POLITECNICO MILANO 1863

School of Industrial Engineering
Department of Mechanical Engineering

Probabilistic estimate at completion of a project through the integration of the Kalman filter with the Earned Value Management system methodology

Advisor: Prof. Franco CARON

Student:
Fabio Garzanti
Student number 876114

Academic Year 2017-2018

Acknowledgments

Prima di tutto vorrei esprimere la mia gratitudine al professor Franco Caron per il suo continuo supporto, per il prezioso aiuto e l'incoraggiamento dimostratomi durante lo svolgimento di questa tesi.

Vorrei ringraziare i miei amici, capaci di rendere meno amare le sessioni di esame e sempre al mio fianco nelle occasioni di festa.

Un merito particolare va a Patrizia, che mi ha supportato e sopportato lungo questo percorso.

Infine ringrazio la mia famiglia per l'amore e il sostegno dimostratomi in questi cinque anni. I miei genitori che mi hanno dato la possibilita' di vivere questa esperienza e mia sorella Giulia sempre pronta a guidarmi con i suoi preziosi consigli.

Contents

List of Figures

List of Tables

Summary

Although the Earned Value Management system (EVM) is common practice in the evaluation of project cost performance, this method runs into estimate errors early in the project, due to its small sample size and, besides, its outcomes are deterministic and do not consider that some errors could affect input. In order to give a new perspective, a new forecasting method is here developed, based on the Kalman filter and the EVM frameworks: it is able to offer probabilistic outlooks of the project final cost and it may be applied since the very beginning without lower degrees of correctness. The algorithm is then applied in three oil and gas projects to evaluate its performances.

Keywords: Estimate at completion, Cost variance, Kalman filter, Probabilistic approach.

Introduction

An accurate estimate at completion from the onset of a project execution is a step of primary importance which tends to a serviceable and proactive project governance: indeed, it allows the adoption of corrective actions highly affecting the projects performances and crucial to pursue the cost, time and quality objective. The project, owing to its inherent uncertainty and unicity, along with

a great number of stakeholders, makes the adoption of more performing models able to achieve reliable estimate at completion an imperative issue. However, the nowadays techniques show not few shortcomings in guaranteeing faultless results in the starting project phases. Additionally, the matter is more troubling taking in to account that, though the articulate process to assess the data necessary for the implementation and the absence of a guideline, the possibility that the information used may presents errors is not considered. Therefore, the ultimate aim of this dissertation is to develop an estimate at completion method reflecting on the possible error presence and refining the accuracy of the outcomes from the beginning of the project execution. Thus, a new model for the estimate at completion (K-EAC) is presented, based on the integration of measurements and an analytical system model through the Kalman filter, delivering a probabilistic approach which acknowledges the possible errors both in the measurement and in the planning phase. Lastly, this method is exploited into three projects in the Oil and Gas sector: the main features of these projects are extreme complexity and uncertainty which lead to a testing situation. The outcomes attest improvements in accuracy, evaluated as mean absolute percentage error (MAPE), with respect to the traditional technique (Earned value management system with three-periods moving average). According to the analysis, the accuracy gain is achieved throughout the initial project phase and shows an improvement in timeliness as well. Furthermore, the technique added value is the probabilistic project status description, the cost variance is no longer given as a punctual value, but with its distribution; the latter is then used to assess the estimate at completion, provided with three values: more probable, optimistic and pessimistic. The estimate is thus enhanced with the

indication of how reliable it can be, improving the quality of the information available to the decision maker.

Literature review

The role of forecasting

Forecasting has a key role in project management and it highly affects crucial phases like planning, controlling and risk monitoring. Before the project execution, forecast is used in planning in order to develop the project baseline plan, that is to say the guide to achieve the work on time and within the budget. During the project execution, instead, the performances are monitored and employed to rectify the estimate of the work remaining, so that they may reveal whether any corrective actions are needed and to what extent. It assumes great importance when the future is so unpredictable that a few aspects of the project may introduce a degree of uncertainty that cannot be tested; that is why, on these occasions forecasting proves to be the most important instrument to rely on for a decision-making process. This is a common condition in project management where, although the project objective is intelligible since the onset, the final outcomes will be available just at the end of the project. Indeed, uncertainty, as it is stated in the PMBOK [1], is an intrinsic feature of any project: frequent changes, unexpected events, several players involved and stakeholders' needs gradually evolving create the environment in which the project manager must take decisions which are rich in variability and unpredictability. Unfortunately, probabilistic forecasting approaches are not very diffused and, what is worse, experts are not highly knowledgeable in this field as they should be, because of the lack of easy-to-implement models providing good performances and not requiring huge amount of data.

Evaluation criteria for forecasting methods

The comparison between forecasting methods, both in case of deterministic and probabilistic, is a tough challenge for forecasters. The problem has occupied researchers and sector experts since 1969, when Bates and Granger [2] published their seminal work. So far several studies have focused on the identification of the best parameters for the comparison. In order to take hold in common use, the method needs to meet several requirements. Firstly, the method should require neither hard nor expensive input data to be collected. Secondly, the method should be simple in the implementation and in the interpretation of the outcomes. Thirdly, the method has to guarantee good performances with regard to the following elements: accuracy, timeliness, stability, flexibility and absence of systematic errors. Since these features are hard to be achieved simultaneously, a trade-off between them is highly recommended.
Whereas stability and flexibility are laborious to quantify, a quantitative method will be used to compare accuracy to timeliness. Teicholz [3], collating 121 construction projects, sets forth a new indicator as a measure of accuracy. Along with the classical statistical methods, as MAPE, the accuracy could be defined by the area between the actual final cost and the path of the estimate at completion plotted against the percentage progress. Moreover, Teicholz probed a procedure to evaluate the timeliness, still today the real challenge of forecasting method: the author reviewed it as the accuracy accomplished in the first half of the project.

Earned Value Management

Introduction to EVM

Nowadays the commonest technique in cost performance forecasting is the "Earned Value Management". The strong point of the method is the simple implementation: indeed, a correct application demands two only requirements, that is to say a project status constant monitoring of the cost bore and the progress succeeded in the project; the second requirement is the uniformity in measurements and the selection of common criteria able to grant the compatibility with results. Over the years lots of formulations have been presented to reach a better estimate quality, especially the most employed version bases the forecasting on monthly values to catch trends which may be difficult to be discerned, improving the outcomes in terms of accuracy.

Formulation

The technique is based on three elements which are fundamental components of all the metrics:

- BCWS budget cost of work scheduled;
- ACWP actual cost of work performed;
- BCWP budget cost of work performed or earned value. It represents the planned cost to perform the activities completed up to now.

Monitoring at regular intervals these elements, an assessment of the project status is attainable. Thus, two indicators are used to summarize it: *cost variance* and *cost performance index*.

$$CV_{TN} = BCWP_{TN} - ACWP_{TN} \qquad CPI_{TN} = \frac{BCWP_{TN}}{ACWP_{TN}}$$

The subscript TN remarks upon the fact that the element is evaluated at time now, so it stresses that all the values are time dependent. In the following formulas the subscript will be omitted according to the notation in literature. While the cost variance could be determined as the difference in cost respect to what was planned, the cost performance index could be seen as a measure of cost efficiency [4]. To better understand the nomenclature, a list of all the acronyms is presented in table 1.

The standard EVM forecasting is grounded in the hypothesis that cumulative performance indices (CPI_c and SPI_c calculated with cumulative value of $BCWS$, $BCWP$, $ACWP$) will not only be indices of the past but of the future performances as well. The latent assumption is that the detected variances are caused by structural problems that will be present until the project ends. Knowing that, the estimate is assessed by summing the already bore cost plus the work remaining adjusted after considering the performances.

$$EAC = ACWP + \frac{BCWR}{PF}$$

CPI_c, as cumulative value, after the 20% of the development will not vary more than 10% tending to stability. Over the years a lot of alternatives of the

$ACWP$		Actual cost of work performed
$BCWS$		Budget cost of work scheduled
$BCWP$		Budget cost of work performed
$BCWR$	$= BAC - BCWP$	Budget cost of work remaining
BAC		Budget at completion: initial cost quote
ETC		Estimate to completion: forecast at time now of the project cost to be sustained from time now to the projects end.
PAC		Planned at completion: initial duration quote.
EAC	$= ACWP + ETC$	Estimate at completion: total final cost forecast at time now.
CV	$= BCWP - ACWP$	Cost variance
SV	$= BCWP - BCWS$	Schedule variance
CPI	$= BCWP/ACWP$	Cost performance index
SPI	$= BCWP/BCWS$	Schedule performance index
PF		Performance factor

Table 1: EVM nomenclature

standard formula have been presented: Anbari [4] and Christensen [5] conducted an extensive research on their applicability. Among the various techniques, the most used in real practice sees as performance factor the cost performance index evaluated as a moving average over the last three periods. The choice comes from the good results achieved that reflect a well-balanced trade-off between the too stable CPI_c and the too sensible CPI_{pm}.

Earned Value Management theory is too extended to be fully explained here and a more complete introduction can be found in other sources (Flemming and Koppelman [6]).

Critics to EVM

Earned value management (EVM) has provided methods for final outlooks and, largely, these have never undergone any attempts of improvement since their formulation. Consequently, the available methods are often oversimplified, based on inconsistent hypotheses and they sometimes provide unreliable and thus not usable results. Three main problems are recognisable: the methodology does not consider the presence of possible errors in the project description, the high sensibility during the project initial phase and the prolongated effect of errors in the estimate.

Firstly, the prior EVM restriction is the deterministic nature of the outcomes, evaluated without pondering the presence of possible errors. The EAC is evaluated by using input values which are measured on the fieldwork but this does not necessarily mean no margin of error. Although assessing the expended money up to a given moment is effortless, it is rather a demanding task to detect the value of the Budget Cost for Work Performed ($BCWP$), given that the usual way to calculate it is to multiply the project budget by the achieved progress. The process complexity, and thus the errors, arise when the overall performed work needs to be quantified and synthetized in a percentage value. The presence of countless activities that require different amount of efforts and resources, the combination of several disciplines and the absence of a strict guideline will never lead to a univocal and flawless percentage progress evaluation. The situation got worsen since the percentage progress evaluation is based on the comparison with the project baseline, that could not reflect the real work needed to complete the project. In fact, errors, a lack of experience, and more often a political and competitive pressure which is usual for companies working in a strongly competitive sector to underestimate risks and overestimate opportunities in order to win the bidding could cause an inaccurate planning. Because of the intrinsic project uncertainty and the possible estimation errors, a probabilistic result is needed because it could be very beneficial for the decision maker to know how reliable the estimate is.

Secondly, another problem is related to the low level of accuracy, especially in the early stage of the project, when the small sample size of data the estimate is based on do not allow to assess a statistical reliable forecasting [7]. The basic EVM estimate idea is that the future performances will be equal to the detected ones hereinbefore; afterwards, they will be used to modify all the remaining work. It is easy to understand that, when the project is at an embryo stage, the remaining work is an important part of the whole so, despite a small drop in performances, even not systemic, a huge impact on the final estimate will occur. This problem is more significant at the very first stages of the project where little observation makes the performance estimator too responsive.

Thirdly, the last issue is about the prolongated effect of errors in the estimate. As said before, the EAC is evaluated adding at the present sustained cost the one that has still to be sustained but modified by a cost performance factor. In this condition the effect of a short cost performance fluctuation is enormous, and the problem is even more serious since the fluctuation could be caused by a measurement error as explained before. In this case, even if the performance index is correctly assessed in the following observation, the error is taken in to account still for other two periods, cause the three moving average system.

Kalman estimate at completion model

A new probabilistic forecasting method, Kalman estimate at completion (K-EAC), is shown below to supervise the project performance and foresee the project estimate at completion with the related probability distribution. As a novelty, the model is built through the implementation of the classical Kalman filter formulation in the project control field exploiting the frameworks of the EVM model. As a consequence, the application of this new tool provides several advantages: first of all, it provides the results, both the cost variance and the estimate at completion, not as punctual values but with the related distribution probabilities, leading the decision maker to the way the outcomes could be trusted. Second, the method conveys to high levels of accuracy and timeliness. Third, it takes into account a few essential problems, such as the quality and reliability of the input data.

The Kalman filter

Kalman filter, also known as linear quadratic estimation, is an algorithm that uses noisy observation to estimate the true but hidden state of the system. The filter was named after Rudolf E. Kalman, one of the first developers of the algorithm in 1960. Since its publication, the applications covered a large range of fields from technology to finance and its common uses are for guidance, navigation and control of vehicles, particularly in the aircraft and spacecraft fields [8]. Furthermore, it is widely applied in the time series analysis area especially in signal processing or econometrics.

The filter works iteratively creating a learning loop composed of three main recursive steps which are well explained in Figure 1: prediction, measurement and posterior estimation.

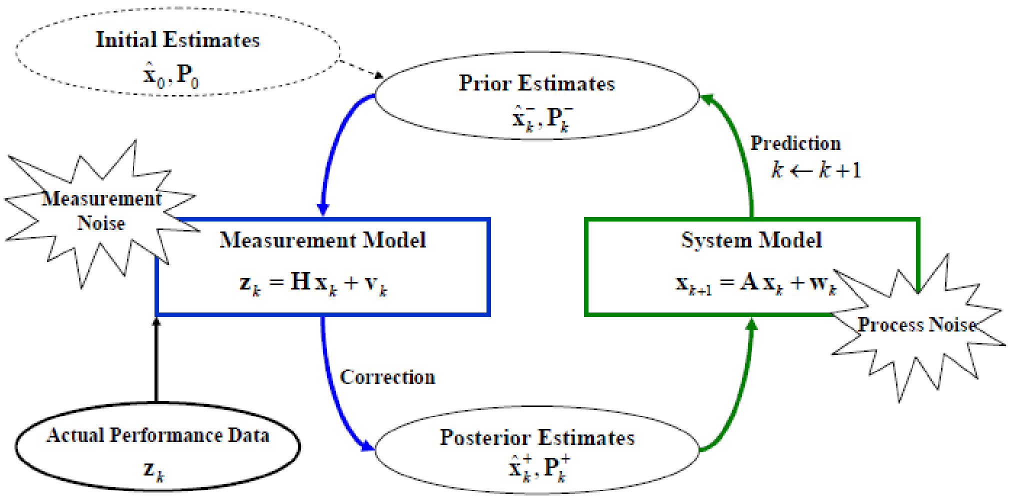

Figure 1: Recursive algorithm of Kalman filter

The state of a dynamic system is described by two sets of variables: the state variables and the error covariance variables. The state variables directly represent the system parameters while the error covariance is the indicator of the estimate uncertainty.

The states and covariance matrix are updated by two stochastic models, the measurement model updates the previous estimate under the evidence of the observation and, instead, the system model foretells the future system state at the following time step.

Item by item, during the prediction step the algorithm produces a prior estimate of the current state variables and their uncertainty. Later, the measure of the state variables is performed, unfortunately the outcomes will be necessarily corrupted by instrument noise and measurement errors. The last passage is the posterior estimate, performed as a weighted average between the a priori estimation and the measure, giving more importance to the less uncertain factor. An updating phase is necessarily required before restarting the loop. The algorithm is recursive and can run real time using the present measure and the previously evaluated state as the only input.

The Kalman filter theory is too wide-ranging to be explained here; a good introduction could be found in essays (Zarchan and Musoff [8], Brookner [9], Welch and Bishop [10]). The framework has been extensively studied all over the world and many notations are currently in use. In order to avoid confusion, this dissertation follows the Welch and Bishop one [10].

The model

Kalman EAC general framework

Based on the general theory of Kalman filter, the application to forecast the project EAC has been developed by analogy with the missiles tracking application. The main steps of the algorithm are synthetized in the block diagram in Figure 2.

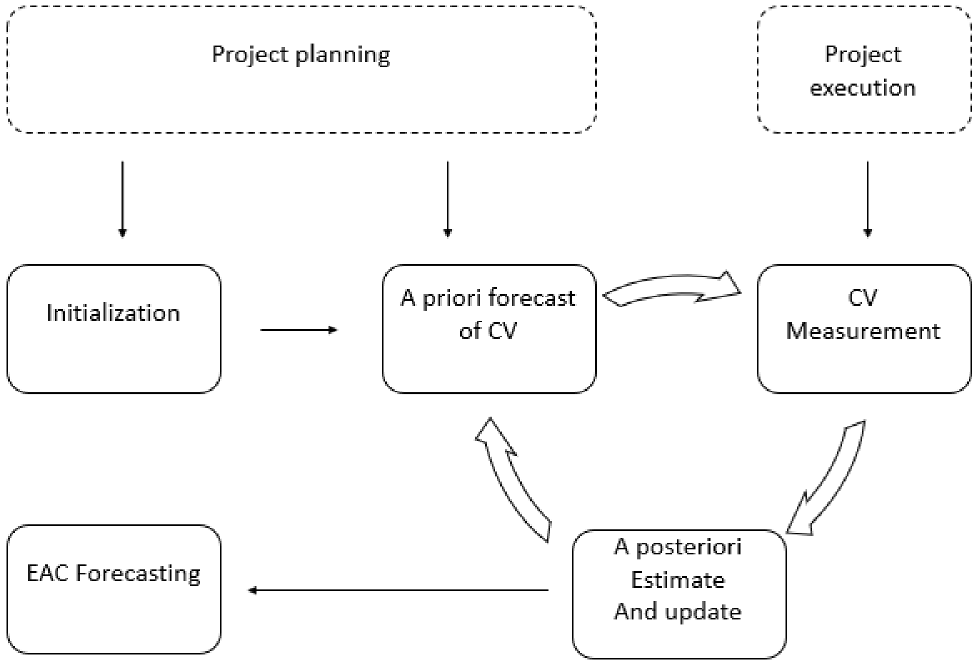

Figure 2: Kalman EAC forecasting algorithm

The Kalman estimate at completion (K-EAC) is based on a recursive learning cycle that aims to detect the real but hidden states of the ongoing project, the output information is then used while performing a probabilistic estimate at completion. At every time the algorithm is applied to detect the real distribution of the cost variance by using a trade-off between a prior estimate from a linear system model and the detected measure corrupted by errors. This balance is weighted depending on how reliable the two parameters can be, considering both the uncertainty of the model and the errors in the measurement (problems previously described). This consideration will result in a posterior state estimate and its probability distribution, used according to the EVM technique for detecting a probabilistic EAC. The algorithm works following some key points:

during the project planning phase, it is important to develop the required input:the baseline progress curve using project plans, and the prior probability distribution of the final project cost;
prior system state estimate: a system model is developed using the baseline curve to predict the state and covariance variables at the next reporting time;
measurement: during the execution, the project performances are measured and periodically accounted as cumulated progress;
posterior system estimate: prior estimate and measurement are used for the estimate of the real state of the system and its probability distribution;
forecast: the obtained probabilistic results are used with EVM technique for the assessment of the EAC distribution;
update: the variables are updated, and the algorithm is ready to be applied on the next time instant.

Required input

Like all the forecasting techniques, K-EAC requires some input information too, briefly shown in Figure 3. In addition to the actual data measured at every iteration, that are the actual cost spent and the cumulated progress which are both vital to assess the value of the cost variance, the algorithm requires prior information about the project and the environment in which it is developed as the baseline curve, the planned at completion (PAC), the budget allocated to the project (BAC) and a prior distribution of the final cost.

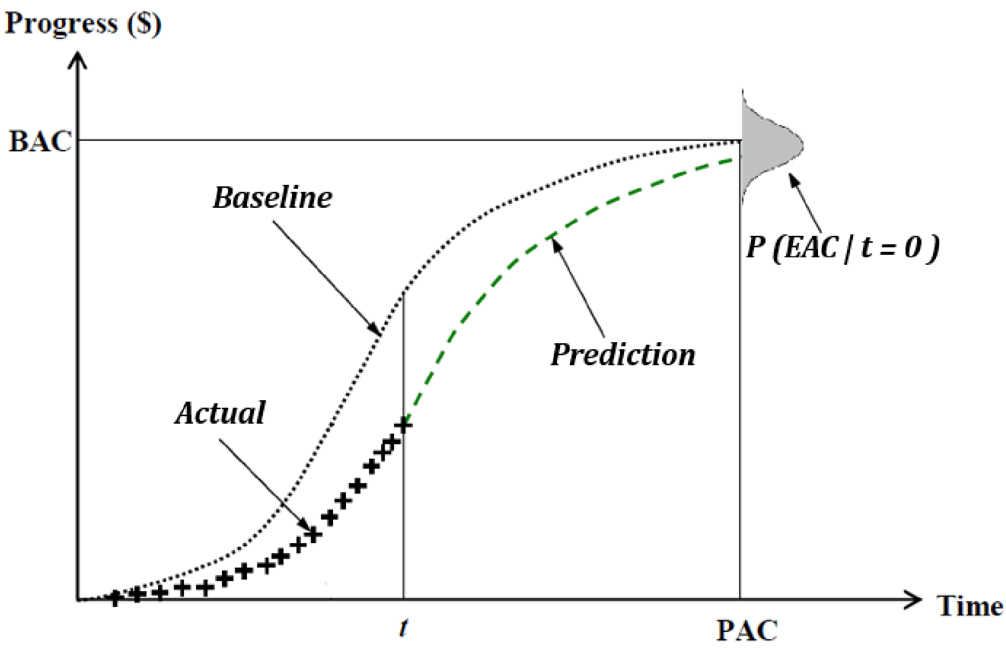

Figure 3: Required inputs

The need for the final cost prior distribution at the very beginning of the project originates from the probabilistic nature of the Kalman filter. It could be detected in several ways, but the easiest one is to approximate a triangular distribution based on more probable, optimistic and pessimistic cost.

The model

The main element and phases of the K-EAC model are shown in Table 2.

Components	Equations	Descriptions
State vector	$x_k = \begin{Bmatrix} CV_k \\ \frac{dCV_k}{dt} \end{Bmatrix}$	CV_k is the cost variation and it is defined as the Earned value minus the actual cost at the k iteration.
Dynamic system model	$x_k = A_k \cdot x_{k-1} + w_{k-1}$ $A_k = \begin{bmatrix} 1 & \Delta T_k \\ 0 & 1 \end{bmatrix}$	A_k is the transition matrix.w_{k-1} is the random process noise vector acting on $\frac{dCV_k}{dt}$
Measurement model	$z_k = H \cdot x_k + v_k$ $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$	H is the measurement matrix, v_k is a vector representing the measurements noise.
Prediction process	$\hat{x}_k^- = A_k \cdot \hat{x}_{k-1}^+$ $\hat{P}_k^- = A_k \cdot \hat{P}_k^- \cdot \hat{A}_k^T + Q_{k-1}$	Calculation of the prior estimate $\hat{x}_k^-$, and of the prior error covariance matrix $\hat{P}_k^-$. Q_{k-1} is the process noise covariance matrix
Kalman gain	$K_k = \frac{\hat{P}_k^- \cdot H^T}{H \cdot \hat{P}_k^- \cdot H^T + R_k}$	K_k is the Kalman gain at the k iteration, which is determined to minimize the posterior error covariance matrix. R_k is the measurement error covariance matrix.
Updating process	$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H \cdot \hat{x}_k^-)$ $\hat{P}_k^+ = [I - K_k \cdot H] \cdot \hat{P}_k^-$	Calculation of the posterior estimate.

Table 2: K-EAC model main components

State vector

The state vector is the objective of the estimating process and describes the status of the project costs:

$$x_k = \begin{Bmatrix} x_{k,1} = CV_k \\ x_{k,2} = \frac{dCV_k}{dt} \end{Bmatrix}$$

It is crucial to consider that x_k does not represent only the real but also the hidden state of the system; our knowledge will be limited to its estimate that will be indicated by $\hat{x}_k$. In particular, two estimates are performed at every time instant: $\hat{x}_k^+$ the prior estimate based on the system model and $\hat{x}_k^+$ the posterior estimate conducted after the measurement process.

Filter initialization

Since the iterative nature of the filter, some parameters and the initial variables values have to be set before the first iteration. At first, the initial state vector and the covariance error matrix values are set equal to zero. In addition, the values of the process noise covariance matrix and the measurement error matrix have to be estimated in advance. This process, usually called "filter initialization", is a very challenging phase since it is performed when no observation from the project is available. The importance of this phase has been widely highlighted in literature by many studies [11] [12] [13] [14], which could provide an extensive dissertation; in this section only the initialization process related to the K-EAC is going to be discussed.

First, the initial variables, both the state vector and the error covariance matrix, are set to zero.

$$\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \hat{P}_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

When the project begins, the method has or should have a well-defined starting point concerning the work which is supposed to be done, the starting time and the initial cost that is reasonable to be set equal to zero, thus there cannot be any chance to have cost variance and the state vector is initialized to zero. Following the same line of reasoning, there is no doubt about the state status, since no progress are achieved, and no money are spent, the absence of uncertainty is reflected into a null error covariance matrix.

The process noise covariance matrix, acting directly in the Kalman gain (K_k), takes into account the system model uncertainty due to lack of information or presence of errors [8]. The matrix is modelled to act on the CV_k derivative over one interval [9]. The process noise covariance matrix Q_k is evaluated as the covariance of the process noise vector w_k:

$$Q_k = Cov[w_k] = \overline{\begin{bmatrix} 0 \\ w_k \end{bmatrix} \begin{bmatrix} 0 & w_k \end{bmatrix}} = [w_k][w_k]^T$$

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & \overline{w_k^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_w^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

The diagonal terms represent the variances of the state variables, the extra diagonal instead represents the covariance values. If only random errors of each state variables are considered, the off-diagonal terms are zero [15] according to the hypothesis developed in the previous chapter. Indeed, q represent the process noise variance, measure of the model uncertainty, and directly acts on the filter convergence. The more the value approaches the zero, the more the system model makes correct estimates of the future; conversely, a high value accords with an increase of the uncertainty affecting the process. The variances are supposed to be constant, not for any rational or empirical results but because there is no information supporting an alternative interpretation.

The value of q is assessed in order to make the model uncertainty coherent with the users prior estimate of the project final cost distribution. The user provides as input the expected project duration (PAC) and the distribution of the expected final cost expressed with the mean μ_c and the variance of the distribution σ_c^2. These are used in an inverse Kalman forecasting algorithm to determine the value of q. In detail, the algorithm, which is based only on the system model, works equivalently to set the gain K equal to zero for all the project duration. In the analysed case where measurements are performed at a constant time interval, the resulting equations are:

$$x_k^- = Ax_{k-1}^+$$

$$P_k^- = AP_{k-1}^+A^T + Q_{k-1}$$

$$K_k = 0$$

$$x_k^+ = x_k^-$$
$$P_k^+ = P_k^-$$

Provided that the project lasts as planned (PAC) and the model uncertainty at the beginning of the forecast process is equal to the users prior estimate of the project final cost variance:

$$P_{k=PAC}^+(1,1) = P_{k=PAC}^-(1,1) = \sigma_c^2$$

It is possible to evaluate q since it is the only variable.

The last element to be set is the measurement error matrix R_k. It represents the accuracy of the measurement and it is expressed as the covariance matrix of the measurement noise vector v_k.

$$R_k = Cov[v_k]$$
$$R_k = \overline{[v_k]\,[v_k]^T}$$
$$R_k = \overline{[v_k]^2} = [\sigma_v^2] = [r]$$

Where r is the measurement error variables and takes into consideration the variance of the measurement error σ_v^2 , r influences the forecasting method sensibility, in particular if it approaches zero, the Kalman gain will increase and, as a consequence, the measured quantity will have higher impact on the posterior estimate. Vice versa, high r will decrease the gain when making the posterior state estimate trusting more the prior estimate than the measured performance. In order to set the value, the program evaluation review technique (PERT) is used [16] [17] [18] and a three-point estimate for the measurement error. The user needs to define the maximum possible measuring error, the variance of the error is evaluated thanks to the PERT technique.

$$Maximumerrorv_k = Emax$$

$$Minimumerrorv_k = -Emax$$

$$\sigma_v^2 = [\frac{Emax - (-Emax)}{6}]^2 = \frac{Emax^2}{9}$$

The value could be adjusted by the project manager to correctly fit different types of project that are developed in different environment. If need be, modifying r is possible to attach either great or little importance to the measured performance with respect to the model estimate.

Application of the forecasting model to three oil and gas projects

In this chapter it is going to be presented the application of the K-EAC model to three real cases in the oil & gas sector. The three projects have already been accomplished, thus they are going to be thoroughly described by focusing on the scope of work, the physical and economical progress achieved during their execution. After that, the forecasted results will be analysed and to understand the model added value they will be compared to the EVMS (Earned Value Management System) methodology outputs that represent the state of the art in the field of forecasting techniques. More precisely, the three-month moving average EAC will be exploited because of its remarkable frequency: nowadays, seeing as how it succeeds in achieving the best outcomes, it is the most used EVM version by project managers. The evaluation focuses its attention on the most challenging forecasting requirements, accuracy and promptness, which are all the key points for a flawless decision-making process. As previously highlighted, a recent aspect of the method is the probabilistic behaviour: this enables a better project status description supporting the project manager choices in the project execution.

The data in the K-EAC implementation are provided by a company working worldwide in the oil & gas sector. The firm usually operates in foreign countries, where the reservoirs of hydrocarbons are located, creating a partnership with the host country. Not only do they acquire a research and extraction authorization upon payment of royalties, but they directly let the host country share in the production earnings: in this scenario the company is seen as an administrator of the oil reservoir, still owned by the host country. Nevertheless, it is noteworthy that this sector represents a few limiting cases, because of its huge size and highly intrinsic complexity. In addition to these difficulties, from the point of view of the project environment, it is recognizable that some hard criticalities can increase the uncertainty level, the high number of stakeholders involved and the political, financial and climate related factors that highly influence the project performances. Thus, the activities that concern the oil exploitation need a high-level planning and risk management system. This is the reason why the push towards more performing techniques in the risk minimization requires the introduction of specific, more complex and diversified methodologies that working in synergy and in an integrated manner to reproduce a model which can be as close as possible to reality. In any oil & gas company the control of the physical and economical progress has a central role to understand when the plant is ready to start up the production, that is to say to realize the moment it will begin making profits out of its activity. Considering the huge financial capital deployed in the project development, it is of great importance to have a control system that allows the decision maker to pursue effective and prompt decisions.

Typical oil & gas sector projects

In the oil & gas sector it is possible to identify three clusters of comparable in terms of size and scope of work: offshore, onshore and subsea. A brief introduction of the type in order to contextualize the model application will be shown below.

