Computers & Industrial Engineering 180 (2023) 109261

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Explainable machine learning for project management control

José Ignacio Santos [a], María Pereda [b,c], Virginia Ahedo [d], José Manuel Galán [e,*]

[a] Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Avenida Cantabria S/N, 09006 Burgos, Spain
[b] Grupo de Investigación Ingeniería de Organización y Logística (IOL), Departamento Ingeniería de Organización, Administración de empresas y Estadística, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain
[c] Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain
[d] Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Avenida Cantabria S/N, 09006 Burgos, Spain
[e] Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Avenida Cantabria S/N, 09006 Burgos, Spain

A R T I C L E I N F O

Keywords:
Project management
Stochastic project control
Earned value management
Shapley values
Explainable machine learning
SHAP

A B S T R A C T

Project control is a crucial phase within project management aimed at ensuring —in an integrated manner— that the project objectives are met according to plan. Earned Value Management —along with its various refinements— is the most popular and widespread method for top-down project control. For project control under uncertainty, Monte Carlo simulation and statistical/machine learning models extend the earned value framework by allowing the analysis of deviations, expected times and costs during project progress. Recent advances in explainable machine learning, in particular attribution methods based on Shapley values, can be used to link project control to activity properties, facilitating the interpretation of interrelations between activity characteristics and control objectives. This work proposes a new methodology that adds an explainability layer based on SHAP —Shapley Additive exPlanations— to different machine learning models fitted to Monte Carlo simulations of the project network during tracking control points. Specifically, our method allows for both prospective and retrospective analyses, which have different utilities: forward analysis helps to identify key relationships between the different tasks and the desired outcomes, thus being useful to make execution/replanning decisions; and backward analysis serves to identify the causes of project status during project progress. Furthermore, this method is general, model-agnostic and provides quantifiable and easily interpretable information, hence constituting a valuable tool for project control in uncertain environments.

1. Introduction

Project control consists of monitoring project progress and performance, controlling the expected output(s), and taking the necessary corrective actions when deviations from the original plan occur. This role is the cornerstone of any project manager and is key to project success (Pellerin & Perrier, 2019).

Project control management methods typically aim to quantify project progress and predict the final outcome if no corrective actions are taken. This prediction should be made as soon as possible so that the range of corrective measures available is as wide as possible. Integrated project management and control systems generally consist of three elements: a baseline schedule, periodic progress data, and a set of analysis techniques capable of identifying potential problems —and perhaps opportunities— in the project (Vanhoucke, 2019). Based on this information, if the control system indicates any possible difficulty to meet the project objectives, the project manager, depending on the context and available options, should try to put the project back on the right track.

The best-known and most popular project control method is probably the Earned Value Management (EVM) (Anbari, 2003; Fleming & Koppelman, 2010; Vanhoucke, 2010). This method, together with its extension, the Earned Schedule (Lipke, 2003, 2004), allows for integrated cost and schedule management during project progress. Importantly, its popularity has promoted different refinements and adaptations that enrich the methodology (Song et al., 2022; Vanhoucke, 2019; Willems & Vanhoucke, 2015).

Although in the standard version of the EVM, the duration and costs of activities are considered deterministic, one of the key aspects of project management is precisely the management of uncertainty. Notably, there is still no complete consensus in the literature on the definition of project uncertainty or project risk. Within the scope of this contribution, under project uncertainty, we will consider stochastic

* Corresponding author.
E-mail address: jmgalan@ubu.es (J.M. Galán).

https://doi.org/10.1016/j.cie.2023.109261
Received 5 July 2022; Received in revised form 5 March 2023; Accepted 19 April 2023
Available online 23 April 2023

variability of task durations, named *aleatoric uncertainty* in previous works (Acebes et al., 2021; Elms, 2004; Frank, 1999; Shafer, 1976). In this sense, our conceptualisation of risk would align with Hillson's (2009) "uncertainty that matters", i.e., that affects project objectives.

Project duration/cost uncertainty arises from the intrinsic variability of task durations and the complex and non-linear interactions between tasks. To assess these complexities, the triad method (Acebes et al., 2014b, 2015) extends the EVM to stochastic contexts using two tools: (i) Monte Carlo simulation, which provides a practical resource to generate a universe of virtual possible project realisations that faithfully represent project variability; and (ii) machine learning models that take as input the data obtained from Monte Carlo simulations (specifically the time, cost and EV of each realisation) to estimate the expected project outcome conditional on its progress at the time of control.

In this paper, we present a novel method for project control with a threefold objective: (i) to estimate the project consequences when no corrective actions are taken; (ii) to pinpoint the most relevant tasks on which to take corrective actions conditional on the current project situation; and (iii) to attribute and understand the causes of the current project status. As in the triad method, our proposal uses Monte Carlo simulation and machine learning models, but, in our case, to map task features to the expected outcome. Notably, in machine learning, if we use sufficiently flexible algorithms —which act as universal approximators— the fitted model can be approximated as closely as desired to the expected stochastic plan by simply increasing the number of simulations. Notwithstanding, one of the main problems of machine learning models —especially in non-linear contexts such as project networks— is that they are subject to a trade-off between interpretability and accuracy. Typically, the type of models that adequately capture project relationships between tasks and outcome are black-box models, which are highly accurate but hardly interpretable. In the context of this paper, we consider interpretability as the degree to which a human can understand the cause of a decision (Miller, 2019). From a management perspective, interpretability is a crucial element. Understanding the relationship between performance elements and outputs is fundamental for correct decision-making.

In recent years, advances in explainable artificial intelligence (XAI) (Samek et al., 2019) —also known as informed machine learning (von Rueden et al., 2021), interpretable machine learning (Molnar, 2022; Molnar et al., 2020) or intelligible intelligence (Weld & Bansal, 2019), among other terms— have set as one of their main goals to provide insights and understanding of how models make their decisions for a human target audience (Murdoch et al., 2019; Roscher et al., 2020).

In this paper, we specifically use some of the most fundamental advances in interpretable machine learning to illustrate how they can assist project managers in making informed decisions on project progress and control. In particular, our approach revolves around the Shapley value decomposition of a model. This choice is motivated by several factors. First, it is a model-agnostic method —i.e., it can be applied to any type of machine learning model— conceived to identify the decision mechanisms that underlie the model. This feature ensures both the generality of the approach and an easy future adaptation to new developments in the fertile field of automatic regression and classification. Second, it is theoretically well-founded, as it is rooted in the concept of Shapley values from cooperative game theory (Shapley, 1953), which is characterised by its fairness and several desirable properties (efficiency, symmetry, null player, additivity). Third, within the framework of project management and, more specifically, in Program Evaluation Review Technique (PERT) networks, Shapley's rule has been shown to exhibit consistency with respect to the cost distribution of potential delays (Bergantiños et al., 2018). Interestingly, the adaptation of the Shapley framework to machine learning models —SHAP (SHapley Additive exPlanations)— allows a fair calculation of the marginal contribution of each feature value to each prediction, both in aggregate and case-level analyses (Fryer et al., 2021; Lundberg & Lee, 2017). On the one hand, case-level analyses explain each individual prediction by computing the individual contribution of each feature value to the prediction. On the other hand, aggregate analyses give insights into the global influence of each variable on the model as a whole. A disadvantage and barrier to the adoption of the SHAP method was that it is computationally expensive. Nevertheless, it has recently included new developments for tree-based and deep-learning models that significantly reduce computation times (Lundberg & Lee, 2017; Molnar, 2022).

The methodology proposed in this article allows project managers to understand the causes and consequences of deviations during project execution and provides a tool to help make corrective re-scheduling decisions. To illustrate the method, the article is organised as follows: The next section provides a literature review of recent developments related to top-down project control and XAI applications in project management. It concludes by framing the contribution of this work and its relation to previous research. The following background section explains the rationale and the context in which the methodology is rooted; to that end, a succinct introduction to the EVM, the triad method and Shapley values is presented. Then, the problem formulation in the context of project control is provided, explaining the type of analysis at which the method is aimed. Subsequently, a simple case study is presented to illustrate the application of the method and its advantages; notably, the selected case study allows exploring different scenarios and, hence, contrasting different interpretations. Finally, in the discussion section, we reflect on both the methodological and managerial implications of our approach, and, to conclude, the main conclusions of the work are outlined.

2. Literature review

2.1. Related literature

This section provides an overview of the scientific literature related to the latest developments and extensions of the Earned Value Method (EVM), a project monitoring technique widely used in both academia and industry (Aramali et al., 2021, 2022; Fleming & Koppelman, 2010; Nizam & Elshannaway, 2019). Additionally, this review covers recent applications of explainable machine learning in project management. Finally, we present and position our contribution to the scientific field within the framework of this literature review.

EVM involves several tasks, such as periodically measuring project progress, comparing the project status with the project plan, taking corrective actions if deviations exceed certain acceptable tolerance limits, and forecasting project time and cost based on current project performance information. Due to its usefulness and interest, the methodology has undergone significant developments, refinements and extensions that have adapted its use to different circumstances and improved its performance from various perspectives (Pellerin & Perrier, 2019; Willems & Vanhoucke, 2015).

To facilitate the review and comparison of the articles in the scientific literature, we have defined the eight characteristics shown in Table 1. The first characteristic identifies whether the work is a refinement, study, or analysis directly related to the EVM methodology. In this regard, note that we have also included some relevant articles on the earned duration management (EDM) methodology (Khamooshi & Golafshani, 2014), a time-based approach to project control based on the concept of earned duration rather than earned schedule (Lipke, 2003). Notwithstanding, in these cases, we have explicitly specified when an article is related to EDM rather than EVM. The second relevant aspect of the analysis is the functionalities offered by the contribution according to three categories: (i) whether the methodology includes tools that make project time/cost forecasting more sophisticated beyond the simple projection techniques of the original method; (ii) whether the authors present methodologies capable of establishing appropriate thresholds or control limits that act as warning signals to the project manager to put corrective measures in place and get the project back on track; and (iii) whether the method establishes the relationship,

Table 1
Literature review.

Authors	Control	XAI	Monte Carlo	Uncertainty	Key approach/methodology	Forecasting	Task attribution/sensitivity activity index	Tolerance limits/action
(Noori et al., 2008)	EVM			x	Fuzzy theory			x
(Pewdum et al., 2009)					ML (Artificial neural networks)	x		
(Bagherpour et al., 2010)	EVM			x	Fuzzy theory			
(Naeni et al., 2011)	EVM			x	Fuzzy theory	x		
(Moslemi Naeni & Salehipour, 2011)	EVM			x	Fuzzy theory	x		
(Ponz-Tienda et al., 2012)	EVM			x	Fuzzy theory	x		
(Acebes et al., 2013)	EVM		x	x	Risk buffer management			x
(Mortaji et al., 2013)	EVM			x	Fuzzy theory (LR numbers)	x		
(Aliverdi et al., 2013)	EVM			x	Statistical quality control charts			x
(Colin & Vanhoucke, 2014)	EVM		x	x	Quantile tolerance limits			x
(Acebes et al., 2014b)	EVM		x	x	Quantile tolerance limits			x
(Czemplik, 2014)	EVM				Schedule forecast indicator	x		
(Wauters & Vanhoucke, 2014)	EVM		x	x	ML (Support vector regression)	x		
(Moslemi Naeni et al., 2014)	EVM			x	Fuzzy theory	x		
(Salari et al., 2014)	EVM			x	Z-numbers	x		
(Colin et al., 2015)	EVM		x	x	Multivariate/PCA			x
(Colin & Vanhoucke, 2015)	EVM		x	x	Critical chain/buffer management			x
(Acebes et al., 2015)	EVM		x	x	ML (model selection)	x		x
(Willems & Vanhoucke, 2015)	EVM			x	Comparative analysis/review			
(Batselier & Vanhoucke, 2015b)	EVM		x		Comparative analysis/review	x		
(Batselier & Vanhoucke, 2015a)	EVM				Comparative analysis/review	x		
(Hu et al., 2016)	EVM			x	Critical chain/buffer management		x	x
(Wauters & Vanhoucke, 2016)	EVM		x	x	ML (model selection)	x		x
(Vanhoucke & Colin, 2016)	EVM			x	Multivariate regression			x
(Wauters & Vanhoucke, 2017)	EVM		x	x	ML (Nearest Neighbour)	x		
(Moradi et al., 2017)	EVM			x	Risk factors with interval-valued fuzzy set	x		
(Martens & Vanhoucke, 2017b)	EVM		x	x	Critical chain/buffer management			x
(Martens & Vanhoucke, 2017a)	EVM		x	x	Scarce resources			x
(Batselier & Vanhoucke, 2017)	EVM				Exponential smoothing	x		
(Khamooshi & Abdi, 2017)	EDM				Exponential smoothing	x		
(Wajdi Hammad et al., 2018)	EVM			x	Critical chain/buffer management			x
(Nadafi et al., 2019)	EVM			x	Grey Theory	x		
(Abdel Azeem et al., 2014)	EVM			x	Kalman Filter	x		
(Zohoori et al., 2019)	EVM			x	Gain Scheduling Fuzzy Control	x		
(Hadian & Rahimifard, 2019)	EVM			x	Multivariate T2 charts			x
(Cheng et al., 2019)					ML (Artificial neural networks)	x		
(Ballesteros-Pérez et al., 2019)	EVM				Comparative analysis/review	x	x	x
(Vanhoucke, 2019)	EVM				Comparative analysis/review			x
(Eshghi et al., 2019)	EVM			x	Fuzzy theory (IT2FSs)	x		
(Hendiani et al., 2020)	EVM			x	Z-numbers	x		
(Votto et al., 2020)	EVM		x	x	Multivariate T2 charts			x
(Z. Chen et al., 2020)	EVM		x	x	Bayesian analysis			x
(Martens & Vanhoucke, 2020)	EVM				Exponential smoothing	x		
(Sackey et al., 2020)	EVM				Regression analysis and smoothing (DEAC model)	x		
(Radhakrishnan & Jaurez, 2021)		x	x	x	ML (model selection)	x		
(Mahmoudi, Bagherpour, et al., 2021)	EVM			x	Grey Theory	x		
(Mahmoudi, Javed, et al., 2021)	EDM			x	Grey Theory	x		x
(Mortaji et al., 2021)	EVM			x	Fuzzy theory (DEVM)	x		x
(Acebes et al., 2022)	EDM		x	x	ML (model selection)	x		x
(Kuchta & Zabor, 2022)	EVM			x	Z-numbers	x		

(*continued on next page*)

Table 1 (*continued*)

Authors	Control	XAI	Monte Carlo	Uncertainty	Key approach/methodology	Forecasting	Task attribution/sensitivity activity index	Tolerance limits/action
(Song et al., 2022)	EVM		x	x	Resource-constrained project control			x
(Barrientos-Orellana et al., 2021)	EVM				Comparative analysis/review	x		
(Wang et al., 2022)		x			ML (Deep neural networks)	x		
(Ghorbany et al., 2022)		x			ML (CBN/XGBoost)			

dependence, or influence between project tasks and performance projections, integrated into the control system.

The original EVM assumes deterministic task durations and costs during the planning phase, which is often unrealistic in real-world projects. Therefore, many studies aim to improve the method by including uncertainty and stochasticity in project planning. In our review, we evaluate whether the methodology addresses the uncertainty problem and whether it uses Monte Carlo simulation, which is a common approach to modelling uncertainty. It is worth noting that Monte Carlo experiments can be used not only to analyse uncertainty but also as a tool to simulate the performance of deterministic methods. Additionally, we considered the methodology on which the extension or refinement is based and whether it includes an explainability layer (XAI) since the inclusion of XAI is one of the novel elements of our contribution.

The written detail of our literature review is provided below. For a general overview, please refer to Table 1.

In terms of deterministic forecasting, pioneering work on the use of machine learning methods to predict the final budget and duration of a project was carried out by Pewdum et al. (2009). They used an artificial neural network fed with data from other similar projects and compared it with the results of EVM projections. More recently, Cheng et al. (2019) used a neural network-long short-term memory to estimate the schedule to completion, achieving better results than the basic EVM and EDM methods. Batselier and Vanhoucke (2015a) used native EVM data to make predictions in deterministic contexts, comparing the accuracy of three different projection techniques and proposing a mutually combined approach. The predictive capability of several of these methodologies was empirically tested by Batselier and Vanhoucke (2015b) using a real-life project database supported by Monte Carlo simulation. Their study shows promising results of EVM not only for cost but also for time forecasting, especially in serialised projects. Forecasting indexes, such as the schedule forecast indicator (Czemplik, 2014), have also been proposed as complementary methods to evaluate the progress of critical activities during project execution.

