] TURKSEN I.B., *Measure of membership functions and their applications*, Fuzzy Sets and Systems, 1991, 4, 5–38.

[14] VICKNAYSON T., MAWDESLEY M.J., *Perception of human risk factors in construction projects: An exploratory study*, International Journal of Project Management, 2004, 22 (2), 131–137.

[15] WANG J., YUAN H., *Factors affecting contractors' risk attitudes in construction projects: Case study from China*, International Journal of Project Management, 2011, 29 (2), 209–219.