Offshore projects involve operations of plants construction and installation for

drilling and hydrocarbons extraction in the sea. The adopted platforms are characterized depending on the sea depth and the structure typology, floating or fixed. The projects include also sea lines installation, ducts for the transport of the extracted material to the storage units in the land.
Onshore projects, instead, are characterized by the construction and installation of plants for directly drilling and extracting on land, with the installation of underground pipelines for the material transportation. At first, the oil extracted from the drilling well is stocked in loco and later it is transported by pipelines to the refinery, only after being sent to a treatment plant. Gaseous hydrocarbons follow the same pattern until the treatment plant after being directly sent to the user by methane pipeline. Projects belonging to this cluster present some criticality during their execution. The main reasons are:

- Authorization and permissions are given with higher difficulties than it is generally for the other two categories, in the country soil, authorities are more reluctant to authorize plants that could be result invasive.

- Local subcontractors are far less reliable and skilled labour is quite difficult to source. Political agreements with the host country bind the company to exploit exclusively unskilled local manpower: as a consequence, the quality plummets, the costs exceed and the delays run and grow.

Nevertheless, that criticalities are often underestimated. The company usually adopts an aggressive planning behaviour to win the contract bidding.
Subsea projects concern the construction of submarine extraction plants. This solution is adopted when the offshore platform cannot be used both for technical and economical unfeasibility. In case of multiple submarine extraction points,

they have to be linked with flowlines, the same connections have to be installed also to connect the wells to the storage platforms or the onshore stoking plants. Compared to the two other categories, this project type is generally characterized by a better financial exposure since the few companies performing this kind of work employ skilled manpower, more reliable in terms of cost, time and quality of the work performed than the local labour.

Project description

The presented forecasting model is applied to three real projects, one for each of the identified category. Analysing already completed cases is possible to assess the physical and economic project life before the project application. In order to respect the company privacy, the cases of study will not be named with the real project name, but with one established by convention:

- Subsea cluster: Case A;
- Offshore cluster: Case B;
- Onshore cluster: Case C.

The A Case

The first analysed case is part of the subsea category. The project consists in the setup of an FPSO (*Floating Production Storage and Offloading Unit*), a ship whose aim is to extract, to stack and to perform the oil preliminary treatment operation.

The objective of work is composed by three main elements shown in Figure 4.6

- the installation of the FPSO, a permanently moored ship, is needed to stock the drilled oil and to start the preliminary treatment phases;
- the installation of flexible sea-lines that will connect the ship to three existing wells;
- the installation of a submarine umbilical control system that regulates the wells and the valves that enable the oil to flow from the seabed to the FPSO, in order to prevent oil leakages.

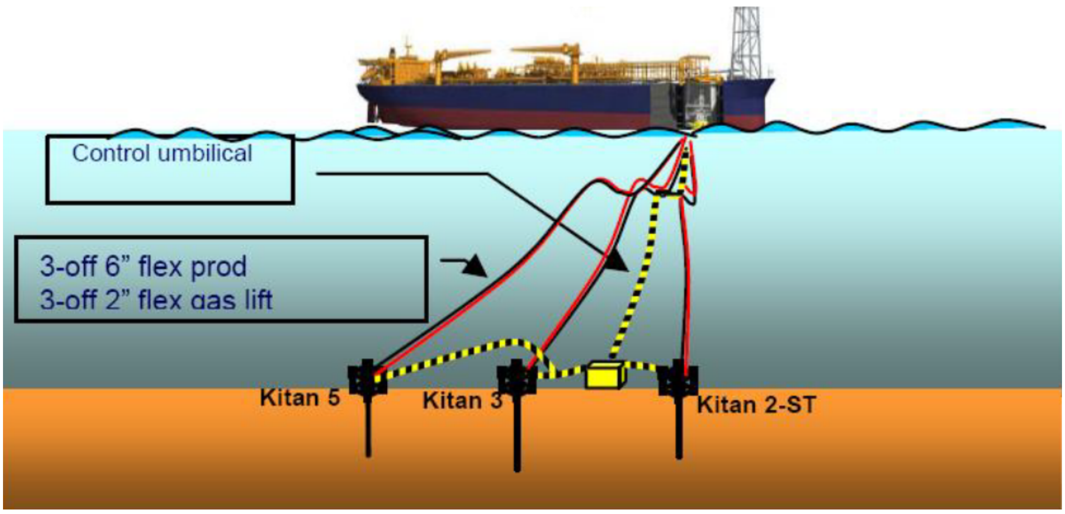

Figure 4: Scope of work project A

The working site is in the Australian North Sea, an area where the company has never worked before: this inexperience of the area conveys more uncertainty to the project. In this case the company does not known the standard subcontractors performances. This could be an issue during the concept definition where the time and the cost of activities have to be estimated to meet the contract obligations.

The A Case is an accomplished project, all the data at our disposal are used to run the model. In Figure 5 and Figure 6 the project history is presented, with both the physical and the economical progress.

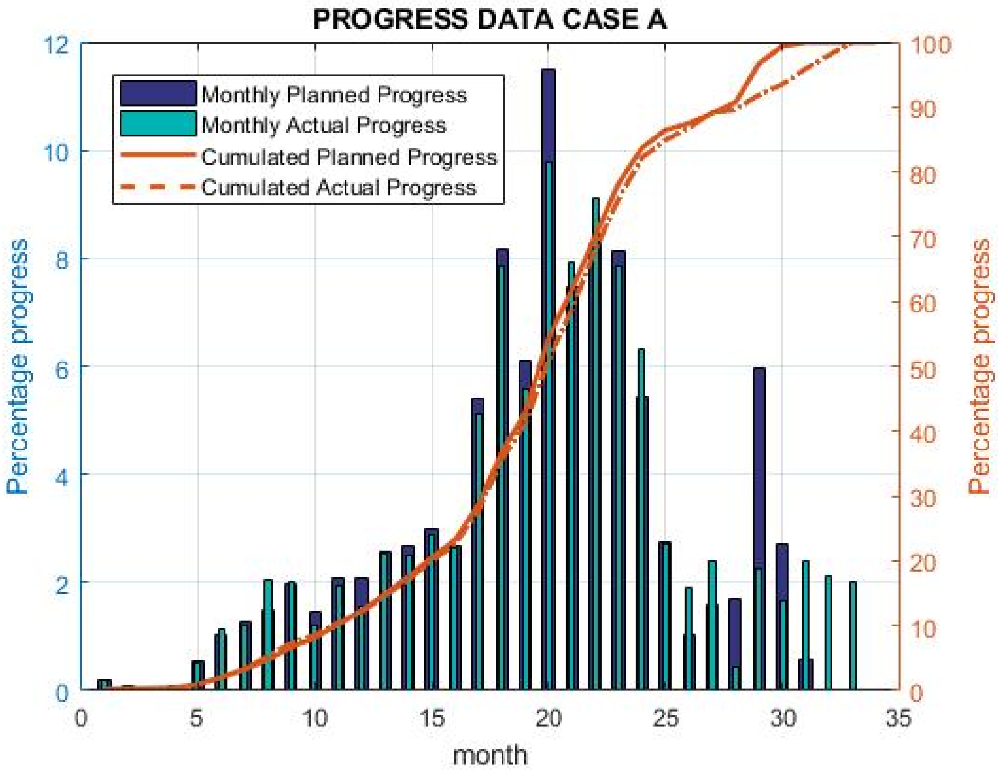

Figure 5: Case A physical progress description

In Figure 5 the project is described through its physical point of view: the red lines represent the cumulated progress, the dotted one is the actual cumulated progress, thus the progress achieved during the execution. The continuous line shows the planned progress, it is the project baseline established during the concept phase. In the bar chart there are the same information but in monthly values. The project starting date is March 2009 and, according to the scheduling, it should last 32 months; against these expectations, the end of the work is achieved only after 34 months with a two-month delay. The performances are periodically checked every month: for the sake of clarity, the project duration is identified by the number of the months after the starting date, March 2009. As shown, the actual progress overlaps the baseline until the twenty-eighth month where the delay is collected.

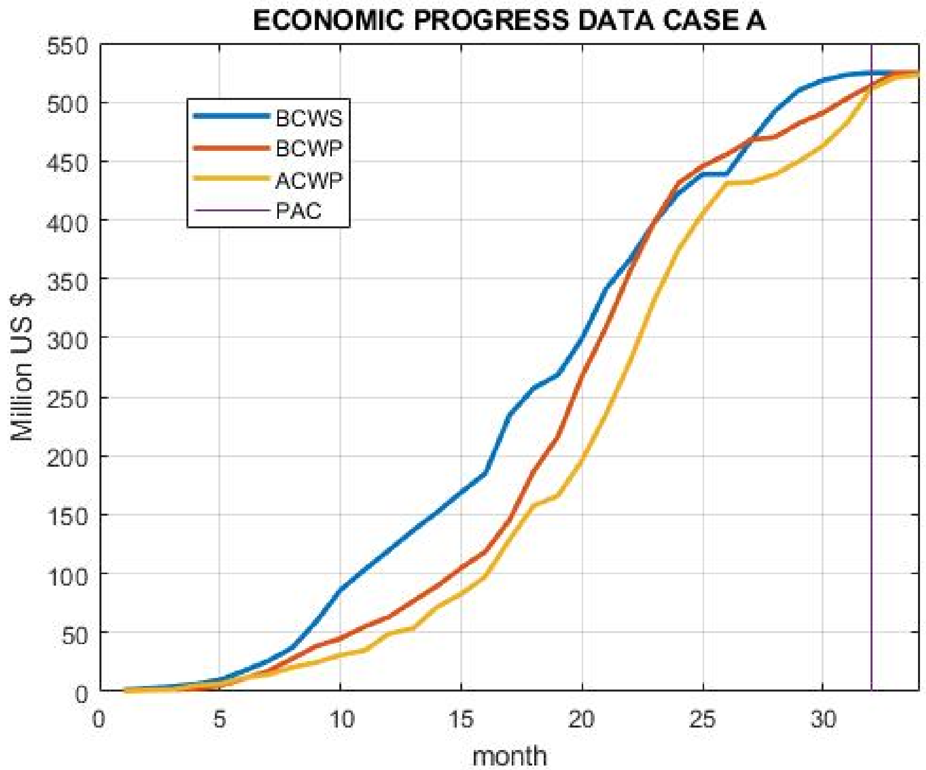

Figure 6: Case A economical progress description

Figure 6 shows the economical project life. The three represented quantities are identified as follows: the actual cost of work performed is marked in yellow, whereas the cost baseline (budget cost of work scheduled) is highlighted in blue and, finally, the budget cost of work performed obtained by the multiplication of the actual physical progressed by the project budget is identified in red. With regard to costs, the budget at completion amounts to 525 million US dollars but, although the project was carried out in an unprecedented area for the company, the final cost is 523 million US \$, less than what expected. All along the project duration the costs do not entail as much money as the money planned to be spent.

The B Case

The second case of the study is part of the offshore cluster. The project area is located in the Gulf of Suez, 77 metres under the Red Sea level. The project development in this site started in 2003 searching the best technical solutions to extract the oil and the optimal place to stock it. The concept selection and its definition last two years: later, in 2005 the execution phase began.
The selected solution consists in the creation of two oil wells completed with a tie-back production system, that is the connection of new wells to the already existing refinery structure on shore with an already present FPSO. The aim of the work includes also the installation of the needed sealines and pipelines.

The C Case

The last case is part of the onshore cluster. The scope of work consists in the realization of a power plant with two 150 MW gas turbines and in the related distribution station in the Congolese southern border. The project includes also the renovation of a power grid and the installation of a plant for the treatment, stocking and compression of the gas. From the treatment centre the gas will be sent in the above mentioned power plant and in two other plants nearby. The pipelines which are required for the transportation have to be put in place. The project starts in February 2008 with a planned duration of 30 months and a budget amounting to 320 million US dollars.

Model application

Uniqueness, as states the PMBOK [16], is one of the intrinsic features of the project. Every project is unique in its objective, in the scope of work, in the plan to achieve its goals and in the adopted management decisions. It is easy to understand that the performances of a forecasting method are not constant in all its application but vary from a project to another depending on specific scenarios [15]. At the same time, it is clear that different forecasting techniques may lead to various results, thus they convey to alternative implementing actions. Despite the same available input data, a few differences may occur in forecasting results due to the intrinsic feature of the applied method. For this reason, the evaluation of the forecasting performances of different techniques is a hard challenge and making an objective comparison may be even more laborious. The problem arises since a project is inherently influenced by the management choices which are based on the forecasting method results, that may be different if a change of the forecasting technique is operated [19]. All the while, a project is unique and cannot be repeated under the guidance of any other different method. The best way to approach this issue is to apply the model to real cases and compare the outcomes to the ones achieved by standard techniques. This section includes the application of the Kalman-EAC model to the three cases presented in the previous chapter. The achieved output is analysed and then compared to the state of the art in forecasting performances represented by the EVM technique in its three-month moving average formulation. The algorithm is applied using MATLAB software: its implementation is effortless and does not require a high power computational software. For this purpose, Excel or any other similar software may succeed in providing the same results.

Case A model application

As described in the K-EAC algorithm presentation, the implementation starts with an initialization phase. The two first elements to assess are the model variables: the state vector and the covariance error matrix. Given that the control is performed from the beginning phase, both elements are set composed by null elements. This is a self-evident choice because at the beginning of the project a cost variance has no chance to be different from zero since no cost is sustained and no work is performed yet. Nonetheless, there is no doubt or uncertainty in this situation, thus confirming the null matrix choice.
The second initialization phase aims to identify the q value, index of the system model uncertainty, it is assessed to make the model uncertainty congruent with the users prior estimate of the project final cost distribution. The user provides in input the expected project duration (PAC) and the distribution of the expected final cost, expressed with μ_c and σ_c^2, the mean and the distribution variance. These elements are used in an inverse Kalman forecasting algorithm to determine the q value. More specifically, the algorithm, solely based on the system model, works equivalently to set the gain K equal to zero for all the project duration. Consistently, as it occurs in the baseline plan, the PAC lasts 32 months and the final cost distribution amounts approximately to a Normal with the mean equal to the budget and a variance equal to 10% of the budget. The last initialization step is the measurement error matrix R_k. r is the mea-

surement error variables and takes into account the variance of the measurement error, σ^2. In order to set the value, the program evaluation review technique (PERT) [16] [17] [18] and a three-point estimate for the measurement error are employed. The user has to define the maximum possible measuring error, thus the error variance is evaluated with the PERT technique.

In this case the value is chosen by setting the max error as 1% of the budget, selected since the $BCWP$ is evaluated as the percentage progress multiplied by BAC. After the initialization the algorithm is ready to run, fed at every stage on the measured cost variance value. The results of the entire project duration are collected in Figure 7.

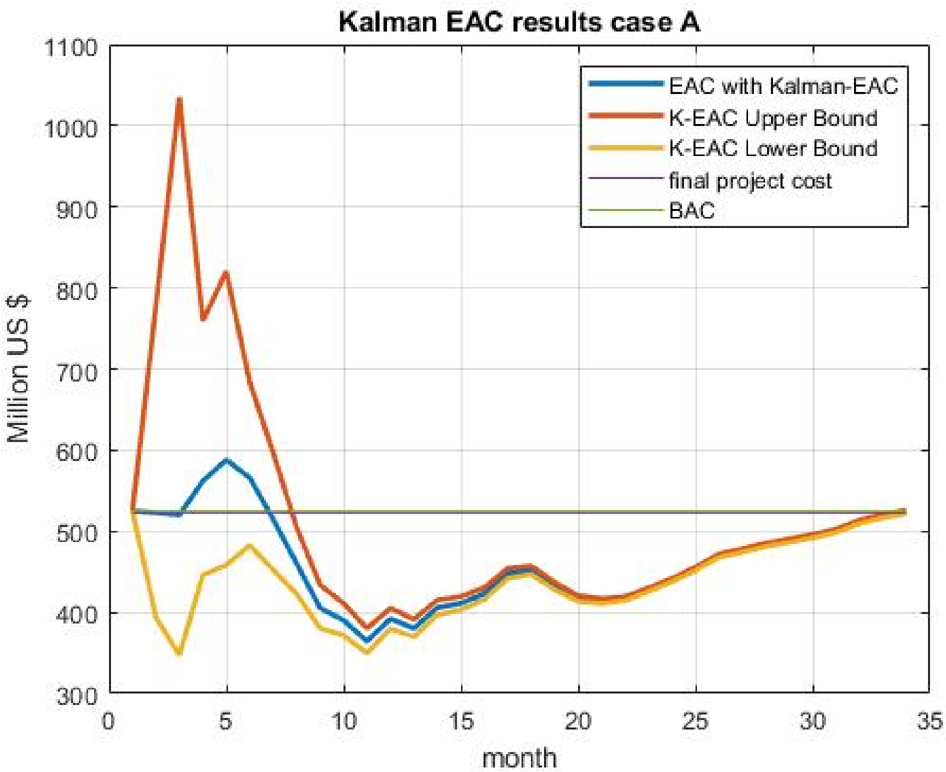

Figure 7: Case A Kalman EAC results, the chart shows the forecasted EAC at every iteration

The chart gathers the forecasted EAC evaluated at each iteration. Since the estimate is obtained projecting the cost variance distribution, the output is presented as three lines: the central blue one represents the mean EAC value obtained projecting the CV distribution mean, the yellow and the red lines represent the optimistic and pessimistic values obtained through the projection of the fifth and the ninety-fifth cost variance distribution percentile. In the first time instants, the algorithm presents a rump up period, the few information at disposal and the huge amount of work remaining makes the distribution estimate too wide. As the project makes headway, more data are available, and the distribution becomes narrower thanks to the more reliable estimations. On the whole, the project steps forward by strictly following the plan in the first months, and, except for a small peak on the fifth month bringing the EAC to almost 600 million US dollars, it is completely underbudget thanks to the optimum economic performances achieved in the central phase. During the final phase the project final cost estimate rises up to 523 million, almost achieving the capped budget. The same graph is presented in Figure 8, where the result of the standard EVM technique in the three-month moving average version is added, that represents the commonest technique nowadays: it is drawn in purple, while the blue line represents the central value of the K-EAC estimate distribution as in the graph above.

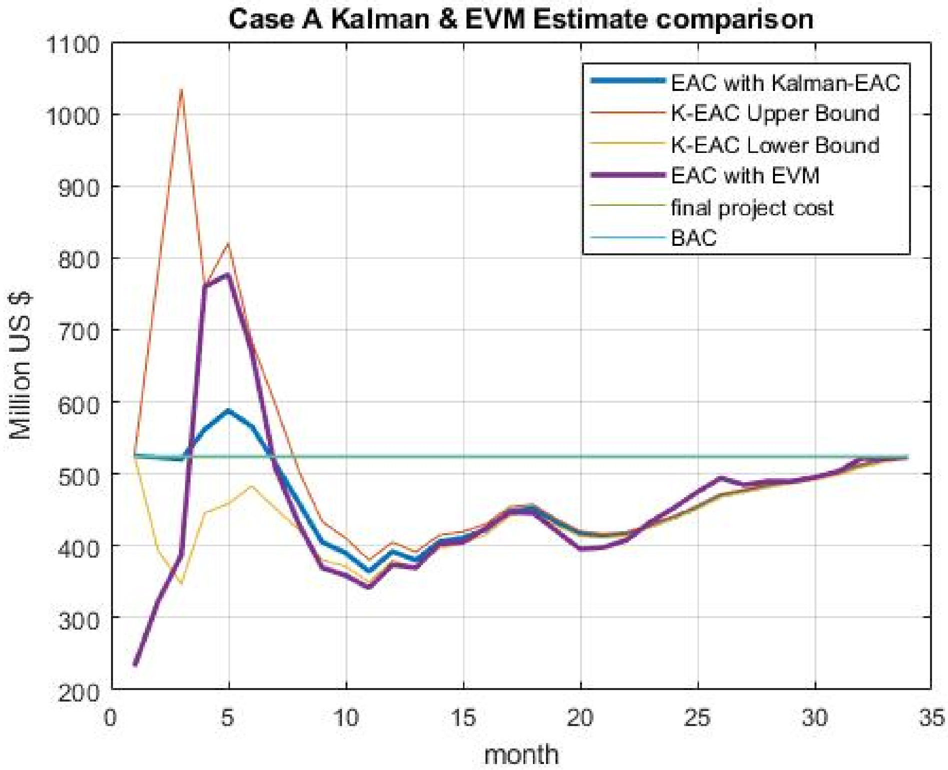

Figure 8: Case A Kalman and EVM Estimate comparison

Overall, the algorithms results maintain the same path, index of coherence in the estimates, and indicates a peak during the fifth month. Here the techniques have different trends: both results record a peak, but the EVM estimate is three times as high as the K-EAC one. Such a pronounced estimate variation is caused by a slight downturn on the cost variance only during the fourth month, whose trend is described in Figure 9. Both algorithms detect the cost variance fluctuation presence, but the interpretation of the phenomenon is different. The traditional EVM estimate raised immediately the EAC to about 800 million US dollars, and, even if the performance deflection at the following time instant is completely recovered, the algorithm continues maintaining a high final cost. This behaviour originates from a latent hypothesis: EVM considers every variation in cost performance as structural, so it will be featured in the project until its conclusion. The second issue derives from the mathematical evaluation of the EVM estimate; the performance factor that modifies the remaining work is based on the average over three periods, so should the oscillation disappear, it would affect the project performance in the following two evaluations. Differently, the K-EAC detects the performance loss presence but, since it is far from the performance trend, it slightly raises the EAC during the fourth period. Focusing back on the EAC during the execution, in the second half of the project, the two estimates have very similar behaviour, whereas it is important to highlight the result obtained in the first part. Here the K-EAC estimate is more stable with respect to the EVM one: this latter is evaluated adding to the cost sustained up to this moment the planned cost of the remaining work, modified by a performance factor measured over the last periods. It is intelligible that, when the project is in its early phase, the work which still needs to be performed is an observation of some weight in the estimate given that even a tiny performance fluctuation could have a huge impact on the final result. It is important to bear this issue in mind since not all the cost variance fluctuations are due to structural causes that will occur from this moment on, or even worse, they could be brought about by measurement errors. This issue is mitigated in the Kalman-EAC model where the obtained cost variances are the results of the combination of two factors, the measurements and the system model. The cost variance obtained with the K-EAC is cleansed by noisy fluctuations, when the algorithm needs to face a shortage of measurements, it does not tend to immediately trust in measurements, especially if their values are far from the one of the system model. The trend of the cost variance during the project execution is presented in Figure 9, where the primary difference between the methods is noteworthy: the probabilistic approach. The K-EAC provides a distribution of the cost variance accounting for the errors present in both the model and the measurement process. This is recognizable in the EAC that is provided with a central, an optimistic and a pessimistic value. The EVM technique provides, instead, a punctual result: in this situation the project manager has no indication about the information quality, thus how reliable they could be. This is a crucial point since the estimates represent the most important supporting tool for the project manager who has to select the corrective actions to implement and their intensity.

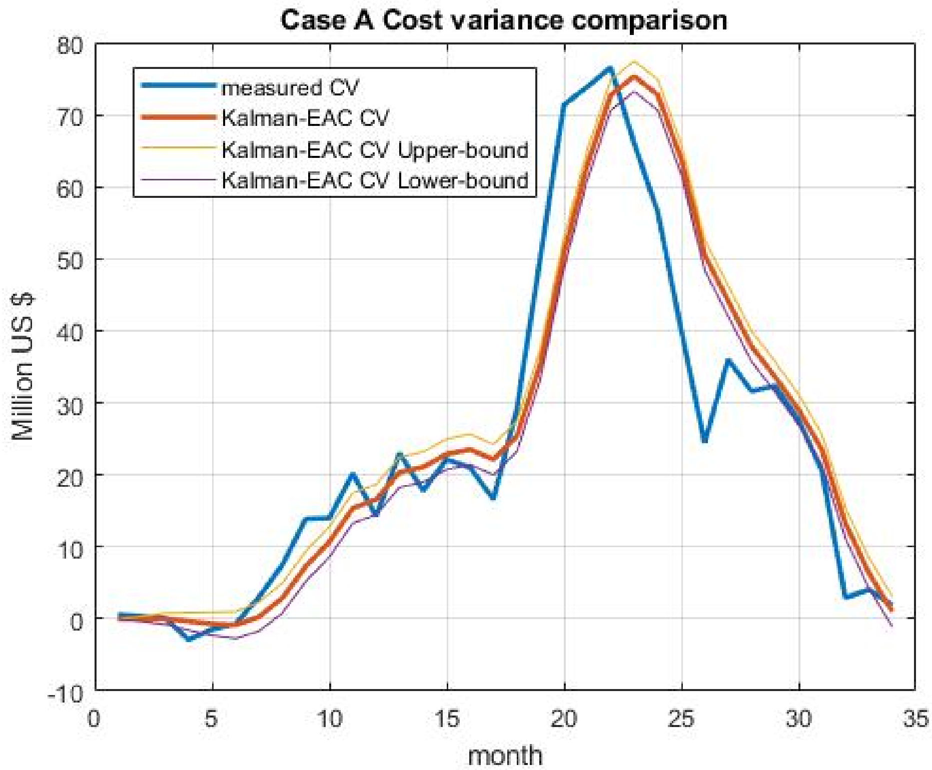

Figure 9: Case A Cost variance comparison

The graph shows the measured cost variances in blue, while the cost variance distribution evaluated with K-EAC model is expressed with the central value, the red line, and the fifth and the ninety-fifth percentiles are respectively the purple and the yellow lines. The above described trend is here clearly deduced, the cost variance distribution estimated with the Kalman filter is more smoothed than the measured one, that is used in the EVM evaluation. The trend is the same, but the short oscillations disappear. A second thing that should be noticed is that the estimated CV follows the same trend followed by the measured one but slightly on the right; this phenomenon must be ascribed to the same cause, that is the algorithm does not tend to directly trust a trend far from the system model result and it is supported by little evidence.

Now the comparison must concern the two most important aspects of a forecasting technique: accuracy and timeliness, in this phase the selection of the right criteria for the evaluation is crucial. In the literature of forecasting techniques, the accuracy is reported as the most commonly used parameter among professionals and researchers [20]. As a measure of accuracy, Vanhoucke and Zwikael [21] [22] submitted in their works well-known statistical measure of errors as the Mean Absolute Percentage Error (MAPE). Regrettably, the literature about the forecasting method evaluation is extremely lacking, but most authors tend to centre their research on these kinds of indicators. Accuracy is measured as the average deviation between the forecast and the actual value over a certain time period. In our case the MAPE is evaluated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{ActualFinalCost - EAC(t)}{ActualFinalCost}\right|$$

The MAPE value is evaluated not only with the K-EAC result, but with the

EVM three-month moving average EAC as well. The proposed method reduces MAPE from 19.816 to 12.836 with an improvement of more than 35%. This result is due to the high difference in the first part of the project, where the EVM estimate is too responsive to performance fluctuations and, even worse, every performance variation is considered as an outcome of a structural cause and affects all the remaining life of the project. The K-EAC model does not trust in performances deviations without evidence, especially if far from the system model output, thus reducing the risk of misleading interpretation.

A second way for the accuracy measurement was brought out by Teicholz [3], who draws attention to a new indicator after comparing 121 construction projects. Together with the classical statistical methods, as mean square error, the accuracy could be represented by the absolute area between the actual final cost and the path of the estimate at completion plotted against the percentage progress: Figure 10 shows the application of this method to the A case. This technique introduces some advantages: firstly, it does not suffer from biasedness if the measurement intervals are not constant; secondly, it offers a visual information about the achievement of the best accuracy results.

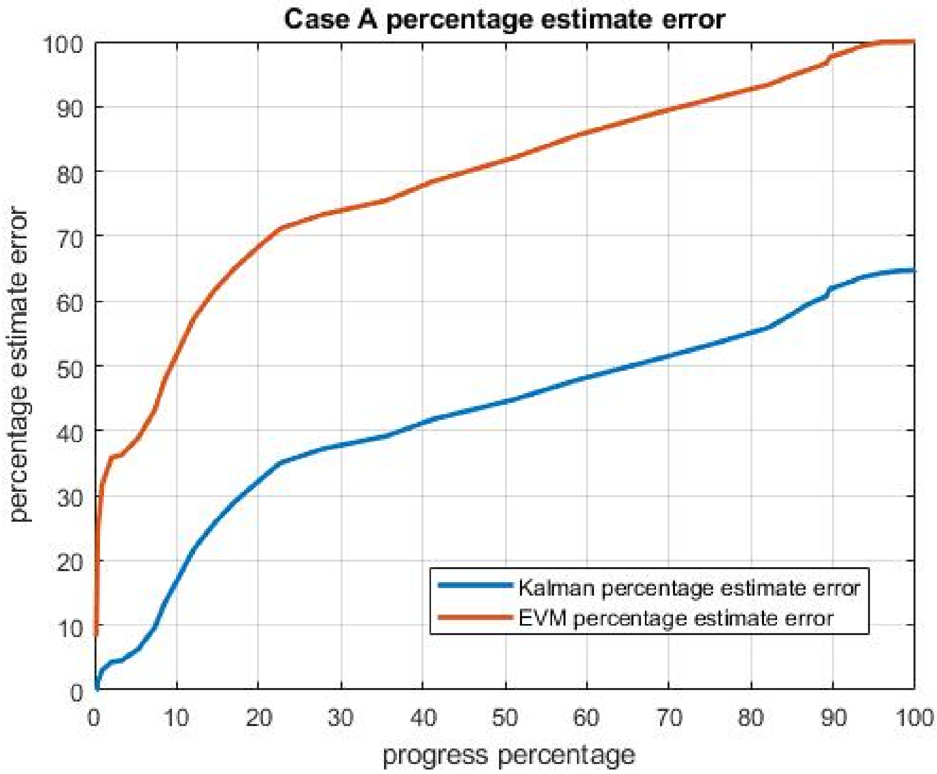

Figure 10: Case A percentage estimate error

On the x-axis the project progress is displayed in percentage value, whereas on the y-axis the percentage estimate error is evaluated as the cumulated area between the estimate at completion and the final cost value. In order to compare the two methods, the areas are normalized using the area of the EVM estimate at completion as reference given that it is the reference method. As above mentioned, there is quite a considerable difference on the first part of the project duration that is constant after its first third, in other words after the two lines head for a similar trend due to the fact that the two estimates have similar tendencies. Besides, the graph introduces a second important aspect for a forecasting method: timeliness. Timeliness is here identified as the ability of the method to provide reliable outcomes over the short term. This is a very significant issue for the project manager who needs to take decisions from the project early stage since correct results are needed as soon as possible. Further in his work, Teicholz figured out a procedure to evaluate the timeliness and which is still nowadays the substantial challenge of forecasting method: the author defined it as the accuracy achieved in the first half of the project. From this viewpoint, it is possible to measure the percentage error of the two estimates at completion in the first half of the project. The A case results are presented in Table 3.