In a later study, Batselier and Vanhoucke (2017) integrated the EVM methodology with the exponential smoothing forecasting approach, resulting in significantly more accurate predictions than the methodologies they compared with. Khamooshi and Abdi (2017) also included the exponential smoothing forecasting technique in EDM, also achieving improvements in predictive capacity. Subsequently, Martens and Vanhoucke (2020) refined the methodology by including corrective actions in the forecasting process. In the same year, Sackey et al. (2020) presented the DEAC (Duration Estimate At Completion)-model, which is based on progress in time units rather than cost; in particular, they proposed a method for estimating project completion time using time series and exponential smoothing. Recently, several studies have conducted comprehensive analyses and comparative reviews of different deterministic methods of project duration forecasting in EVM (Ballesteros-Pérez et al., 2019; Barrientos-Orellana et al., 2021). Notably, the work proposed by Ballesteros-Pérez et al. (2019) includes activity-level metrics in deterministic contexts that are useful for predictions and have the potential for activity prioritisation and resource allocation.

To explicitly address uncertainty in forecasting processes, there are two main branches in the literature. One is based on the incorporation of multi-valued logic in the EVM framework, while the other uses probability theory, typically through Monte Carlo simulation and statistical and machine learning techniques. Some of the early advances in the first branch were made by incorporating time-related uncertainty through fuzzy logic and fuzzy control charts, but, curiously, using EVM in the context of production control instead of in project management (Bagherpour et al., 2010). Subsequently, Moslemi Naeni et al. (Moslemi Naeni et al., 2014; Moslemi Naeni & Salehipour, 2011; Naeni et al., 2011) pioneered a complete fuzzy approach to EVM, allowing fuzzy indices and estimates to be obtained. Ponz-Tienda et al. (2012) incorporated all feasible schedules and the accuracy of estimated values into previous contributions, while Mortaji et al. (2013) formalised EVM in a vagueness environment using L-R fuzzy numbers. Moradi et al. (2017) proposed a new evaluation model that combines EVM with risk analysis to improve the forecasting of future project performance using linguistic variables represented by interval-valued triangular fuzzy numbers. Meanwhile, Zohoori et al. (2019) integrated EVM with gain scheduling fuzzy control to design an adaptive monitoring system to support real-time and production time control. Eshghi et al. (2019) proposed a new forecasting approach (IT2F-EVM) for project cost and schedule estimation in megaprojects based on the integration of interval type 2 fuzzy sets (IT2FSs) with EVM, considering several factors affecting project success, including quality, stakeholder satisfaction, safety, and risk. To reduce computational complexity and increase the ability to express project-specific dynamics, Mortaji et al. (2021) introduced Directed Earned Value Management (DEVM), which uses ordered fuzzy numbers to address uncertainties and captures consistent and interpretable information on the trend of project progress.

In recent years, several studies (Mahmoudi, Bagherpour, et al., 2021; Nadafi et al., 2019) have proposed integrating EVM with the grey system theory, resulting in the development of Grey Earned Value Management (EVM-G). Furthermore, this approach has also been adapted for integration with EDM (Mahmoudi, Javed, et al., 2021). According to the case studies analysed by the original authors, EVM-G has advantages over fuzzy-EVM in terms of information needs and performance.

In 2014, Salari et al. (2014) proposed integrating *Z*-numbers with EVM as a way to express uncertainty and incorporate reliability in fuzzy reasoning. This idea has recently been implemented to establish the *Z*-Earned Value Management (ZEVM) framework, which uses *Z*-numbers to improve the accuracy of cost-duration tracking in project management (Hendiani et al., 2020). In addition, Kuchta and Zabor (2022) have proposed the ZG-EVM model, which also integrates Z-numbers into EVM. However, their version is only used to model the cost that will still be incurred.

The other branch that deals with project uncertainty is focused on probability theory and involves explicitly modelling the probability distributions of task durations and costs within a project. Although probabilistic methodologies such as the Kalman filter have been proposed (Abdel Azeem et al., 2014; Kim et al., 2010), the most common approach involves the use of Monte Carlo simulations to sample possible realisations of the project within the range of its stochastic definition. The resulting data can be analysed in various ways to extract relevant project information, with machine learning models being the most commonly used tools due to their ability to capture complex interactions and nonlinear behaviours. In the context of EVM, Wauters and Vanhoucke (2014) proposed one of the earliest contributions in this line. They used Support Vector Regression models trained on intermediate earned value metrics obtained through Monte Carlo simulation to forecast project progress. In a subsequent study, the authors extended their analysis by incorporating pre-processing through Principal Component Analysis (PCA) and comparing the performance of five well-known machine learning techniques, which again yielded promising forecasting results (Wauters & Vanhoucke, 2016).

In 2015, Acebes et al. (2015) introduced the triad method, which integrates Monte Carlo simulation with the EVM. The resulting simulation data were used to build classification and regression models (supervised learning methods) to predict the achievement of time and cost targets. The study also included unsupervised anomaly detection methods to establish tolerance limits. To implement this approach, the authors proposed training various machine-learning models and using nested cross-validation for model selection. Remarkably, the triad method has been recently adapted for integration into the EDM framework (Acebes et al., 2022).

Wauters and Vanhoucke (2017) further explored the application of the Nearest Neighbour algorithm in EVM for two purposes: first, to evaluate its predictive capability, and second, to investigate its potential as a tool for hybridising with other machine learning techniques by reducing the number of training instances. Additionally, although not directly related to project control, Radhakrishnan and Jaurez (2021) proposed using Monte Carlo simulations for machine learning model selection and evaluating feature importance through impurity metrics or permutation tests in the context of estimating completion times in project management.

As noted at the beginning of this literature review, EVM and its various extensions are concerned with making predictions of time and cost to completion, but also with setting tolerance limits or warning signals that can alert project managers to potential problems and prompt corrective actions. The most basic methods for establishing these limits are the so-called rules-of-thumb, which rely on static and somewhat arbitrary values that, when exceeded, indicate potential problems in project execution. However, in the probabilistic framework of project definition, significant advances have been made in establishing statistical tolerance limits that indicate whether a project is progressing within the planned parameters.

Within this statistical approach, two types of statistical process control methods have emerged (Colin & Vanhoucke, 2014; Vanhoucke, 2019). The first, statistical process control for projects (SPC-PC), involves the establishment of limits as the project progresses. In the second approach, statistical project control using statistical tolerance limits (SPC-STL), Monte Carlo simulation is used to establish *a priori* acceptable variation parameters, i.e., confidence intervals, before any progress has been made on the project. These established limits are then monitored during the execution phase to ensure that progress occurs within the previously set limits.

From the SPC-PC perspective, one of the most recent contributions is that of Aliverdi et al. (2013). They propose individual quality control charts as a monitoring tool for EVM indices. However, in recent years, SPC-STL has received more attention. Acebes et al. (2013) presented a graphical method that integrates Monte Carlo simulation of expected project variability and the risk baseline evolution indices proposed by Pajares and López-Paredes (2011). Subsequently, Acebes et al. (2014b) proposed a method that establishes time and cost limits separately based on percentiles obtained through simulation, adjusted to the project control point. Similarly, Colin and Vanhoucke (2014) proposed statistical tolerance limits and their analysis through X and R charts. However, individual cost and time analyses can generate wrong alarm signals, including false positives and negatives. Colin et al. (2015) proposed two multivariate project schedule control metrics (Hotelling's T^2 and squared prediction error) combined with PCA to address this issue. For the same problem, Acebes et al. (2015) proposed using anomaly detection techniques on simulation data.

Inspired by the critical chain/buffer management method, Colin and Vanhoucke (2015) proposed two methods integrating EVM with multiple control points (EVM-FPB and EVM-SNB). Vanhoucke and Colin (2016) further extended the multivariate approach by using matrix decomposition, kernel variant, and partial least-squares regression. Hadian and Rahimifard (2019) also focused on the possible correlations between EVM control indicators to establish the control system by multivariate Hotelling's T^2 control chart. Their study proposed several multivariate process indices to describe the capability of project performance. Votto et al. (2020) proposed a similar approach to generalise the methodology not only to EVM indicators but also to those provided by EDM. Chen et al. (2020) proposed an algorithm from a Bayesian perspective to optimise tolerance limits from conditional distributions of inputs.

In addition to static and statistical project control methods, there is a third approach to establishing tolerance limits for EVM extensions, which involves analytical tolerance limits (Vanhoucke, 2019). This method aims to set appropriate control thresholds, rather than relying on arbitrary values, based solely on the project baseline schedule and basic EVM performance indexes to maintain computational simplicity. Within this framework, Hu et al. (2016) proposed buffer management as a useful control tool to monitor schedule deviations during project execution. Notably, they used an integrated schedule monitoring system that includes an activity cruciality index and a strategic expediting procedure to support more accurate decision-making. Martens and Vanhoucke (2017b) subsequently proposed a control method that assigns a predefined buffer to different project phases based on EVM metrics. The allowable consumption of the buffer in each phase is used as a threshold to indicate possible overruns in the project duration. Building on their work, Martens and Vanhoucke (2017a) incorporated resource information in defining tolerance limits by using the availability and needs of scarce resources to create tighter limits on project phases that are more likely to experience resource conflicts and delays. Wajdi Hammad et al. (2018) proposed a straightforward method for schedule contingency management using the theory of constraints and EVM, which includes two new metrics: the buffer performance index and buffer variance. These metrics measure the expected and actual remaining buffer and set limits that generate warning signals and initiate corrective activities. Finally, Song et al. (2022) propose an extension of resource-constrained project control approaches that allow for the analysis of project progress and the establishment of tolerance limits based on resource constraints.

Although, to our knowledge, there is still no work that includes explainable machine learning in the framework of integrated project control, recently, applications of XAI in project management (from a more general perspective) have emerged. For example, in addition to the work of Radhakrishnan and Jaurez (2021) discussed earlier, Wang et al. (2022) analysed the economic factors that influence the estimation of project construction cost using deep neural networks. They then used an explainability layer to interpret their impact. Similarly, Ghorbany et al. (2022) used literature reviews, expert judgment, and questionnaires to extract key performance indicators (KPI) for public–private partnership projects. They then used copula Bayesian networks and explainable machine learning methods to optimise the network and determine the causal structure of KPIs.

2.2. Research objective and contribution

Our work extends the EVM to account for uncertainty under the hypothesis of explicitly modelling the probability distributions of task durations and costs. While previous research has explored the use of surrogate machine learning models to predict project outcomes from Monte Carlo data, our proposal offers several novel aspects that advance and complement these methodologies. Specifically, we incorporate an explainability layer into the fitted models, which provides information that no other approach addresses under the EVM framework.

Our approach provides prediction-level local explanations at control tracking points. These explanations attribute project outcomes to the specific activities comprising the project, which is highly insightful information for understanding the project's current situation and its future expectations. Additionally, our method employs model-level global explanations to obtain task-level sensitivity metrics, which are integrated into the control method itself. These metrics offer information on the size and direction of the task marginal contributions and the interaction between different tasks. These explanations have desirable axiomatic properties in all cases, including null sets, additivity, symmetry, and efficiency. Notably, our proposal represents one of the first attempts to integrate XAI into a top-down integrated project control management method.

3. Background.

3.1. EVM and triad method

Earned value management (EVM) is one of the most widespread methods in project management for controlling and evaluating project progress. In essence, the EVM method is based on three variables: i) the budgeted cost for work scheduled (BCWS) or planned value (PV); ii) the actual cost for work performed (ACWP) or actual cost (AC); and iii) the budgeted cost for work performed (BCWP) or earned value (EV), which gives its name to the method. The planned value curve is assumed to be known from the beginning of the project and its endpoint determines: (1) the expected project completion time on the x-axis; and (2) the Budget at Completion (BAC) —i.e., the planned cost of the project— on the y-axis. The Actual Cost and Earned Value curves are obtained throughout the project execution and are assumed to be known until the time of control.

From these three primary measures —PV[BCWS], AC[ACWP] and EV [BCWP]— and given an Actual Time (AT) —the time at which project control is performed, defined as the time elapsed since the beginning of the project— it is possible to obtain the indicators that quantify the project progress in terms of time and cost. In the traditional EVM method, the Cost Variance ($CV = EV[BCWP] - AC[ACWP]$) estimates whether the project is under budget (positive CV) or over budget (negative CV), and the Schedule Variance ($SV = EV[BCWP] - PV[BCWS]$), similarly, indicates whether the project is ahead (positive SV) or behind schedule (negative SV).

In the example of Fig. 1, the EV at the actual time is lower than the PV and AC, representing a project that would be delayed and cost overrun. Variance values are measured in cost and are absolute differences. To obtain relative measures, Performance Indexes are defined: Cost Performance Index ($CPI = EV/AC$) and Schedule Performance Index ($SPI = EV/PV$). In this case, to assess the state of the project, instead of considering the sign, we analyse whether the index is greater than 1, in which case the project is performing satisfactorily, or less than 1, a situation that would indicate a potential problem in cost and/or schedule, depending on the index in which it occurs.

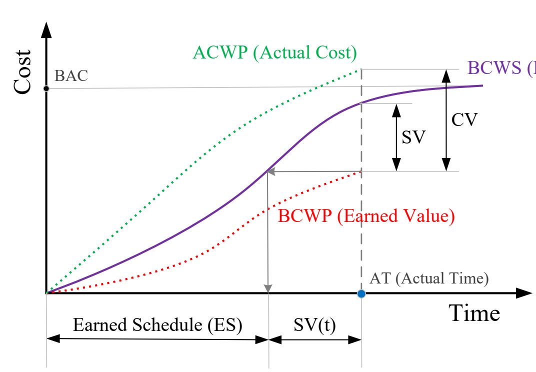

Fig. 1. Outline and main variables of Earned Value and Earned Schedule Management methods.

Notwithstanding the above, the EVM method presents some inconsistencies and interpretation problems in schedule management. Consequently, Lipke (2003, 2004) introduced a new measure for its evaluation that refines and complements the EVM control system: the Earned Schedule (ES). The ES is calculated by projecting the EV on the Planned Value curve. From this new measure, the Schedule Variance can be redefined as the difference $SV(t) = ES - AT$, which gives the project advance or delay in time units instead of cost units as in the original definition. An equivalent ratio can also be defined as performance index $SPI(t) = ES/AT$.

It is important to note that traditional EVM assumes certainty about the duration and costs of project activities. This assumption is usually extremely strong since project management is characterised precisely by its non-repetitive nature and by the management of uncertainty, which is difficult to eliminate given the uniqueness and context-dependence of each project.

A common way to incorporate project uncertainty is to model the variability of task cost and task duration as probability distributions (Colin & Vanhoucke, 2016; Pérez et al., 2016; Vanhoucke & Batselier, 2019). Then, the usual approach is to analyse projects from stochastic networks in which precedence relationships between tasks are included.

Based on these two ideas —EVM and stochastic networks— Acebes et al. (2014b, 2015) defined the triad method to account for uncertainty in project management. Briefly, this approach uses Monte Carlo simulation to obtain a representative sample of the universe of possible project realisations according to its stochastic definition. For each of the simulations, the value of time t and cost c at which the simulation reached the value EV is registered, constituting the triad (EV, t, c). From this information, the method uses advanced statistical learning methods to obtain relevant information on the progress of the project. On the one hand, it uses unsupervised learning algorithms for anomaly detection to determine whether the project is within the range of what could be expected from its stochastic definition or not. On the other hand, through supervised learning techniques and using the final time or cost of the simulation as target, the method answers two different questions: (i) the probability of project completion in time and/or in cost (classification problem); and (ii) the estimation of the expected cost and/or time at completion conditional on the current state of the project (regression problem). Recently, the triad method has been adapted (Acebes et al., 2022) to the Earned duration management (EDM) method (Khamooshi & Golafshani, 2014), an alternative to EVM in which the value of activities is expressed as work periods. Remarkably, the method presented in our paper is defined so that project progress is measured through EVM; notwithstanding, it could be used similarly with project progress measured through EDM.