K-EAC	EVM
47,74	82.02

Table 3: Case A, percentage estimate error on the first project half

The value stands for the total error performed in the first half of the project duration. Both indices are comparable because they were normalized by using as reference the error made by the EVM estimate at completion, that represents the state of the art. It emerges that the Kalman estimate at completion produces more accurate results at the beginning of the project than the traditional technique does. This can be explained, as cited previously, by the twofold nature of the Kalman filter that, using measurement and the analytical system model, could filter the cost variances fluctuations not due to structural causes. A third aspect which requires consideration is the probabilistic estimate introduced by the K-EAC model. In this case, to perform a comparison is not possible since the EVM system performs only punctual estimates. The presence of a probability distribution introduces several advantages: the most valuable is that it gives a measure of how certain the estimate can be. This means valuable knowledge which allows the decision maker to introduce actions that are supposed to deflect the course of the project development.

The B and C cases model application

The model application to the B and C cases and the model previously described are alike: the outcomes collected in the following chapter show a trend which is similar to the A case results, namely, a headway with accuracy and promptness respect to the EVM technique. Phenomena marked in the C project application: in this case, the little progress achieved during the first period begets an incredibly high estimate and affects it for two more periods. Conversely, the K-EAC estimate is lower: owing to the little available data, the algorithm relies more on the system model and this eventually results in a sizable MAPE reduction.

Conclusions

Comments over results

The application of the K-EAC method to the three cases presented above gives as output two pivotal pieces of information: the cost variance and the estimate at completion.

The cost variance detected by the algorithm aims at the reduction of the misleading effect of errors, present both in measurement process and in the input data. It is provided at every iteration with its probability distribution and assumes a central role in the forecasting contest since, in addition to describing the progress of the project, it is used during the estimate at completion process.

Comparing the trend of the estimated cost variance throughout the project execution to the measured one, in all three cases of study two aspects are easily highlighted: firstly, they have a very similar trend, which means an index of coherence on the results; secondly, the estimated CV smooths the short peaks detected by the measurement system but, the unitary period duration considered, they are seen as fluctuations which do not describe the state of the project but which are due to measurement errors. The algorithm is so cautious that it does not immediately trust measured value of cost variance when it is far from the actual trend and not supported by previous observations. Evidently, the algorithm does not completely delete the peak presence; however, it lowers its intensity, placing more trust in the system model results than in the measurement. At the same time, if the trend is still present at the following period, the algorithm gives more confidence to the measurement system by totally regaining the actual CV.

The second output of the algorithm is the estimate at completion, provided with three values: the most probable, optimistic and pessimistic values. In every application, the algorithm presents a warm-up period when the measure of the EAC uncertainty initially grows and later starts to get smaller. The rump-up period in the initial months is due to the little information the algorithm can rely on. The second descending trend is instead in-line with the expectation: with the ongoing project the algorithm has at disposal more information to elaborate the estimate.

The results are then compared to the ones obtained by the technique representing the state of the art in the forecasting field: the three-month moving average version of the EVM estimate at completion. On the whole, the results are con-

sistent since the two estimate trends are similar. The two methods are then evaluated on the aspects which are the most important features for a method, according to the forecasting literature: accuracy and timeliness.
As a measure of accuracy a standard error evaluation technique is proposed: the MAPE (Mean Absolute Percentage Error) between the estimate obtained at each step and the actual final cost, whose results are proposed in Table 4.

CASE	K-EAC	EVM
A CASE	12.836	19.816
B CASE	39.895	55.597
C CASE	18.689	59.305

Table 4: MAPE value of the estimate in the three cases

As it can be noticed, the outcomes are encouraging and the proposed method introduces nosedives in all the evaluated cases. Since MAPE is a mean indicator, in order to evaluate where the improvement is achieved, Teicholz's alternative accuracy measure is also proposed [3], based on the cumulative areas between the estimates and the real final cost lines over the project duration. The graphs show that, for the three projects, the crucial difference is obtained in the first third of the project. The estimate at completion is evaluated adding to the cost sustained until this time the planned cost of the remaining work, modified by a performance factor which has been measured over the last periods: one soon realizes that, when the project is in its early phase, the work still to be performed has some weight in the estimate, thus even a very small performance fluctuation could have a critical impact on the final result. The divergence between the obtained EAC is due to the different performance factors used by the

two methods: on one hand, the K-EAC bases its EAC over the estimated cost variance which, as explained above, lowers the presence of performance random fluctuation; on the other hand, EVM technique directly uses the measurement by maintaining the short peaks, thus considering the fluctuation due to the structural causes which will be affecting all the remaining project life.
A second problem with the EVM estimate is that the fluctuation, though it could be recovered in the following time instant, affects the estimate in the two following periods since the EAC is evaluated through a three-period average performance factor. On the contrary, the K-EAC, once identified the wrong value, bases its estimate mostly on the system model and avoids the misleading EAC.
The second aspect analysed is the timeliness, namely the ability of the method to provide meaningful results in the short term. The evaluation abides by the method indicated by Teicholz, that takes as an indicator the percentage of the error made in the first half of the project progress, the value to be compared are normalized over the EVM estimate error, the indices are reported in Table 5.

CASE	K-EAC	EVM
A CASE	47.74	82.02
B CASE	61.08	89.57
C CASE	18.44	87.44

Table 5: Percentage estimate error on the first project half

With respect to the standard method, the K-EAC shows up as the faster in providing telling results. As mentioned above, the EVM estimate at completion is too responsive during the early phase of the project and this leads to wrong final cost estimation characterized by dramatic peaks. The proposed method lowers the issue combining a system model with the measurement process.

Conclusions and future studies

In the present work a new model for the estimation of the final cost of a project is developed, aiming to contribute to the state-of-the-art technique consisting in the EVM system methodology. The objective is to reduce some issues that a project manager needs to face nowadays during the project execution, such as criticalities which lie on the hypothesis the three-month moving average EVM estimate at completion is based on.

First of all, the EVM technique deterministically describes the project status, without considering the possibility to run into any kind of error, providing punctual indicators. Errors are easy to commit given that the cost performance indices are based on two quantities: the actual sustained cost and the budget cost of work performed. The latter is easily mistakable because, in order to get assessed, the percentage of physical progresses achieved and the planned cost for the activities are needed. It is challenging to synthetize in a single progress percentage value a whole project containing a large number of activities and several different disciplines, and, at the same time, it is hard to forecast the cost of activities and subcontractors before the beginning of the project, even more if the activities are performed in a new site where the company is not used to working. The issue is extremely convoluted since the information has a substantial value for the project manager to base expensive actions on.

The second issue concerns the responsiveness of the algorithm during the starting phase of the project. The phenomenon derives from the way the EAC

is evaluated: the algorithm provides the final estimate adding to the present sustained cost the one which has still to be sustained but modified by a cost performance factor. The responsiveness is due to the large amount of remaining work. In this condition the effect of a short cost performance fluctuation is sizable and the problem is even more severe since the fluctuation could be caused by a measurement error as explained above. In this case, although the performance index is correctly assessed in the following observation by applying the three moving average system, the error keeps being taken into account for two more periods.

Throughout the technique development the attention has been focused on the method performances, namely, accuracy, timelines and friendliness which are considered as key parameters, according to the forecasting literature.

The proposed algorithm aims at the reduction of the previously mentioned shortcomings. Based on the classical Kalman filter framework and the EVM techniques, the algorithm uses both a system model, built over the planned project baseline, and the measured data to evaluate the probability distribution of the real but hidden cost variance, describing the project status and giving also a measure of how trustworthy these values are. The obtained CV distribution is freed from short performance peaks since the K-EAC algorithm does no trust fluctuation far from the trend when it is not supported by evidence. Later, this is used to evaluate the EAC, provided with a central value, in addition to an optimistic and a pessimistic one. Reducing the cost performance index peaks, the method initial responsiveness gets lowered and the problem will avoid happening again over the following periods.

In order to evaluate the method performances, it is applied to three real cases, showing positive results and coherence with the competitor method. A further advantage is achieved: though the measure of the actual performance could not be performed, to perform the estimate is still possible with the system model. As a consequence, the positive results obtained are unfortunately balanced by an increasing algorithm complexity. The computation, composed of matrix multiplications, is straightforward and it can be implemented with an undemanding software as Excel or MATLAB which do not need any sophisticated calculation powers. Even the result interpretation is intuitive because the numerical value of the cost variances and EAC are supported by graphical tools. The raising complexity lies in the introduction of the initialization phase which requires additional input data respect to the EVM technique, specifically the prior cost distribution and the maximum errors of the measurement system, introduced in the model presentation chapter. On the other hand, the initialization phase provides a great adaptability to the method: here the parameters are assessed by helping the technique to fit each time the analysed project avoiding standardization. The parameters change time to time describing the project and the environment it is developed in. This opens a window to extend the method application not only to construction project (where the EVM is developed), but also to new project type.

Based on the obtained result, some suggestions about the possible further study have been made.

Further model verification with a bigger model cluster. Only three projects in the Oil & Gas field cannot represent an adequate and satisfactory collection to guarantee generality to the method so it is necessary to test the algorithm and its accuracy with a wider project number.

In addition, the algorithm results and the estimate reliability have been validated for specific projects, especially for long-term ones and hardly ever exposed to accidents with high impact on project performances like the ones related to the US defence department (the area where these techniques were first developed) [23]. There is the need to spread out the methodology to a wider project range marked by a different duration, operative processes and uncertainty level. Since the initialization phase gives the algorithm high adaptability, it could be tested in new project types.

Integration of the algorithm with other informative sources. The K-EAC has a significant potential for the integration with other informative sources: it provides an extremely flexible framework for combining new state variables and, furthermore, the initialization phase, where prior distributions are needed, could be enriched by expert judgement or similar past project analyses. A powerful method would give the decision maker the chance to add some other information sources to make the outcomes more reliable.

Chapter 1

Introduction

The present thesis springs from a keen interest in the project management discipline which has more and more drawn the author's attention over the last years. This curiosity increased during the course Industrial Project Management which featured the merits and the complexity of engineering projects. The most engrossing aspect is that, despite their power, the splicing of challenging goals, the countless actors involved and the different disciplines, projects have to be managed harmoniously and the right choice needs to be successful in the first attempt of its implementation. More specifically, the growing interest has increasingly focused on the estimate at completion, a key point of the control phase and tremendously essential to reach the project objective. This becomes an extremely challenging point since the project complexity joins the unpredictability of the future. Forecasting has a key role in project management and it highly affects crucial phases like planning, controlling and risk monitoring. Even before the project execution, forecasting is used in planning for developing the project baseline plan, namely the guide to achieve the work on time and within the budget. Although it is during the project execution that forecasting becomes a pivotal step, in this phase the performances are monitored in order to understand the ongoing project and they are employed to rectify the estimate of the work remaining: this is the way they will reveal the necessity of corrective actions. It is on this kind of occasions that a decision-maker relies on the tool of forecasting proves. As a reply to this issue, nowadays the commonest technique in performance monitoring and forecasting is the "Earned Value Management" (EVM): more precisely, it becomes widespread in its three-period moving average version that, according to experimental results, it has been identified as the better technique in terms of accuracy. Even though the EVM is employed also in the schedule performances area, the research will focus on the economic point of view through the analysis of the cost performances. This observation takes into consideration two main aspects: the project economic status and the estimate at completion. The first one is used for identifying if the work performed overspends respect to what was planned and to describe it the indices are the cost variance (CV) and the cost performance index (CPI), whereas the estimate at completion (EAC) is the estimate of the final project cost that considers the project performances up to a given moment. The EVM has provided methods for nal outlooks and, largely, these have never undergone any attempts of improvement since their formulation. Consequently, the available methods are often oversimplified, based on inconsistent hypotheses and they sometimes provide unreliable and thus not usable outcomes. It is possible to detect three main problems affecting the EVM system technique: firstly, the methodology does not consider the presence of possible error; secondly the method is characterized by strong sensibility during the project initial phase making it overresponsive and, thirdly, the prolongated effect of measurement errors in the evaluated estimate. The EVM technique describes the project status in a deterministic manner, that is to say not considering the possibility to run into errors, only providing punctual indicators. These errors are easy to commit because of the way cost performance indices are evaluated. Based on two quantities, the actual sustained cost and the budget cost of work performed, it is exactly in the latter that it is easy to make mistakes. To be assessed, the budget cost of work performed, two elements are needed: the percentage of physical progress achieved and the planned cost for the activities. It is a hard challenge to synthetize in a single progress percentage value a whole project containing a large number of activities and several different disciplines also considering the deficiency of a univocal methodology or guidelines. At the same time, it is hard to forecast the cost of activities and subcontractors before the project starts, even bearing in mind that projects could last several years and that activities could be performed in a new site where the company has never worked before and has thus no experience in the area. The second issue concerns the high responsiveness of the algorithm throughout the project starting phase. The phenomenon derives from the way the EAC is evaluated. The algorithm provides the final estimate adding to the present sustained cost the one which still has to be sustained but modified by a cost performance factor. The high responsiveness depends on the large amount of remaining work. In this condition the effect of a cost performance fluctuation is enormous, and the problem is even more serious since the fluctuation could be caused by a measurement error as explained above. The third issue is about the long-term effect of an error in the cost performance index measurement over the estimate at completion. In this case, despite the fact the performance index is correctly assessed in the following observation, the mistake will continue to affect the estimate for two more periods. The shortcoming occurs since the cost performance index which modifies the work remaining cost in the estimate evaluation is assessed with a three-period moving average. The point in question holds a particular importance given that the high value of the information provided, on which the project manager bases expensive manoeuvres which may decide whether the project will succeed or fail. The sizeable sum of capitals involved makes this information important not only for the project but for the company business as well. It is clear that in such a complex and dynamic environment, the decision maker needs correct information as soon as possible and a measure of their accuracy could be of great help in supporting the project manager during the decision process. In addition to this, from the point of view of the domain of the construction project, where these techniques were first developed, it is recognizable that some hard criticalities can increase the uncertainty level, such as the multitude of the stakeholders involved and the political, financial and climate-related factors which have inevitably effects on the project performances. Thus, the activities need a high-level planning and risk management system. This is the reason why a wise and clever nudge towards more performing techniques which require the introduction of more complex, specific and diversified methodologies working together and in a combined manner to reproduce a model which may be as close as possible to reality. As a result, the objective of the research is to develop a new technique that will be able to reduce the previously mentioned shortcomings by taking into account the main features which are required to a working forecasting method. Throughout the technique development, the attention has been focused on the method performances, namely, accuracy, timelines and friendliness which are considered as key parameters, according to the forecasting literature. The timeliness here identified as the ability of the method to provide reliable outcomes over the short term while user-friendliness is needed to guarantee an easy and agile algorithm implementation and results understanding. Based on the classical Kalman filter framework and the EVM techniques, the K-EAC method will provide as output both the cost variance distribution and the estimate at completion. The proposed technique uses both a system model, built over the planned project baseline, and the measured data to evaluate the probability distribution of the real cost variance which is concealed in the previously mentioned errors. This adds value to the information because not only does the cost variance better describe the project status but also gives information about the way these values are grounded. Since the K-EAC algorithm does not trust fluctuations far from the trend unless it is supported by evidence, the evaluated cost variance is freed from the performance fluctuations which do not have a structural meaning, maybe caused by measurement errors as those described few lines above. The evaluated cost variance is then used for evaluating the project EAC, provided with a central, an optimistic and a pessimistic value, giving, also in this case in point, a measure of the value accuracy. The method performances are evaluated through the application to three real projects in the oil and gas sector. The results are then compared to the ones obtained with conventional techniques showing positive outcomes and consistency with the competitor method. After that, the essay moves forward with a description of the way the work is organized. Chapter two starts with a presentation of the role of forecasting in project management and, in detail, of its importance in the monitoring and control process. Subsequently, an excursus over the main characteristics which are noteworthy in a forecasting method and the evaluation criteria to measure them; besides, the mathematical tools which will be used to analyse the model application to real cases will be presented in the second part of chapter 4. The general context of the estimate at completion and its role in the project control are then featured and so is a description of the state of the art in the forecasting field acts as prologue for the presentation of the most used technique in the field, the earned value management system. Nevertheless, the technique is too extensive to be completely discussed, so the classical formulation will be shown exactly as it was first developed followed by the versions mainly in use nowadays. In order to help the reader get by in the large number of acronyms used in the area of the Earned Value Management, a glossary will list their meanings through a thorough explanation. More lines will be granted to illustrate the indicators applied to the measurement of the system performances and their evaluation because they will be the crucial features in the new proposed method as well. Lastly, some criticism to the method and its main disadvantages will be described. In the third chapter the K-EAC model will be widely shown. Starting with a brief introduction of the Kalman filter, initially with a contextualization about its origins and its current use and besides a description of its framework will be shown by focusing not in the mathematics beyond, too extensive to be exhaustively covered here, but through the explanation of the main filter working phases. Soon after, the attention will be drawn to the proposed model: firstly, his functioning will be described, that is to say the way the Kalman filter is adapted in the estimate at completion process. Secondly, the reader will find each algorithm component and phase to describe their functioning. A particular attention will be given to clarify the algorithm needed inputs and the filter initialization phase whose role is of prime importance in order to achieve the forecast outcomes. The fourth chapter describes the model application to three projects in the oil and gas field. In the first part a brief excursus of the way a typical project in the oil and gas sector is developed and the typical project cluster along with their features to introduce the reader to the criticalities present in the field. A description of three cases of study will follow. Each of them will be described by focusing on the scope of work and, since the projects have already been completed, and their development, in both economic and physical terms, by highlighting whether the project runs out of their objective in terms of time and cost. Finally, the K-EAC method application will be fully shown: the obtained results are then analysed and compared with two other outcomes obtained during the application of the standard technique (earned value management system, in the three-period moving average estimate at completion). A comparison will be made by focusing on accuracy and timeliness, considered by the forecasting literature two of the main aspects needed for a functioning and efficient forecasting technique. The comparison uses both standard tools, such as MAPE, and also some more specific ones for the forecasting field. Lastly, before spotlighting the conclusions of this thesis and especially the future studies about the mentioned particular issues, in the fifth chapter a recap about the achieved outcomes will be provided and the advantages and disadvantages of the model will be carefully evaluated. In the three cases it is applied to, the proposed algorithm introduces improvement in the estimate accuracy: the gain is obtained in the first project phase, also denoting an improvement in timeliness. With regard to the pros and cons of the technique, in the first category the enhanced information coming from a probabilistic approach and the project adaptability could be effortlessly included, whereas in the second group the increased algorithm complexity is a matter of fact. The dissertation will come to its end through a speculation on a few possible future studies: the need to test the algorithm with a wider project cluster and some tips related to the possibility to include in the algorithm some more sources of information.

Chapter 2

Literature review

2.1 The role of forecasting

Forecasting has a key role in project management and it highly affects crucial phases like planning, controlling and risk monitoring. Before the project execution, forecast is used in planning in order to develop the project baseline plan, that is to say the guide to achieve the work on time and within the budget. During the project execution, instead, the performances are monitored and employed to rectify the estimate of the work remaining, so that they may reveal whether any corrective actions are needed and to what extent. It assumes great importance when the future is so unpredictable that a few aspects of the project may introduce a degree of uncertainty that cannot be tested; that is why, on these occasions forecasting proves to be the most important instrument to rely on for a decision-making process. This is a common condition in project management where, although the project objective is intelligible since the onset, the final outcomes will be available just at the end of the project. Indeed, uncertainty, as it is stated in the PM-Book [1], is an intrinsic feature of any project: frequent changes, unexpected events, several players involved and stakeholders' needs gradually evolving create the environment in which the project manager must take decisions which are rich in variability and unpredictability. In the event of a perfectly known scenario, where every factor conditioning an event is certain, not only could the future be deterministically predicted or even controlled, but the estimate will be useless as well. Evidently, the absence of uncertainty is a utopic situation; so decisions should be taken after considering the overall risks and their related occurrence probability. Unfortunately, probabilistic forecasting approaches are not very diffused and, what is worse, experts are not highly knowledgeable in this field as they should be, because of the lack of easy-to-implement models providing good performances and not requiring huge amount of data. The issue is not new and has been addressed over more than half a century [24] [25] but no substantial improvement has been done since then.

2.2 Forecasting in project control process

The project management is the response to the intrinsically uncertain nature of the project aiming to manage the continuously evolving project behaviour. As the PM-Book states [24]:

> *"Project Management is the application of knowledge, skills, tools and techniques to project activities to meet project requirements".*

More specifically, its frame is composed by the management process that, according to the Project management institute identification are: start up, planning, monitoring/executing, control and close out. These processes could be applied in each of the knowledge areas in which the project management is articulated, from integration to human resources. Nevertheless, there are a few applications which may be critical: among them, there are surely time, quality and cost management; these parameters are used to pinpoint exactly the constraints to be respected and the results. Over the years, several studies have been conducted in order to define the parameters that strongly affect the project success: literature suggests that a well performed planning and an efficient control are crucial to determine a good result [26] [27] [28]. Results confirmed also by Salazar-Aramayo [29] in the dissertation about the processes success in the Oil & Gas field. The centrality of the control phase during the project evolution cannot be disputed and it may be considered as an essential procedure to be carefully followed to meet the expected outcomes; since it directly involves forecasting, it is consequential to describe its purpose and implementing rules. The objective is to detect and, if necessary, carry out corrective actions if there is an offset between the planned performances and the current ones in order to meet the contract obligations [16].

The control system is not univocal and varies according to the project type, size and involved players, but in all cases could be figured out the fundamental steps [5]:

- Define the most important factors under observation: a priority must be given to parameters that have higher impact on the project results, aim and strategy. At a later time, the chosen variables will be examined to understand whether the current performances are positive or corrective actions are suitable. In this phase, it is important to take into account the contract restrictions and obligations. Typically, the most influent factors are costs, quality and time.

- Define boundaries to figure out the acceptable variations: considering for example a project related to the Health, Safety and Environment (HSE) group, the tolerance here is set to zero, whereas it is possible to have positive or negative variances for time and costs in others projects type. While positive variations are hard to exploit, negatives tend to accumulate and propagate; in those cases, an accurate and more focused monitoring from their early appearance has to be performed. A threat level has to be set: if it reveals itself exceeding, a corrective action will be taken place. This threshold level is different from every project and depends on the related risks magnitude.

- Measure: the measurement is crucial for multiple reasons. It must be reliable and performed at the very right time. Needless to say, to perform corrective actions coherent with the situation, an accurate measure must be required, even though a too precise measurement will be expensive and may turn out useless as well; therefore, a careful weighting between cost and quality of the measure is essential.

Furthermore, the second basic decision to be taken into consideration is the timing for the measures: it would be reasonable if they were frequent enough to discover negative behaviour as soon as possible, but, at the same time, no so often detected so that the risk of negative influence on the project team work may be bypassed. Surely, a rapid measurement is useless if not followed by a prompt analysis and response.

- Forecast future performances: the main task of the forecasting phase is to perceive an early warning signal to perform proactive actions.

- Implement, if necessary, corrective actions: the problem in this phase is to understand the better action to be performed and act rapidly. The process is followed by a re-planning phase.

.

The governance technique must be proactive: it cannot be sufficient to amend problems when they occur but it is necessary to project the current performances in the future to avoid forthcoming deviations from plans. The control system could be schematized by a feed-forward loop as in figure 2.1. It depends by , the offset between the output of the operative processes and the target objective.

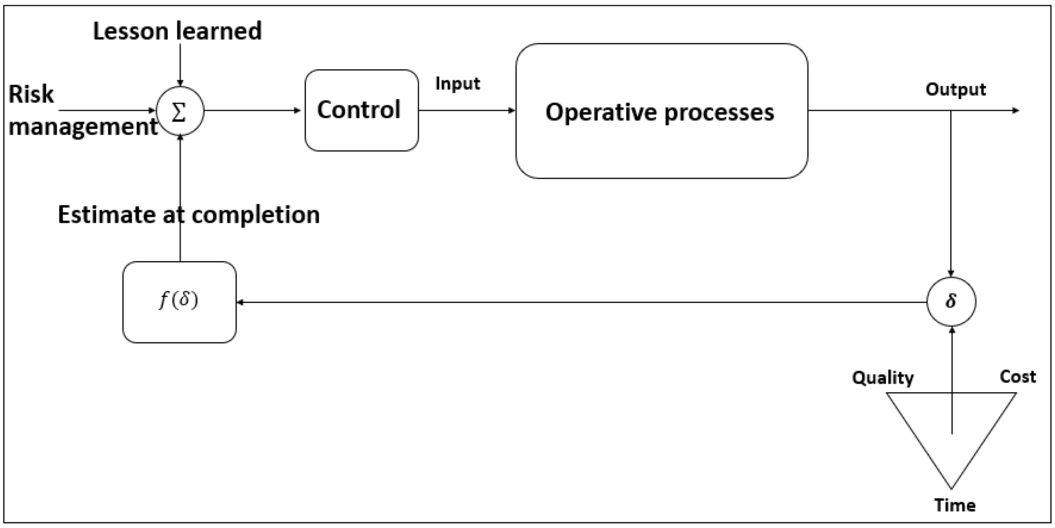

Figure 2.1: control feedforward loop

It is not difficult to understand that correctly forecasting the future is of considerable importance for the final project results. At the end, the control process is completed through a phase of ongoing lessons learned over the work done in order to improve the project management system and a culture of risk management during the entire development.

2.3 Evaluation criteria for forecasting methods

The comparison between forecasting methods, both in case of deterministic and probabilistic, is a tough challenge for forecasters. The problem has occupied researchers and sector experts since 1969, when Bates and Granger [2] published their seminal work. So far several studies have focused on the identification of the best parameters for the comparison. In order to take hold in common use and to be taken into consideration by project managers, the method needs to meet several requirements, such as the quality of results and the ability to be user friendly both in the application and in the interpretation of the outcomes. The main principles are shown below to draw a guideline. Firstly, the method should require neither hard nor expensive input data to be collected. The project could differ in terms of size, type and budget: this will grant the method applicability. Secondly, the method should be simple in the implementation and in the interpretation of the outcomes; it goes without saying that a complex or murky algorithm providing complicated results will be put aside and never be used in practice. Thirdly, the method has to guarantee good performances. A desirable method should consider:

- Accuracy: forecasted estimate should prove to be close to the final value;
- Timeliness: the method should have a short ramp-up period. The results have to be available for use and reliable as soon as possible. Timeliness needs to be maintained also during the project execution in order to avoid any delays in trend detection;
- Avoid the presence of systematic errors, both over- and under-estimating real values;
- Stability: forecasts with highly variable outcomes will not be as trusted as stable ones;
- Flexibility: the ability of the filter to correctly fit the different phases of the project life, maybe characterized by different progress rates, or to be used in heterogeneous project, not only with different size but with different project type as well.

In light of the above, it is important to note that no method could achieve perfect accuracy or stability until there is no knowledge of future events and but only a few hypotheses about future project conditions. A trade-off between the listed features is required: for example, high stability could hide relevant trends over recent past causing a delay in detection of problems that could affect the forecast by reducing timeliness and accuracy; conversely, a sensible method will be unstable and overreactive. Another important aspect not mentioned above is the quality of the outcomes, more precisely how reliable they are. Most of the methods used today are deterministic and do not express the reliability of the results and providing a measure of their soundness, they will increase the trust on the method and help the decision process. On the contrary, a low degree of belief in the forecasting results may be a serious problem that could lead the user to take decisions overlooking the method output. While stability and flexibility are hard to quantify, for a comparison between accuracy and timeliness will be used a quantitative method. Teicholz [3], comparing 121 construction projects, proposed a new indicator as measure of accuracy. Along with the classical statistical methods, as mean square error, the accuracy could be represented by the area between the actual final cost and the path of the estimate at completion plotted against the percentage progress. This technique shows an advantage: it does not suffer for biasedness if the measurement intervals are not constant. Further in his work, Teicholz figured out a procedure to evaluate the timeliness, still nowadays the real challenge of forecasting method: the author evaluated it as the accuracy achieved in the first half of the project. These tools will be used in the dissertation to evaluate and compare the analysed techniques.

2.4 Conventional method for project forecasting

2.4.1 Estimate at completion

As previously described, estimate at completion are a crucial part of the control system since the project is characterized by unicity that makes it a non-repetitive process. The most basic governance technique consists in a systematic comparison between measured and planned results, determined by contract limits. In addition to these control tools, others are as lessons learned or risk management which have limitations that make estimate at completion the most effective control method. The integration of the lesson learned in the everyday practice has two main shortcomings: it is not instantaneous and it is even hard to formalize and to integrate in the control system. Despite the presence of a dedicated section in the close-out document, it is not easy for project managers to read, elaborate and assimilate it; more easily every project manager will acquire experience from projects performed by themselves, losing, as such, a remarkable part of information. The second issue concerns formalization, since nowadays there is a lack of tools which may be able to systematically integrate the lesson learned in the control system. Considering instead risk management, even if extensively analysed in literature, it is hard to apply in a systematic way. It demands huge efforts for the ongoing risk register updating and needs time consuming preliminary analysis that will compromise the control system efficiency, that is to say the ratio between the obtained results over the effort in input. In addition, it requires an overall view hard to acquire in an environment that include multiple disciplines and players. Estimate at completion instead requires a limited effort and, using traditional methodologies as Earned Value Management, does not need data series for the implementation. To a given instant, called time-now, is demanded to determine the amount of money or time still necessary to complete the activity of the scope of work. As of this moment, time separates the completed work from the work remaining. This research focuses its attention on the cost forecasting problem, estimate at completion (EAC) is the forecasted final cost of the project, as the project progresses. Following the PMBook [1] guide can be evaluated in three ways:

- Actual cost at time-now plus the remaining project budget modified by a performance factor: this approach is commonly used when the current variances are typical of the future ones. This approach is the only one studied in the dissertation, the remaining part of the work is not seen as a stand-alone project but it will take into account a correlation between the past and the future performances;

- Actual cost at time-now plus a new estimate for all the work remaining. This approach is most often used when past performances show that the original estimating assumptions were fundamentally flawed, or when they are no longer relevant because of a few changes in conditions;

- Actual cost at time-now plus the budget of the work remaining. This approach is used when actual variances are atypical; as of this moment, the project management team expects that such measured variances will not occur again at any time in the future.

Even estimate at completion presents some criticalities: the accuracy, the possibility to integrate multiple information sources and deterministic behaviour. The further the project advances, the higher accuracy of the estimate, as shown in figure 2.2, but, at the same time, the fewer possibilities to influence the final outcomes.

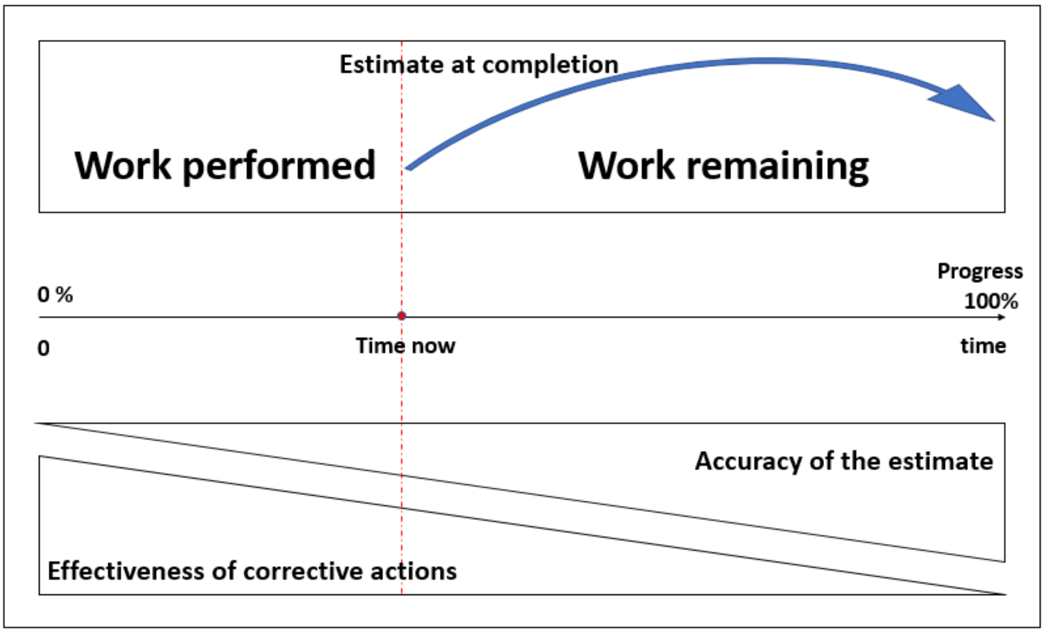

Figure 2.2: progress and control effectiveness relationship

As the project draws to a close, the work remaining where a manoeuvre margin is still possible is reduced, so the most accurate estimates are present but the control is less effective. It is clear the importance to have good estimate since the very first stages of the project where the control system is powerful. The second issue concerns the information sources used to forecast; in traditional method like EVM system, an evaluation of the estimates is based on past performances indicator, extrapolated by the work already done. It could be useful to develop a system that, when other information sources are available, allows their integration to improve the results quality. Other typical information sources are, for instance, the experts subjective judgement, a power instrument tested for a long time. Another source could be historical data coming from similar projects; surely, in this case the company has to keep trace of the past project. In the end, the last point is the deterministic behaviour of the estimate, current technique as the EVM system provides as results a point estimate. Unfortunately, the environment in which the project is developed is plenty of uncertainty, more suitable for a probabilistic approach than for a deterministic one. The uncertainty level is a mixture of several factors that change their impact in every project (human, organizational, political, etc.). The situation gets worsened because the measurement and control system is not free from errors, thus, it could cause distortion of the estimate. What appears clear is the need of a probabilistic value that give the decision maker a wider view of the future possible scenarios.

2.5 Earned Value Management

2.5.1 Introduction to EVM

Nowadays the commonest technique in cost performance forecasting is called "Earned Value Management". This concept was first expressed by the USA Defence Department (DoD) in 1967 and was imposed as Cost-Schedule Control System Criteria to every private joining governmental project. At first presented as a collection of criteria to follow in order to match meaningful results, in 1996 some private companies under the supervision of the DoD developed the Earned Value Management system [30]. Since that time EVM has taken place in common practice. That is also because it is endorsed by the Project Management institute (PMI) and has been used by private company and public agency like NASA, American Department of Energy and department of Defence. As PM-Book guide [1] states:

> *"Earned value analysis in its various forms is the most commonly used method of performance measurement. It integrates scope, cost and schedule measures to help the project management team assess project performance".*

The relevant advantage of the technique is to be perfectly applicable to project of any type, size, duration or complexity because all the project could be described by three quantities: the planned value (BCWS budget cost of work scheduled), the earned value (BCWP budget cost of work performed) and the actual cost (ACWP actual cost of work performed).

A correct EVM application demands almost two requirements, first of all a project status constant monitoring, of the cost bore and the progress achieved in the project. Besides, the second requirement is the uniformity in measurements, the adoption of common criteria that guarantee the compatibility on results. EVM took hold in common practice because it is easy to implement and does not require many data which are, on top of that, easy to collect during the process. During the years many formulations were presented to improve the estimate quality. In particular the most used version bases the forecasting on monthly values in order to catch trends which may be hard to be noted, achieving good results in terms of accuracy. The final goal of the technique is to provide an early reliable warning signal about the cost performance of a project [6]. EVM system is used in schedule performance as well with the introduction of the earned schedule, but this aspect will not be explained in this section since cost performances are the main topic.

2.5.2 Formulation

The general EVM framework is presented in the following paragraph. All the technique versions are based on three elements, fundamental components of all the metrics:

- BCWS budget cost of work scheduled, also known as planned value, is the planned evolution of cumulated costs during the project lifecycle. The BCWS reach the budget value in the planned date of project end;

- ACWP actual cost of work performed, represent the sustained cost to perform the activities completed until time now;
- BCWP budget cost of work performed or earned value. Represents the planned cost to perform the activities completed until time now. It corresponds with the project budget when the project reaches the real work completion.

Monitoring at regular intervals those elements, is possible to assess the project status, and summarize it with two indicators: variances and performance indexes.
Based on those above, the element cost and the scheduled performance are obtained as:

- CV, cost variance:

$$CV_{TN} = BCWP_{TN} - ACWP_{TN}$$

- SV, schedule variance:

$$SV_{TN} = BCWP_{TN} - BCWS_{TN}$$

The measurement units of the variances, both of time and cost, is monetary. The subscript TN underlines that the element is evaluated at time now, so emphasizes that all the values are time dependent. In the following formulas the subscript will be omitted according with the notation presents in literature.

Cost Variance measures the difference between actual and planned cost of the work performed. As shown above, schedule variance is the difference between the value of work performed and the one planned at time now; it is a measure of the compliance with what has already been done and what should be done.

Likewise, the same comparison could be expressed with performance indexes. The two presented indices are obtained as ratio so that it is dimensionless. The used elements are the same of the variances, but working as ratio the meaning changes: they can be seen as measure of efficiency in terms of cost and time [4].

- Cost performance index

$$CPI = \frac{BCWP}{ACWP}$$

- Schedule performance index

$$SPI = \frac{BCWP}{BCWS}$$

On the one hand, if CV is positive and CPI is higher than one, the work performed will not cost as much as planned; on the other hand, if CV is negative and CPI is lower than one, the work performed will cost more than planned. A similar analysis could be performed for schedule indices: indeed, if SV is positive and SPI is higher than one, there will be an anticipation in the scheduling; failing that, the project will be late. The Table 2.1 displayed below may be a useful recapitulation to understand the meaning of the indices as all the possible scenarios are investigated.

		Schedule		
		SV >0 & SPI >1	SV = 0 & SPI = 1	SV <0 & SPI <1
Cost	CV >0 & CPI >1	Ahead of schedule and under budget	On schedule and under budget	Behind schedule and under budget
	CV = 0 & CPI = 1	Ahead of schedule and on budget	On schedule and on budget	Behind schedule and on budget
	CV <0 & CPI <1	Ahead of schedule and over budget	On schedule and over budget	Behind schedule and over budget

Table 2.1: EVM performance indicators

To better understand the nomenclature, a list of all the acronyms is presented in table 2.2.

The standard EVM forecasting is grounded in the hypothesis that cumulative performance indices (CPI_c and SPI_c calculated with cumulative value of $BCWS$, $BCWP$, $ACWP$) will not only be indices of the past but of the future performances as well. The latent assumption is that the detected variances are caused by structural problems that will be present until the project ends. The Estimate is done by summing the already bore cost plus the work remaining adjusted after considering the performances.

$$EAC = ACWP + \frac{BCWR}{PF}$$

CPI_c, as cumulative value, after the 20% of the development will not

$ACWP$		Actual cost of work performed
$BCWS$		Budget cost of work scheduled
$BCWP$		Budget cost of work performed
$BCWR$	$= BAC - BCWP$	Budget cost of work remaining
BAC		Budget at completion: initial cost quote
ETC		Estimate to completion: forecast at time now of the project cost to be sustained from time now to the projects end.
PAC		Planned at completion: initial duration quote.
EAC	$= ACWP + ETC$	Estimate at completion: total final cost forecast at time now.
CV	$= BCWP - ACWP$	Cost variance
SV	$= BCWP - BCWS$	Schedule variance
CPI	$= BCWP/ACWP$	Cost performance index
SPI	$= BCWP/BCWS$	Schedule performance index
PF		Performance factor

Table 2.2: EVM nomenclature

vary more than 10% tending to stability. Over the years a lot of alternatives of the standard formula have been presented: Anbari [4] and Christensen [5] conducted an extensive research on their applicability. Researchers focus on cost performances, the EAC is evaluated considering the work remaining not as a stand-alone project. Past and future variances are instead supposed linked: the most used versions are presented in table 2.3.

Type	Performance Factor	Description
Original	$PF = 1$	When past performances are not a good indicator and the project is supposed to follow the plan
Standard	$PF = CPI_c$	Standard formula
Moving average	$PF = CPI_m(t)$	Moving average of incremental CPI over the last 'm' intervals
Last period	$PF = CPI_{pm}$	Value of CPI registered on the most recent period
Weighted	$PF = w1 \cdot CPI_c$	Composite index, w_1 and w_2 are weight and their sum has to be 1, the value is a project manager decision.
% Complete	$+w_2 \cdot SPI_c$	Is a modified version of the Weighted formula where weights change according to the percentage of project completion
Cost schedule index SCI	$PF = \%C \cdot CPI_c$	Combines the effect of inefficiency and delay supposing that sooner or later temporal delay will be transformed in a cost raising to recover the delay. It is a prudential approach.

Table 2.3: Performance factors

Among the previous indices, the most used in real practice sees PF as the cost performance index evaluated as a moving average over the last three periods. The choice comes from the good results achieved that reflect a well-balanced trade-off between the too stable CPI_c and the too sensible CPI_{pm}. Earned Value Management theory is too extended to be fully explained here and a more complete introduction can be found in other sources (Flemming and Koppelman [6]).

2.6 Critics to EVM

Earned value management (EVM) has provided methods for final outlooks and, largely, these have never undergone any attempts of improvement since their formulation. Consequently, the available methods are often oversimplified, based on inconsistent hypotheses and they sometimes provide unreliable and thus not usable results. Three main problems are recognisable: the methodology does not consider the presence of possible errors in the project description, the high sensibility during the project initial phase and the prolongated effect of errors in the estimate.

Firstly, the first EVM limitation is the deterministic nature of the provided results, without considering the presence of possible errors; the project is developed in an environment which presents a large amount of variabilities that makes the estimate uncertain and subjected to instability within the project progress. Moreover, the EAC is evaluated by using input values that are measured on the field but this does not mean there cannot be any margin of error. Despite the fact that assessing the expended money up to a given moment is easy, it is never a simple task to find the value of the Budget Cost for Work Performed ($BCWP$).

Since the $BCWP$ is nothing but the amount of money that should be spent in order to achieve the actual progress, the typical way to find it is to multiply the project budget by the real achieved progress. The process complexity, and thus the errors, arise when the overall performed work needs to be quantified and synthetized in a percentage value. The presence of countless activities that require different amount of efforts and resources, the combination of several disciplines and the absence of a strict guideline will never lead to a univocal and flawless percentage progress evaluation. The situation got worsen since the percentage progress evaluation is based on the comparison with the project baseline, that could not reflect the real work needed to complete the project. In fact, errors, a lack of experience, and more often a political and competitive pressure which is usual for companies working in a strongly competitive sector to underestimate risks and overestimate opportunities in order to win the bidding could cause an inaccurate planning. Because of the intrinsic project uncertainty and the possible estimation errors, a probabilistic result is needed because it could be very beneficial for the decision maker to know how reliable the estimate is.

Secondly, another problem is related to the low level of accuracy, especially in the early stage of the project, when the small sample size of data the estimate is based on do not allow to assess a statistical reliable forecasting [7]. The basic EVM estimate idea is that the future performances will be equal to the detected ones hereinbefore; afterwards, they will be used to modify all the remaining work. It is easy to understand that, when the project is at an embryo stage, the remaining work is an important part of the whole so, despite a small drop in performances, even not systemic, a huge impact on the final estimate will occur. This problem is more significant at the very first stages of the project where little observation makes the performance estimator too responsive.

Thirdly, the last issue is about the prolongated effect of errors in the estimate. As said before, the EAC is evaluated adding at the present sustained cost the one that has still to be sustained but modified by a cost performance factor. In this condition the effect of a short cost performance fluctuation is enormous, and the problem is even more serious since the fluctuation could be caused by a measurement error as explained before. In this case, even if the performance index is correctly assessed in the following observation, the error is taken in to account still for other two periods, cause the three moving average system.

Chapter 3

Kalman estimate at completion model

3.1 Introduction

A new probabilistic forecasting method, Kalman estimate at completion (K-EAC), is shown below to supervise the project performance and foresee the project estimate at completion with the related probability distribution. As a novelty, the model is built through the implementation of the classical Kalman filter formulation in the project control field exploiting the frameworks of the EVM model. As a consequence, the application of this new tool provides several advantages: first of all, it provides the results, both the cost variance and the estimate at completion, not as punctual values but with the related distribution probabilities, leading the decision maker to the way the outcomes could be trusted. Second, the method conveys to high levels of accuracy and timeliness. Third, it takes into account a few essential problems, such as the quality and reliability of the input data.

The project control is an unceasing process of monitoring the status and the performances of the ongoing project in order to detect early signals of swerve from the plan and, if needed, find and implement the more appropriate corrective actions. The efficiency and effectiveness of the process depend on multiple factors, such as the time it takes to detect a bad behaviour, the choice of the correct responsive action and the quickness of its implementation. A second grouping of variables which highly affect the process concerns the quality of data, more specifically the baseline plan viability and the accuracy of the measured performances and project progress; nevertheless, this latter category is a real problem in project management since both planning and measuring tools are not perfect and, thus, exempt from errors. As an example, during the planning phase, the baseline cost is developed by making assumptions about the availability and the cost of materials, services and subcontractors which could be actually different or affected by price fluctuation; at the same time, it is difficult to access the real achieved progress and synthetize it in a percentage value. All things considered, the huge project dimension, the high number of different activities involved, the presence of not-homogeneous tasks with different physical properties or measurement units, often overlapping with each other, make a perfect evaluation impossible.

The objective of this research is to improve the capability of project managers in reaching decisions by providing a reliable forecasting method for the final cost that not only will it supply the punctual EAC, but it will define its distribution as well. Kalman filter combined with EVM technique can be used as of the onset of the project without the risk of a loss of accuracy; its implementation is straightforward and requires input such as available data that could be integrated with various kinds of prior information, such as data series or expert experience to improve its accuracy.

3.2 The Kalman filter

Kalman filter, also known as linear quadratic estimation, is an algorithm that uses noisy observation to estimate the true but hidden state of the system. The filter was named after Rudolf E. Kalman, one of the first developers of the algorithm in 1960. Since its publication, the applications covered a large range of fields from technology to finance and its common uses are for guidance, navigation and control of vehicles, particularly in the aircraft and spacecraft fields [8]. Furthermore, it is widely applied in the time series analysis area especially in signal processing or econometrics. It plays a central role in robotic motion planning and control and it is sometimes included in trajectory optimization, images processing and tracking in radar technology.

The filter works iteratively creating a learning loop composed of three main recursive steps which are well explained in Figure 3.1 prediction, measurement and posterior estimation.

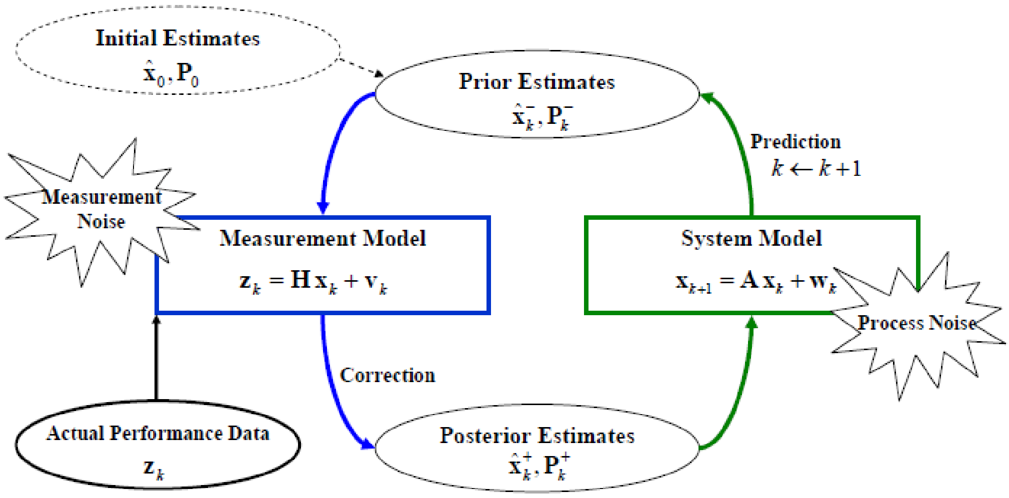

Figure 3.1: Recursive algorithm of Kalman filter

The state of a dynamic system is described by two sets of variables: the state variables and the error covariance variables. The state variables directly represent the system parameters while the error covariance is the indicator of the estimate uncertainty.
The states and covariance matrix are updated by two stochastic linear models, the measurement model updates the previous estimate under the evidence of the observation and, instead, the system model foretells the future system state at the following time step.

Item by item, during the prediction step the algorithm produces a prior estimate of the current state variables and their uncertainty. Later, the measure of the state variables is performed, unfortunately the outcomes will be necessarily corrupted by instrument noise and measurement errors. The last passage is the posterior estimate, performed as a weighted average between the a priori estimation and the measure, giving more importance to the less uncertain factor. An updating phase is necessarily required before restarting the loop. The algorithm is recursive and can run real time using the present measure and the previously evaluated state as the only input.

The Kalman filter theory is too wide-ranging to be explained here; a good introduction could be found in essays (Zarchan and Musoff [8], Brookner [9], Welch and Bishop [10]). The framework has been extensively studied all over the world and many notations are currently in use. In order to avoid confusion, this dissertation follows the Welch and Bishop one [10].

3.3 The model

3.3.1 Kalman EAC general framework

Based on the general theory of Kalman filter, the application to forecast the project EAC has been developed by analogy with the missiles tracking application. The main steps of the algorithm are synthetized in the block diagram in Figure 3.2.

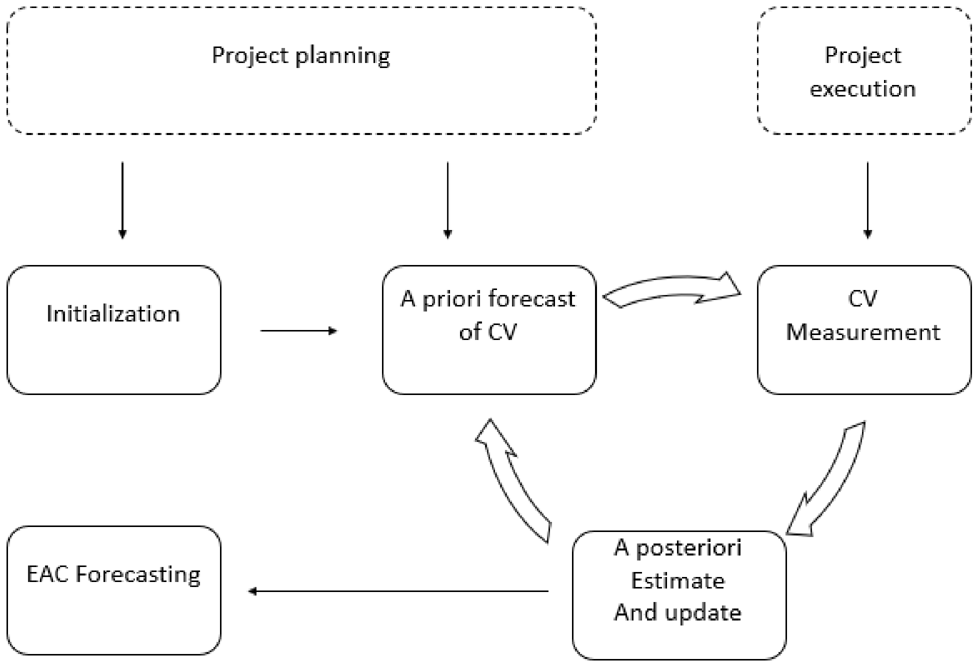

Figure 3.2: Kalman EAC forecasting algorithm

The Kalman estimate at completion (K-EAC) is based on a recursive learning cycle that aims to detect the real but hidden states of the on-going project, the output information is then used while performing a probabilistic estimate at completion. At every time the algorithm is applied to detect the real distribution of the cost variance by using a trade-off between a prior estimate from a linear system model and the detected measure corrupted by errors. This balance is weighted depending on how reliable the two parameters can be, considering both the uncertainty of the model and the errors in the measurement (problems previously described). This consideration will result in a posterior state estimate and its probability distribution, used according to the EVM technique for detecting a probabilistic EAC. The algorithm works following some key points.

During the project planning phase it is important to develop the required input: the baseline progress curve using project plans, and the prior probability distribution of the final project cost. As common to all the algorithm, the quality of the results depends on the input correctness, here the input information has a double use, to initially set the filter and after that to derive the system model to be applied during all the project life. Projects are usually developed under the guidance of the baseline plan that includes all the project prior information such as the schedule, the work breakdown structure, the resources plan and the budget. The baseline plan has got a central role in the project executions as it represents the approved time phased plan; it is used for being compared with the actual performances and for detecting the presence of deviations in terms of cost and schedule.

Graphically, a typical way to represent the project progress level is the S-curve: the baseline progress curve is a representation that shows every time the cumulative values of progress planned to be achieved until a given time, the budget provided to complete the work, while the slope indicates his time derivative. The curve is used to derive a state space equation which shows the knowledge of the future project status. It is clear that talking about future will be characterized by uncertainty. The second required input is a probabilistic distribution of the final project cost: this need arises from the probabilistic nature of the filter and will be treated in the initialization phase. To correctly evaluate, it is important to consider all the possible risks according to the project environment.

Prior system state estimate: a system model is developed using the baseline curve to predict the state and covariance variables at the next reporting time. As previously described, the cumulative project progress during the execution are assumed to follow a dynamic state space equation, representing the project baseline curve. The equation proves the knowledge of the future status of the project and so it is affected by uncertainty and modelling errors. Seeing as how the model is focused on the cost estimate, it is completely described by two quantities and their related uncertainties: the cost variance and its rate of change. These variables are used in the system model for performing a prior estimate of the state status at the next time interval.

Measurement: During the execution, the project performances are measured and periodically accounted as cumulated progress. During the project execution a continuous monitoring is performed and the actual performances are measured. Every time instant the sustained costs and the achieved progress percentage are evaluated. These values are needed when assessing the project performance indicators that are systematically compared with the plan for the purpose of detecting the presence of deviations. However, the real project status is not possible to acquire because it is concealed and corrupted by measurement errors.

Posterior system estimate: prior estimate and measurement are used for the estimate of the real state of the system and its probability distribution. The posterior estimate is evaluated in order to consider the errors presence, both in the model and in the measure, because neither planning tools nor measuring ones are free from errors. During the planning phase for example, the baseline cost is developed by making assumptions about the availability and the cost of materials, services and subcontractors that, during the project execution, could be different or have price fluctuation. Meanwhile, it is hard to access the real achieved progress and synthesize it in a percentage value, the huge dimension of the project, the high number of different activities involved, not-homogeneous tasks with different physical properties or measurement units that often overlap with each other: all these factors convey a high degree of uncertainty to any attempt for a perfect evaluation. This step results in the distribution probability of the index able to describe the current status of the project.

Forecast: the obtained probabilistic results are used with EVM technique for the assessment of the EAC distribution. So as to perform a more accurate estimate at completion with a probability distribution, the parameter could be applied soon after the removal of the effect of noise and random fluctuation.

Update: the variables are updated, and the algorithm is ready to be applied on the next time instant. The filter proceeds by repeating the prediction process in a recursive learning cycle until the project completion.

3.3.2 Required input

Like all the forecasting techniques, K-EAC requires some input information too, briefly shown in Figure 3.3. In addition to the actual data measured at every iteration, that are the actual cost spent and the cumulated progress achieved, the algorithm requires prior information about the project and the environment in which it is developed as the baseline curve, the planned at completion (PAC), the budget allocated to the project (BAC) and a prior distribution of the final cost.

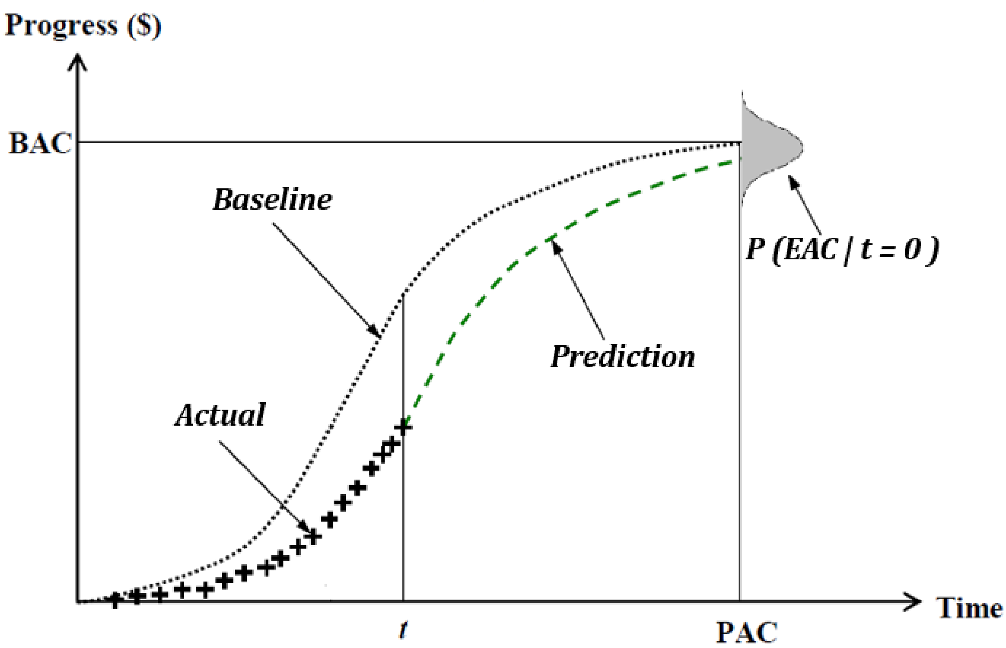

Figure 3.3: Required inputs

At every iteration the measurement process requires all the data to assess the current system status. Since in the proposed formulation the ongoing project is described by the cost variance, the needed elements are the $ACWP$ and the $BCWP$. Unfortunately, the budget cost of work performed is not directly accessible, and to assess the value the easiest technique is to multiply the BAC by the percentage progress value. With regards to the prior information, PAC and BAC are easy to find because they are included in the project definition and represent together the project objective. The need for the prior distribution of the final cost at the very beginning arises from the probabilistic nature of the Kalman filter. This distribution will consider the risk probability and impact magnitude present in the project environment that could lead to a final cost which will not meet the expectations which were foreseen in the budget. The prior final cost distribution could be found in several ways: the easiest one to be implemented is to approximate a triangular distribution based on optimistic and pessimistic cost. If need be, the model could be integrated with other information sources as experts opinions or data from similar completed projects. A tailored prior distribution allows the achievement consistent outcomes more quickly, thus reducing the warm up period. Finally, also the project baseline is required to perform comparisons with the actual values and the detect variances. The prevailing way to represent it is an S-curve, a model showing cumulative costs, labour hours and some other quantities plotted over time. The name derives from the S-shape of the curve, flatter at the beginning and at the end while steeper in the middle, typical in project where a gradual start, an acceleration in the middle phase and a slower tail is an ordinary trend are common [1]. The most widespread approach refers the curve to cumulative progress thus it is possible to recognize the amount of work that should be done any time. In the following section all the K-EAC components will be discussed in detail.

The main components are shown in table 3.1.

Components	Equations	Descriptions
State vector	$x_k = \begin{Bmatrix} CV_k \\ \frac{dCV_k}{dt} \end{Bmatrix}$	CV_k is the cost variation and it is defined as the Earned value minus the actual cost at the k iteration.
Dynamic system model	$x_k = A_k \cdot x_{k-1} + w_{k-1}$ $A_k = \begin{bmatrix} 1 & \Delta T_k \\ 0 & 1 \end{bmatrix}$	A_k is the transition matrix, and w_{k-1} is a vector representing the random process noise.
Measurement model	$z_k = H \cdot x_k + v_k$ $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$	H is the measurement matrix, v_k is a vector representing the measurements noise.
Prediction process	$\hat{x}_k^- = A_k \cdot \hat{x}_{k-1}^+$ $\hat{P}_k^- = A_k \cdot \hat{P}_k^- \cdot \hat{A}_k^T + Q_{k-1}$	Calculation of the prior estimate $\hat{x}_k^-$, and of the prior error covariance matrix $\hat{P}_k^-$.
Kalman gain	$K_k = \frac{\hat{P}_k^- \cdot H^T}{H \cdot \hat{P}_k^- \cdot H^T + R_k}$	K_k is the Kalman gain at the $k$iteration, it will indicate how the posterior estimate is weighted.
Updating process	$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H \cdot \hat{x}_k^-)$ $\hat{P}_k^+ = [I - K_k \cdot H] \cdot \hat{P}_k^-$	Calculation of the posterior estimate.

Table 3.1: K-EAC model main components

3.3.3 State vector

The state vector is the objective of the estimating process and describes the status of the project costs. In cost control the commonest tool for monitoring the ongoing project is the cost variance, that is to say the difference between the budget cost of work performed and the actual cost of work performed.

$$CV_k = BCWP_k - ACWP_k$$

As presented in the EVM chapter, cost variance estimates the difference between the actual and the planned cost of the work performed in monetary units. The K-EAC focuses on two states which evolve over time:

$$\left\{ \begin{matrix} x_{k,1} = CV_k \\ x_{k,2} = \frac{dCV_k}{dt} \end{matrix} \right\}$$

Whereas $x_{k,1}$ is the cost variance at the time k, the second element represents its time derivative and so is the CV incremental. It is a measure of the way CV changes along the interval. If the time between the measurements is constant, for example in a periodic monitoring model, the second element is calculated as follows:

$$x_{k,2} = x_{k,1} - x_{(k,1),1}$$

It is crucial to consider that x_k does not represent only the real but also the hidden state of the system; our knowledge will be limited to its estimate that will be indicated by $\hat{x}_k$. In particular, two estimates are performed at every time instant: $\hat{x}_k^-$ the prior estimate based on the system model and $\hat{x}_k^+$ the posterior estimate conducted after the measurement process.