Notably, the method proposed in this contribution extends the functionality of the triad method in two relevant ways. First, it allows to analyse the individual influence of each project task t in terms of both risk and uncertainty in a comprehensive and intuitive manner. And second, it enables the fair attribution to the project activities of the causes that led to the current state of the project.

3.2. Shapley values

As pointed out in the introduction, the machine learning community has become increasingly interested in explainability and attribution in recent years. In particular, the most recurrent question in terms of attribution is the following: given a machine learning model that has already been fit to the data, if we use it to predict a new instance, how does each feature contribute to the prediction?

The answer to this question is straightforward for linear regression models, where the contribution of each feature value to the response is determined by its coefficient in the fitted model. However, for more complex and sophisticated models, obtaining the contribution of each explanatory variable to the response requires alternative and more complex approximations. Remarkably, the Shapley value approach is one of the most used. This approximation came from the field of cooperative game theory and was originally devised in the context of a game in which players cooperate in a coalition to obtain a certain profit from their cooperation. Specifically, it was developed to determine how to distribute fairly the surplus among the different players according to their contribution to its attainment.

Formally, a coalitional game can be defined as a finite set N of n players ($N = \{1, 2, \cdots, n\}$) who can decide whether to cooperate or not. Cooperation among these players is formalised through coalitions, where a coalition S is a subset of N, i.e., an element of 2^N —the set of all possible subsets of N $\left(S \in 2^N\right)$. Let us denote as $\mathscr{N}$ the set of non-empty coalitions of N, and refer to N as the grand coalition. Additionally, let us consider a value function *val* that maps the different possible coalitions of players to their corresponding payoff $\left(val : 2^{|N|} \rightarrow \mathbb{R}, \text{ with } val(\varnothing) = 0\right)$. In this scenario, the Shapley value of a player is defined as her contribution to the payoff, weighted and summed over all possible coalitions of players.

$$\phi_j(val) = \sum_{S \in \mathscr{N}: j \in S} \frac{(|S| - 1)!(|N| - |S|)!}{|N|!} (val(S) - val(S \backslash \{j\})) \tag{1}$$

Expression (1) can be interpreted as follows. Let us assume that the coalition is composed of all players, i.e., $S \equiv N$, and is formed sequentially according to an input order. Each time a player $j \in N$ joins the already formed coalition $S \backslash \{j\}$, she is assigned a payoff equal to her marginal contribution to the payoff, i.e., she receives $val(S) - val(S \backslash \{j\})$. Notably, there are $(|S| - 1)!$ possible entry orders (permutations) for players in subset $S \backslash \{j\}$. Once j joins the coalition $S \backslash \{j\}$, the rest of the players belonging to $N \backslash S$ continue to enter the coalition. There are $(|N| - |S|)!$ possible entry orders for them. Thus, there are $(|S| - 1)!(|N| - |S|)!$ possible entry orders for players in $N \backslash \{j\}$ to join the coalition and, hence, player j has $(|S| - 1)!(|N| - |S|)!$ possibilities to receive the payoff $val(S) - val(S \backslash \{j\})$.

If we repeat this calculation on all the possible coalitions of N that contain j, and then divide by the total number of possible entry orders of the players (namely $|N|!$), we obtain the average of the marginal contribution of player j to the coalitions of N.

Alternatively, the Shapley value equation can be expressed as follows:

$$\phi_j(val) = \sum_{S \subseteq \{1, \cdots, n\} \backslash \{j\}} \frac{|S|!(n - |S| - 1)!}{n!} (val(S \cup \{j\}) - val(S)) \tag{2}$$

Importantly, the Shapley value is the only attribution technique that has proved to satisfy the axioms (desirable properties) of efficiency, symmetry, dummy and additivity, which together constitute the definition of a fair payoff (Shapley, 1953).

The **Efficiency** axiom refers to the fact that the sum of the Shapley values of all players equals the payoff of the grand coalition, i.e., the payoff when everybody cooperates. Formally:

$$\sum_{j \in N} \phi_j(val) = val(N) \tag{3}$$

Regarding the **Symmetry** axiom, it states that if two players j and k contribute equally to all possible coalitions, then their contributions (Shapley values) should be the same. That is:

$$\begin{aligned} &\text{If } val(S \cup \{j\}) = val(S \cup \{k\}) \forall S \subseteq N \\ &\quad \rightarrow \phi_j(val) = \phi_k(val) \end{aligned} \tag{4}$$

The **Dummy/Null player** axiom implies that if a player j does not change the payoff regardless of the coalition of players in which it is included, then it must have a Shapley value of 0.

$$\begin{aligned} &\text{If } val(S \cup \{j\}) = val(S) \forall S \subseteq N \backslash \{j\} \\ &\quad \rightarrow \phi_j(val) = 0 \end{aligned} \tag{5}$$

Lastly, the **Additivity/Linearity** axiom says that if two coalitional games with gain functions v and w are combined, then the fair distribution of payoffs should correspond to the sum of the payoffs derived from both games separately:

$$\phi_j(v + w) = \phi_j(v) + \phi_j(w) \ \forall j \in N \tag{6}$$

Also, for any real number a:

$$\phi_j(av) = a\phi_j(v) \ \forall j \in N \tag{7}$$

At this point, it is important to note that Shapley's original characterisation did not explicitly incorporate the marginalist criterion as an *a priori* desirable property of Shapley values. Nevertheless, it was subsequently introduced into the axiomatics by other researchers (Ghintran, 2011), as it seemed straightforward because marginalism was already present in the definition of the Shapley value itself, which states that to ensure fairness in the distribution of the payoff among the cooperating players, each of them must be assigned the average of her marginal contributions to all possible coalitions.

In view that back in the 50s, the Shapley value was shown to be the only fair way to distribute payoff among players (as it is the only attribution method that satisfies the above four properties of efficiency, symmetry, null player, and additivity), when the machine learning community began to deal with interpretability and attribution, they quickly turned to Shapley's work and adapted it to the problems addressed by machine learning. In overall terms, the translation from game theory to interpretable machine learning was enacted as follows:

- A *game* is the issuance of a prediction for a single instance.
- The *gain* is the difference between the prediction for the considered instance and the average prediction for all instances.
- The *players* are the values of the explanatory variables, that can be viewed as cooperating to issue the prediction.
- The *Shapley value* is the average marginal contribution of a feature value to the prediction across all possible combinations of the other feature values.
- The sum of the Shapley values of all features yields the difference between the actual prediction (for the instance) and the average prediction (for the dataset).

It follows from the foregoing that in the framework of machine learning, the Shapley value is used to explain the difference between a particular prediction and the average prediction. Importantly, the Shapley value should not be confused with the difference in the predicted value that would result from eliminating that explanatory variable from the model.

More formally, and coming back to Eq. (2), its translation to machine learning would read as follows:

$$\phi_j(val) = \sum_{S \subseteq \{1, \cdots, n\} \backslash \{j\}} \frac{|S|!(n - |S| - 1)!}{n!} (val(S \cup \{j\}) - val(S)) \tag{2}$$

There is a finite set N of n feature values ($N = \{1, 2, \cdots, n\}$) that can be used together (or not) to predict a given output. The different groups of feature values (subsets of N) that can be used for prediction are denoted by S (remember that $S \in 2^N$, the total number of possible subsets). In this context, the value function *val* maps the different combinations of feature values to the prediction obtained using them. In this regard, please note that the *val* function can be a machine learning model of any type, rendering the Shapley value method a model-agnostic tool. With all that in mind, the estimated Shapley value of a feature value is its contribution to the difference between the actual prediction and the average prediction for the data set, weighted and summed over all possible feature value combinations.

In other words, let X be the data matrix with dimensions $m \times n$, i.e., m rows (instances) and n columns (features). For instance i and feature j, the Shapley value $\left(\phi_{ij}\right)$ for feature value x_{ij} is interpreted as follows: for instance i the value of the j^{th} explanatory variable (x_{ij}) contributed (on average) ϕ_{ij} to the prediction of the response y_i compared to the average prediction for the dataset.

In terms of computation, to calculate the exact Shapley value of a given feature value x_{ij}, all possible combinations of feature values have to be evaluated with and without the j^{th} feature. Given that for N feature values the total number of possible feature combinations is 2^N, the computation time to obtain the exact Shapley values increases exponentially with the number of features: $O\left(2^{|N|}\right)$. Consequently, in most real-world scenarios, only estimations can be calculated. As a result, multiple less computationally expensive approximations were proposed to estimate the Shapley values (Lundberg et al., 2020; Lundberg & Lee, 2017; Štrumbelj & Kononenko, 2010, 2014).

We opted for Lundberg and Lee's approximation in the present contribution because of its multiple advantages (detailed below).

3.3. SHAP (SHapley Additive exPlanations)

The SHAP (SHapley Additive exPlanations) method (Lundberg & Lee, 2017) was originally conceived as an approximation to compute the Shapley values of the different feature values in individual predictions. Notably, SHAP also incorporates global interpretation methods based on the aggregation of individual Shapley values.

Let us begin with the calculation of the Shapley values for individual predictions and how it is enacted in SHAP. Remarkably, one key innovation in SHAP is that the Shapley value decomposition is represented by a linear model, that is, an additive feature attribution method. Specifically, for a dataset X with dimensions $m \times n$, i.e., m rows (instances) and n columns (features), the explanation of the prediction obtained for a particular instance x is formulated as follows:

$$g(z') = \phi_0 + \sum_{j=1}^{N} \phi_j z'_j \tag{8}$$

Where g is the explanatory model; z' is a binary vector (also known as coalition vector) that indicates whether the feature values participate in the individual prediction (1) or not (0); specifically, $z' \in \{0,1\}^N$ with N being the maximum coalition size, i.e., the total number of feature values; and ϕ_j are the Shapley values.

Please note that this linear representation in (8) is nothing but an artifice for the computation of the Shapley values.

As detailed in the previous section, the Shapley values are the only fair way to distribute the difference between the individual prediction and the average prediction among all the feature values involved, since they simultaneously satisfy the axioms of efficiency, symmetry, null player, and additivity. As for the SHAP approximation, in Lundberg & Lee (2017) some discrepancies between Shapley properties and SHAP properties are detailed. Notwithstanding, the three main desirable properties of SHAP are:

1. Local accuracy, which is the result of expressing the Shapley efficiency property in terms of the explanatory model g (instead of the *val* function) and the coalition vector z'. Concretely, for our instance of interest x:

$$g(x) = g\left(z'_x\right) = \phi_0 + \sum_{j=1}^{N} \phi_j z'_x j \tag{9}$$

If we now assume that all feature values are present, i.e., we set the coalition vector z'_x to all 1s, we get:

$$g(x) = g\left(z'_x\right) = \phi_0 + \sum_{j=1}^{N} \phi_j z'_x j = \phi_0 + \sum_{j=1}^{N} \phi_j \tag{10}$$

Which is a reformulation of equation (3):

$$\sum_{j \in N} \phi_j(val) = val(N) \equiv \phi_0 + \sum_{j=1}^{N} \phi_j = g\left(z'_x\right) = g(x) \tag{11}$$

2. Missingness ensures that features that are missing in the coalition vector receive an attribution of 0:

$$z'_x j = 0 \rightarrow \phi_j = 0 \tag{12}$$

This property is not within the original properties of Shapley values. However, it was included in SHAP to avoid artefacts in the results, since, given that in the additive linear expression —equation (9)— the Shapley values are multiplied by 0 or 1 according to the coalition vector z', it could be the case that an absent feature had an arbitrary Shapley value without affecting the local accuracy property —note that since it is multiplied by 0, whatever its value, it has no impact on equation (9).

3. Consistency, which is formulated in terms of two different explanatory models $\widehat{f}'$ and $\widehat{f}$ as follows:

$$\text{If } \ \widehat{f}'_x(z') - \widehat{f}'_x(z' \backslash \{j\}) \geq \widehat{f}_x(z') - \widehat{f}_x(z' \backslash \{j\})$$

$$\text{then } \ \phi_j(\widehat{f}', x) \geq \phi_j(\widehat{f}, x)$$

Hence, consistency refers to the fact that if in an alternative explanatory model $(\widehat{f}')$ the marginal contribution of a feature value increases or remains the same with respect to a previous explanatory model $(\widehat{f})$, regardless of the other feature values, then the corresponding Shapley values obtained with the alternative model also increase or remain the same accordingly. Importantly, Lundberg and Lee (2017) proved that from this consistency property, the Shapley axioms of linearity, null player and symmetry also hold.

The name given by Lundberg and Lee (2017) to their Shapley value estimation procedure is KernelSHAP. Specifically, it consists of five main steps: (i) sampling K feature value coalitions; (ii) obtaining the corresponding predictions for each coalition vector (for which it is necessary first to map coalition vectors to actual feature values); (iii) computing the weight for each coalition according to the equation of the SHAP kernel (eq. 14); (iv) fitting a weighted linear regression model in which the sampled coalition vectors are the input, the SHAP kernel is the weighting scheme, and the prediction for the different coalitions is the output; and (v) returning the coefficients of the linear model, which are the Shapley values.

The detail of the coalition weighting scheme (eq. 14) is the following:

$$\pi_x(z') = \frac{(N-1)}{\binom{N}{|z'|}|z'|(N-|z'|)} \tag{14}$$

Where z' is the coalition vector for instance x, $|z'|$ is the number of features present in coalition z' (i.e., those with a 1 in the coalition vector), and N is the total number of features in the dataset, i.e., the maximum coalition size. Notably, this weight is conceived so that for each sampled coalition, it corresponds to the weight that the coalition would get in the Shapley value estimation. Hence, both small (majority if 0 s) and large (majority of 1 s) coalitions obtain the largest weights. The rationale behind it is that to learn the most about individual features, the best approach is to study them in isolation. Remarkably, since better Shapley value estimates are obtained if large-weight coalitions are used as inputs for the linear regression model, KernelSHAP is implemented so that coalitions are considered in the sampling scheme in descending order of weight.

In addition, it should be noted that in KernelSHAP, feature values absent in the coalition vector are replaced by the value of that feature in a randomly sampled data instance. This mechanism is technically known as *sampling from the marginal distribution* and implies ignoring the possible correlations and dependence structures between features included and not included in the coalition(s).

4. Method formulation

Let be an acyclic directed graph that defines the precedence relationships —i.e., the constraints and temporality— of the p activities of a given project. Two types of points in time are considered in the method:

1. *The planning time*. It is the moment at which the duration of each of the activities is defined by a stochastic distribution —considered to be known— and which can be dependent or independent of other previous activities. It is worth noting that no assumptions are made about the type of the stochastic distribution. Regarding cost, a planned (stochastic or deterministic) cost function is assumed for each activity. Typically, the planning stage is carried out before project execution starts, i.e., EV $= 0$. Still, the method can be applied without loss of generality for later replanning conditional on a partially completed project history (EV $= a$).
2. *The actual time (analysis time)*. It is posterior to the planning time and is defined by a given earned value, EV $= b$.

The first step of the method takes place at the planning time and consists of generating a dataset compatible with the project schedule —i.e., with the stochastic distributions of task durations— via Monte Carlo simulation of the associated project PERT/CPM network (see top part of Fig. 2). For this purpose, the project is simulated n times. For each Monte Carlo simulation, the completion progress in terms of the duration of all activities at analysis time EV $= b$ is collected. These observations are points in a p-dimensional space $\chi \equiv \mathscr{R}^p$ with $x_i^b \in \chi$, where x_i^b denotes the vector of task durations for the i^{th} simulation of project realisation at the time in the simulation when EV $= b$. The dataset is completed with a response variable y_i for each vector x_i^b. Depending on the type of analysis to be performed later with the method, y_i can be either a continuous or a categorical variable. Typically, it takes one of the following four values:

1. $y_i = D_i^{BAC}$, where D_i^{BAC} represents the total duration of the project at the end of simulation i.
2. $y_i = T_i^b$, where T_i^b represents the simulation time in which the simulation i reached EV $= b$.
3. $y_i = C_i^{BAC}$, where C_i^{BAC} represents whether the simulation i finished earlier or later than the planned time in the project plan.
4. $y_i = C_i^b$, where C_i^b denotes whether the simulation time at which simulation i reached the value EV $= b$ is ahead or behind what would be expected for that EV value.