3.3.4 Transition process

The transition process is based on the system model and represents the future project evolution. The underlying idea wants the cost variance at the following time instant to be equal to the one previously modified by a variable quantity. Specifically, the quantity changes according to the time interval elapsing between the two observations and the changing rate of the previous interval. The process uncertainty is taken into account by a term representing the noise. A detailed explanation follows.

$$x_k = A_k \cdot x_{k-1} + w_{k-1}$$

A_k is the transition matrix and it is obtained under three hypotheses:

- constant rate between two consecutive observations: the cost variance increment will be the identical to the previous interval;
- linear approximation of $x_{k,1}$;
- Constant interval between two measurements: this hypothesis is

not necessary but leads to an easier model: A_k will be the same every time interval ΔT. This is the real situation for a periodic monitoring model.

$$x_{k,1} = x_{(k-1),1} + \Delta T_k \cdot x_{(k-1),2}$$

$$x_{(k),2} = x_{(k-1),2}$$

Therefore, it is possible to gather the transition matrix as follows:

$$A_k = \begin{bmatrix} 1 & \Delta T_k \\ 0 & 1 \end{bmatrix}$$

w_{k-1} is the vector of random process noise representing the error and the uncertainty which are present in the system model. The noise is assumed to be Gaussian, zero mean, and white (and so not correlated in time). The noise has a covariance matrix Q_k defined as:

$$E[w_i \quad w_k^T] = \begin{cases} q & \text{for } i = k \\ 0 & \text{for } i \neq k \end{cases}$$

The magnitude of the introduced noise depends on the quality of the data used to build the system model and so how it respects the project behaviour.

3.3.5 Measurement process

The state variables represent the project status. However, in a real project the project status could be assessed only through a performance measurement process, and anyway it is corrupted by measurement errors. The process is represented by:

$$z_k = H \cdot x_k + v_k$$

z_k is the measurement vector and stands for the outcome of the measured CV_k; H is the observation matrix and it is required to pass from the vector of the state matrix to the single value of the measure. Since the cost variance is the only measured performance, z_k is one-dimensional and the matrix H becomes the following:

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Adding the random noise vector v_k, in the measurement process, the uncertainty and the measurement errors are taken into consideration. The random noise term is assumed to be Gaussian, zero-mean, not correlated with the process noise, white and with R_k covariance matrix defined as shown below:

$$E[v_i \quad v_k^T] = \begin{cases} R_k & \text{for } i = k \\ 0 & \text{for } i \neq k \end{cases}$$

Because of the one-dimensional feature of the measurement vector, R_k is a scalar too. The magnitude of the introduced noise depends on the quality of the measurement process and of the used tool, it has to be estimated by experts or with the help of similar completed processes.

3.3.6 State estimate and error covariance

Since the real system state is hidden by models or measurement related uncertainty, its knowledge is represented by the estimate of the state vector $\hat{x}_k$, and P_k error covariance matrix. Both the values are estimated at every instant k before and after the measurement: they will be marked by a minus sign or a plus sign at the apex. As a result, prior estimate of state vector and error covariance will be represented by $\hat{x}_k^-$ and P_k^-, posterior ones instead with $\hat{x}_k^+$ and P_k^+. The error covariance is defined starting from the state error estimates. The error vector e_k is nothing but the difference between the real but hidden state and the estimate performed. If there are both prior and posterior estimate, there will be two vectors for the state errors as well.

$$e_k^- = x_k - \hat{x}_k^-$$

$$e_k^+ = x_k - \hat{x}_k^+$$

As from these values it is possible to define the covariance error matrix:

$$P_k^- = E[e_k^- \quad e_k^{-T}]$$

$$P_k^+ = E[e_k^+ \quad e_k^{+T}]$$

3.3.7 Prior estimate process

In every period k a prior estimate is conducted through the transition process and based on the posterior estimate of the previous period:

$$\hat{x}_k^- = A_k \cdot \hat{x}_{k-1}^+$$

Along with the state vector, also the covariance error matrix has to be estimated: to do that, it is useful to start from the definition of the prior error covariance obtained from the definition expressed before:

$$\begin{aligned}
e_k^- &= x_k - \hat{x}_k^- \\
e_k^- &= A_k \cdot x_{k-1} + w_{k-1} - A_k \cdot \hat{x}_{k-1}^+ \\
e_k^- &= A_k \cdot v_{k-1} - \hat{x}_{k-1}^+ \\
e_k^- &= A_k \cdot \hat{e}_{k-1}^+ + w_{k-1}
\end{aligned}$$

An evaluation of the error covariance matrix is now possible starting from its definition:

$$P_k^- = E\,[e_k^- \quad e_k^{-T}]$$
$$P_k^- = E\,[\{A_k\, e_{k-1}^+ + w_{k-1}\}\,\{A_k\, e_{k-1}^+ + w_{k-1}\}^T]$$
$$P_k^- = E\,[A_k\, e_{k-1}^+\, e_{k-1}^{+T}\, A_k^T] + E\,[A_k\, e_{k-1}^+ w_{k-1}^T] + E\,[w_{k-1}\, e_{k-1}^{+T}\, A_k^T] + E\,[w_{k-1} + w_{k-1}^T]$$

Since the error vector and the system noise vector could never be correlated, the final expression of the error covariance matrix becomes:

$$P_k^- = E\,[A_k\, e_{k-1}^+\, e_{k-1}^{+T}\, A_k^T] + E\,[w_{k-1} + w_{k-1}^T]$$
$$P_k^- = A_k\, P_{k-1}^+\, A_k^T + Q_{k-1}$$

Defining $Q_(k-1)$ the process noise matrix, a term which means the system model uncertainty.

3.3.8 Kalman gain evaluation and Posterior estimate process

In the Kalman filter formulation the posterior estimate $\hat{x}_k^+$ is evaluated as a linear combination between the prior estimate $\hat{x}_k^-$ and the weighted difference between the performed measure z_k and the predicted measurement $H\hat{x}_k^-$.

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H \cdot \hat{x}_k^-)$$

The Kalman gain K_k rules the trade-off between the prior estimate and the measurement considering how much they can be trusted. The gain is evaluated in order to minimize the posterior error covariance. The result is provided by:

$$K_k = \frac{P_k^- \, H^T}{H \, P_k^- \, H^T + R_k}$$

It could be important to highlight that when the transition matrix A and the measurement matrix H are constant over time, as in the previous simplification hypothesis, the Kalman gain will depend only on the initial error covariance P_0, identified at the onset of the project, and noise covariance matrix R_k and Q_k. It is easy to understand the relations behind the gain: if the magnitude of the measurement covariance matrix increases, which is a symptom of a bed performed measurement, the denominator of the gain will grow too conveying to a lower K_k. The closer the gain is to zero, the more the posterior estimate will be mainly based on the analytic model prevision. The inverse relation is intuitively correct as the effect of a new observation decreases as long as the uncertainty on the measure increases.

Once obtained the Kalman gain, the posterior error covariance is evaluated as follows:

$$P_k^+ = [I - K_k \cdot H] \cdot P_k^-$$

The control of the Kalman gain is not only useful when measurement has low reliability but also when it is not reliable at all. In the latter case, the Kalman gain will be set equal to zero and the algorithm will provide an outlook by only considering the system model.

3.3.9 Forecast

The output of the Kalman filter at each step is a gaussian distribution of the cost variance. Making only linear transformation over variables, all gaussian for, the cost variance will have a gaussian distribution with average equal to the first element of the state vector and a variance equal to the first element of the covariance error matrix.

$$P_k^+(1,1) = \sigma^2$$
$$\hat{x}_k^+(1) = \mu$$

At every time instant, the reliability of the posterior state vector is estimated and projected in order to forecast the EAC and its distribution. EAC formula is derived from a general EVM definition. The final cost is the sum of the already spent money plus the cost of the work remaining: the latter addend has to be modified by a performance factor to be consistent with the current cost performances.

$$EAC = ACWP + \frac{BCWR}{CPI}$$

$$EAC = ACWP + \frac{BAC - BCWP}{\frac{BCWP}{ACWP}}$$

$$EAC = ACWP + \frac{BAC \cdot ACWP}{BCWP} - ACWP$$

$$EAC = \frac{BAC \cdot ACWP}{ACWP}$$

$$EAC = \frac{ACWP}{ACWP + CV} \cdot BAC$$

3.3.10 Updating

The algorithm is iterative, this implies that an updating step is required at every interval. The algorithm stops once the work is completely performed.

3.4 Filter initialization

Since the iterative nature of the filter, some parameters and the initial variables values have to be set before the first iteration. At first, the initial state vector and the covariance error matrix values are set equal to zero. In addition, the values of the process noise covariance matrix and the measurement error matrix have to be estimated in advance. This process, usually called "filter initialization", is a very challenging phase since it is performed when no observation from the project is available. The importance of this phase has been widely highlighted in literature by many studies [11] [12] [13] [14], which could provide an extensive dissertation; in this section only the initialization process related to the K-EAC is going to be discussed.

First, the initial variables, both the state vector and the error covariance matrix, are set to zero.

$$\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\hat{P}_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

When the project begins, the method has or should have a well-defined starting point concerning the work which is supposed to be done, the starting time and the initial cost that is reasonable to be set equal to zero, thus there cannot be any chance to have cost variance and the state vector is initialized to zero. Following the same line of reasoning, there is no doubt about the state status, since no progress are achieved, and no money are spent, the absence of uncertainty is reflected into a null error covariance matrix.

The process noise covariance matrix, acting directly in the Kalman gain (K_k), takes into account the system model uncertainty due to lack of information or presence of errors [8]. The matrix is modelled to act on the CV_k derivative over one interval [9]. The process noise covariance matrix Q_k is evaluated as the covariance of the process noise vector w_k:

$$Q_k = Cov[w_k] = \overline{\begin{bmatrix} 0 \\ w_k \end{bmatrix} \begin{bmatrix} 0 & w_k \end{bmatrix}} = [w_k][w_k]^T$$

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & \overline{w_k^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_w^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

The diagonal terms represent the variances of the state variables, the extra diagonal instead represents the covariance values. If only random errors of each state variables are considered, the off-diagonal terms are zero [15] according to the hypothesis developed in the previous chapter.

Indeed, q represent the process noise variance, measure of the model uncertainty, and directly acts on the filter convergence. The more the value approaches the zero, the more the system model makes correct estimates of the future; conversely, a high value accords with an increase of the uncertainty affecting the process. The variances are supposed to be constant, not for any rational or empirical results but because there is no information supporting an alternative interpretation.
The value of q is assessed in order to make the model uncertainty coherent with the users prior estimate of the project final cost distribution. The user provides as input the expected project duration (PAC) and the distribution of the expected final cost expressed with the mean μ_c and the variance of the distribution σ_c^2. These are used in an inverse Kalman forecasting algorithm to determine the value of q. In detail, the algorithm, which is based only on the system model, works equivalently to set the gain K equal to zero for all the project duration. In the analysed case where measurements are performed at a constant time interval, the resulting equations are:

$$\begin{aligned}
x_k^- &= Ax_{k-1}^+ \\
P_k^- &= AP_{k-1}^+A^T + Q_{k-1} \\
K_k &= 0 \\
x_k^+ &= x_k^- \\
P_k^+ &= P_k^-
\end{aligned}$$

Provided that the project lasts as planned (PAC) and the model uncertainty at the beginning of the forecast process is equal to the users prior estimate of the project final cost variance:

$$P^{+}_{k=PAC}(1,1) = P^{-}_{k=PAC}(1,1) = \sigma_c^2$$

It is possible to evaluate q since it is the only variable.

The last element to be set is the measurement error matrix R_k. It represents the accuracy of the measurement and it is expressed as the covariance matrix of the measurement noise vector v_k.

$$\begin{aligned} R_k &= Cov[v_k] \\ R_k &= \overline{[v_k]\,[v_k]^T} \\ R_k &= \overline{[v_k]^2} = [\sigma_v^2] = [r] \end{aligned}$$

Where r is the measurement error variables and takes into consideration the variance of the measurement error σ_v^2 , r influences the forecasting method sensibility, in particular if it approaches zero, the Kalman gain will increase and, as a consequence, the measured quantity will have higher impact on the posterior estimate. Vice versa, high r will decrease the gain when making the posterior state estimate trusting more the prior estimate than the measured performance. In order to set the value, the program evaluation review technique (PERT) is used [16] [17] [18] and a three-point estimate for the measurement error. The user needs to define the maximum possible measuring error, the variance of the error is evaluated thanks to the PERT technique.

$$Maximumerrorv_k = Emax$$

$$Minimumerrorv_k = -Emax$$

$$\sigma_v^2 = [\frac{Emax - (-Emax)}{6}]^2 = \frac{Emax^2}{9}$$

The value could be adjusted by the project manager to correctly fit different types of project that are developed in different environment. If need be, modifying r is possible to attach either great or little importance to the measured performance with respect to the model estimate.

Chapter 4

Application of the forecasting model to three oil and gas projects

4.1 Introduction

In this chapter it is going to be presented in detail the application of the K-EAC model to three real cases. The oil & gas sector will be the core field of the analysed cases; the reason of this choice is to better understand the complex dynamics present in this field. To follow, a brief presentation of the typical project structure and its features. The three projects have already been accomplished, thus they are going to be thoroughly described by focusing on the scope of work, the physical and economical progress achieved during their execution. After that, the forecasted results will be analysed and to understand the model added value they will be compared to the EVMS (Earned Value Management System) methodology outputs that represent the state of the art in the field of forecasting techniques. More precisely, the three-month moving average EAC will be exploited because of its remarkable frequency:

nowadays, seeing as how it succeeds in achieving the best outcomes, it is the most used EVM version by project managers. The evaluation focuses its attention on the most challenging forecasting requirements, accuracy and promptness, which are all the key points for a flawless decision-making process. As previously highlighted, a recent aspect of the method is the probabilistic behaviour: this enables a better project status description supporting the project manager choices in the project execution.

The data in the K-EAC implementation are provided by a company working worldwide in the oil & gas sector. The firm usually operates in foreign countries, where the reservoirs of hydrocarbons are located, creating a partnership with the host country. Not only do they acquire a research and extraction authorization upon payment of royalties, but they directly let the host country share in the production earnings: in this scenario the company is seen as an administrator of the oil reservoir, still owned by the host country. Nevertheless, it is noteworthy that this sector represents a few limiting cases, because of its huge size and highly intrinsic complexity. In addition to these difficulties, from the point of view of the project environment, it is recognizable that some hard criticalities can increase the uncertainty level, the high number of stakeholders involved and the political, financial and climate related factors that highly influence the project performances. Thus, the activities that concern the oil exploitation need a high-level planning and risk management system. This is the reason why the push towards more performing techniques in the risk minimization requires the introduction of specific, more complex and diversified methodologies that working in synergy and in an integrated manner to reproduce a model which can be as close as possible to reality. In any oil & gas company the control of the physical and economical progress has a central role to understand when the plant is ready to start up the production, that is to say to realize the moment it will begin making profits out of its activity. Considering the huge financial capital deployed in the project development, it is of great importance to have a control system that allows the decision maker to pursue effective and prompt decisions. These results are achievable only in case the decision process is supported by accurate information obtained through analytical and mathematical models rather than simple and subjective methodologies.

4.2 The project development process

Before describing the model application, it is useful to introduce a brief description of a typical project in the oil & gas sector in order to understand the main characteristics in this field. Broadly speaking, the project evolution process could be schematized in the following phases listed in Figure 4.1.

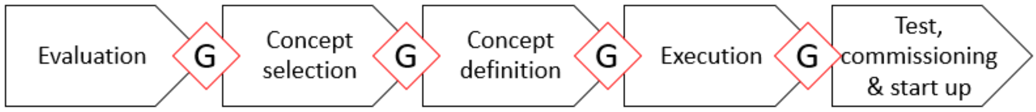

Figure 4.1: Project development process phases

- Evaluation: after an oilfield presence is detected through research & exploration activities, a feasibility study is performed to pinpoint the possible settlement sites and to verify the alignment with the company policy;
- Concept Selection: several possibilities for the energetic source to be exploited are developed. Technical and economical assessments are performed in order to detect the one that maximizes the project economical value;
- Concept definition: a detailed analysis of the selected solution is performed to develop the project plans which will be core elements during the execution phase;
- Execution: implementation of the planned project trying to respect the contract constraints in terms of time, budget and quality;
- Commissioning, start-up and performance test: tests are performed to assess whether the performances are compliant with the ones previously established in the contract. This is the time the plant starts the production.

The rhombuses in Figure 6 represent the process gates, key steps where an evaluation of project characteristics occurs. If they are in line with the established requirements, the process will proceed to the next phase, otherwise it determines the project closure. The K-EAC model will be applied at the execution phase, core of the project, from the start of the operative activities until their closure that conventionally coincide with the first day of production (first oil) [31] [32]. The model objective is the evaluation of the final cost at the end of this phase. Evidently, in order to identify the cost deviations at each time now, the values obtained during the execution phase will be compared to what was planned in the concept definition phase.

4.2.1 Typical oil & gas sector projects

In the oil & gas sector it is possible to identify three clusters of comparable in terms of size and scope of work: offshore, onshore and subsea. A brief introduction of the type in order to contextualize the model application will be shown below.

Offshore projects involve operations of plants construction and installation for drilling and hydrocarbons extraction in the sea. The adopted platforms are characterized depending on the sea depth and the structure typology, floating or fixed (see figure 4.2).

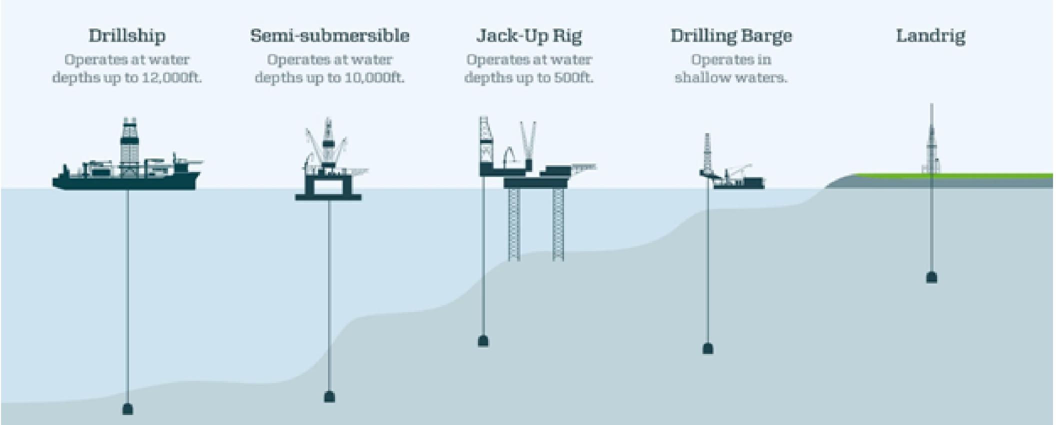

Figure 4.2: Type of offshore platform

The projects include also sea lines installation, ducts for the transport of the extracted material to the storage units in the land. A typical example is shown in Figure 4.3.

Figure 4.3: Offshore fixed platform

Onshore projects, instead, are characterized by the construction and installation of plants for directly drilling and extracting on land, with the installation of underground pipelines for the material transportation.

At first, the oil extracted from the drilling well is stocked *in loco* and later it is transported by pipelines to the refinery, only after being sent to a treatment plant. Gaseous hydrocarbons follow the same pattern until the treatment plant after being directly sent to the user by methane pipeline. Examples are presented in Figure 4.4. Projects belonging to this cluster present some criticality during their execution. The main reasons are:

- Authorization and permissions are given with higher difficulties than it is generally for the other two categories, in the country soil, authorities are more reluctant to authorize plants that could be result invasive.

- Local subcontractors are far less reliable and skilled labour is quite difficult to source. Political agreements with the host country bind the company to exploit exclusively unskilled local manpower: as a consequence, the quality plummets, the costs exceed and the delays run and grow.

Nevertheless, that criticalities are often underestimated. The company usually adopts an aggressive planning behaviour to win the contract bidding.

Figure 4.4: Onshore extracting well

Subsea projects concern the construction of submarine extraction plants. This solution is adopted when the offshore platform cannot be used both for technical and economical unfeasibility. In case of multiple submarine extraction points, they have to be linked with flowlines, the same connections have to be installed also to connect the wells to the storage platforms or the onshore stoking plants. An example could be seen in Figure 4.5 . Compared to the two other categories, this project type is generally characterized by a better financial exposure since the few companies performing this kind of work employ skilled manpower, more reliable in terms of cost, time and quality of the work performed than the local labour.

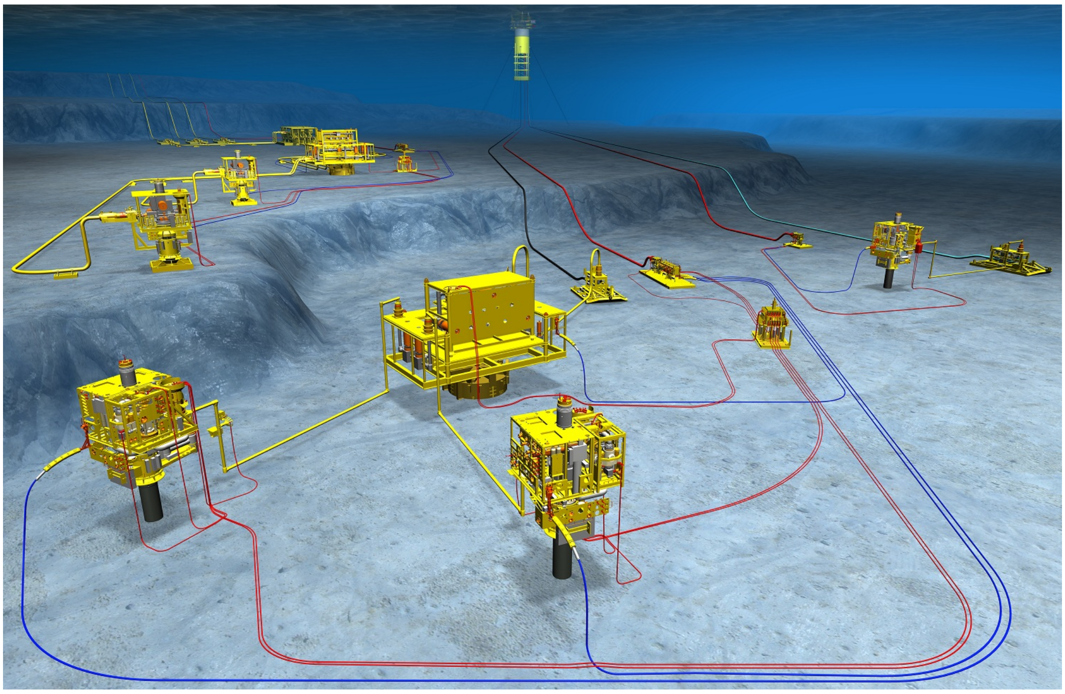

Figure 4.5: Subsea extraction plants

4.3 Project description

The presented forecasting model is applied to three real projects, one for each of the identified category. Analysing already completed cases is possible to assess the physical and economic project life before the project application. In order to respect the company privacy, the cases of study will not be named with the real project name, but with one established by convention:

- Subsea cluster: Case A;
- Offshore cluster: Case B;
- Onshore cluster: Case C.

4.3.1 The A Case

The first analysed case is part of the subsea category. The project consists in the setup of an FPSO (*Floating Production Storage and Offloading Unit*), a ship whose aim is to extract, to stack and to perform the oil preliminary treatment operation.

The objective of work is composed by three main elements shown in Figure 4.6

- the installation of the FPSO, a permanently moored ship, is needed to stock the drilled oil and to start the preliminary treatment phases;
- the installation of flexible sea-lines that will connect the ship to three existing wells;
- the installation of a submarine umbilical control system that regulates the wells and the valves that enable the oil to flow from the seabed to the FPSO, in order to prevent oil leakages.

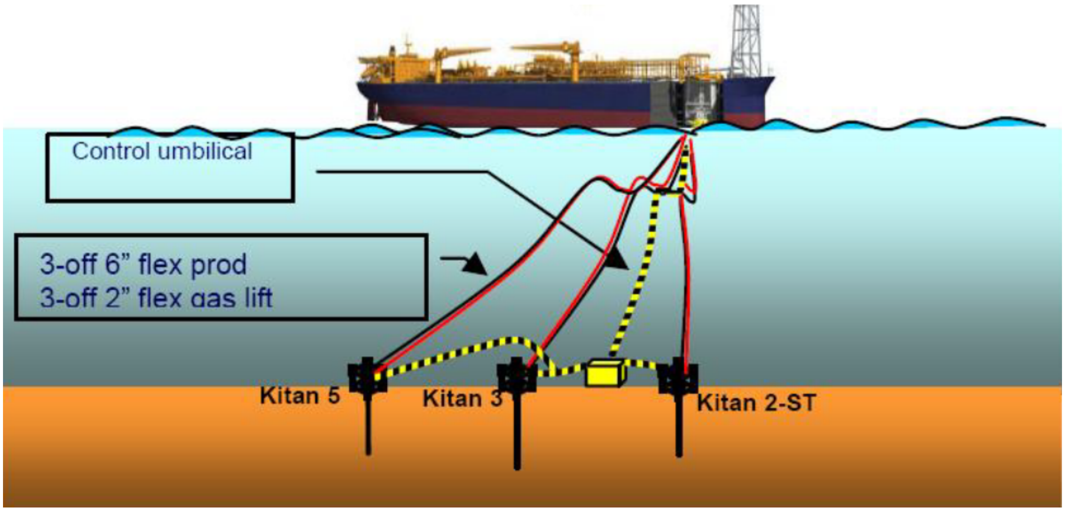

Figure 4.6: Scope of work project A

The working site is in the Australian North Sea, an area where the company has never worked before: this inexperience of the area conveys more uncertainty to the project. In this case the company does not known the standard subcontractors performances. This could be an issue during the concept definition where the time and the cost of activities have to be estimated to meet the contract obligations.

The A Case is an accomplished project, all the data at our disposal are used to run the model. In Figure 4.7 and Figure 4.8 the project history is presented, with both the physical and the economical progress.

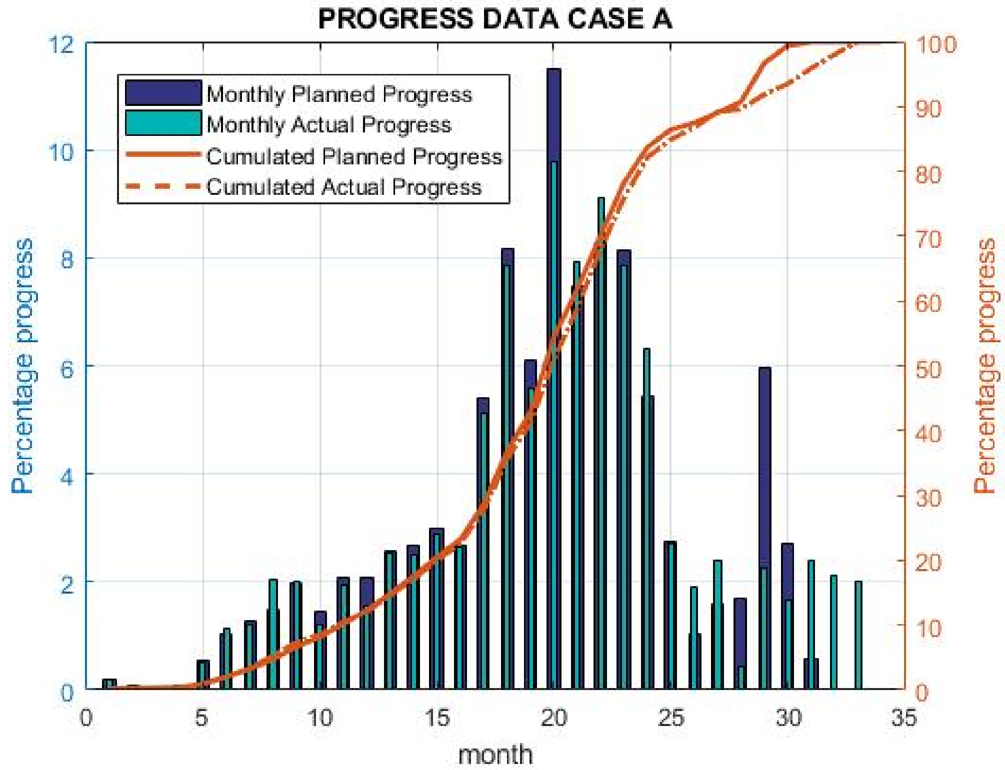

Figure 4.7: Case A physical progress description

In Figure 4.7 the project is described through its physical point of view: the red lines represent the cumulated progress, the dotted one is the actual cumulated progress, thus the progress achieved during the execution. The continuous line shows the planned progress, it is the project baseline established during the concept phase. In the bar chart there are the same information but in monthly values. The project starting date is March 2009 and, according to the scheduling, it should last 32 months; against these expectations, the end of the work is achieved only after 34 months with a two-month delay. The performances are periodically checked every month: for the sake of clarity, the project duration is identified by the number of the months after the starting date, March 2009. As shown, the actual progress overlaps the baseline until the twenty-eighth month where the delay is collected.

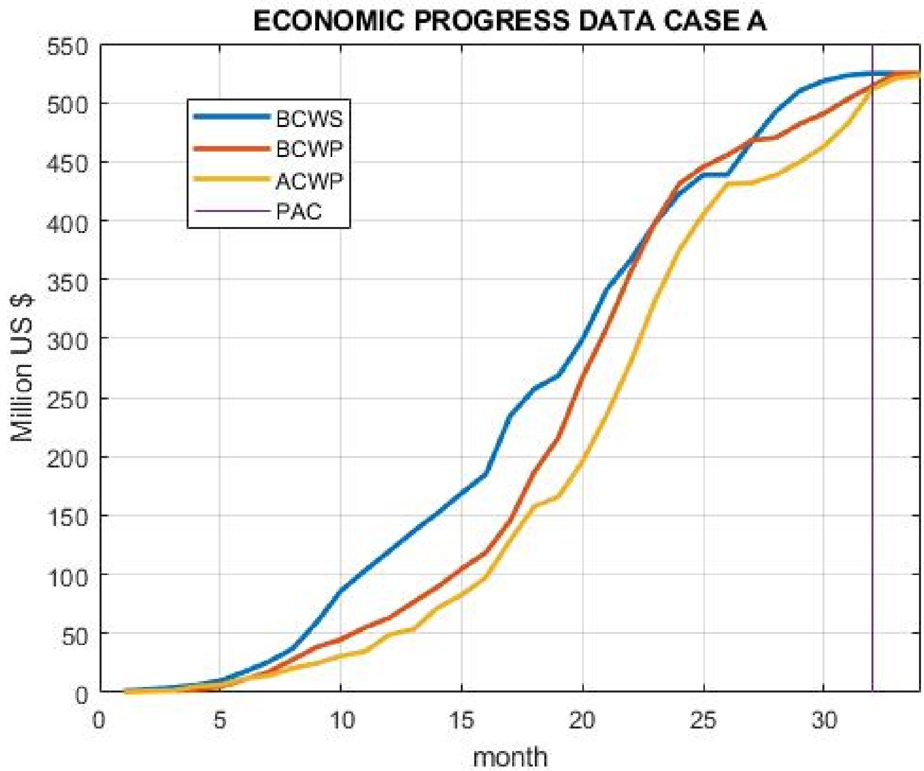

Figure 4.8: Case A economical progress description

Figure 4.8 shows the economical project life. The three represented quantities are identified as follows: the actual cost of work performed is marked in yellow, whereas the cost baseline (budget cost of work scheduled) is highlighted in blue and, finally, the budget cost of work performed obtained by the multiplication of the actual physical progressed by the project budget is identified in red. With regard to costs, the budget at completion amounts to 525 million US dollars but, although the project was carried out in an unprecedented area for the company, the final cost is 523 million US $, less than what expected. All along the project duration the costs do not entail as much money as the money planned to be spent.