In the machine learning literature, the first two cases define regression problems and the next two define classification problems. Let us consider case 1) to illustrate the formulation of the method.

In this case, we are looking for a model f that captures the relationship $f : \chi \rightarrow \mathscr{R}$ in the best possible way. Specifically, we are interested in the distribution of the dependent variable Y given X as explanatory variables, where Y is the vector of all the responses y_i (duration of the project at the end of simulation i) and X denotes the $n \times p$ matrix of all the x_i^b vectors for the n simulations. The main assumption of the method is that $f(x)$ is a reasonable approximation of $E_{Y|x}(Y)$, i.e., that $E_{Y|x}(Y) \approx f(x)$. Notably, the quality of fit will typically increase as a function of the size of the dataset (the larger the dataset, the better the quality of fit), and the ability of the model in terms of bias-variance to fit the problem (Hastie et al., 2009). In our approach, we propose to consider a sufficiently flexible set of models and to perform model selection by nested-cross validation (stratified in the case of classification). Remarkably, if more efficient machine learning models are developed in the future, these could become part of the set of models considered for model selection. As for nested cross-validation, its need arises from the fact that many of the best-performing algorithms in both regression and classification are parametric, so an outer loop is required to identify the family of best-performing algorithms, and an inner loop is needed to adjust $\widehat{\theta} \in \Theta$, where Θ is the space of all the possible model parameters

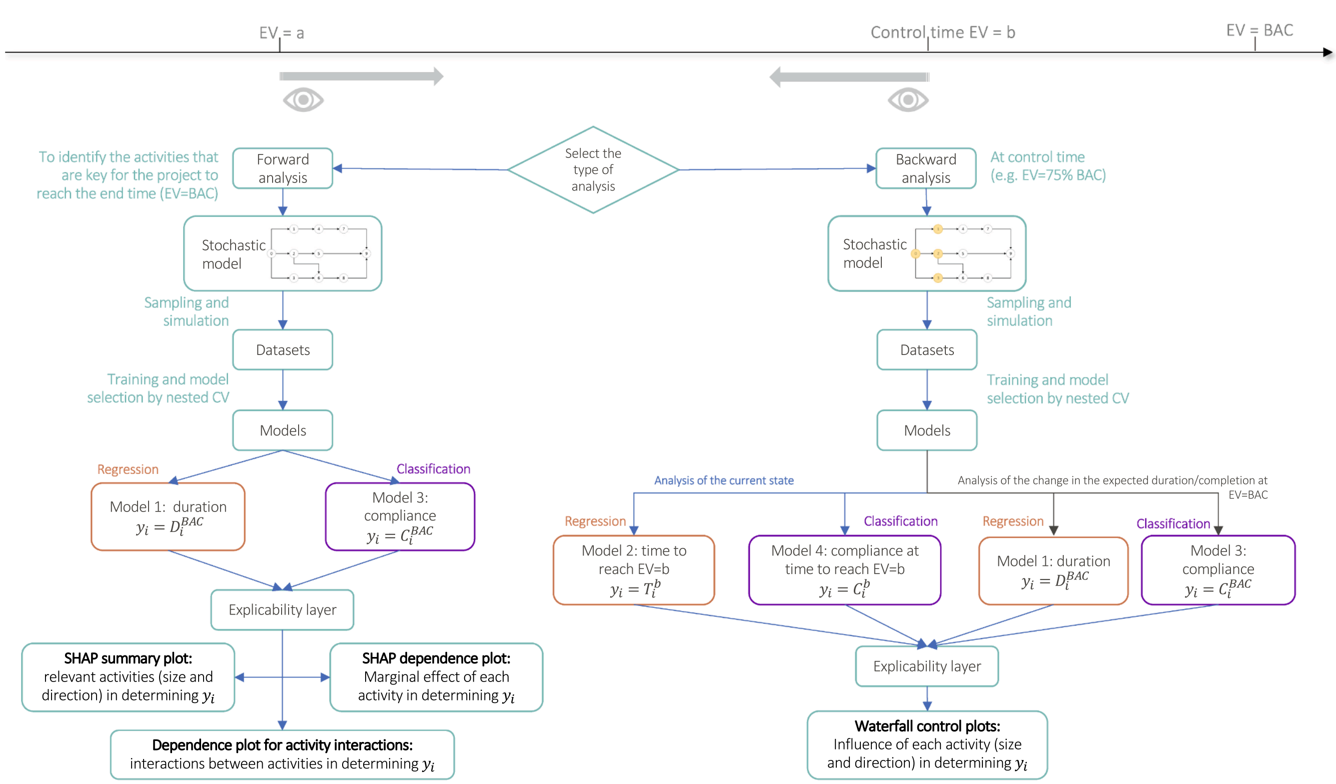

Fig. 2. Methodology flowchart.

and $\widehat{\theta}$ are the parameters selected. The $\widehat{\theta}$ search process consists of selecting the set of model coefficients that minimises some predefined loss function $L()$: $\widehat{\theta} = \mathrm{arg}min_{\widehat{\theta}\in\Theta} L[\{Y, f(\theta; X)\} + \lambda(\theta)]$, where y is the response vector, $f(\theta; X)$ are the predictions made by model f when coefficients θ are used, and $\lambda(\theta)$ is a regularisation term typically included to control model complexity (Zou & Hastie, 2005).

Once a model has been selected by nested cross-validation, we have a relationship $y = f(x)$. As noted above, the meaning and interpretation of y will depend on the target value selected from the four options available [1.- 4.].

Interestingly, the exploration of this function $f(x)$ by means of Shapley Additive Explanations allows for two types of analysis depending on the time point at which it is performed:

1. Forward analysis. Assume we are at a time point in which the progress of the project is given by EV = a (see left part of Fig. 2). We are interested in identifying the key activities that allow reaching a later time —specifically, the time when EV = b— in the best possible way —less duration and cost. To that end, we can use the Shapley values provided by the SHAP package (Lundberg & Lee, 2017). More precisely, the SHAP summary plots of the model function combine the importance of the input variables (the project activities in this case) with their impact on the response variable y. It is an aggregated analysis —since each point in the summary plot corresponds to the Shapley value for a given feature value in a particular instance— that shows the comparative influence of the different tasks on the output, and the relationships between the value of a feature —for instance, the duration of one of the activities— and the prediction. Besides, information about each activity can also be disaggregated using SHAP Dependence Plots and SHAP Interaction Values. SHAP Dependence Plots allow identifying the importance of each task in the prediction (measured through Shapley values) according to its feature value(s). As for SHAP Interaction Values, they quantify the effect of the interaction between pairs of variables once their individual contributions have been discounted.

In the framework of Explainable Machine Learning, this type of analysis of the model function $f(x)$ corresponds to a global interpretation of the model and can be helpful for the project manager to make prospective decisions. With respect to traditional methods in project management, our approach provides richer and more complete information than, for instance, criticality indices, and helps to better understand the relationships between variables/project activities and their implications for decision-making.

2. Backward analysis is conducted at a given control time (actual time when EV = b) (see right part of Fig. 2). In this context, understanding the contribution of each task to the current situation —i.e., how we got to where we are and who are the main responsible parties for it—is also relevant. In these cases, the Shapley values for the current stage of the project at the time of control serve to fairly distribute the responsibility for the result among the different tasks. In Explainable Machine Learning terms, this type of analysis of the model function $f(x)$ corresponds to a local model interpretation and can be extremely useful for both the project manager and the project sponsor in attributing responsibilities.

In the case when $y_i = D_i^{BAC}$ and control is performed at EV = b, with the model trained on the information obtained when EV = a = 0, backward analysis could be conducted to determine, at the time of control (EV = b), the impact that the duration of each task has had on the final prediction of the total duration of the project. At that stage, a replanning could be performed considering EV = b as the planning time, training a new model conditioned to the current situation, and using EV = BAC as the analysis time. Once fit, we could perform a forward analysis of the remaining project.

For the other possible values of the output variable y_i, the interpretation changes and, therefore, the method offers complementary information. In particular, for $y_i = T_i^b$ (case 2), the analysis is conducted with respect to the actual time (EV = b) instead of the project completion time (EV = BAC), thus serving to identify the most influential tasks with respect to project duration at the actual time (EV = b). As for $y_i = C_i^{BAC}$ (case 3) and $y_i = C_i^b$ (case 4), the interpretation of the analyses is based on the relevance of the tasks for meeting or not meeting the target duration in the planned completion time or the actual time, respectively. Remarkably, while the interpretation is made on the absolute value of duration in the first two cases, the results are probabilistic in the other two cases.

5. Case study

5.1. Description

In this section, we provide an application example of the methodology proposed. In particular, we have chosen the project example from Lambrechts et al. (2008) and used by Acebes et al. (2014b, 2015) to illustrate the triad method. Specifically, we incorporated project uncertainty into the case study through Monte Carlo simulation, and modelled the project planning and control as follows. The directed acyclic graph in Fig. 3 depicts the dependence relationships between activities. Activity durations are assumed to follow a known probability distribution, a normal distribution characterised by the mean and the variance of the activity duration (Table 2). In addition, we assumed a fixed variable cost for each task since, although the focus of our analyses is on the time dimension of the project, costs are necessary to compute the earned value. On top of that, we introduced a change in the probability distribution of the duration of activity A5 conditional on the occurrence of a given event in activity A2. The details of this change are explained in the next section. The code for the Monte Carlo simulations and the Shapley value analyses of this case study is publicly available in the repository https://github.com/jismartin/sheva.

5.2. Simulation scenarios

The stochasticity of the project is explored via Monte Carlo simulation (50,000 iterations). The control time is set at the project state when the earned value (EV) of the project activities is equal to 75% of the planned value at end of the project. In each simulation iteration, activity durations are drawn from their respective probability distributions of duration. In addition, the duration of activities at 75% EV, and the final duration of the project were collected too.

Among the possible sources of uncertainty and risk that can affect projects (Curto et al., 2022; Hazır & Ulusoy, 2020; Mentis, 2015; Orangi et al., 2011), environmental factors can have a very relevant impact in some industries. For example, in the construction sector, weather and meteorological conditions are one of the leading causes of project delays and changes in planning (Ballesteros-Pérez et al., 2015, 2018; Durdyev & Hosseini, 2019; Ibbs & Kang, 2018), being environmental factors also relevant in marine operations (Gudmestad, 2019; Kubacka et al., 2021). To illustrate the usefulness of the method in this respect —while keeping

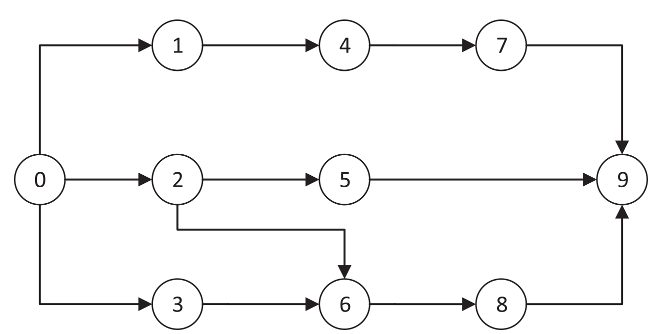

Fig. 3. Project network of our case study. The duration of each activity follows a normal distribution whose parameters are shown in Table 2.

Table 2
Parametrisation of the activities for the case study.

Activity	Mean duration	Variance	Variable cost
1	2	0.15	755
2	4	0.83	1750
3	7	1.35	93
4	3	0.56	916
5	6	1.72	34
6	4	0.28	1250
7	8	2.82	875
8	2	0.14	250

a simple design— we assumed that the probability distribution of the duration of activity A5 changes as a function of its starting time. Specifically, if activity A2 is delayed with respect to its mean duration by one standard deviation or more (4.91), then the probability distribution of A5 changes from $N(6, 1.72)$ to $N(18, 1.72)$, i.e., its mean duration increases drastically. It should be noted that this modification is exaggerated so that the effect can be illustrated more clearly in the analyses. In particular, the objective is to simulate an increase in the average duration of one activity triggered by the duration of another. The rationale behind changing the probability distribution of the duration depending on the starting time of the activity is that many tasks are affected by meteorological conditions whose probability depends on the month or season in which they are undertaken. For example, in this same line, Acebes et al. (2014a) modelled the seasonal risk derived from the probability of frost days for each month in the city of Valladolid (Spain), with a potential negative impact on some tasks. Because of this phenomenon, there is an interaction effect between the activity affected by its start time (A5) and its precedent activity (A2), which also determines the beginning of subsequent tasks. To analyse whether the machine learning model selected can capture this interaction effect and whether the SHAP method correctly attributes both responsibility and importance to activity A2, we compared this seasonal risk scenario —which we called the *conditional interaction scenario*— with a baseline in which the probability distribution of the duration of task A5 changes randomly with a frequency identical to that of the *conditional interaction scenario*.

5.3. Model selection

Once the dataset for the case study was created via Monte Carlo simulation (50,000 simulations of the project) we used four machine learning algorithms based on decision tree ensembles to try to capture the relationship $f : \chi \rightarrow Y$ for the different values of Y considered to be of interest. Remarkably, the four of them are state-of-the-art algorithms with high predictive performance in different contexts (Feng et al., 2020; Fernández-Delgado et al., 2014; Gómez-Ríos et al., 2017; Martin et al., 2022). As discussed in the Method formulation section, we used nested-cross validation —specifically 5-fold nested cross-validation— for the analyses. In the inner loop of the nested cross-validation scheme, the hyperparameters of the algorithms were optimised through a grid search. In the outer loop, model selection was performed based on the predictive error obtained (mean squared error, MSE) for the case of regression. Eventually, the best algorithm according to the above 5-fold nested cross-validation was fitted to the full dataset and parameterised using grid search 5-fold cross-validation.

In particular, of the four ensemble algorithms used in this analysis, three were based on boosting for regression: Adaptive Boosting (AdaBoost) (Freund & Schapire, 1996), Gradient Boosting (Friedman, 2001), and eXtreme Gradient Boosting (XGBoost) (T. Chen & Guestrin, 2016), and the remaining one on bootstrap aggregation (bagging): Random Forest (Breiman, 2001a). As for the more technical details, boosting consists in obtaining a strong learner from the sequential combination of weak base learners (Gómez-Ríos et al., 2017). In the case of AdaBoost, in each iteration, the weight of each data point changes such that the next learner preferentially focuses on those points that had a higher error in the previous iterations. In Gradient Boosting, instead of modifying the weights of the different instances, the algorithm tries to optimise the loss function of the previous learner. With respect to XGBoost, it shares many theoretical elements with Gradient Boosting; however, it implements parallelisation and incorporates different mechanisms to efficiently exploit memory resources and significantly reduce computation times. Lastly, the random forest algorithm combines the results of multiple classification or regression trees that are trained on different bootstrapped samples from the original training set. Importantly, it also uses the random subspace method at each split of the decision trees, thus achieving greater decorrelation between the estimators and better results than simply using the bagging technique. The regression results obtained for our case study are shown in Table 3. For more details on the computation procedures and/or the results, please refer to: https://github.com/jismartin/sheva.

For the classification approach, the model selection and model fitting procedures were very similar to the ones for regression, with the exception that we used accuracy (instead of MSE) as the quality metric for model selection, and that the versions of the algorithms were adapted for binary classification. The classification results obtained for our case study are shown in Table 4.

5.4. Forward analysis

In the forward analysis, as previously stated, the aim is to identify the key activities for the project to reach the end time (EV = BAC) while keeping the duration and cost as low as possible. We summarise this analysis with SHAP summary plots and SHAP dependence plots.

5.4.1. SHAP summary plots

Fig. 4 shows the SHAP summary plots of the two scenarios considered: the *conditional interaction scenario* and the baseline. This plot simultaneously combines the importance of each project activity with the feature effects. The different project tasks are ordered on the y-axis according to their importance (average SHAP value, provided in brackets to the right of the task name). On its part, the position on the x-axis represents the Shapley value for a project activity and a simulation (instance). More precisely, for each task on the vertical axis, the different SHAP values obtained for each of the simulations are plotted as a data point along the horizontal axis, adding some vertical jitter to avoid overlapping. The colour of the data points represents the variability of task durations, where an intense red indicates a high task duration relative to its average duration.

Several insights can be drawn from this plot: (1) the average SHAP value shows the relative importance of each project activity with respect to the response variable Y. Since in Fig. 4, the analysis was conducted for $y_i = D_i^{BAC}$, the top-ranked tasks are the most relevant for determining the final duration of the project; (2) the colour of the data points provides additional information on the duration value of a given task for a specific realisation (each point is coloured according to its duration value between red for long durations and blue for short durations); specifically, colours allow to interpret an individual value of task duration within the range of variability of the duration of this task, and to assess its effect on the final prediction.

Table 3
Results obtained via 5-fold nested cross-validation for regression model selection. The model with the best performance metric, gradient boosting, has been used.

Model	MSE (mean)	MSE (stdv)
Gradient Boosting Regressor	9.1939	0.2734
Random Forest Regressor	9.2052	0.2514
XGBoost Regressor	9.2168	0.2687
AdaBoost Regressor	11.0620	0.1232

Table 4
Results obtained via 5-fold nested cross-validation for classification model selection. The model with the best performance metric, gradient boosting, has been used.

Model	Accuracy (mean)	Accuracy (stdv)
Gradient Boosting Classifier	0.8570	0.0020
XGBoost Classifier	0.8564	0.0023
Random Forest Classifier	0.8544	0.0035
AdaBoost Classifier	0.8537	0.0028

Fig. 4 also shows the dependence between activity A2 and activity A5 in the scenario with seasonal risk (*conditional interaction scenario)* when compared with the baseline scenario. Specifically (and as originally intended), when A2 has a high duration, it causes an increase in the duration of A5 due to its environmental constraints —which we modelled as a change in the probability distribution of its duration from $N(6, 1.72)$ to $N(18, 1.72)$. On the contrary, if its duration remains at average or low duration values, the impact of A2 on the final project duration is minimal since, as it is not triggering an increase in the duration of task A5, other alternative paths in the project network become the most relevant. These results contrast with those of the baseline (i.e., the model without interaction between tasks A2 and A5), where, because of the independence of their durations, the impact and relative importance of the path through A2 and A5 is split between the two activities (instead of being mostly assigned to A2). In addition, the relevance of A5 is much higher than in the *conditional interaction scenario* due to its polarised statistical behaviour.

Notably, the information that can be extracted from the SHAP summary plots is very useful for a project manager, since it complements task information with other global importance measures, such as the criticality index or the cruciality index (Williams, 2002), but with the advantages that in this case: (i) task importance is combined with its level of uncertainty; (ii) interactions between tasks are captured; (iii) individual information (i.e., the particular value of task duration for a given simulation) is provided, which allows conducting case-level analysis; and (iv) by aggregating the individual interpretations of the case-level analyses, it is possible to perform global analyses of the different activities from a joint perspective of the entire project network.

5.4.2. SHAP dependence plots

While SHAP summary plots provide general and comparative information between the different tasks that integrate the project, SHAP dependence plots provide more detailed information on the individual importance of each of the activities. This type of plot shows the marginal effect that a particular predictor has on the final prediction of the fitted model. Specifically, for each individual simulation, the horizontal axis represents the duration value of the task, and the vertical axis represents its corresponding SHAP value. Therefore, variance (in importance) can be observed on the y-axis. In terms of interpretation, the prediction can be regarded as starting from the average prediction for the dataset and then being affected by the effect of each task duration in each simulation; this effect is quantified by the corresponding SHAP value and can either contribute to increase the final prediction (positive SHAP value)

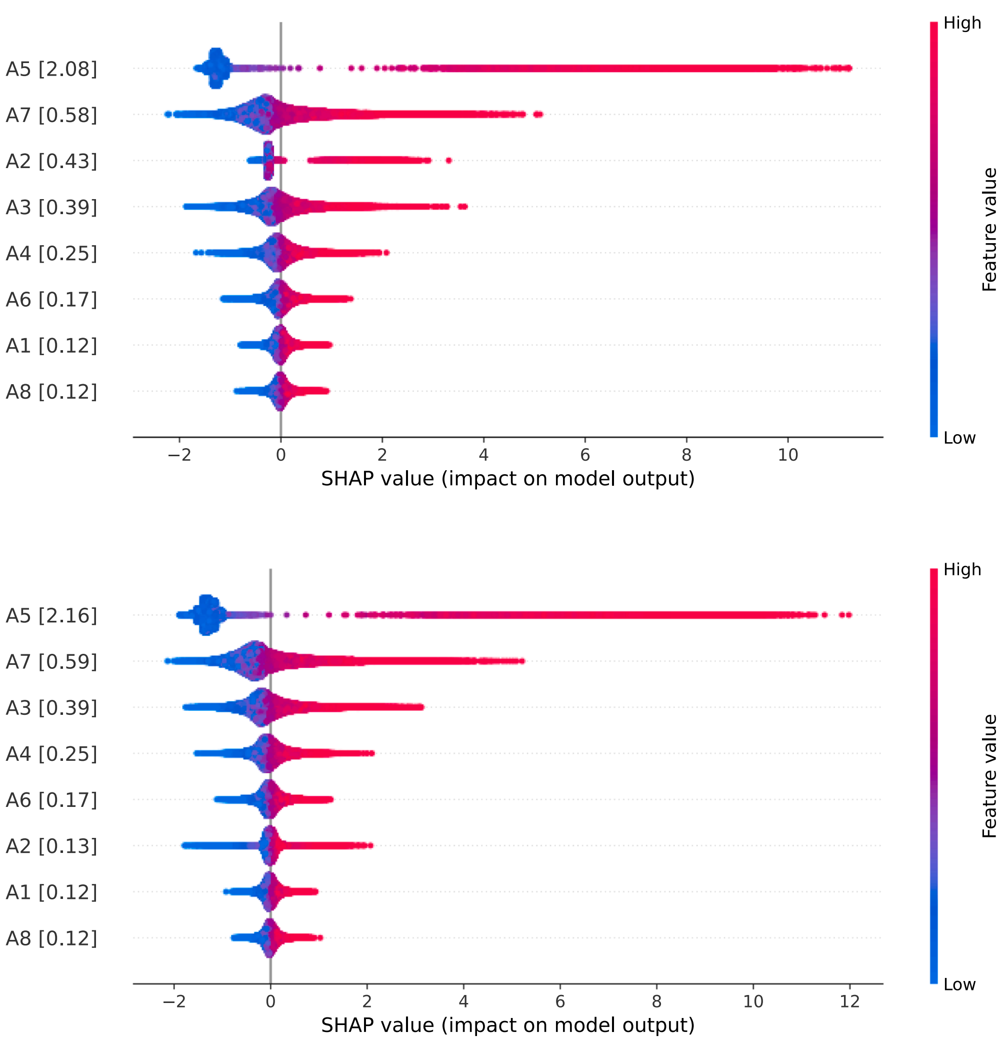

Fig. 4. SHAP summary plots of the two scenarios considered: *conditional interaction scenario* (top) and baseline scenario (bottom). Each subfigure represents the relative importance of the project tasks (their duration) on the total project duration. The vertical axis shows the project tasks ordered from top to bottom according to their average SHAP value (in brackets to the right of the task name). For each task, the SHAP values obtained for each individual simulation are shown as a data point (with vertical jitter to avoid overlapping) and coloured as a function of the relative task duration.

or to decrease it (negative SHAP value). As for the global interpretation of each subfigure in Fig. 5, it shows how the duration value of a given project activity relates to the model output —in this case, the total project duration, which is indicated by means of the colour scale. In other words, negative SHAP values indicate an impact on expected project duration capable of reducing it, while positive SHAP values indicate the potential for delaying it.

In our case study, for each of the two scenarios considered (*conditional interaction scenario* and baseline), since the influence is measured on the same scale, the figures allow for a detailed comparison of the relative impact of the different activities as a function of their duration. Furthermore, a high variance or vertical spread of SHAP values given an activity duration value indicates possible interaction and non-linear effects with the rest of the project activities, e.g., the possible existence of parallel paths.

The results in Fig. 5 are very revealing to understand the impact of each of the project activities in the two scenarios. Firstly, one can see that the distribution of SHAP values of activity A2 is notoriously different in each case. In the scenario with seasonal risk (top subfigure), the duration of activity A2 is key in determining the total project duration (note that there is a yellowish cloud of points in the A2 subplot corresponding to those simulations in which a delay in A2 caused an

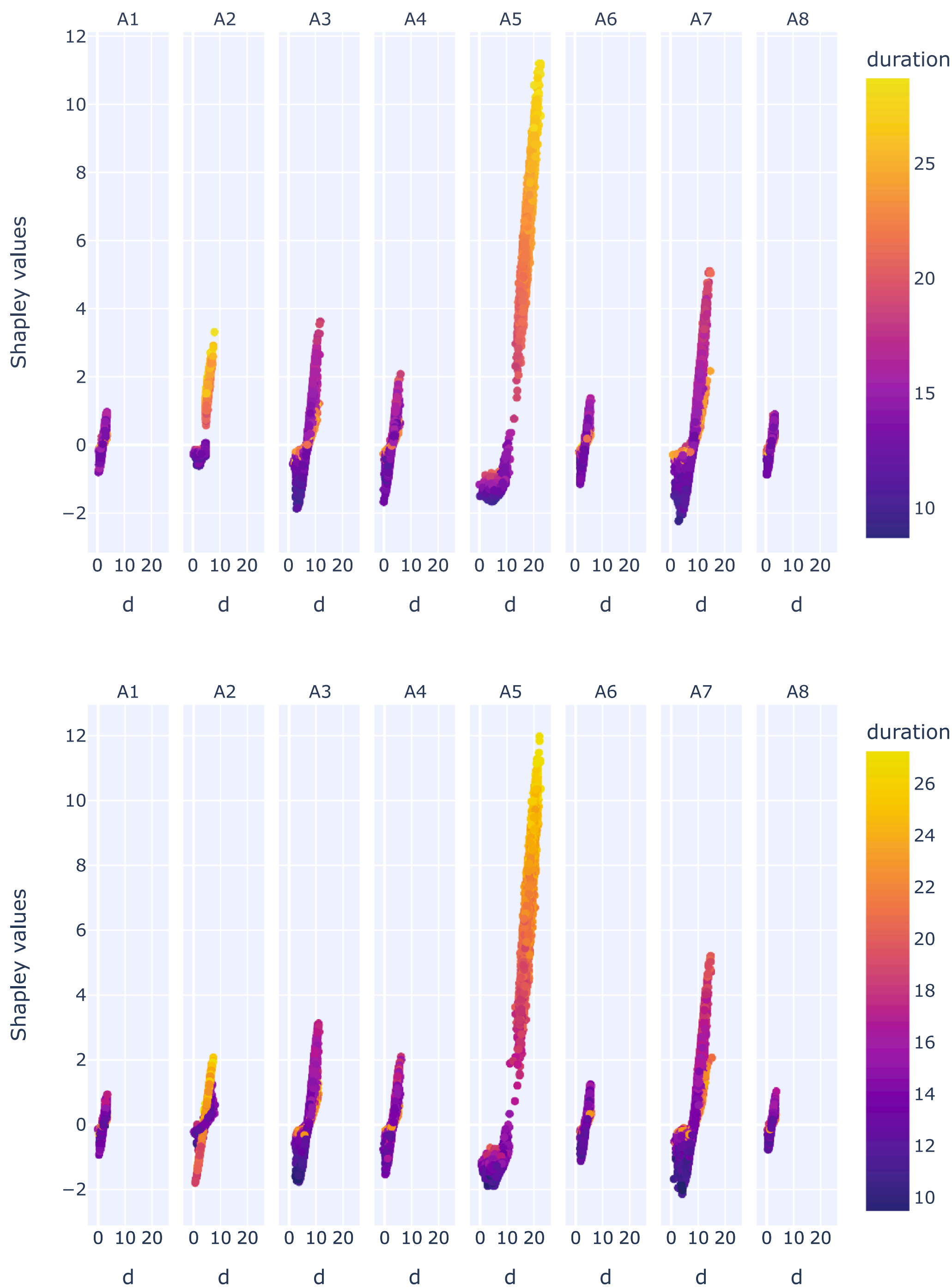

Fig. 5. Dependence plot of the different project activities for the two scenarios: *conditional interaction scenario* (top) and baseline (bottom). Each of the individual plots shows the marginal effect of the duration of each activity on the range of variation of the duration of each task and its impact on the expected total project duration (given by the colour scale).

increase in the final duration of the project). As already known, this effect is due to the seasonal character of activity A2 in this scenario. More precisely, when A2 duration exceeds the seasonal threshold (4.91, defined by design in the initial conceptualisation), A5 is affected by an increase in its average duration that translates into the final duration of the project jumping from expected values within the range 10–15 to expected values close to 25. In contrast, the impact of activity A2 in the baseline scenario has a very different behaviour that will be better understood with the next type of plot: the dependence plot for activity interactions.

5.4.3. Dependence plot for activity interactions

This type of plot constitutes an additional analysis tool that allows a better understanding of the interrelation between project tasks. In Fig. 6, we illustrate this interrelation effect for our case study by monitoring activities A2 and A5 and comparing their behaviour in the two scenarios: *conditional interaction scenario* and baseline —i.e., the scenario without forced interrelation between A2 and A5. In the scenario with seasonal risk (Fig. 6, left), we see that the dependence plot for activity interactions perfectly captures the effect that A2 has on A5. Specifically, and as intended by design, we can see that when A2 is delayed, A5 is also delayed due to the interrelation between the two —remember that if A2 is delayed, A5 follows a distribution with a higher average value. Consequently, two clusters of points appear in the interdependence plots, one for smaller durations of both A2 and A5 activities and the other for longer durations of the two. Moreover, we observe that the yellowish data points —which correspond to the cluster with longer activity durations— have positive SHAP values and present a positive correlation between the duration of activity A2 and its SHAP values; such correlation can be interpreted as a tendency towards increasing the duration of A5 and, eventually, the final duration of the project. Remarkably, this effect is caused by the criticality of activity A2 since it becomes part of the critical path when it is delayed, so its duration directly impacts the project duration. As for the baseline scenario, it has two clouds of data points as well; however, their interpretation is different. Yellowish points correspond to high A5 durations and purplish points correspond to low A5 durations. Remarkably, the points from both clusters extend along the entire horizontal axis, which evidences that there is no relationship between the durations of A2 and A5. In the yellowish cluster, i.e., when A5 durations follow a distribution with a higher average value, we see an increasing linear relationship (positive correlation) between the duration of A2 and its SHAP values; this linear relationship extends from negative to positive values on both axes, thus covering almost the entire plot. As for its interpretation, it would be as follows: in the yellowish cloud of points activity A2 is always part of the critical path, so an increase/decrease in its duration directly affects the total project duration. In other words, the duration of A2 is added almost directly and linearly (one-to-one) to determine the global SHAP value of activity A2, which is nothing but the contribution of A2 to the final duration of the project.

On the other hand, for low duration values of A5 (purplish cluster in Fig. 6 right), A2 SHAP values have a higher vertical spread, which indicates that the path along A2 may be critical, but other alternative paths also exist; this implies that increases in the duration of A2 are not directly transferred (in a one-to-one fashion as before) to the duration of the project.

5.5. Backward analysis

One of the main advantages of the SHAP framework is that it allows for both global and local interpretability. In the forward analysis performed in the previous sections, the aim was to understand the global behaviour of the model that encapsulates the stochastic behaviour of the project network. This information is relevant for future decision-making on project completion. Notwithstanding, understanding the reasons behind the current state of the project (case-level analysis) is also very relevant. This is what backward analysis is all about. Specifically, for a given project realisation, it uses SHAP to explain the individual contributions of the model inputs to the current situation.