4.3.2 The B Case

The second case of the study is part of the offshore cluster. The project area, shown in Figure 4.9, is located in the Gulf of Suez, 77 metres under the Red Sea level. The project development in this site started in 2003 searching the best technical solutions to extract the oil and the optimal place to stock it. The concept selection and its definition last two years: later, in 2005 the execution phase began.

Figure 4.9: Case B location

The selected solution consists in the creation of two oil wells completed with a tie-back production system, that is the connection of new wells to the already existing refinery structure on shore with an already present FPSO. The aim of the work includes also the installation of the needed sealines and pipelines. The environment described is represented in Figure 4.10.

Figure 4.10: Case B site

As follows, the physical and cost progress, respectively in Figure 4.11 and Figure 4.12. As in the A case description, Figure 4.11 shows the physical progress both in monthly values and cumulated. The execution starts in June 2005 and, according to the plan, it should last 32 months. Throughout the project, especially during the central phase, the physical achieved progress is lower than the planned one, unavoidably bringing about a delay of five months to the completion.

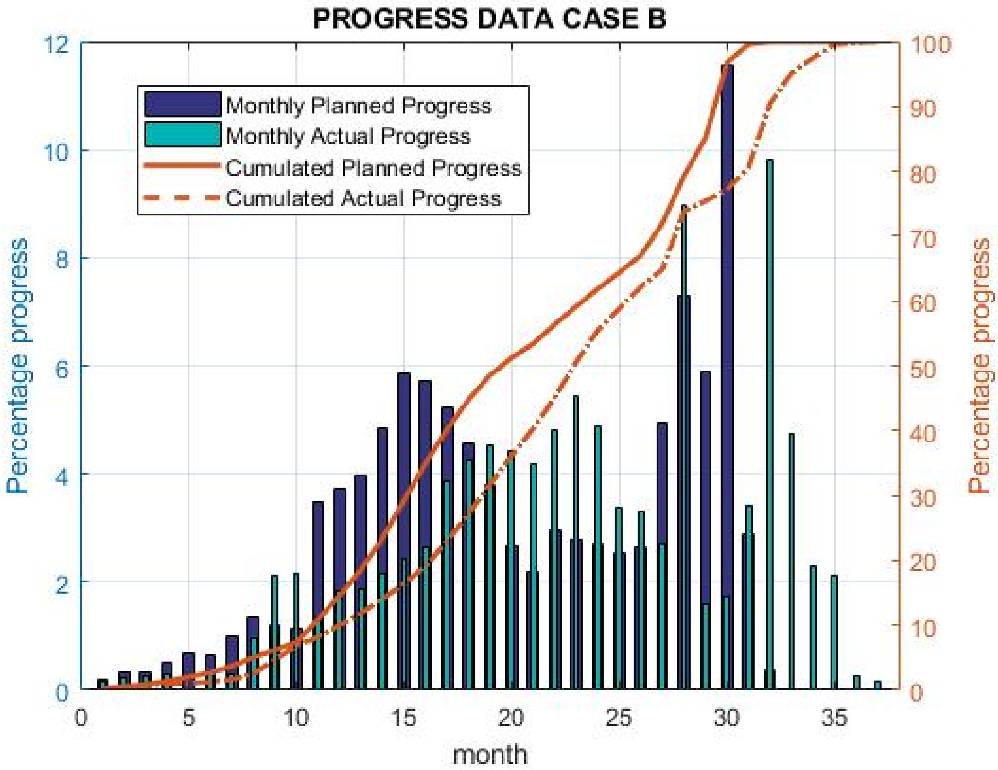

Figure 4.11: Case B physical progress description

The economical situation described in Figure 4.12 is more complex. During the first project phase, despite the overlapping of the actual and scheduled costs, the situation is not simple due to the delay in physical progress. Having a glance at the BCWS, it is clear that the sustained cost and the planned one are alike, but fewer activities than planned have been carried out up to this moment. Throughout the central phase the project development is fitting, the actual costs are lower than the planned ones and even lower than the BCWP. The performances worsen at the thirtieth month when actual costs start rising and overcome the BCWP. As a result, the beginning BAC whose amount was 80 million US dollars turned out to be insufficient: indeed, at the end of the project the final cost exceeded by 7 million US dollars.

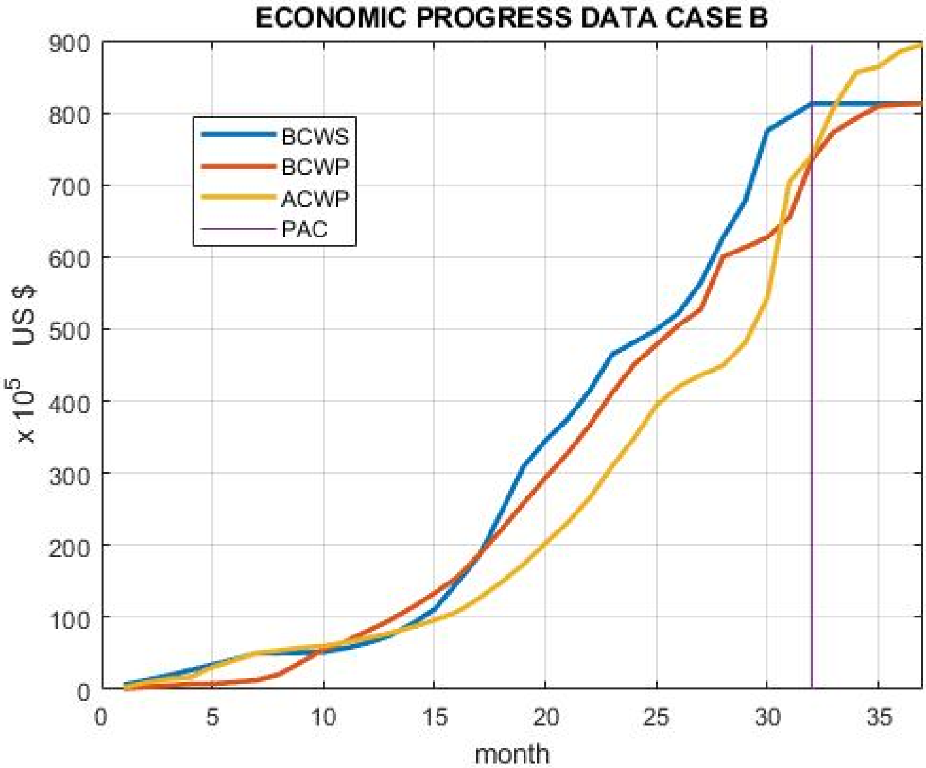

Figure 4.12: Case B economical progress description

4.3.3 The C Case

The last case is part of the onshore cluster. The scope of work consists in the realization of a power plant with two 150 MW gas turbines and in the related distribution station in the Congolese southern border. The project includes also the renovation of a power grid and the installation of a plant for the treatment, stocking and compression of the gas. From the treatment centre the gas will be sent in the above mentioned power plant and in two other plants nearby. The pipelines which are required for the transportation have to be put in place. A provisional draft of the project is presented in Figure 4.13. The project starts in February 2008 with a planned duration of 30 months and a budget amounting to 320 million US dollars.

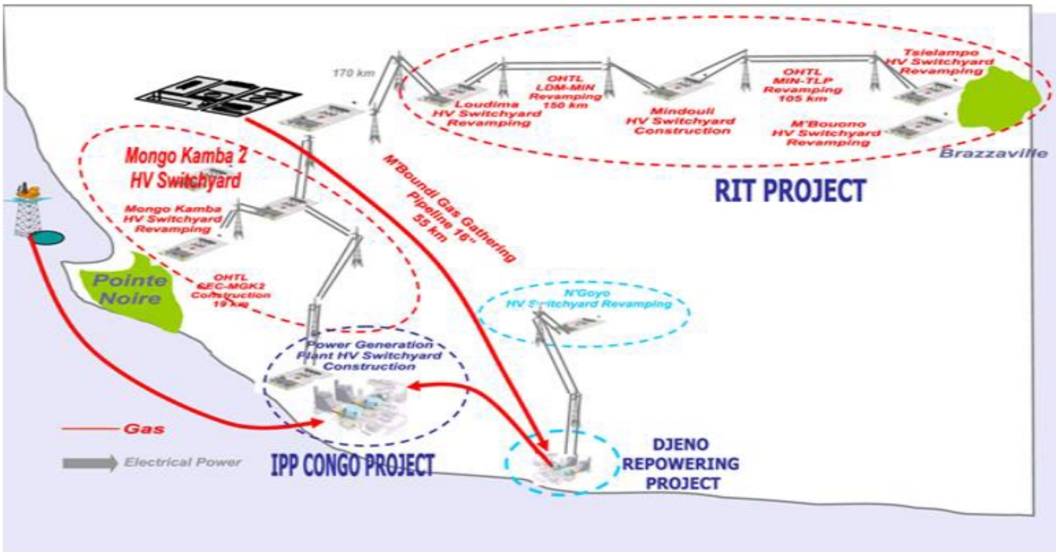

Figure 4.13: Case C scope of work

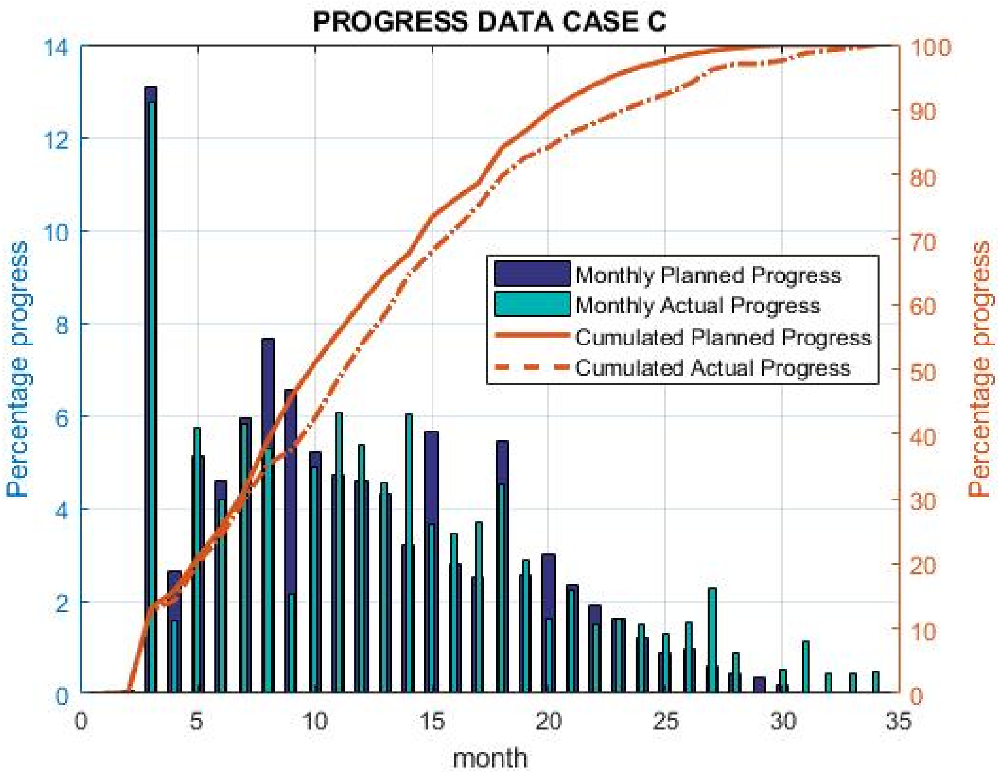

Figure 4.14: Case C physical progress description

The physical progress development is shown in Figure 4.14, during the first third of the project: the progress achieved meets what was in the planning. From the seventh month to the end of the work, the performances are lower than planned and this causes a four-month delay in the completion. From the point of view of the cost, halfway through the project, the actual performances are similar to the planned ones, after the actual cost dramatically rises bringing about a further outlay amounting to 171 million US dollars, equal to 53% of the budget, as shown in Figure 4.15.

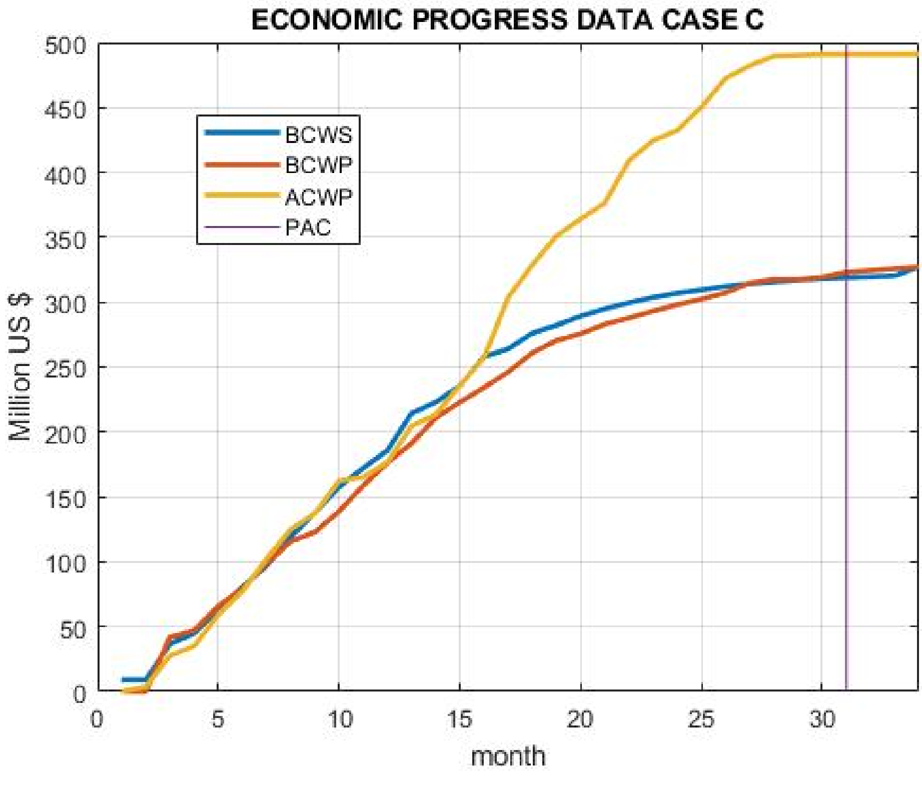

Figure 4.15: Case C economical progress description

4.4 Model application

Uniqueness, as states the PMBOK [16], is one of the intrinsic features of the project. Every project is unique in its objective, in the scope of work, in the plan to achieve its goals and in the adopted management decisions. It is easy to understand that the performances of a forecasting method are not constant in all its application but vary from a project to another depending on specific scenarios [15]. At the same time, it is clear that different forecasting techniques may lead to various results, thus they convey to alternative implementing actions. Despite the same available input data, a few differences may occur in forecasting results due to the intrinsic feature of the applied method. For this reason, the evaluation of the forecasting performances of different techniques is a hard challenge and making an objective comparison may be even more laborious. The problem arises since a project is inherently influenced by the management choices which are based on the forecasting method results, that may be different if a change of the forecasting technique is operated [19]. All the while, a project is unique and cannot be repeated under the guidance of any other different method. The best way to approach this issue is to apply the model to real cases and compare the outcomes to the ones achieved by standard techniques. This section includes the application of the Kalman-EAC model to the three cases presented in the previous chapter. The achieved output is analysed and then compared to the state of the art in forecasting performances represented by the EVM technique in its three-month moving average formulation. The algorithm is applied using MATLAB software: its implementation is effortless and does not require a high power computational software. For this purpose, Excel or any other similar software may succeed in providing the same results.

4.4.1 Case A model application

As described in the K-EAC algorithm presentation, the implementation starts with an initialization phase. The two first elements to assess are the model variables: the state vector and the covariance error matrix. Given that the control is performed from the beginning phase, both elements are set composed by null elements.

$$\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\hat{P}_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This is a self-evident choice because at the beginning of the project a cost variance has no chance to be different from zero since no cost is sustained and no work is performed yet. Nonetheless, there is no doubt or uncertainty in this situation, thus confirming the null matrix choice. The second initialization phase aims to identify the q value, index of the system model uncertainty.

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

q value is assessed to make the model uncertainty congruent with the users prior estimate of the project final cost distribution. The user provides in input the expected project duration (PAC) and the distribution of the expected final cost, expressed with $_c$ and σ^2, the mean and the distribution variance. These elements are used in an inverse Kalman forecasting algorithm to determine the q value. More specifically, the algorithm, solely based on the system model, works equivalently to set the gain K equal to zero for all the project duration. Consistently, as it occurs in the baseline plan, the PAC lasts 32 months and the final cost distribution amounts approximately to a Normal with the mean equal to the budget and a variance equal to 10% of the budget.
The last initialization step is the measurement error matrix R_k.

$$R_k = [r]$$

r is the measurement error variables and takes into account the variance of the measurement error, σ^2. In order to set the value, the program evaluation review technique (PERT) [16] [17] [18] and a three-point estimate for the measurement error are employed. The user has to define the maximum possible measuring error, thus the error variance is evaluated with the PERT technique.

$$Maximumerrorv_k = Emax$$
$$Minimumerrorv_k = -Emax$$

$$\sigma_v^2 = [\frac{Emax - (-Emax)}{6}]^2 = \frac{Emax^2}{9}$$

In this case the value is chosen by setting the max error as 1% of the budget, selected since the $BCWP$ is evaluated as the percentage progress multiplied by BAC. Nevertheless, the progress is challenging to be correctly measured: multiple activities, where multiple disciplines are involved and enormous size makes very tough any attempt of synthesis of the whole performed work in a percentage number. After the initialization the algorithm is ready to run, fed at every stage on the measured cost variance value. The results of the entire project duration are collected in Figure 4.16.

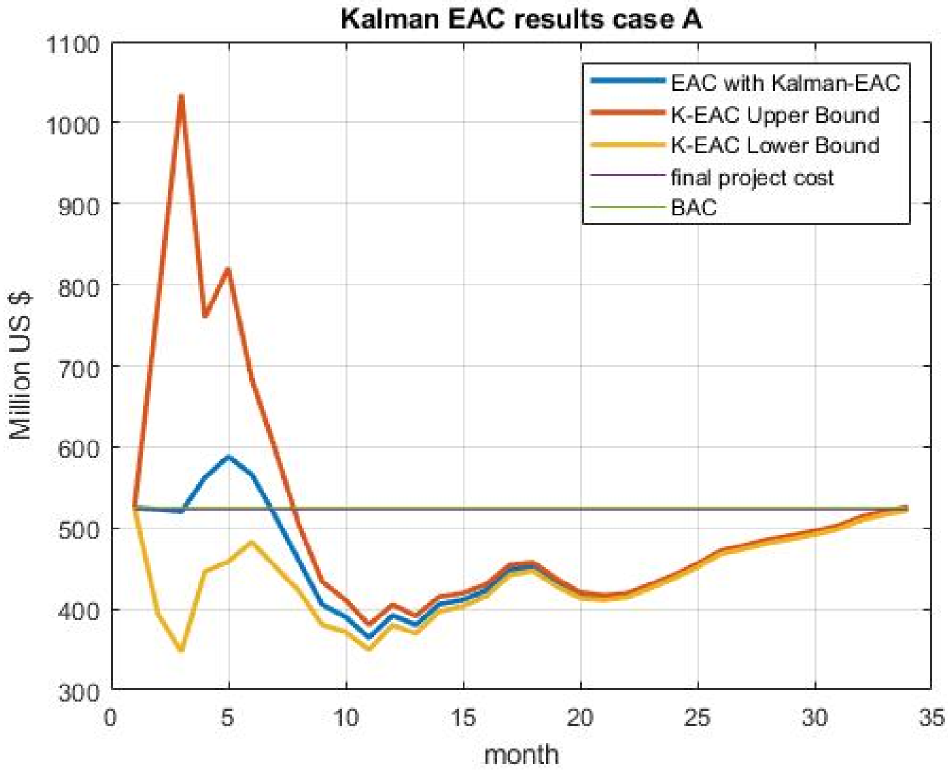

Figure 4.16: Case A Kalman EAC results, the chart shows the forecasted EAC at every iteration

The chart gathers the forecasted EAC evaluated at each iteration. Since the estimate is obtained projecting the cost variance distribution, the output is presented as three lines: the central blue one represents the mean EAC value obtained projecting the CV distribution mean, the yellow and the red lines represent the optimistic and pessimistic values obtained through the projection of the fifth and the ninety-fifth cost variance distribution percentile. The two straight lines represent the budget and the final cost achieved: in this case the values are overlapped. In the first time instants, the algorithm presents a rump up period, the few information at disposal and the huge amount of work remaining makes the distribution estimate too wide. As the project makes headway, more data are available, and the distribution becomes narrower thanks to the more reliable estimations. On the whole, the project steps forward by strictly following the plan in the first months, and, except for a small peak on the fifth month bringing the EAC to almost 600 million US dollars, it is completely underbudget thanks to the optimum economic performances achieved in the central phase. During the final phase the project final cost estimate rises up to 523 million, almost achieving the capped budget. The same graph is presented in Figure 4.17, where the result of the standard EVM technique in the three-month moving average version is added, that represents the commonest technique nowadays: it is drawn in purple, while the blue line represents the central value of the K-EAC estimate distribution as in the graph above.

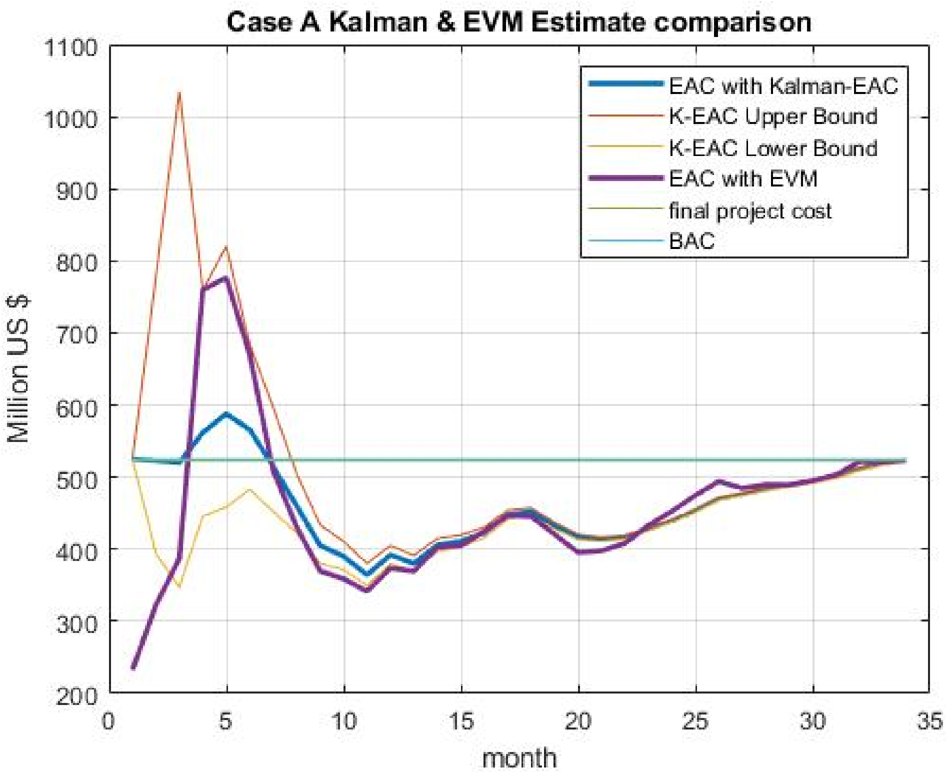

Figure 4.17: Case A Kalman and EVM Estimate comparison

Overall, the algorithms results maintain the same path, index of coherence in the estimates, and indicates a peak during the fifth month. Here the techniques have different trends: both results record a peak, but the EVM estimate is three times as high as the K-EAC one. Such a pronounced estimate variation is caused by a slight downturn on the cost variance only during the fourth month, whose trend is described in Figure 4.18. Both algorithms detect the cost variance fluctuation presence, but the interpretation of the phenomenon is different. The traditional EVM estimate raised immediately the EAC to about 800 million US dollars, and, even if the performance deflection at the following time instant is completely recovered, the algorithm continues maintaining a high final cost. This behaviour originates from a latent hypothesis: EVM considers every variation in cost performance as structural, so it will be featured in the project until its conclusion. The second issue derives from the mathematical evaluation of the EVM estimate; the performance factor that modifies the remaining work is based on the average over three periods, so should the oscillation disappear, it would affect the project performance in the following two evaluations. Differently, the K-EAC detects the performance loss presence but, since it is far from the performance trend, it slightly raises the EAC during the fourth period. Focusing back on the EAC during the execution, in the second half of the project, the two estimates have very similar behaviour, whereas it is important to highlight the result obtained in the first part. Here the K-EAC estimate is more stable with respect to the EVM one: this latter is evaluated adding to the cost sustained up to this moment the planned cost of the remaining work, modified by a performance factor measured over the last periods. It is intelligible that, when the project is in its early phase, the work which still needs to be performed is an observation of some weight in the estimate given that even a tiny performance fluctuation could have a huge impact on the final result. It is important to bear this issue in mind since not all the cost variance fluctuations are due to structural causes that will occur from this moment on, or even worse, they could be brought about by measurement errors. This issue is mitigated in the Kalman-EAC model where the obtained cost variances are the results of the combination of two factors, the measurements and the system model. The cost variance obtained with the K-EAC is cleansed by noisy fluctuations, when the algorithm needs to face a shortage of measurements, it does not tend to immediately trust in measurements, especially if their values are far from the one of the system model. The trend of the cost variance during the project execution is presented in Figure 4.18, where the primary difference between the methods is noteworthy: the probabilistic approach. The K-EAC provides a distribution of the cost variance accounting for the errors present in both the model and the measurement process. This is recognizable in the EAC that is provided with a central, an optimistic and a pessimistic value. The EVM technique provides, instead, a punctual result: in this situation the project manager has no indication about the information quality, thus how reliable they could be. This is a crucial point since the estimates represent the most important supporting tool for the project manager who has to select the corrective actions to implement and their intensity.

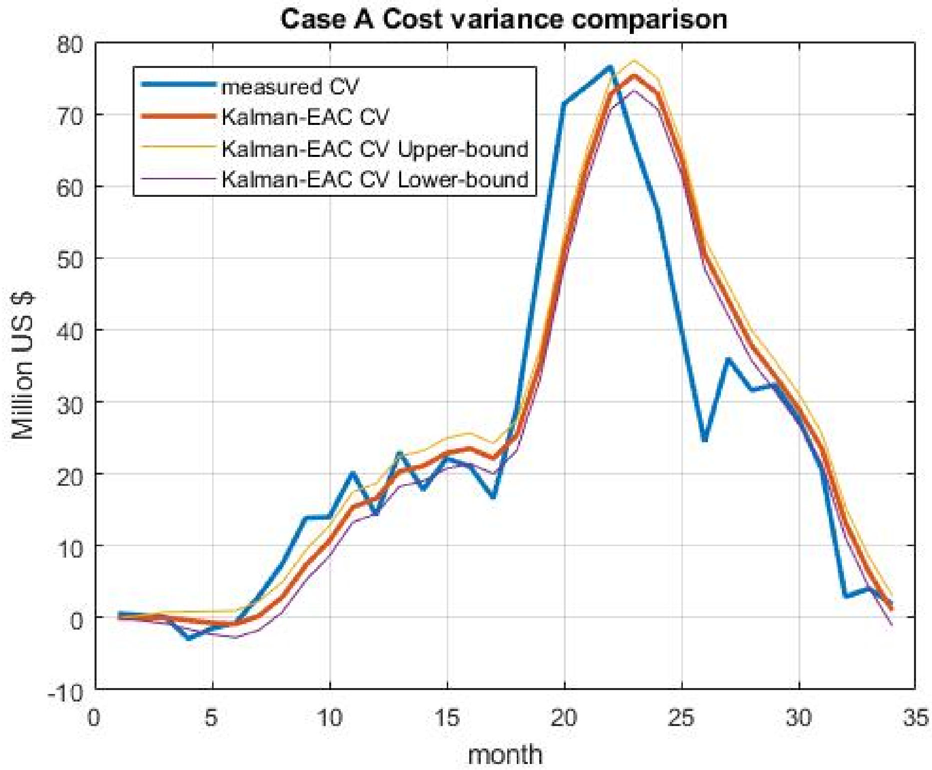

Figure 4.18: Case A Cost variance comparison

The graph shows the measured cost variances in blue, while the cost variance distribution evaluated with K-EAC model is expressed with the central value, the red line, and the fifth and the ninety-fifth percentiles are respectively the purple and the yellow lines. The above described trend is here clearly deduced, the cost variance distribution estimated with the Kalman filter is more smoothed than the measured one, that is used in the EVM evaluation. The trend is the same, but the short oscillations disappear. A second thing that should be noticed is that the estimated CV follows the same trend followed by the measured one but slightly on the right; this phenomenon must be ascribed to the same cause, that is the algorithm does not tend to directly trust a trend far from the system model result and it is supported by little evidence.

Now the comparison must concern the two most important aspects of a forecasting technique: accuracy and timeliness, in this phase the selection of the right criteria for the evaluation is crucial. In the literature of forecasting techniques, the accuracy is reported as the most commonly used parameter among professionals and researchers [20]. As a measure of accuracy, Vanhoucke and Zwikael [21] [22] submitted in their works well-known statistical measure of errors as the Mean Absolute Percentage Error (MAPE). Regrettably, the literature about the forecasting method evaluation is extremely lacking, but most authors tend to centre their research on these kinds of indicators. Accuracy is measured as the average deviation between the forecast and the actual value over a certain time period. In our case the MAPE is evaluated as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{ActualFinalCost - EAC(t)}{ActualFinalCost} \right|$$

The MAPE value is evaluated not only with the K-EAC result, but with the EVM three-month moving average EAC as well: the results are shown in Figure 4.19.