To explain the usefulness of backward analysis, we first simulated the project until the moment when it reached an EV = 75% of the BAC and, for that time point, we registered the progress of each project activity in terms of duration. This information is shown in Fig. 7., where the duration of the activities is represented on the *y*-axis of each of the subplots. It should be noted that the subplots on the left correspond to the *conditional interaction scenario*, while those on the right correspond to the baseline scenario. Concretely, each subfigure is a waterfall plot in which the influence of each task on the output variable is represented by a coloured flag; the length of the flag is proportional to the magnitude of the influence, and its colour is given by the sign of the influence: red if it is positive and blue if it is negative. This influence value (in numbers inside the flag) indicates how the task modifies the output under

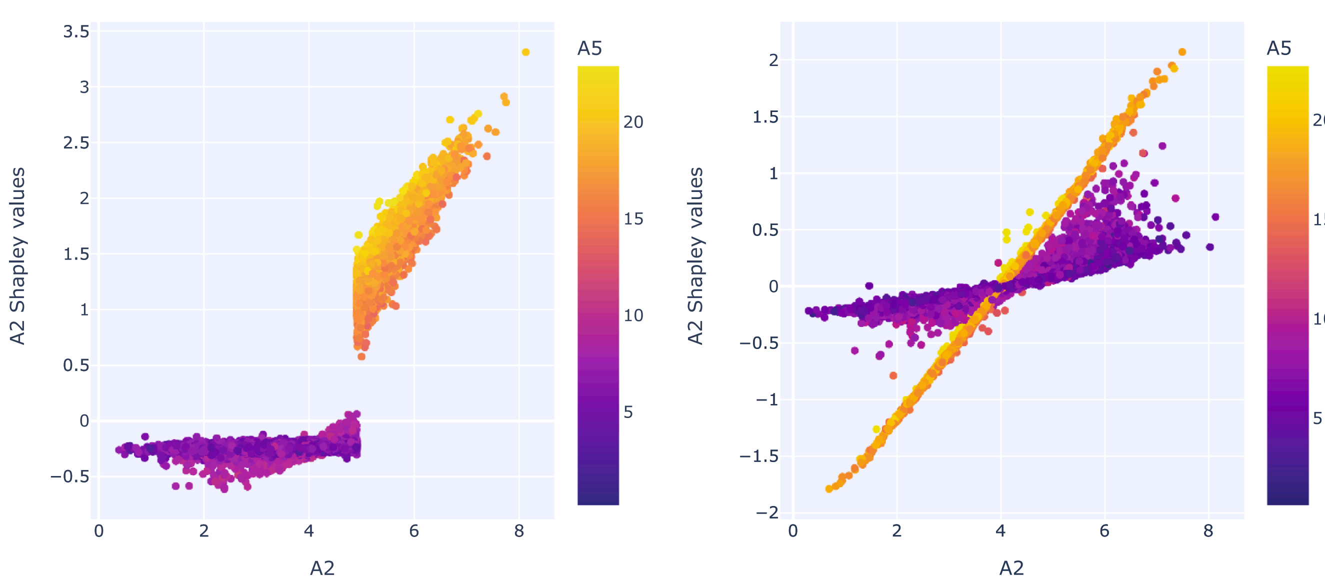

Fig. 6. Dependence plot of the interactions between activities A2 and A5 for the two scenarios: *conditional interaction scenario* (left) and baseline scenario (right). These plots show the Shapley value of activity A2 for each Monte Carlo simulation as a function of the durations of activity A2 (horizontal axis), and the duration of activity A5 (indicated by the colour scale).

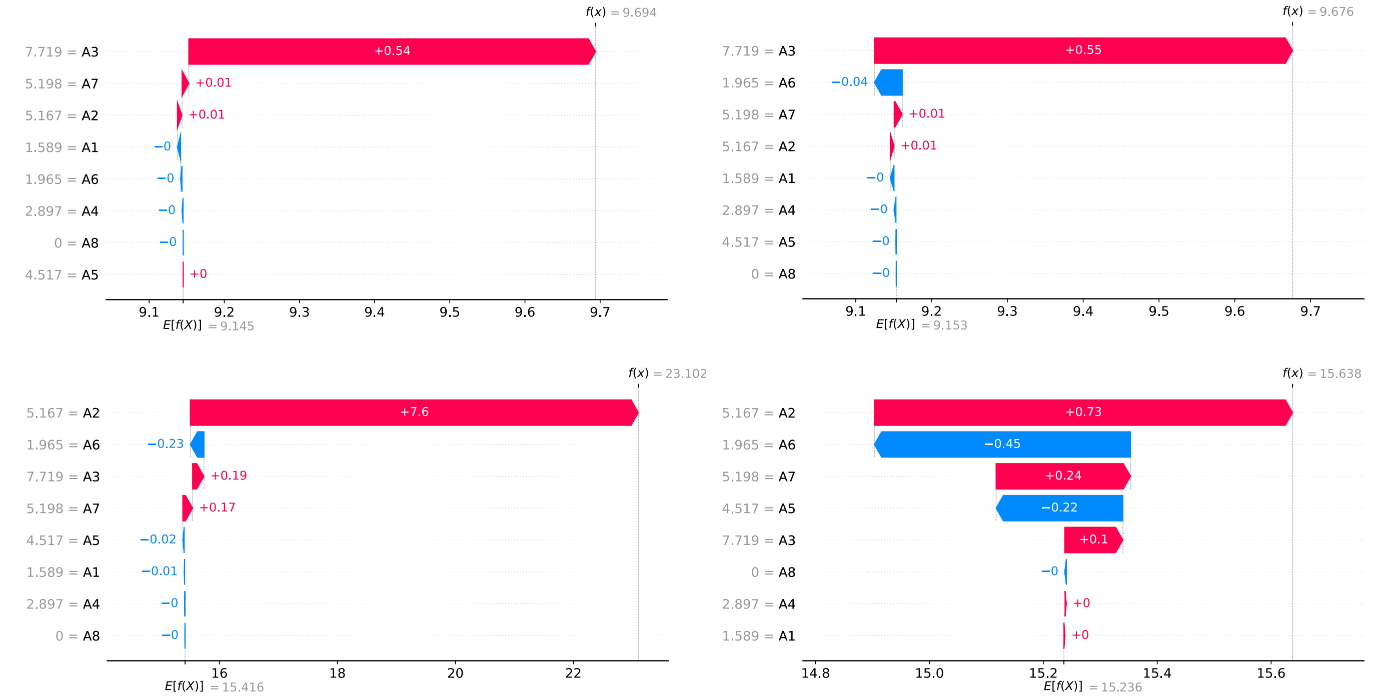

Fig. 7. Waterfall control plots for the analysis of the duration of a given realisation of our example project at EV = 75% BAC. Two scenarios considered: *conditional interaction scenario* (left) and baseline (right), and two possible outputs analysed: the duration of the project at the current state: $y_i = T_i^b$ (top) and its total duration until project end: $y_i = D_i^{BAC}$ (bottom). The value E[f(X)] on the *x*-axis represents the average predicted value across all Monte Carlo simulations, hence constituting the expected predicted value. The value f(x) shown at the right top of each subplot represents the new prediction for this instance as a consequence of the progress of the tasks. Specifically, the SHAP values indicate how each task contributed (positively or negatively) to the instance prediction compared to the average prediction. High absolute SHAP values indicate a significant impact of the task on the change in prediction.

scrutiny. Please recall that, as explained in the forward analysis, SHAP starts the prediction from the average prediction for the dataset (expected value —$E[f(X)]$); in Fig. 7., this value is provided below the horizontal axis. On its part, the final prediction of the output variable for the project realisation considered is presented in the right top corner —$f(x)$.

In the top subfigures, we analyse how the current state of the project (the activity durations at the time of control, i.e., when the EV = 75% BAC) relates to the duration of the project $y_i = T_i^b$(for b = 75% BAC). In this case, we are interested in explaining the current state of the project in terms of its current duration, namely $y = 9.684$. This project duration is slightly longer than expected for the level of earned value considered ($E[f(X)] = 9.15$). Nonetheless, the models for both scenarios give similar values around 9.68 (see $f(x)$ values on the right top corner of the subfigures in Fig. 7.), which implies that both predict a project delay compared to the expected duration. Notably, both top subfigures show equivalent explanations for this delay, as both point to activity A3 as the responsible. In particular, activity A3 has slightly delayed the time at which we reach an EV = 75% BAC, causing an increase of 0.5 time units.

In the bottom subfigures, the fitted models associate the current project status with the expected total duration of the project: $y_i = D_i^{BAC}$. Therefore, in this case, the question to answer is different: we want to explain why and how the expectation of the total project duration has changed due to its current state (progress). The results obtained are very interesting and clearly reflect the usefulness of the method. In the *conditional interaction scenario*, the prediction for the project duration has changed from 15.638 (its expected value) to 23.102. The trigger for this increase in the predicted value is the duration of activity A2; the other variables influence minimally. In particular, the delay in A2 has changed the total duration expectation of activity A5 (which has started but is still unfinished), so the project is likely to be significantly delayed. This contrasts with the prediction for the baseline scenario (bottom right subfigure), for which the model prediction (15.638) is only slightly longer than the expected total project duration given the current progress. More precisely, although in the baseline scenario the progress of activities A2 and A7 worsens the total duration predicted, the progress of A5 and A6 compensates for this delay (decreases in time of 0.45 and 0.22 time units, respectively). Consequently, the reading to be made from the baseline is that, since there are no interactions between tasks in this scenario, the behaviour is very different, and the attributions of responsibility for the expected completion time are also distinct.

To complete the backward analysis and illustrate how further information can be obtained for control purposes, Fig. 8 shows the analysis of the project at EV = 75% of the BAC —i.e., at the same time point as Fig. 7— but now using $y_i = C_i^{BAC}$ as the output variable in the model. C_i^{BAC} represents whether the simulation *i* finished before or after the final duration foreseen in the project plan (planned value). The left subfigure in Fig. 8 shows the results for the *conditional interaction scenario*, while the right subfigure corresponds to the baseline scenario. To determine whether the total duration of the project is delayed or not, we compared it to the final duration obtained for the deterministic PERT of the project —knowing *a priori* that it is probably a very optimistic value under stochastic conditions (Acebes et al., 2014b; Klingel, 1966; MacCrimmon & Ryavec, 1964; Schonberger, 1981). (Please recall that the results would be different if we used another PV curve). Specifically, in our analysis, the model results confirm that the baseline probability of finishing the project with delay $E[f(X)]$ is very high in both cases: 0.945 and 0.927, respectively. Importantly, since we know the progress of the tasks up to the control time, we can determine whether the probability of completing the project has changed with respect to the baseline planned value and the reasons for the change. In our example, the results show that the chances of finishing on time are low. In the *conditional interaction scenario* (left panel of Fig. 8), the duration of activity A2 has worsened the forecast of on-time project completion by 0.03, and the duration of activities A3 and A7 by an additional 0.01. This means that

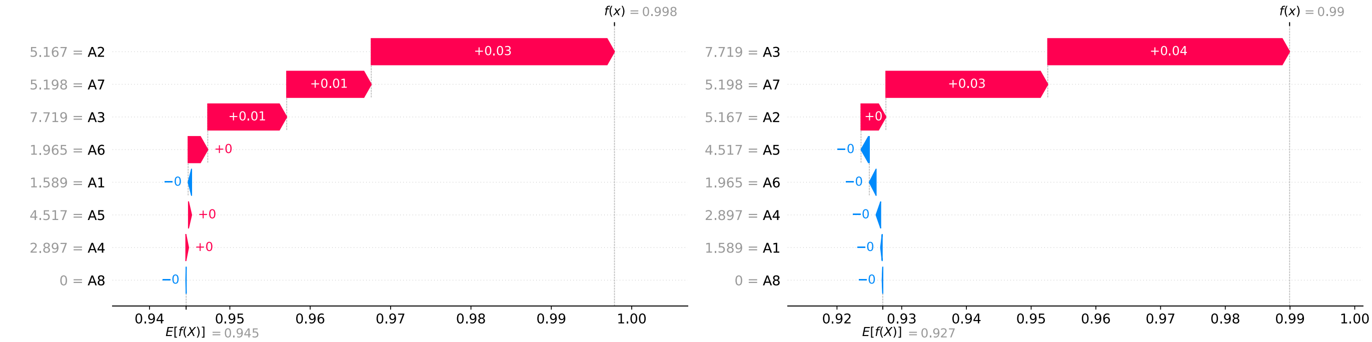

Fig. 8. Waterfall control plots for probability analysis. The individual instance analysed corresponds to a project in which the task durations are those indicated on the *y*-axis, with an earned value advance equal to 75% of the BAC. The figures correspond to a model fitted with $y_i = C_i^{BAC}$. The figure on the left shows the analysis for the *conditional interaction scenario*, while the baseline scenario is comparatively analysed on the right. The value E[f(X)] on the *x*-axis represents the average predicted probability computed across all Monte Carlo simulations to complete the project after the time limit planned. The value of f(x) at the top of each subfigure represents the new prediction (for the project realisation considered) of the probability of delay given the progress of the activities.

the model has captured the effect that A2 will have on the probability that A5 will be long, thus attributing to it a relevant impact on the likelihood of completing the project on time. In the baseline scenario (right of Fig. 8.), the worsening of the forecast is primarily attributed to the durations of activities A3 and A7. These results are consistent with those in Fig. 7, but with the added value that they complete the analysis from a probabilistic perspective (recall that in Fig. 7, we had absolute duration values instead of probabilities).

A noteworthy aspect of these plots, and which differs from the typical use of SHAP, is that in Fig. 8., task durations are provided as probabilities. In particular, in classification models, the SHAP package returns by default the influence of variables in terms of log odds. However, the interpretation of log odds and of their changes is not as intuitive as that of probabilities, which is what we provide in Fig. 8. To perform this transformation, the probabilities have been calculated from the log odds as the estimate of the Shapley values of the logit function. In the case of a project as simple as our case study, this transformation can be computed without difficulty, and, thus, the exact values can be obtained. However, the process can be computationally expensive (its computational complexity is exponential). In the repository associated with this work, we show how we use the method of Castro et al. (2009) to calculate the transformation. It is an unbiased polynomial method to estimate the Shapley values based on sampling theory. We use this as an alternative to SHAP because of the unintuitive nature of log odds.

6. Discussion

6.1. Methodological discussion

Our proposal is mainly methodological, consisting in a XAI extension of the stochastic earned value analysis, originally presented by Acebes et al. (2015) through the triad method. The case study in the previous section has served to illustrate the proposed methodology in detail.

The main contributions of this study may be summarised as follows:

- This paper presents a novel stochastic earned value technique based on the inclusion of a SHAP-value-based explainability layer. Such incorporation substantially increases the information available and enables a deeper understating of the project and its interrelations, thus allowing for more informed decision making than previous approaches.
- From a retrospective analysis perspective, given the current earned value, our method allows to: (i) compare whether the current project time and/or cost are as expected; (ii) to explain how each task has contributed to the current situation; and (iii) to analyse the change in completion expectations (in time and/or budget) through SHAP instance-level explanations (single prediction explanations).
- Prospectively, given a project control point, the use of SHAP global model explanations provides quantitative information on the overall effect and the relative importance of the different project tasks on the project outcome. Notably, such information is provided at different levels of detail depending on the type of analysis chart.

Notwithstanding, the development of our method implied several decisions, choices and assumptions that may be of interest to discuss.

First, our method takes both the probability distributions of task durations (or costs) and the project network as inputs. In this regard, note that both the probability distributions and the project network can be of any type but are assumed to be known. Regarding probability distributions, as shown in the case study, they may account for interactions with other tasks, risks, processes or external probabilistic events that can be of interest to model in the project plan. As for the project network, even though our description of the method is focused on CPM/PERT networks (Kelley, 1961; Malcolm et al., 1959), other networks may be used. In this line, further refinements may be necessary to evaluate its effectiveness in projects defined by GERT networks (Pritsker, 1966).

Second, our method requires sampling the probabilistic network of the project, which implies instantiating the probabilistic events and simulating them to generate the different analysis datasets. To that end, we employed Monte Carlo simulation as the sampling method, obtaining both the datasets and the expected values. Importantly, this technique is straightforward and facilitates experiment design, even in the case of interactions, as previously illustrated. However, other statistical sampling and design of experiments techniques may be more efficient in obtaining a dataset that covers the space of possible combinations, particularly in situations with limited time constraints.