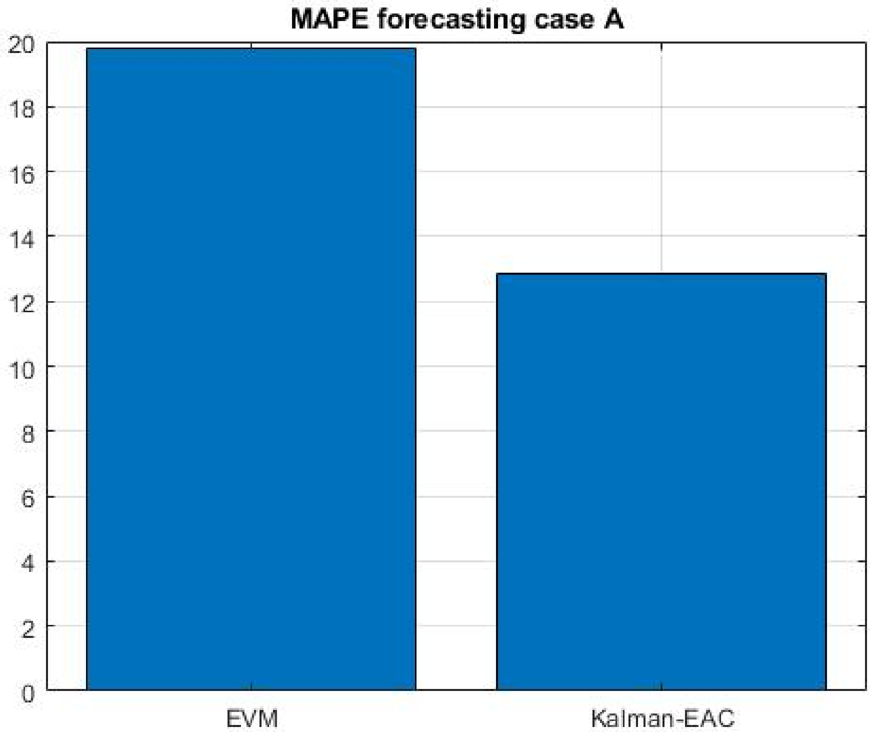

Figure 4.19: MAPE forecasting case A

The proposed method reduces MAPE from 19.816 to 12.836 with an improvement of more than 35%. This result is due to the high difference in the first part of the project, where the EVM estimate is too responsive to performance fluctuations and, even worse, every performance variation is considered as an outcome of a structural cause and affects all the remaining life of the project. The K-EAC model does not trust in performances deviations without evidence, especially if far from the system model output, thus reducing the risk of misleading interpretation. Figure 4.20 presents the difference between the two estimates at completion and the final actual cost, to appreciate the initial K-EAC advantage, whose central value is plotted in the graph. In accordance with what has been proved so far, in this first application the K-EAC model reaches good performance in terms of accuracy.

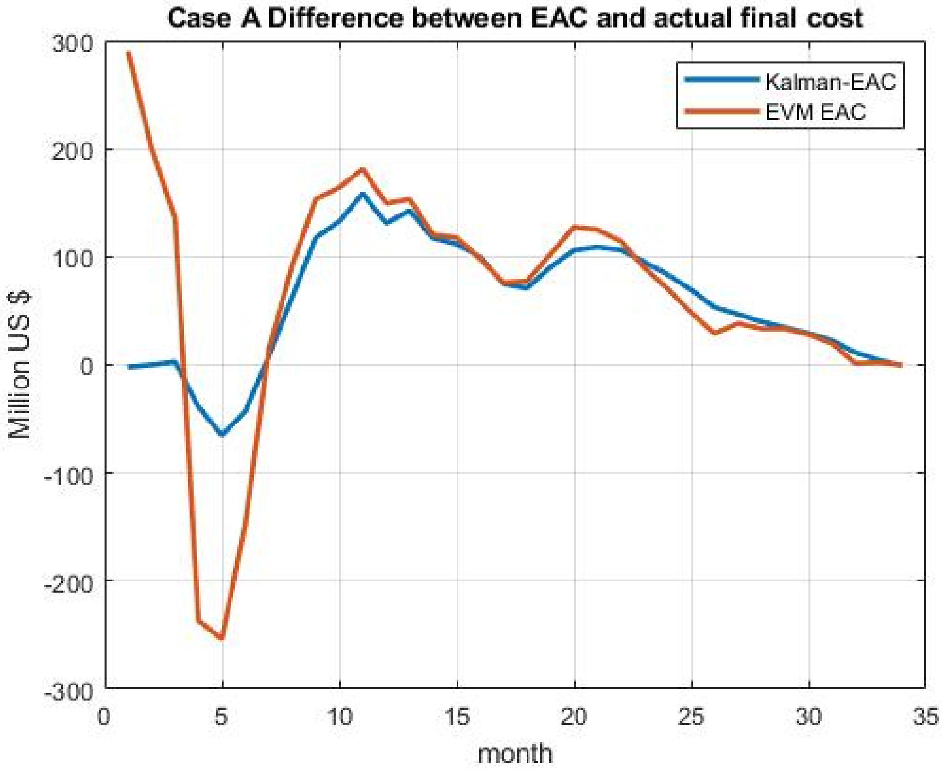

Figure 4.20: Case A Difference between EAC and actual final cost

A second way for the accuracy measurement was brought out by Teicholz [3], who draws attention to a new indicator after comparing 121 construction projects. Together with the classical statistical methods, as mean square error, the accuracy could be represented by the absolute area between the actual final cost and the path of the estimate at completion plotted against the percentage progress: Figure 4.21 shows the application of this method to the A case. This technique introduces some advantages: firstly, it does not suffer from biasedness if the measurement intervals are not constant; secondly, it offers a visual information about the achievement of the best accuracy results.

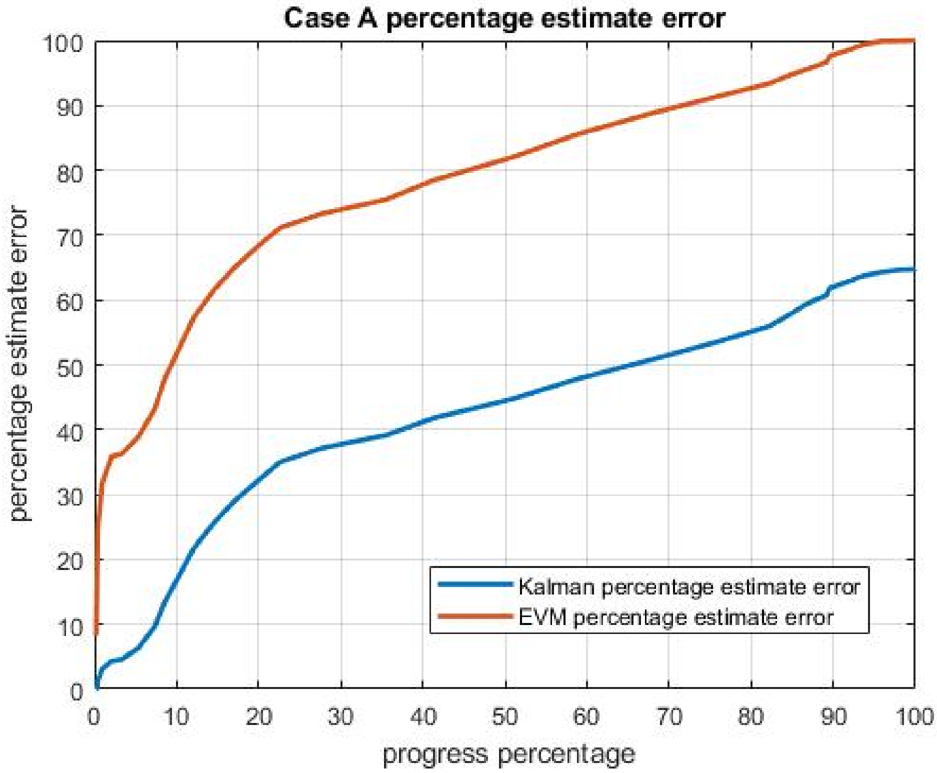

Figure 4.21: Case A percentage estimate error

On the x-axis the project progress is displayed in percentage value, whereas on the y-axis the percentage estimate error is evaluated as the cumulated area between the estimate at completion and the final cost value. In order to compare the two methods, the areas are normalized using the area of the EVM estimate at completion as reference given that it is the reference method. As above mentioned, there is quite a considerable difference on the first part of the project duration that is constant after its first third, in other words after the two lines head for a similar trend due to the fact that the two estimates have similar tendencies. Besides, the graph introduces a second important aspect for a forecasting method: timeliness.

Timeliness is here identified as the ability of the method to provide reliable outcomes over the short term. This is a very significant issue for the project manager who needs to take decisions from the project early stage since correct results are needed as soon as possible. Further in his work, Teicholz figured out a procedure to evaluate the timeliness and which is still nowadays the substantial challenge of forecasting method: the author defined it as the accuracy achieved in the first half of the project. From this viewpoint, it is possible to measure the percentage error of the two estimates at completion in the first half of the project. The A case results are presented in Table 4.1.

K-EAC	EVM
47,74	82.02

Table 4.1: Case A, percentage estimate error on the first project half

The value stands for the total error performed in the first half of the project duration. Both indices are comparable because they were normalized by using as reference the error made by the EVM estimate at completion, that represents the state of the art. It emerges that the Kalman estimate at completion produces more accurate results at the beginning of the project than the traditional technique does. This can be explained, as cited previously, by the twofold nature of the Kalman filter that, using measurement and the analytical system model, could filter the cost variances fluctuations not due to structural causes. A third aspect which requires consideration is the probabilistic estimate introduced by the K-EAC model. In this case, to perform a comparison is not possible since the EVM system performs only punctual estimates. The presence of a probability distribution introduces several advantages: the most valuable is that it gives a measure of how certain the estimate can be. This means valuable knowledge which allows the decision maker to introduce actions that are supposed to deflect the course of the project development.

4.4.2 Case B model application

Also for the B case the first step for the K-EAC is the initialization phase. The two first elements to assess are the model variables: the state vector and the covariance error matrix. Given that the control is performed from the beginning phase, both elements are set composed by null elements.

$$\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\hat{P}_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This is a self-evident choice because at the beginning of the project a cost variance has no chance to be different from zero since no cost is sustained and no work is performed yet. Nonetheless, there is no doubt or uncertainty in this situation, thus confirming the null matrix choice. The second initialization phase aims to identify the q value, index of the system model uncertainty.

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

q value is assessed to make the model uncertainty congruent with the users prior estimate of the project final cost distribution. The user provides in input the expected project duration (PAC) and the distribution of the expected final cost, expressed with μ_c and σ^2, the mean and the distribution variance. These elements are used in an inverse Kalman forecasting algorithm to determine the q value. More specifically, the algorithm, solely based on the system model, works equivalently to set the gain K equal to zero for all the project duration. Consistently, as it occurs in the baseline plan, the PAC lasts 32 months and the final cost distribution amounts approximately to a Normal with the mean equal to the budget and a variance equal to 10% of the budget. The last initialization step is the measurement error matrix R_k.

$$R_k = [r]$$

r is the measurement error variables and takes into account the variance of the measurement error, σ^2. In order to set the value, the program evaluation review technique (PERT) [16] [17] [18] and a three-point estimate for the measurement error are employed. The user has to define the maximum possible measuring error, thus the error variance is evaluated with the PERT technique.

$$Maximumerrorv_k = Emax$$
$$Minimumerrorv_k = -Emax$$

$$\sigma_v^2 = [\frac{Emax - (-Emax)}{6}]^2 = \frac{Emax^2}{9}$$

In this case the value is chosen by setting the max error as 1% of the budget, selected since the $BCWP$ is evaluated as the percentage progress multiplied by BAC. Nevertheless, the progress is challenging to be correctly measured: multiple activities, where multiple disciplines are involved and enormous size makes very tough any attempt of synthesis of the whole performed work in a percentage number. After the initialization the algorithm is ready to run, fed at every stage on the measured cost variance value. The results of the entire project duration are collected in Figure 4.22.

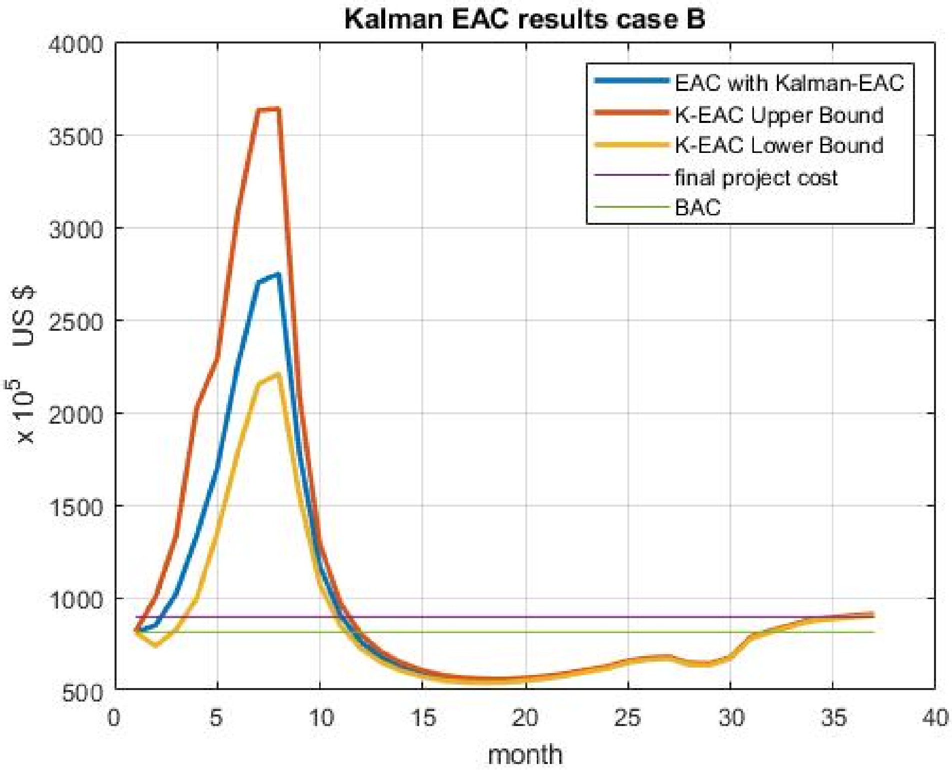

Figure 4.22: Case B Kalman EAC results

The chart gathers the forecasted EAC evaluated at each iteration. Since the estimate is obtained projecting the cost variance distribution, the output is presented as three lines: the central blue one represents the mean EAC value obtained projecting the CV distribution mean, the yellow and the red lines represent the optimistic and pessimistic values obtained through the projection of the fifth and the ninety-fifth cost variance distribution percentile. The straight lines display the budget in green and the final project cost in purple. The obtained EAC path is seesawing: during the first eight months the achieved progress is more limited than expected, generating a negative cost variance trend and reflected in the raising of the final estimated cost. Throughout the project central phase, the improvement of the obtained performances begets an increasing cost variance trend and leads the EAC to values which are lower than the budget. During the last phase, the cost variance becomes negative again, making the final estimate raise. In the first time instants, the algorithm presents a rump-up period, where the few information at disposal and the huge amount of work remaining make the distribution estimate too wide. As the project moves forward, more data are available, and the distribution becomes narrower, since the estimates are more reliable. The same graph is presented in Figure 4.23 with the integration of the estimate performed with the EVM technique in the three-month moving average that represents the most used technique nowadays, drawn in purple.

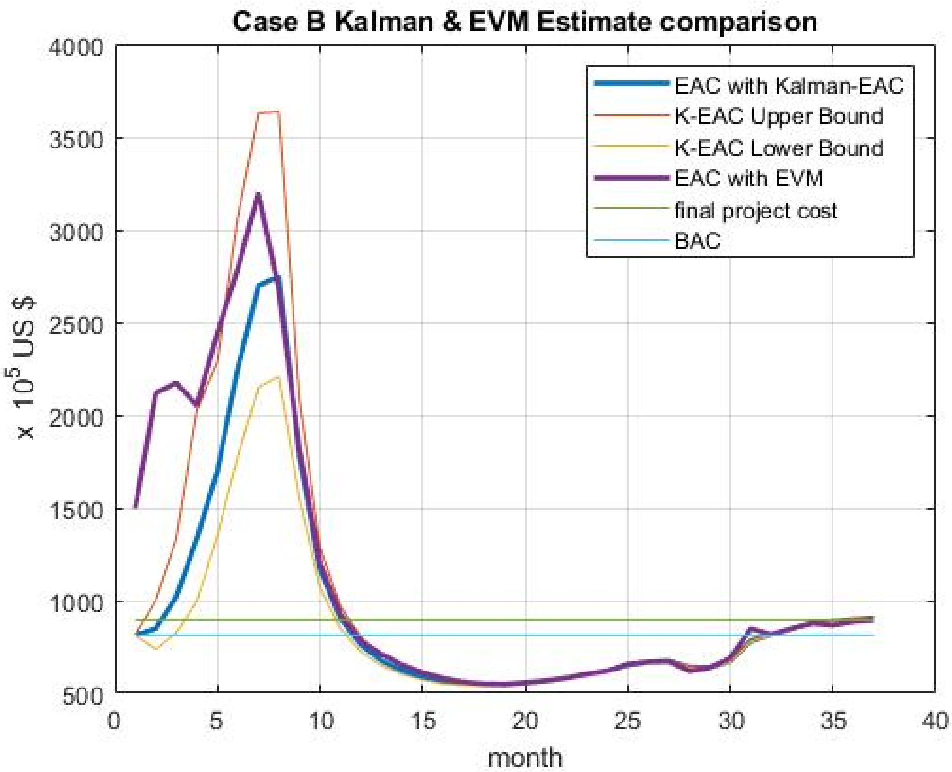

Figure 4.23: Case B Kalman and EVM Estimate comparison

In this application the two methods results are quite alike, suggesting the coherence in the methods, specifically after the tenth month, they overlap almost completely. Some differences are visible during the project early phase, where low project progress achieved makes EVM estimate raise instantly; Kalman-EAC grows at a slower pace. As previously introduced, the proposed method does not directly rely on the performance index if these elements are too far away from the system model results and not supported by previous observation, that is what happened during the two first months, after the outcomes of the two algorithms got closer. The likeness of the two outputs derives from the high stability of the detected cost variance, as displayed in Figure 4.24. The detected and the estimated ones are very close though the filter action is still visible during the thirty-first and the thirty-fourth month where single peaks of short duration decrease. Also in this case, it must be underlined that the probabilistic distribution of the cost variance takes into account the possible presence of measurement errors and the quality of data provided as input, and besides gives origin to an EAC presented with an optimistic and a pessimistic value. In this situation the project manager has indication of how certain the information could be.

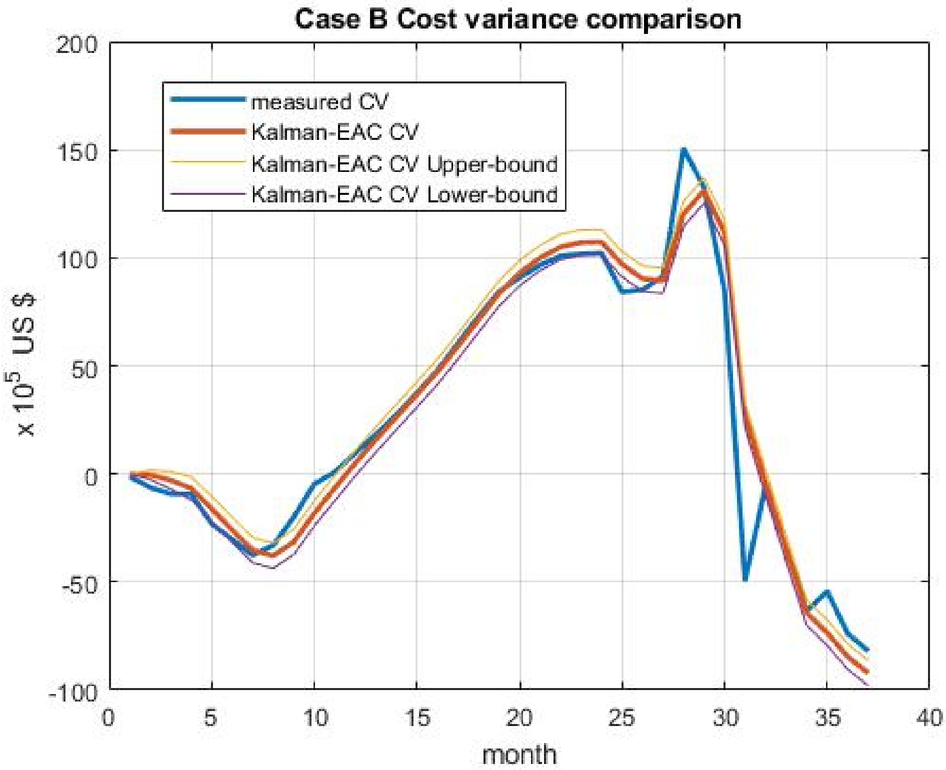

Figure 4.24: Case B Cost variance comparison

The Kalman obtained cost variance is exposed in Figure 4.24 the red line represents its central value, the purple shows the fifth and the yellow the ninethy-fifth percentage of his distribution, while the blue line is the measured one.

Now the comparison must consider the two most important aspects of a forecasting technique: accuracy and timeliness. In this phase it is crucial to select the right criteria for the evaluation. As previously explained, the accuracy will be evaluated through MAPE as a standard measure of error and through the new method estimated by Teicholz [3]. MAPE is evaluated as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{ActualFinalCost - EAC(t)}{ActualFinalCost} \right|$$

The MAPE value is evaluated with both the K-EAC result and the EVM three-month moving average EAC: their results are shown in Figure 4.25.

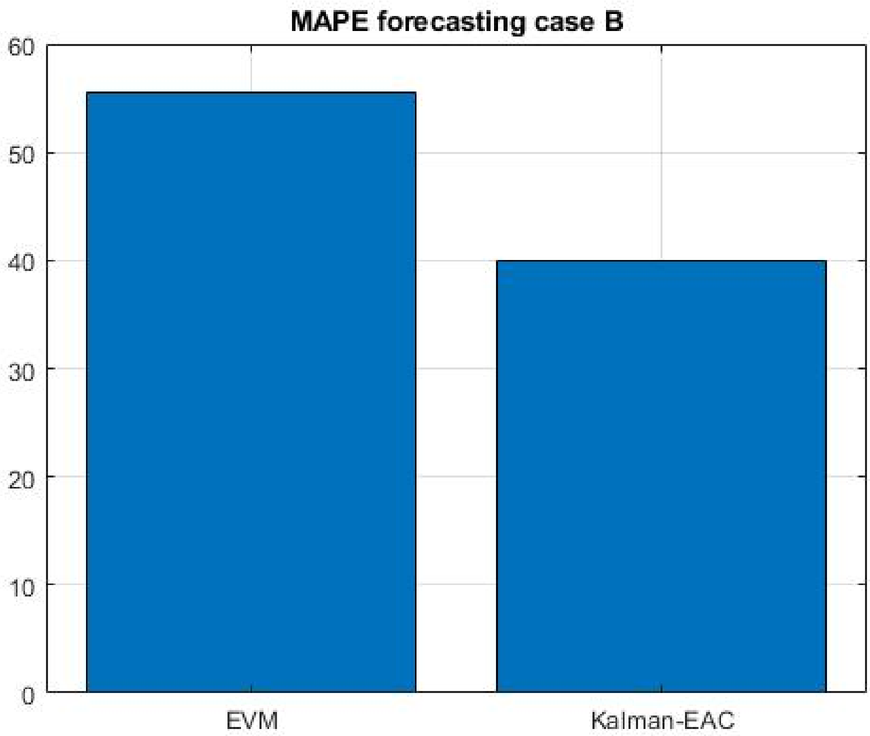

Figure 4.25: MAPE forecasting case B

As shown in the previous graph, the proposed method reduces the MAPE from 55.597 to 39.895 with a reduction which oversteps more than 28%. The result is positive but lower than in the A case since the project does not present as much random fluctuation as in the cost variance. A second way for the accuracy measurement was addressed by Teicholz

[3], who, by comparing 121 construction projects, proposed a new indicator. Along with the classical statistical methods, as mean square error, the accuracy could be represented by the absolute area between the actual final cost and the path of the estimate at completion plotted against the percentage progress. Figure 4.26 shows its application to the B case. The proposed accuracy measure introduces two advantages: it does not suffer from biasedness if the measurement intervals are not constant and, besides allows to better realize where the better accuracy results are achieved; on the contrary, the MAPE is a mean index. The graph shows that the great part of the accuracy improvement is achieved throughout the early project phase, as previously described.

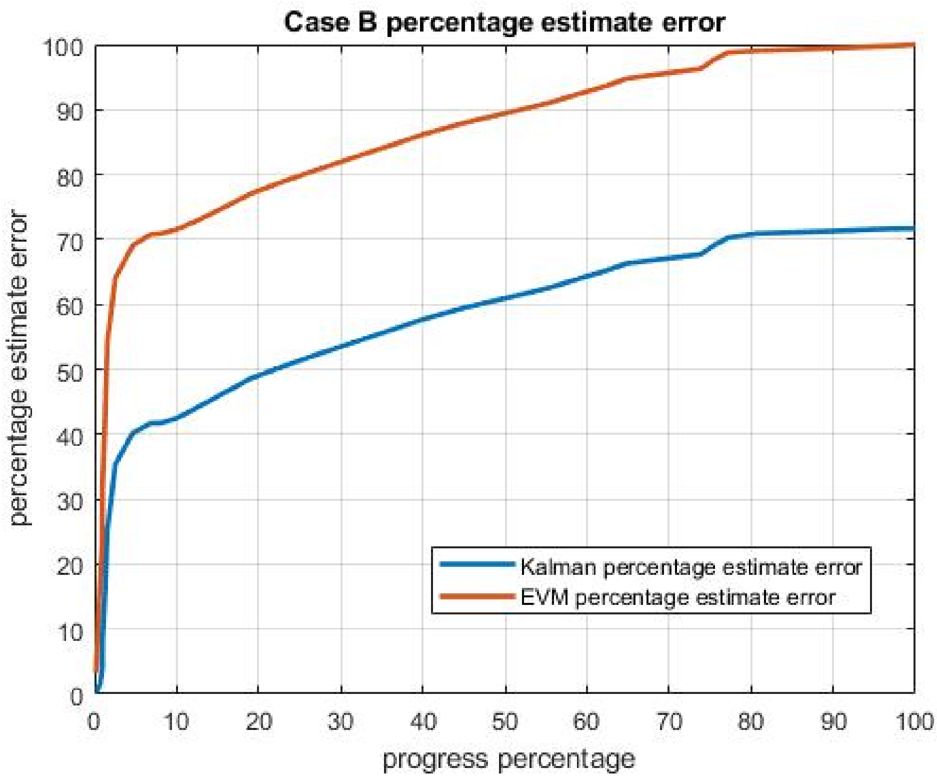

Figure 4.26: Case B percentage estimate error

On one hand, the project duration is reported in percentage value on the x-axis, on the other hand the y-axis indicates the percentage estimate error evaluated as the cumulated area between the estimate at completion and the final cost value. In order to compare the two methods, the areas are normalized using the area of the EVM estimate at completion as reference since it represents the state of the art. The graph introduces a second important aspect for a forecasting method: timeliness.

Timeliness is here identified as the ability of the method to provide reliable outcomes over the short term. This is a very significant issue for the project manager who needs to take decisions from the project early stage since correct results are needed as soon as possible. Further in his work, Teicholz figured out a procedure to evaluate the timeliness and which is still nowadays the substantial challenge of forecasting method: the author defined it as the accuracy achieved in the first half of the project. It is possible to measure the percentage error of the two estimates at completion in the first half of the project, the B case results are presented in Table 4.2.

K-EAC	EVM
61.08	89.57

Table 4.2: Case B, percentage estimate error on the first project half

The value is a measure of the total error in the first half of the project execution. The two indices are comparable because both was normalized using as reference the error made by the EVM estimate at completion, since it represents the state of the art. This can be explained, as previously mentioned, by the twofold nature of the Kalman filter that, using measurement and the analytical system model, could filter the cost variances fluctuations which are not due to structural causes. Although the magnitude of the peak is lowered, the presence is detected.

A third aspect to be considered is the probabilistic estimate introduced by the K-EAC model. In this case a comparison is not possible to perform since the EVM system performs exclusively punctual estimate. The presence of a probability distribution introduces several advantages, whose the most valuable is that it gives a measure of how certain the estimate can be. This means valuable knowledge which allows the decision maker to introduce actions that are supposed to deflect the course of the project development.

4.4.3 Case C model application

As described in the K-EAC algorithm presentation, its implementation to the C case starts with an initialization phase. The two first elements to assess are the model variables: the state vector and the covariance error matrix. Given that the control is performed from the beginning phase, both elements are set composed by null elements.

$$\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\hat{P}_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This is a self-evident choice because at the beginning of the project a cost variance has no chance to be different from zero since no cost is sustained and no work is performed yet. Nonetheless, there is no doubt or uncertainty in this situation, thus confirming the null matrix choice. The second initialization phase aims to identify the q value, index of the system model uncertainty.

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

q value is assessed to make the model uncertainty congruent with the users prior estimate of the project final cost distribution. The user provides in input the expected project duration (PAC) and the distribution of the expected final cost, expressed with μ_c and σ_c^2, the mean and the distribution variance. These elements are used in an inverse Kalman forecasting algorithm to determine the q value. More specifically, the algorithm, solely based on the system model, works equivalently to set the gain K equal to zero for all the project duration. Consistently, as it occurs in the baseline plan, the PAC lasts 32 months and the final cost distribution amounts approximately to a Normal with the mean equal to the budget and a variance equal to 10% of the budget. The last initialization step is the measurement error matrix R_k.

$$R_k = [r]$$

r is the measurement error variables and takes into account the variance of the measurement error, ^2_v. In order to set the value, the program evaluation review technique (PERT) [16] [17] [18] and a three-point estimate for the measurement error are employed. The user has to define the maximum possible measuring error, thus the error variance is evaluated with the PERT technique.

$$Maximumerrorv_k = Emax$$

$$Minimumerrorv_k = -Emax$$

$$\sigma_v^2 = [\frac{Emax - (-Emax)}{6}]^2 = \frac{Emax^2}{9}$$

In this case the value is chosen by setting the max error as 1% of the budget, selected since the $BCWP$ is evaluated as the percentage progress multiplied by BAC. Nevertheless, the progress is challenging to be correctly measured: multiple activities, where multiple disciplines are involved and enormous size makes very tough any attempt of synthesis of the whole performed work in a percentage number. After the initialization the algorithm is ready to run, fed at every stage on the measured cost variance value. The results of the entire project duration are collected in Figure 4.27.