Subsequently, our proposal involves fitting machine learning models to the different datasets previously obtained via Monte Carlo sampling. In this regard, our method is model-agnostic because it proposes to use several models and select the one with the best predictive performance on previously unseen samples. Such model selection approach is based on the important current line of research in automated machine learning (Chauhan et al., 2020; Hutter et al., 2019). Specifically, we chose as performance metrics the mean squared error (MSE) for regression and accuracy for classification. However, other metrics may be more appropriate depending on the project and its conditions. For instance, in a classification problem in which the cost of false positives or false negatives is not equivalent (it is not the same cost to predict that the project will finish on time and be wrong as to predict that it will not finish on time and be wrong), alternative metrics that weigh the confusion matrix differently could be used. Once the performance metrics are calculated, the selection of the best-performing algorithm is done by means of nested-cross validation (NCV). Remarkably, because of the model-agnostic nature of our method, it will allow for the inclusion of new and better algorithms that may emerge in the future.

Another important aspect to highlight is that the entire methodology relies on the correct stochastic definition of the project. Values that fall outside of the expected ranges imply that the predictive methods require extrapolation, which means that the results should not be trusted. Depending on the objectives of the analysis, such as risk management, it might be better to train the datasets on uniform distributions or to use other sampling methods that include extreme values more often, even if this means that the expected values differ from those of the project.

It is also noteworthy that, since our method allows to generate as much simulation data as needed and may use universal approximator models (or directly explainable techniques without mediating through surrogated models), it facilitates approximating the learning model to the stochastic project model at convenience. Nevertheless, it should be noted that model fitting and the calculation of SHAP values can be computationally expensive in very complex projects, hence posing challenges and relevant questions in the application scale of our method. Specifically, while optimised SHAP value calculation techniques are already available, these are specific for certain machine learning models such as tree-based models (Tree SHAP) or deep learning ones (Deep SHAP), hence potentially compromising the generality of the approach. In this regard, since for application purposes the computational efficiency of all simulation and fitting techniques will be crucial, more research will be needed to move from academic solutions to professional real-world ones. In upcoming years, if future research reveals that certain machine learning algorithms are systematically more time-efficient than others and capable of consistently obtaining superior predictive results, then the set of models considered for selection could be reduced, thereby improving the computational burden of the overall process.

Besides, it should be noted that different models may produce similar results when fitted on the same data due to the different internal structures and assumptions underlying each one, resulting in the well-known Rashomon Effect (Breiman, 2001b). Therefore, to increase the robustness of the analyses and explanations with a lower simulation burden, using Rashomon sets, or directly using interpretable models (Rudin, 2019) (depending on their performance) may be interesting avenues for future research.

Lastly, in relation to the explainability tools implemented in our method, we chose SHAP because we believe that the axiomatic properties that it inherited from Shapley's solution make it the most suitable approach for the problem addressed in comparison to other alternatives such as Break Down (Robnik-Sikonja & Kononenko, 2008) and Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016). In particular, such properties are especially desirable for the framework of prediction-level local explanations that underpins our backward analysis approach.

Nevertheless, SHAP is not the only recent XAI tool that can provide relevant information on prediction models from a prospective perspective. Hence, future research may explore the use of XAI tools complementary to SHAP. Such tools may offer alternative views on the relationship between task-sensitivity metrics and outcomes at the level of global model explanations. For example, from a model-agnostic perspective, Permutation Feature Importance (Fisher et al., 2019), Individual Conditional Expectations (Goldstein et al., 2015), and Accumulated Local Effects (Apley & Zhu, 2020), among other techniques, may potentially provide valuable and different insights into the influence of task duration and cost on the final project outcome(s). Furthermore, various machine learning models, such as decision trees and ensemble techniques, provide specific tools to conduct variable importance analyses (Wei et al., 2015), which could enrich the overall picture particularly in complex projects.

6.2. Managerial implications

The methodology proposed in this paper has significant implications for project management under uncertainty. Thanks to the XAI layer it incorporates to the framework of stochastic earned value management, it can help project managers to improve project control, manage complexity risk, and make better decisions.

By means of a proactive, data-driven approach, our methodology provides more detailed information on the relationship between project activities, progress and expected vs. obtained results. By leveraging simulation techniques, machine learning models and XAI, project managers can better understand the non-linear impact of the project network and its different tasks on project performance, which will translate into better decisions.

In particular, the use of SHAP as the explainability layer allows to quantify how different project activities and decisions would affect the overall project outcomes (time and cost). Consequently, these prospective analyses can help project managers make better decisions/adjustments on project execution so that time and cost targets are met in later project phases.

Besides, SHAP also allows to retrospectively analyse the causes of project deviations from the planned value curve (both in time and cost), which has significant implications as well. Specifically, it allows for a certain level of accountability attribution between the parties involved in the project. Apart from that, from a knowledge and experience management perspective, our method helps to identify possible biases in planning and provides clues to improve future planning in projects with similar activities.

Eventually, as previously stated, a relevant advantage of our methodology is its generalisability. The approach is general and model-agnostic, making it a versatile tool that can be used in different sectors and project types. Notwithstanding, since the computational burden of some of the analyses can be high, project managers (and future research) may need to assess the computational resources required to apply this methodology to larger projects, and determine whether the advantages of general approaches to model selection outweigh the computational costs.

7. Conclusions

In this paper, we have proposed a new methodology for the advanced control of projects subject to stochastic variability based on the EVM and the triad methods. The inputs for our method are: (i) the primary elements of the EVM —earned value (EV), planned value (PV), and actual cost (AC); (ii) the stochastic definition of activity durations; and (iii) the relationships between project activities (and eventually costs) according to the project network. From this information, two types of analyses: *forward* and *backward*, can be conducted.

To perform forward analysis, the proposed method consists of three phases: (i) the obtention of Monte Carlo simulations compatible with the stochastic definition of the project, (ii) the selection of the machine learning model that best encapsulates the behaviour of the project from the simulation data, and (iii) the exploitation of the model results through interpretable machine learning techniques, in particular, through SHAP. This forward analysis framework allows for a global understanding of the influence of the different activities on the final project duration and/or of the probability of finishing the project on time. Through SHAP summary plots, the relative impact of each activity can be analysed in terms of the duration values of each task. This information can be enriched and detailed by means of Shap dependence plots and plots for activity interactions, in which the combined effect of the activity durations on the performance measures of interest can be analysed.

Alternatively, by feeding the machine learning model with the durations of the tasks performed up to the actual time (control time), the backward analysis framework can be used to explore how the project predictions have changed conditional on the current project status and to identify the reasons for such change.

The proposed methodology provides two types of analysis that can be helpful for a project manager. Forward analysis provides more descriptive richness than classic task importance analysis techniques but keeps the overall project information condensed and easily understandable. Furthermore, understanding the non-linear relationship between inputs-outputs —such as those defined by project structures— allows for improved decision-making by acting on the decision levers with the most significant impact. In turn, the information from the backward analyses provides the project manager with essential control information. It allows to know how the future expectations of the project have changed conditional on the current situation and provides her with a tool to fairly attribute the responsibility for the progress of the project up to the moment of analysis (control).

The method presented has been illustrated with a case study based on the EVM and project duration analysis. Still, it is a very general framework, easily adaptable to other integrated control techniques such as EDM, or to the study of costs instead of durations —which could be conducted in an identical way. Furthermore, it is also possible to feed models based on task durations and progress rates, which may allow a better fit to the project structure in small projects.

Eventually, the method proposed does also have some limitations related to the choices and decisions made to formulate it, such as the computational burden for large-scale projects, the possibility that other stochastic sampling techniques may also provide good results, the use of CPM/PERT networks instead of GERT ones, etc. Nonetheless, these limitations constitute promising lines for future research that will be worth exploring.

CRediT authorship contribution statement

José Ignacio Santos: Software, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Conceptualization. **María Pereda:** Software, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Conceptualization. **Virginia Ahedo:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Conceptualization. **José Manuel Galán:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code for the Monte Carlo simulations and the Shapley value analysis of this case study is publicly available in the repository https://github.com/jismartin/sheva

References

Abdel Azeem, S. A., Hosny, H. E., & Ibrahim, A. H. (2014). Forecasting project schedule performance using probabilistic and deterministic models. *HBRC Journal, 10*(1), 35–42. https://doi.org/10.1016/j.hbrcj.2013.09.002

Acebes, F., Pajares, J., Galán, J. M., & López-Paredes, A. (2013). Beyond earned value management: A graphical framework for integrated cost, schedule and risk monitoring. *Procedia - Social and Behavioural Sciences, 74*, 231–239. https://doi.org/10.1016/j.sbspro.2013.03.027

Acebes, F., Pajares, J., Galán, J. M., & López-Paredes, A. (2014a). Exploring the influence of seasonal uncertainty in project risk management. *Procedia - Social and Behavioural Sciences, 119*, 329–338. https://doi.org/10.1016/j.sbspro.2014.03.038

Acebes, F., Pajares, J., Galán, J. M., & López-Paredes, A. (2014b). A new approach for project control under uncertainty. Going back to the basics. *International Journal of Project Management, 32*(3), 423–434. https://doi.org/10.1016/j.ijproman.2013.08.003

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. https://doi.org/10.1016/j.ijproman.2015.06.012

Acebes, F., Poza, D., González-Varona, J. M., & López-Paredes, A. (2022). Stochastic earned duration analysis for project schedule management. *Engineering, 9*, 148–161. https://doi.org/10.1016/j.eng.2021.07.019

Acebes, F., Poza, D., González-Varona, J. M., Pajares, J., & López-Paredes, A. (2021). On the project risk baseline: Integrating aleatory uncertainty into project scheduling. *Computers & Industrial Engineering, 160*, Article 107537. https://doi.org/10.1016/j.cie.2021.107537

Aliverdi, R., Moslemi Naeni, L., & Salehipour, A. (2013). Monitoring project duration and cost in a construction project by applying statistical quality control charts. *International Journal of Project Management, 31*(3), 411–423. https://doi.org/10.1016/j.ijproman.2012.08.005

Anbari, F. T. (2003). Earned value project management method and extensions. *Project Management Journal, 34*(4), 12–23. https://doi.org/10.1177/875697280303400403

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

Aramali, V., Gibson, G. E., el Asmar, M., & Cho, N. (2021). Earned value management system state of practice: identifying critical subprocesses, challenges, and environment factors of a high-performing EVMS. *Journal of Management in Engineering, 37*(4). https://doi.org/10.1061/(asce)me.1943-5479.0000925

Aramali, V., Sanboskani, H., Gibson, G. E., el Asmar, M., & Cho, N. (2022). Forward-looking state-of-the-art review on earned value management systems: The disconnect between academia and industry. *Journal of Management in Engineering, 38*(3). https://doi.org/10.1061/(asce)me.1943-5479.0001019

Bagherpour, M., Zareei, A., Noori, S., & Heydari, M. (2010). Designing a control mechanism using earned value analysis: An application to production environment. *International Journal of Advanced Manufacturing Technology, 49*(5–8), 419–429. https://doi.org/10.1007/s00170-009-2406-z

Ballesteros-Pérez, P., del Campo-Hitschfeld, M. L., González-Naranjo, M. A., & González-Cruz, M. C. (2015). Climate and construction delays: Case study in Chile. *Engineering, Construction and Architectural Management, 22*(6), 596–621. https://doi.org/10.1108/ECAM-02-2015-0024

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Mora-Melià, D., González-Cruz, M. C., Fuentes-Bargues, J. L., & Pellicer, E. (2019). Earned schedule min-max: Two new EVM metrics for monitoring and controlling projects. *Automation in Construction, 103*, 279–290. https://doi.org/10.1016/j.autcon.2019.03.016

Ballesteros-Pérez, P., Smith, S. T., Lloyd-Papworth, J. G., & Cooke, P. (2018). Incorporating the effect of weather in construction scheduling and management with sine wave curves: Application in the United Kingdom. *Construction Management and Economics, 36*(12), 666–682. https://doi.org/10.1080/01446193.2018.1478109

Barrientos-Orellana, A., Ballesteros-Pérez, P., Mora-Melia, D., González-Cruz, M. C., & Vanhoucke, M. (2021). Stability and accuracy of deterministic project duration forecasting methods in earned value management. *Engineering, Construction and Architectural Management.*. https://doi.org/10.1108/ECAM-12-2020-1045

Batselier, J., & Vanhoucke, M. (2015a). Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earned value management. *International Journal of Project Management, 33*(7), 1588–1596. https://doi.org/10.1016/j.ijproman.2015.04.003

Batselier, J., & Vanhoucke, M. (2015b). Empirical evaluation of earned value management forecasting accuracy for time and cost. *Journal of Construction Engineering and Management, 141*(11). https://doi.org/10.1061/(asce)co.1943-7862.0001008

Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management, 35*(1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003

Bergantiños, G., Valencia-Toledo, A., & Vidal-Puga, J. (2018). Hart and mas-colell consistency in PERT problems. *Discrete Applied Mathematics, 243*, 11–20. https://doi.org/10.1016/j.dam.2017.08.012

Breiman, L. (2001a). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3). https://doi.org/10.1214/ss/1009213726

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research, 36*(5), 1726–1730. https://doi.org/10.1016/j.cor.2008.04.004

Chauhan, K., Jani, S., Thakkar, D., Dave, R., Bhatia, J., Tanwar, S., & Obaidat, M. S. (2020). Automated Machine Learning: The New Wave of Machine Learning. *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 205–212. https://doi.org/10.1109/ICIMIA48430.2020.9074859.

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, Z., Demeulemeester, E., Bai, S., & Guo, Y. (2020). A Bayesian approach to set the tolerance limits for a statistical project control method. *International Journal of Production Research, 58*(10), 3150–3163. https://doi.org/10.1080/00207543.2019.1630766

Cheng, M.-Y., Chang, Y.-H., & Korir, D. (2019). Novel approach to estimating schedule to completion in construction projects using sequence and nonsequence learning. *Journal of Construction Engineering and Management, 145*(11). https://doi.org/10.1061/(asce)co.1943-7862.0001697

Colin, J., Martens, A., Vanhoucke, M., & Wauters, M. (2015). A multivariate approach for top-down project control using earned value management. *Decision Support Systems, 79*, 65–76. https://doi.org/10.1016/j.dss.2015.08.002

Colin, J., & Vanhoucke, M. (2014). Setting tolerance limits for statistical project control using earned value management. *Omega, 49*, 107–122. https://doi.org/10.1016/j.omega.2014.06.001

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. https://doi.org/10.1016/j.eswa.2014.12.007

Colin, J., & Vanhoucke, M. (2016). Empirical perspective on activity durations for project-management simulation studies. *Journal of Construction Engineering and Management, 142*(1), 04015047. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001022

D. Curto J. de Antón D. Poza F. Acebes A Review of Tools and Techniques in Uncertainty Management C. Avilés-Palacios M. Gutiérrez Ensuring Sustainability Lecture Notes in Management and Industrial Engineering 2022 Springer 233 243 10.1007/978-3-030-95967-8_21.

Czemplik, A. (2014). Application of earned value method to progress control of construction projects. *Procedia Engineering, 91*, 424–428. https://doi.org/10.1016/j.proeng.2014.12.087

Durdyev, S., & Hosseini, M. R. (2019). Causes of delays on construction projects: A comprehensive list. *International Journal of Managing Projects in Business, 13*(1), 20–46. https://doi.org/10.1108/IJMPB-09-2018-0178

Elms, D. G. (2004). Structural safety–issues and progress. *Progress in Structural Engineering and Materials, 6*(2), 116–126. https://doi.org/10.1002/pse.176

Eshghi, A., Mousavi, S. M., & Mohagheghi, V. (2019). A new interval type-2 fuzzy approach for analyzing and monitoring the performance of megaprojects based on earned value analysis (with a case study). *Neural Computing and Applications, 31*(9), 5109–5133. https://doi.org/10.1007/s00521-018-04002-x

Feng, Y., Wang, D., Yin, Y., Li, Z., & Hu, Z. (2020). An XGBoost-based casualty prediction method for terrorist attacks. *Complex & Intelligent Systems, 6*(3), 721–740. https://doi.org/10.1007/s40747-020-00173-0

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., & Amorim Fernández-Delgado, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research, 15*, 3133–3181.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are wrong, but many are useful: Learning a Variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research : JMLR, 20*, 1–81. http://www.ncbi.nlm.nih.gov/pubmed/34335110.

Fleming, Q. W., & Koppelman, J. M. (2010). *Earned value project management* (4th ed.). Project Management Institute.

Frank, M. V. (1999). Treatment of uncertainties in space nuclear risk assessment with examples from Cassini mission applications. *Reliability Engineering & System Safety, 66*(3), 203–221. https://doi.org/10.1016/S0951-8320(99)00002-2

Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. *Proceedings of the 13th International Conference on Machine Learning*, 148–156. https://doi.org/10.5555/3091696.3091715.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. http://www.jstor.org/stable/2699986.

Fryer, D. V., Strumke, I., & Nguyen, H. (2021). Model independent feature attributions: Shapley values that uncover non-linear dependencies. *PeerJ Computer Science, 7*, e582.

Ghintran, A. (2011). Marginalisme et valeur de shapley. *In Revue d'Economie Politique, 121*(2), 155–177. https://doi.org/10.3917/redp.212.0155. Editions Dalloz Sirey.

Ghorbany, S., Yousefi, S., & Noorzai, E. (2022). Evaluating and optimizing performance of public–private partnership projects using copula Bayesian network. *Engineering, Construction and Architectural Management.*. https://doi.org/10.1108/ECAM-05-2022-0492

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics, 24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Gómez-Ríos, A., Luengo, J., & Herrera, F. (2017). A study on the noise label influence in boosting algorithms: Adaboost, GBM and XGBoost. *Lecture Notes in Computer Science, 10334 LNCS*, 268–280. https://doi.org/10.1007/978-3-319-59650-1_23

Gudmestad, O. T. (2019). Waiting on Suitable Weather to Perform Marine Operations. In K. Murali, V. Sriram, A. Samad, & N. Saha (Eds.), *Proceedings of the Fourth International Conference in Ocean Engineering (ICOE2018). Lecture Notes in Civil Engineering* (pp. 3–12). Springer Singapore. https://doi.org/10.1007/978-981-13-3119-0_1.

Hadian, H., & Rahimifard, A. (2019). Multivariate statistical control chart and process capability indices for simultaneous monitoring of project duration and cost. *Computers and Industrial Engineering, 130*, 788–797. https://doi.org/10.1016/j.cie.2019.03.021

T. Hastie R. Tibshirani J. Friedman The Elements of Statistical Learning (2nd ed.). 2009 Springer New York 10.1007/978-0-387-84858-7.

Hazır, Ö., & Ulusoy, G. (2020). A classification and review of approaches and methods for modeling uncertainty in projects. *International Journal of Production Economics, 223*, Article 107522. https://doi.org/10.1016/j.ijpe.2019.107522

Hendiani, S., Bagherpour, M., Mahmoudi, A., & Liao, H. (2020). Z-number based earned value management (ZEVM): A novel pragmatic contribution towards a possibilistic cost-duration assessment. *Computers and Industrial Engineering, 143*. https://doi.org/10.1016/j.cie.2020.106430

Hillson, D. (2009). *Managing Risk in Projects*. Gower Publishing Limited.

Hu, X., Cui, N., Demeulemeester, E., & Bie, L. (2016). Incorporation of activity sensitivity measures into buffer management to manage project schedule risk. *European Journal of Operational Research, 249*(2), 717–727. https://doi.org/10.1016/j.ejor.2015.08.066

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). In *Automated Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5.

Ibbs, W., & Kang, J. M. (2018). Weather-related delay provisions in public transportation construction contracts. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 10*(3), 04518009. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000259

Kelley, J. E. (1961). Critical-path planning and scheduling: mathematical basis. *Operations Research, 9*(3), 296–320. https://doi.org/10.1287/opre.9.3.296

Khamooshi, H., & Abdi, A. (2017). Project duration forecasting using earned duration management with exponential smoothing techniques. *Journal of Management in Engineering, 33*(1). https://doi.org/10.1061/(asce)me.1943-5479.0000475

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. https://doi.org/10.1016/j.ijproman.2013.11.002

Kim, B.-C., Asce, A. M., Reinschmidt, K. F., & Asce, M. (2010). Probabilistic forecasting of project duration using kalman filter and the earned value method. *Journal of Construction Engineering and Management, 136*(8), 834–843. https://doi.org/10.1061/ASCECO.1943-7862.0000192

Klingel, A. R., Jr. (1966). Bias in pert project completion time calculations for a real network. *Management Science, 13*(4), B194–B201.

Kubacka, M., Matczak, M., Kałas, M., Gajewski, L., & Burchacz, M. (2021). Weather risk management in marine survey campaigns for the offshore investment projects in the Polish Exclusive Economic Zone. *Weather, Climate, and Society, 13*(4), 899–911. https://doi.org/10.1175/WCAS-D-20-0168.1

Kuchta, D., & Zabor, A. (2022). A new approach to z-number based earned value management. *Fuzzy Information and Engineering, 14*(4), 361–378. https://doi.org/10.1080/16168658.2022.2156250

Lambrechts, O., Demeulemeester, E., & Herroelen, W. (2008). Proactive and reactive strategies for resource-constrained project scheduling with uncertain resource availabilities. *Journal of Scheduling, 11*(2), 121–136. https://doi.org/10.1007/s10951-007-0021-0

Lipke, W. (2003). Schedule is different. *The Measurable News, 31*(4), 31–34.

Lipke, W. (2004). Connecting earned value to the schedule. *The Measurable News, 1*, 6–16.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *In NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4768–4777).

MacCrimmon, K. R., & Ryavec, C. A. (1964). An analytical study of the PERT assumptions. *Operations Research, 12*(1), 16–37.

Mahmoudi, A., Bagherpour, M., & Javed, S. A. (2021). Grey earned value management: Theory and applications. *IEEE Transactions on Engineering Management, 68*(6), 1703–1721. https://doi.org/10.1109/TEM.2019.2920904

Mahmoudi, A., Javed, S. A., & Deng, X. (2021). Earned duration management under uncertainty. *Soft Computing, 25*(14), 8921–8940. https://doi.org/10.1007/s00500-021-05782-6

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. https://doi.org/10.1287/opre.7.5.646

Martens, A., & Vanhoucke, M. (2017a). The integration of constrained resources into top-down project control. *Computers and Industrial Engineering, 110*, 277–288. https://doi.org/10.1016/j.cie.2017.05.020

Martens, A., & Vanhoucke, M. (2017b). A buffer control method for top-down project control. *European Journal of Operational Research, 262*(1), 274–286. https://doi.org/10.1016/j.ejor.2017.03.034

Martens, A., & Vanhoucke, M. (2020). Integrating corrective actions in project time forecasting using exponential smoothing. *Journal of Management in Engineering, 36*(5). https://doi.org/10.1061/(asce)me.1943-5479.0000806

Martin, O., Ahedo, V., Santos, J. I., & Galan, J. M. (2022). Comparative study of classification algorithms for quality assessment of resistance spot welding joints from pre and post-welding inputs. *IEEE Access, 10*, 6518–6527. https://doi.org/10.1109/ACCESS.2022.3142515

Mentis, M. (2015). Managing project risks and uncertainties. *Forest Ecosystems, 2*(1), 2. https://doi.org/10.1186/s40663-014-0026-z

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Leanpub.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). *Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges* (Vol. 1323,, 417–431. https://doi.org/10.1007/978-3-030-65965-3_28

Moradi, N., Mousavi, S. M., & Vahdani, B. (2017). An earned value model with risk analysis for project management under uncertain conditions. *Journal of Intelligent and Fuzzy Systems, 32*(1), 97–113. https://doi.org/10.3233/JIFS-151139

Mortaji, S. T. H., Bagherpour, M., & Noori, S. (2013). Fuzzy earned value management using L-R fuzzy numbers. *Journal of Intelligent and Fuzzy Systems, 24*(2), 323–332. https://doi.org/10.3233/IFS-2012-0556
Mortaji, S. T. H., Noori, S., & Bagherpour, M. (2021). Directed earned value management based on ordered fuzzy numbers. *Journal of Intelligent and Fuzzy Systems, 40*(5), 10183–10196. https://doi.org/10.3233/JIFS-201248
Moslemi Naeni, L., & Salehipour, A. (2011). Evaluating fuzzy earned value indices and estimates by applying alpha cuts. *Expert Systems with Applications, 38*(7), 8193–8198. https://doi.org/10.1016/j.eswa.2010.12.165
Moslemi Naeni, L., Shadrokh, S., & Salehipour, A. (2014). A fuzzy approach for the earned value management. *International Journal of Project Management, 32*(4), 709–716. https://doi.org/10.1016/j.ijproman.2013.02.002
Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences, 116*(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116
Nadafi, S., Moosavirad, S. H., & Ariafar, S. (2019). Predicting the project time and costs using EVM based on gray numbers. *Engineering, Construction and Architectural Management, 26*(9), 2107–2119. https://doi.org/10.1108/ECAM-07-2018-0291
Naeni, L. M., Shadrokh, S., & Salehipour, A. (2011). A fuzzy approach for the earned value management. *International Journal of Project Management, 29*(6), 764–772. https://doi.org/10.1016/j.ijproman.2010.07.012
Nizam, A., & Elshannaway, A. (2019). Review of earned value management (EVM) methodology, its limitations, and applicable extensions. *The Journal of Management and Engineering Integration, 12*(1), 59–70.
Noori, S., Bagherpour, M., & Zareei, A. (2008). Applying fuzzy control chart in earned value analysis: A new application. *World Applied Sciences Journal, 3*(4), 684–690.
Orangi, A., Palaneeswaran, E., & Wilson, J. (2011). Exploring delays in victoria-based astralian pipeline projects. *Procedia Engineering, 14*, 874–881. https://doi.org/10.1016/j.proeng.2011.07.111
Pajares, J., & López-Paredes, A. (2011). An extension of the EVM analysis for project monitoring: The cost control index and the schedule control index. *International Journal of Project Management, 29*(5), 615–621. https://doi.org/10.1016/j.ijproman.2010.04.005
Pellerin, R., & Perrier, N. (2019). A review of methods, techniques and tools for project planning and control. *International Journal of Production Research, 57*(7), 2160–2178. https://doi.org/10.1080/00207543.2018.1524168
J.G. Pérez M. del Martín M. L., García, C. G., & Sánchez Granero, M. Á. Project management under uncertainty beyond beta: The generalized bicubic distribution Operations Research Perspectives 3 2016 67 76 10.1016/j.orp.2016.09.001.
Pewdum, W., Rujirayanyong, T., & Sooksatra, V. (2009). Forecasting final budget and duration of highway construction projects. *Engineering, Construction and Architectural Management, 16*(6), 544–557. https://doi.org/10.1108/09699980911002566
Ponz-Tienda, J. L., Pellicer, E., & Yepes, V. (2012). Complete fuzzy scheduling and fuzzy earned value management in construction projects. *Journal of Zhejiang University: Science A, 13*(1), 56–68. https://doi.org/10.1631/jzus.A1100160
Pritsker, A. A. B. (1966). *GERT-Graphical evaluation and review technique*. https://ntrs.nasa.gov/search.jsp?R=19670022025.
Radhakrishnan, B. D., & Jaurez, J. J. (2021). Explainable Artificial Intelligence (XAI) in Project Management Curriculum: Exploration and Application to Time, Cost, and Risk. *2021 ASEE Virtual Annual Conference Content Access*. https://peer.asee.org/37135.
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778.
Robnik-Sikonja, M., & Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering, 20*(5), 589–600. https://doi.org/10.1109/TKDE.2007.190734
Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access, 8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199
Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
Sackey, S., Lee, D. E., & Kim, B. S. (2020). Duration estimate at completion: Improving earned value management forecasting accuracy. *KSCE Journal of Civil Engineering, 24*(3), 693–702. https://doi.org/10.1007/s12205-020-0407-5
Salari, M., Bagherpour, M., & Wang, J. (2014). A novel earned value management model using Z-number. *International Journal of Applied Decision Sciences, 7*(1), 97. https://doi.org/10.1504/IJADS.2014.058037
Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2019). *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Vol. 11700). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6.
Schonberger, R. J. (1981). Why projects are "always" late: A rationale based on manual simulation of a PERT/CPM network. *Interfaces, 11*(5), 66–70.
Shafer, G. (1976). A mathematical theory of evidence. *Princeton University Press*. https://doi.org/10.1515/9780691214696
Shapley, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games (AM-28)* (Volume II, pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018.
Song, J., Martens, A., & Vanhoucke, M. (2022). Using earned value management and schedule risk analysis with resource constraints for project control. *European Journal of Operational Research, 297*(2), 451–466. https://doi.org/10.1016/j.ejor.2021.05.036
Štrumbelj, E., & Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research, 11*, 1–18. http://www.ailab.si/orange/datasets.psp.
Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x
M. Vanhoucke Measuring Time Improving Project Performance Using Earned Value Management Vol. 136 2010 Springer, US 10.1007/978-1-4419-1014-1.
Vanhoucke, M. (2019). Tolerance limits for project control: An overview of different approaches. *Computers & Industrial Engineering, 127*, 467–479. https://doi.org/10.1016/j.cie.2018.10.035
Vanhoucke, M., & Batselier, J. (2019). Fitting activity distributions using human partitioning and statistical calibration. *Computers & Industrial Engineering, 129*, 126–135. https://doi.org/10.1016/j.cie.2019.01.037
Vanhoucke, M., & Colin, J. (2016). On the use of multivariate regression methods for longest path calculations from earned value management observations. *Omega, 61*, 127–140. https://doi.org/10.1016/j.omega.2015.07.013
von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., … Schuecker, J. (2021). Informed Machine Learning – A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering, 1–1*. https://doi.org/10.1109/TKDE.2021.3079836
Votto, R., Lee Ho, L., & Berssaneti, F. (2020). Multivariate control charts using earned value and earned duration management observations to monitor project performance. *Computers and Industrial Engineering, 148*. https://doi.org/10.1016/j.cie.2020.106691
Wajdi Hammad, M., Abbasi, A., & Ryan, M. J. (2018). *Developing a Novel Framework to Manage Schedule Contingency Using Theory of Constraints and Earned Schedule Method*. https://doi.org/10.1061/(ASCE)CO.1943.
Wang, R., Asghari, V., Cheung, C. M., Hsu, S. C., & Lee, C. J. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. *Automation in Construction, 134*. https://doi.org/10.1016/j.autcon.2021.104080
Wauters, M., & Vanhoucke, M. (2014). Support vector machine regression for project control forecasting. *Automation in Construction, 47*, 92–106. https://doi.org/10.1016/j.autcon.2014.07.014
Wauters, M., & Vanhoucke, M. (2016). A comparative study of Artificial Intelligence methods for project duration forecasting. *Expert Systems with Applications, 46*, 249–261. https://doi.org/10.1016/j.eswa.2015.10.008
Wauters, M., & Vanhoucke, M. (2017). A Nearest Neighbour extension to project duration forecasting with Artificial Intelligence. *European Journal of Operational Research, 259*(3), 1097–1111. https://doi.org/10.1016/j.ejor.2016.11.018
Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety, 142*, 399–432. https://doi.org/10.1016/j.ress.2015.05.018
Weld, D. S., & Bansal, G. (2019). The challenge of crafting intelligible intelligence. *Communications of the ACM, 62*(6), 70–79. https://doi.org/10.1145/3282486
Willems, L. L., & Vanhoucke, M. (2015). Classification of articles and journals on project control and earned value management. *International Journal of Project Management, 33*(7), 1610–1634. https://doi.org/10.1016/j.ijproman.2015.06.003
Williams, T. (2002). *Modelling complex projects*. John Wiley and Sons.
Zohoori, B., Verbraeck, A., Bagherpour, M., & Khakdaman, M. (2019). Monitoring production time and cost performance by combining earned value analysis and adaptive fuzzy control. *Computers and Industrial Engineering, 127*, 805–821. https://doi.org/10.1016/j.cie.2018.11.019
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x