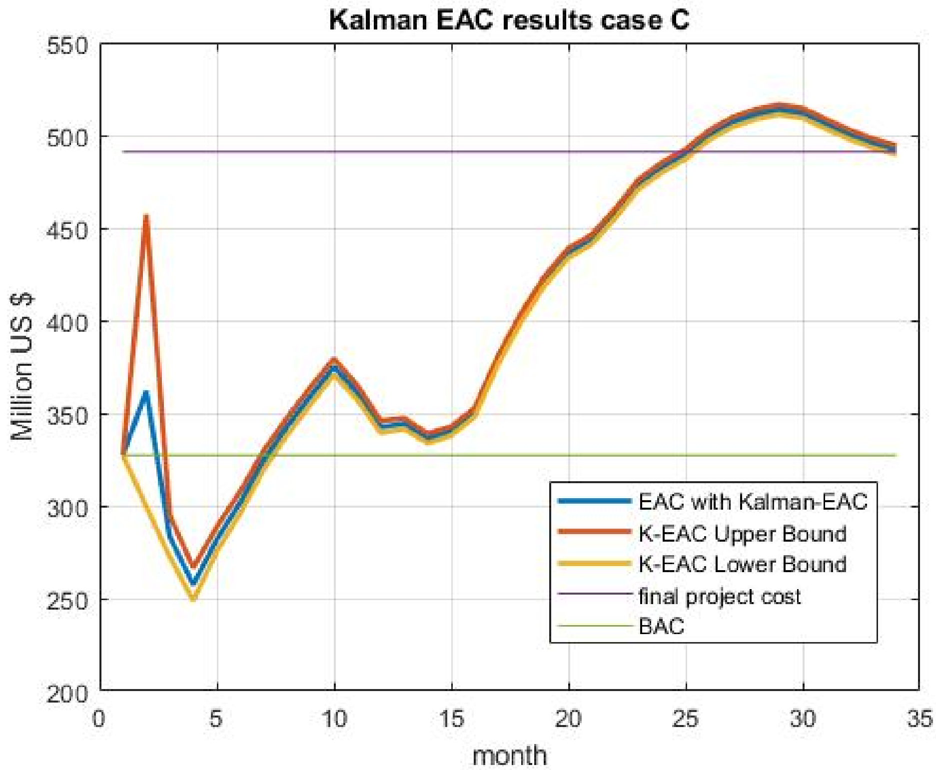

Figure 4.27: Kalman EAC results case C, the chart shows the forecasted EAC at every iteration

The chart gathers the forecasted EAC evaluated at each iteration. Since the estimate is obtained projecting the cost variance distribution, the output is presented as three lines: the central blue one represents the mean EAC value obtained projecting the CV distribution mean, the yellow and the red lines represent the optimistic and pessimistic values obtained through the projection of the fifth and the ninety-fifth cost variance distribution percentile. The two straight lines are the project budget in green and the final cost in purple. In the first time instants, the algorithm presents a rump-up period, the few information at disposal and the huge amount of work remaining make the distribution estimate too wide. As the project steps forward, more data are available, and the distribution becomes narrower thanks to more reliable estimations. The obtained EAC path is unstable, even if increasing and decreasing phases alternate during the execution, the estimate at completion is dominated by a growing trend.

The same graph is presented in Figure 4.28 the purple line is added to represent the estimate results of the EVM technique in the three-month moving average version, that represents the commonest technique nowadays.

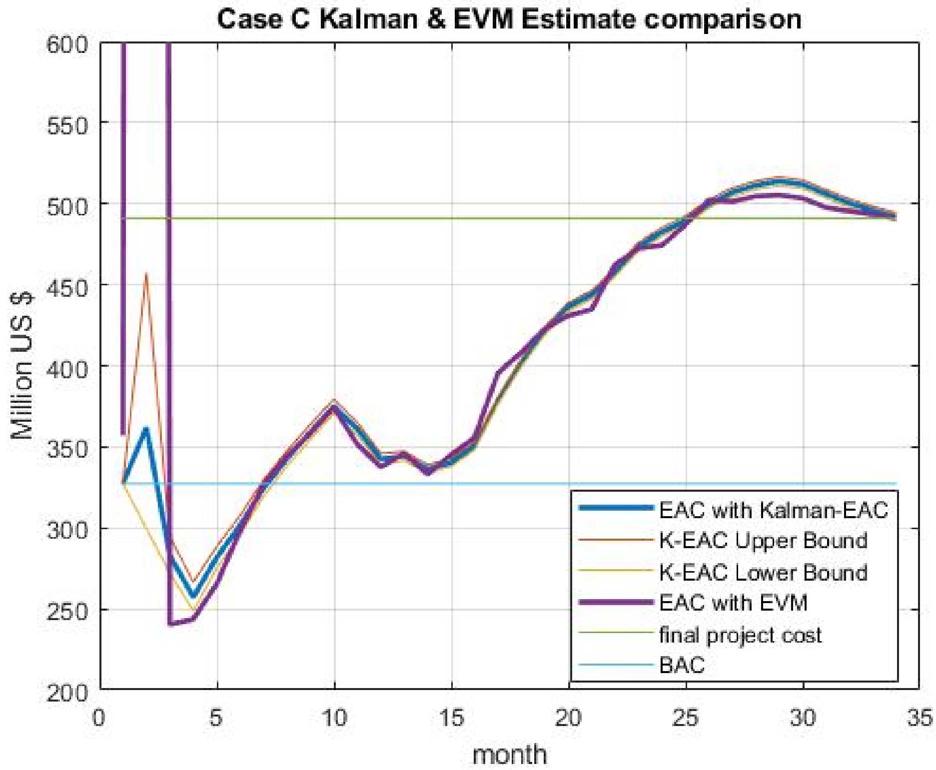

Figure 4.28: Case C Kalman and EVM Estimate comparison

The first thing that stands out is the presence of a high peak during the second month of the project: apart from that the results follow the same trend, showing congruence between the two methods. The algorithm has to face a short cost variance fluctuation during the second month, completely recovered over the following period that is characterized by an extremely positive cost variance. In this case the slightly negative CV makes the EVM EAC assume incredibly high values. A different situation is achieved with the K-EAC model that, vice versa lowers the effect reducing the obtained peak. This is caused by the high sensibility of the EVM technique in the early phase, when the work to be performed is still a great quantity. The estimate at completion is evaluated adding to the cost sustained up to this moment the planned cost of the remaining work, modified by a performance factor measured in the last periods. It is easy to understand that, when the project is in its early phase, the work to be performed is an argument of some weight in the estimate because even a very small performance fluctuation could give a huge impact on the final result. It is important to bear this in mind since not all the cost variance fluctuations are due to structural causes that will occur in the future, or even worse, could be caused by measurement errors. This issue is mitigated in the Kalman-EAC model where the obtained cost variances are the results of the combination of two factors: the measurements and the system model. The cost variance obtained with the K-EAC is cleansed by noisy fluctuations: when the available measurements are deficient, the algorithm does not tend to immediately trust in measurements, especially if their values are far from the one of the system model.

The trend of the cost variance along the project execution is presented in Figure 4.29, where the measured one is shown in blue, while the three others represent the estimated distribution, in red the central value, in purple and yellow respectively the fifth and the ninety-fifth percentile. Apart from a short peak in the third month, the execution is characterized by always negative values.

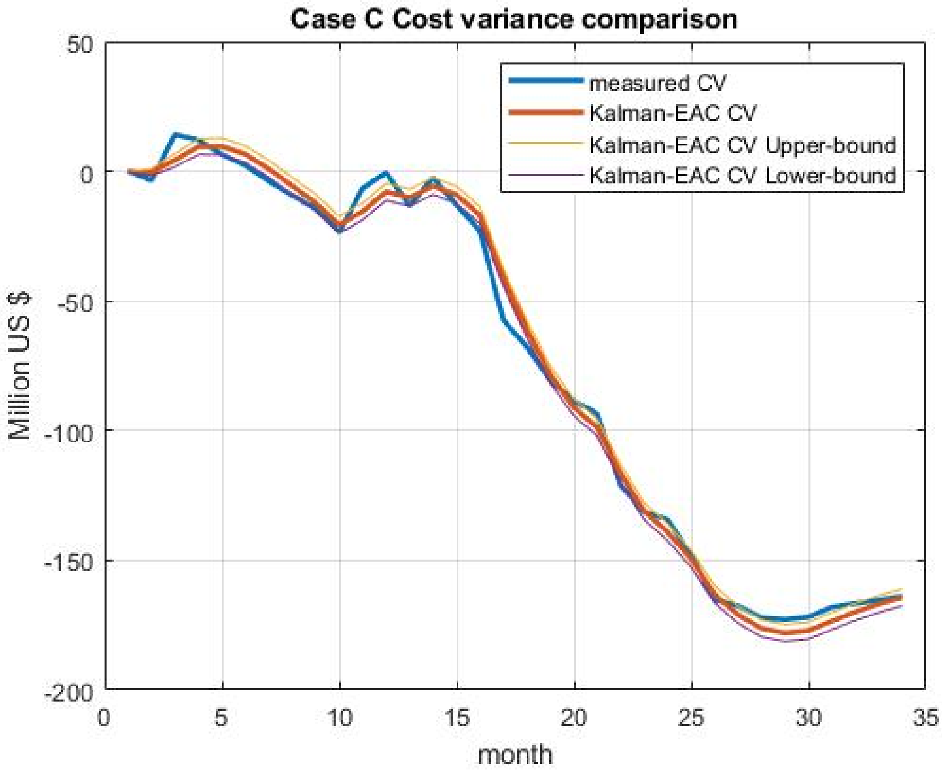

Figure 4.29: Case C Cost variance comparison

The above described trend is here evident: the cost variance estimated with the Kalman filter is more smoothed than the measured one, that is applied in the EVM evaluation. The trend is the same, but the short oscillations disappear. The phenomenon is evident in the second month where a little drop in cost performance leads to an increase of the EVM estimate to 7400 million US dollars; also K-EAC detects the loss of performance despite the fact that it is more conservative. The Kalman EAC is more sceptical about the judgement: that is the reason why it lowers the effect of the drop since it moves away from the project trend.

Now the comparison needs to consider the two most important aspects of a forecasting technique: accuracy and timeliness. Accuracy is measured as the average deviation between the forecast value and the actual one over a definite time period. In our case the MAPE is evaluated as follow:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{ActualFinalCost - EAC(t)}{ActualFinalCost}\right|$$

The MAPE value is evaluated with both the K-EAC result and the EVM three-month moving average EAC. The outcomes are shown below in Figure 4.30.

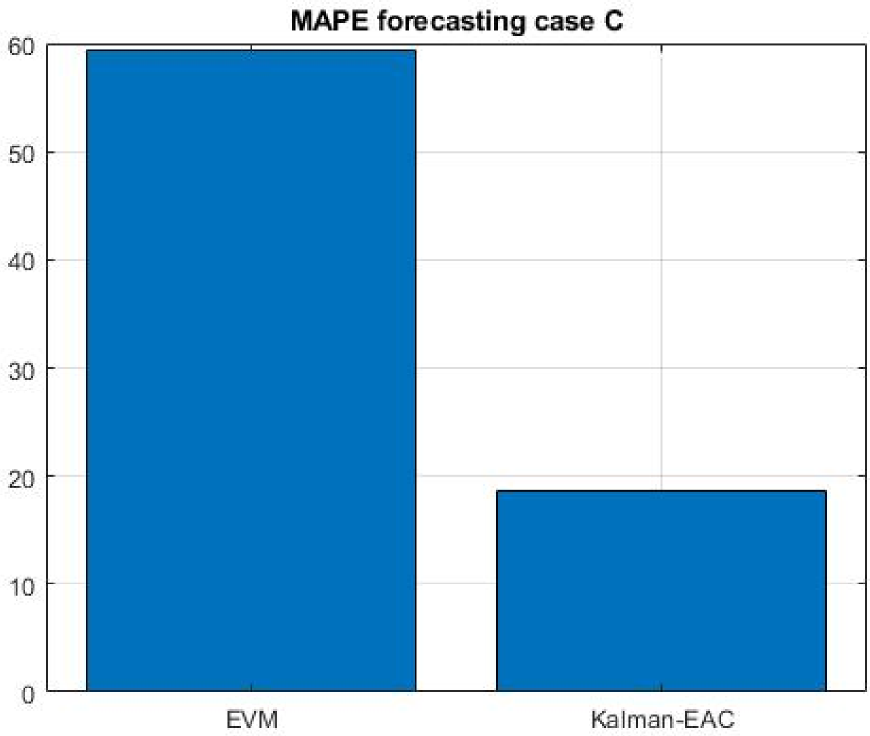

Figure 4.30: MAPE forecasting case C

The proposed method reduces MAPE from 59.305 to 18.689 with an improvement of more than 68%. This result is due to the high difference in the first part of the project, where the EVM estimate is too responsive to performance fluctuations and, even worse, every performance variation is considered as an outcome of a structural cause and affects all the remaining life of the project. The K-EAC model does not trust in performances deviations without evidence, especially if far from the system model output, thus reducing the risk of misleading interpretation.

A second way for the accuracy measurement was brought out by Teicholz [3], who draw attention to a new indicator after comparing 121 construction projects. Together with the classical statistical methods, as mean square error, the accuracy could be represented by the absolute area between the actual final cost and the path of the estimate at completion plotted against the percentage progress: Figure 4.31 shows the application of this method to the C case. This technique introduces some advantages: firstly, it does not suffer from biasedness if the measurement intervals are not constant; secondly, it offers a visual information about the achievement of the best accuracy results.

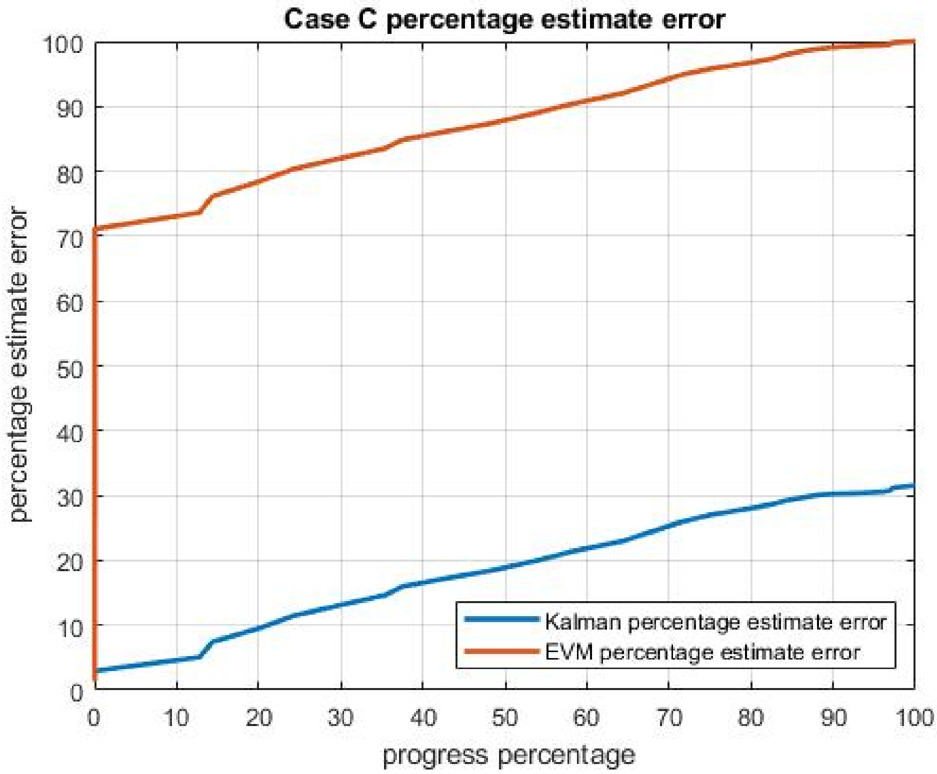

Figure 4.31: Case C percentage estimate error

On the x-axis the project progresses are displayed in percentage values, whereas on the y-axis the percentage estimate error is evaluated as the cumulated area between the estimate at completion and the final cost value. In order to compare the two methods, the areas are normalized using the area of the EVM estimate at completion as reference given that it is the reference method. As above mentioned, there is quite a considerable difference on the first part of the project duration, while after the two lines head for a similar trend due to the fact that the two estimates have similar tendencies. Besides, the graph introduces a second important aspect for a forecasting method: timeliness.

Timeliness is here identified as the ability of the method to provide reliable outcomes over the short term. This is a very significant issue for the project manager who needs to take decisions from the project early stage since correct results are needed as soon as possible. Further in his work, Teicholz figured out a procedure to evaluate the timeliness and which is still nowadays the substantial challenge of forecasting method: the author defined it as the accuracy achieved in the first half of the project. From this viewpoint, it is possible to measure the percentage error of the two estimates at completion in the first half of the project. The C case results are presented in Table 4.3.

K-EAC	EVM
18.44	87.44

Table 4.3: Case C, percentage estimate error on the first project half

The value stands for the total error performed in the first half of the project duration. Both indices are comparable because they were normalized by using as reference the error made by the EVM estimate at completion, that represents the state of the art. It emerges that the Kalman estimate at completion produces more accurate results at the beginning of the project than the traditional technique does. This can be explained, as cited previously, by the twofold nature of the Kalman filter that, using measurement and the analytical system model, could filter the cost variances fluctuations not due to structural causes. A third aspect which requires consideration is the probabilistic estimate introduced by the K-EAC model. In this case, to perform a comparison is not possible since the EVM system performs only punctual estimates. The presence of a probability distribution introduces several advantages: the most valuable is that it gives a measure of how certain the estimate can be. This means valuable knowledge which allows the decision maker to introduce actions that are supposed to deflect the course of the project development.

Chapter 5

Conclusions

5.1 Comments over results

The application of the K-EAC method to the three cases presented above gives as output two pivotal pieces of information: the cost variance and the estimate at completion.

The cost variance detected by the algorithm aims at the reduction of the misleading effect of errors, present both in measurement process and in the input data. It is provided at every iteration with its probability distribution and assumes a central role in the forecasting contest since, in addition to describing the progress of the project, it is used during the estimate at completion process. Comparing the trend of the estimated cost variance throughout the project execution to the measured one, in all three cases of study two aspects are easily highlighted: firstly, they have a very similar trend, which means an index of coherence on the results; secondly, the estimated CV smooths the short peaks detected by the measurement system but, the unitary period duration considered, they are seen as fluctuations which do not describe the state of the project but which are due to measurement errors. The algorithm is so cautious that it does not immediately trust measured value of cost variance when it is far from the actual trend and not supported by previous observations. Evidently, the algorithm does not completely delete the peak presence; however, it lowers its intensity, placing more trust in the system model results than in the measurement. At the same time, if the trend is still present at the following period, the algorithm gives more confidence to the measurement system by totally regaining the actual CV.

The second output of the algorithm is the estimate at completion, provided with three values: the most probable, optimistic and pessimistic values. In every application, the algorithm presents a warm-up period when the measure of the EAC uncertainty initially grows and later starts to get smaller. The rump-up period in the initial months is due to the little information the algorithm can rely on. The second descending trend is instead in-line with the expectation: with the ongoing project the algorithm has at disposal more information to elaborate the estimate.

The results are then compared to the ones obtained by the technique representing the state of the art in the forecasting field: the three-month moving average version of the EVM estimate at completion. On the whole, the results are consistent since the two estimate trends are similar. The two methods are then evaluated on the aspects which are the most important features for a method, according to the forecasting literature: accuracy and timeliness.

As a measure of accuracy a standard error evaluation technique is proposed: the MAPE (Mean Absolute Percentage Error) between the estimate obtained at each step and the actual final cost, whose results are proposed in Table 5.1.

CASE	K-EAC	EVM
A CASE	12.836	19.816
B CASE	39.895	55.597
C CASE	18.689	59.305

Table 5.1: MAPE value of the estimate in the three cases

As it can be noticed, the outcomes are encouraging and the proposed method introduces nosedives in all the evaluated cases. Since MAPE is a mean indicator, in order to evaluate where the improvement is achieved, Teicholz's alternative accuracy measure is also proposed [3], based on the cumulative areas between the estimates and the real final cost lines over the project duration. The graphs show that, for the three projects, the crucial difference is obtained in the first third of the project. The estimate at completion is evaluated adding to the cost sustained until this time the planned cost of the remaining work, modified by a performance factor which has been measured over the last periods: one soon realizes that, when the project is in its early phase, the work still to be performed has some weight in the estimate, thus even a very small performance fluctuation could have a critical impact on the final result. The divergence between the obtained EAC is due to the different performance factors used by the two methods: on one hand, the K-EAC bases its EAC over the estimated cost variance which, as explained above, lowers the presence of performance random fluctuation; on the other hand, EVM technique directly uses the measurement by maintaining the short peaks, thus considering the fluctuation due to the structural causes which will be affecting all the remaining project life.

A second problem with the EVM estimate is that the fluctuation, though it could be recovered in the following time instant, affects the estimate in the two following periods since the EAC is evaluated through a three-period average performance factor. On the contrary, the K-EAC, once identified the wrong value, bases its estimate mostly on the system model and avoids the misleading EAC.
The second aspect analysed is the timeliness, namely the ability of the method to provide meaningful results in the short term. The evaluation abides by the method indicated by Teicholz, that takes as an indicator the percentage of the error made in the first half of the project progress, the value to be compared are normalized over the EVM estimate error, the indices are reported in Table 5.2.

CASE	K-EAC	EVM
A CASE	47.74	82.02
B CASE	61.08	89.57
C CASE	18.44	87.44

Table 5.2: Percentage estimate error on the first project half

With respect to the standard method, the K-EAC shows up as the faster in providing telling results. As mentioned above, the EVM estimate at completion is too responsive during the early phase of the project and this leads to wrong final cost estimation characterized by dramatic peaks. The proposed method lowers the issue combining a system model with the measurement process.

5.2 Conclusions and future studies

In the present work a new model for the estimation of the final cost of a project is developed, aiming to contribute to the state-of-the-art technique consisting in the EVM system methodology. The objective is to reduce some issues that a project manager needs to face nowadays during the project execution, such as criticalities which lie on the hypothesis the three-month moving average EVM estimate at completion is based on.

First of all, the EVM technique deterministically describes the project status, without considering the possibility to run into any kind of error, providing punctual indicators. Errors are easy to commit given that the cost performance indices are based on two quantities: the actual sustained cost and the budget cost of work performed. The latter is easily mistakable because, in order to get assessed, the percentage of physical progresses achieved and the planned cost for the activities are needed. It is challenging to synthetize in a single progress percentage value a whole project containing a large number of activities and several different disciplines, and, at the same time, it is hard to forecast the cost of activities and subcontractors before the beginning of the project, even more if the activities are performed in a new site where the company is not used to working. The issue is extremely convoluted since the information has a substantial value for the project manager to base expensive actions on.

The second issue concerns the responsiveness of the algorithm during the starting phase of the project. The phenomenon derives from the way the EAC is evaluated: the algorithm provides the final estimate adding to the present sustained cost the one which has still to be sustained but modified by a cost performance factor. The responsiveness is due to the large amount of remaining work. In this condition the effect of a short cost performance fluctuation is sizable and the problem is even more severe since the fluctuation could be caused by a measurement error as explained above. In this case, although the performance index is correctly assessed in the following observation by applying the three moving average system, the error keeps being taken into account for two more periods.

Throughout the technique development the attention has been focused on the method performances, namely, accuracy, timelines and friendliness which are considered as key parameters, according to the forecasting literature.

The proposed algorithm aims to reduce the previously mentioned shortcomings. Based on the classical Kalman filter framework and the EVM techniques, the algorithm uses both a system model, built over the planned project baseline, and the measured data to evaluate the probability distribution of the real but hidden cost variance, describing the project status and giving also a measure of how trustworthy these values are. The obtained CV distribution is freed from short performance peaks since the K-EAC algorithm does no trust fluctuation far from the trend when it is not supported by evidence. Later, this is used to evaluate the EAC, provided with a central value, in addition to an optimistic and a pessimistic one. Reducing the cost performance index peaks, the method initial responsiveness gets lowered and the problem will avoid happening again over the following periods.

In order to evaluate the method performances, it is applied to three real cases, showing positive results and coherence with the competitor method. A further advantage is achieved: though the measure of the actual performance could not be performed, to perform the estimate is still possible with the system model.

As a consequence, the positive results obtained are unfortunately balanced by an increasing algorithm complexity. The computation, composed of matrix multiplications, is straightforward and it can be implemented with an undemanding software as Excel or MATLAB which do not need any sophisticated calculation powers. Even the result interpretation is intuitive because the numerical value of the cost variances and EAC are supported by graphical tools. The raising complexity lies in the introduction of the initialization phase which requires additional input data respect to the EVM technique, specifically the prior cost distribution and the maximum errors of the measurement system, introduced in the model presentation chapter. On the other hand, the initialization phase provides a great adaptability to the method: here the parameters are assessed by helping the technique to fit each time the analysed project avoiding standardization. The parameters change time to time describing the project and the environment it is developed in. This opens a window to extend the method application not only to construction project (where the EVM is developed), but also to new project type.

Based on the obtained result, some suggestions about the possible further study have been made.

Further model verification with a bigger model cluster. Only three projects in the Oil & Gas field cannot represent an adequate and satisfactory collection to guarantee generality to the method so it is necessary to test the algorithm and its accuracy with a wider project number.

In addition, the algorithm results and the estimate reliability have been validated for specific projects, especially for long-term ones and hardly ever exposed to accidents with high impact on project performances like the ones related to the US defence department (the area where these techniques were first developed) [23]. There is the need to spread out the methodology to a wider project range marked by a different duration, operative processes and uncertainty level. Since the initialization phase gives the algorithm high adaptability, it could be tested in new project types.

Integration of the algorithm with other informative sources. The K-EAC has a significant potential for the integration with other informative sources: it provides an extremely flexible framework for combining new state variables and, furthermore, the initialization phase, where prior distributions are needed, could be enriched by expert judgement or similar past project analyses. A powerful method would give the decision maker the chance to add some other information sources to make the outcomes more reliable.

Bibliography

[1] Kenneth H Rose. A guide to the project management body of knowledge (pmbok® guide)fifth edition. *Project management journal*, 44(3):e1–e1, 2013.

[2] John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

[3] Paul Teicholz. Forecasting final cost and budget of construction projects. *Journal of computing in civil engineering*, 7(4):511–529, 1993.

[4] Frank T Anbari. Earned value project management method and extensions. *Project management journal*, 34(4):12–23, 2003.

[5] DS Christensen and DS Christensen. The estimate at complete problem: a review of three studies. *International Journal of Project Management*, 24(1):37–42, 1993.

[6] Quentin W Fleming and Joel M Koppelman. Earned value project management. Project Management Institute, 2016.

[7] David S Christensen and Scott R Heise. Cost performance index stability. *National Contract Management Journal*, 25(1):7–15, 1993.

[8] Howard Musoff and Paul Zarchan. *Fundamentals of Kalman filtering: a practical approach.* American Institute of Aeronautics and Astronautics, 2009.

[9] Eli Brookner. *Tracking and Kalman filtering made easy.* Wiley New York, 1998.

[10] Gary Bishop, Greg Welch, et al. An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8(27599-3175):59, 2001.

[11] Craig F Ansley and Robert Kohn. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics*, pages 1286–1316, 1985.

[12] José Casals, Miguel Jerez, and Sonia Sotoca. Exact smoothing for stationary and non-stationary time series. *International Journal of Forecasting*, 16(1):59–69, 2000.

[13] Piet De Jong. Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association*, 84(408):1085–1088, 1989.

[14] Evgenia Suzdaleva and AV Utia. Initial conditions for kalman filtering: prior knowledge specification. In *Proceedings of the 7th WSEAS International Conference on Systems Theory and Scientific Computation, Athens, Greece*, 2007.

[15] Byung Cheol Kim. *Forecasting project progress and early warning of project overruns with probabilistic methods.* Texas A&M University, 2007.

[16] Franco Caron. *Gestione dei grandi progetti d'ingegneria*. ISEDI De Agostini Scuola, 2009.

[17] C Perry and ID Greig. Estimating the mean and variance of subjective distributions in pert and decision analysis. *Management Science*, 21(12):1477–1480, 1975.

[18] Donald L Keefer and Samuel E Bodily. Three-point approximations for continuous random variables. *Management Science*, 29(5):595–609, 1983.

[19] George A Hazelrigg. Validation of engineering design alternative selection methods. *Engineering Optimization*, 35(2):103–120, 2003.

[20] Robert Carbone and J Scott Armstrong. Note. evaluation of extrapolative forecasting methods: results of a survey of academicians and practitioners. *Journal of Forecasting*, 1(2):215–217, 1982.

[21] Mario Vanhoucke and Stephan Vandevoorde. A simulation and evaluation of earned value metrics to forecast the project duration. *Journal of the Operational Research Society*, 58(10):1361–1374, 2007.

[22] Ofer Zwikael, Shlomo Globerson, and Tzvi Raz. Evaluation of models for forecasting the final cost of a project. *Project Management Journal*, 31(1):53–57, 2000.

[23] Walt Lipke. A re-examination of project outcome prediction-using earned value management methods. *The measurable news*, 2005.

[24] HS Ang Alfredo and WH Tang. Probability concepts in engineering and design, 1975.

[25] James E Spooner. Probabilistic estimating. *Journal of the Construction Division*, 100(c01), 1974.

[26] Terry Cooke-Davies. The real success factors on projects. *International journal of project management*, 20(3):185–190, 2002.

[27] Paul C Dinsmore and Jeannette Cabanis-Brewin. *The AMA handbook of project management.* Amacom Books, 2006.

[28] Harold Kerzner and Harold R Kerzner. *Project management: a systems approach to planning, scheduling, and controlling.* John Wiley & Sons, 2017.

[29] Jesus Leodaly Salazar-Aramayo, Roseane Rodrigues-da Silveira, Mariana Rodrigues-de Almeida, and Tereza Neuma de Castro-Dantas. A conceptual model for project management of exploration and production in the oil and gas industry: The case of a brazilian company. *International journal of project management*, 31(4):589–601, 2013.

[30] NDIA PMSC. Earned value management systems intent guide. 2006.

[31] Beatrice Pierini. Stima a finire di progetto mediante integrazione bayesiana di opinioni di esperti, dati storici relativi a progetti similari e dati di progetto al time now. 2013.

[32] Cristian Borgarucci. Stima a finire mediante integrazione bayesiana tra visione interna ed esterna di progetto. 2011.