International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:12, No:5, 2018

A Construction Management Tool: Determining a Project Schedule Typical Behaviors Using Cluster Analysis

Natalia Rudeli, Elisabeth Viles, Adrian Santilli

Abstract**—**Delays in the construction industry are a global phenomenon. Many construction projects experience extensive delays exceeding the initially estimated completion time. The main purpose of this study is to identify construction projects typical behaviors in order to develop a prognosis and management tool. Being able to know a construction projects schedule tendency will enable evidence-based decision-making to allow resolutions to be made before delays occur. This study presents an innovative approach that uses Cluster Analysis Method to support predictions during Earned Value Analyses. A clustering analysis was used to predict future scheduling, Earned Value Management (EVM), and Earned Schedule (ES) principal Indexes behaviors in construction projects. The analysis was made using a database with 90 different construction projects. It was validated with additional data extracted from literature and with another 15 contrasting projects. For all projects, planned and executed schedules were collected and the EVM and ES principal indexes were calculated. A complete linkage classification method was used. In this way, the cluster analysis made considers that the distance (or similarity) between two clusters must be measured by its most disparate elements, i.e. that the distance is given by the maximum span among its components. Finally, through the use of EVM and ES Indexes and Tukey and Fisher Pairwise Comparisons, the statistical dissimilarity was verified and four clusters were obtained. It can be said that construction projects show an average delay of 35% of its planned completion time. Furthermore, four typical behaviors were found and for each of the obtained clusters, the interim milestones and the necessary rhythms of construction were identified. In general, detected typical behaviors are: (1) Projects that perform a 5% of work advance in the first two tenths and maintain a constant rhythm until completion (greater than 10% for each remaining tenth), being able to finish on the initially estimated time. (2) Projects that start with an adequate construction rate but suffer minor delays culminating with a total delay of almost 27% of the planned time. (3) Projects which start with a performance below the planned rate and end up with an average delay of 64%, and (4) projects that begin with a poor performance, suffer great delays and end up with an average delay of a 120% of the planned completion time. The obtained clusters compose a tool to identify the behavior of new construction projects by comparing their current work performance to the validated database, thus allowing the correction of initial estimations towards more accurate completion schedules.

Keywords**—**Cluster analysis, construction management, earned value, schedule.

Natalia Rudeli is with the Agencia Nacional de Vivienda, Uruguay (e-mail: nataliarudeli@gmail.com).

I. INTRODUCTION

CONSTRUCTION is one of the most widespread and ancient activities of human kind and is positioned as one of the largest industries in the world. Unfortunately, construction projects schedule overruns are a common phenomenon that causes tremendous damage to the global economy [1].

A schedule overrun is defined as additional time required to finalize a construction project beyond its original planned duration [2]. Other authors define schedule overrun as

> "an act or event that extends the time to complete or perform the contract" or as "the time overrun either beyond completion date specified in a contract or beyond the date that the parties agreed upon for delivery of a project" [3], [4].

It is basically a project slipping over its planned schedule and is considered as a common problem in the construction project sector [5].

The construction process can be divided into three main phases: the project conception, project design and project construction. The vast majority of the delays tend to occur during the construction phase, where most unforeseen factors are involved [6]. During this stage, a proper monitoring allows to determine the current situation of a project and to make an educated prediction of their future status, as long as no further variables are introduced [7]. According to [8] and [9], one of the most popular and accepted tools for controlling and monitoring construction projects are schedules. A schedule is a listing of a project's milestones, activities and deliverables, usually with intended start and finish dates [10].

This article present a statistical study to determinate schedule typical behaviors based on a cluster analysis and the use of EVM indexes, validated with a large number of case studies which are also presented. The obtained results will serve as a tool for managers to predict a projects future performance by way of comparing their current situation to the typical project behaviors identified and presented below.

II. LITERATURE REVIEW

Earned Value Analysis (EVA) is an accepted theoretical technique advocated to the control of projects. The EVM represents an organized approach to integration and measurement of costs and time distribution or business achievements, defined by project objectives [11]. Unlike other project management techniques which are mainly focused on control, EVM transcends this form and provides information that can more easily predict the future directions of a project [12].

According to [13], EVM is a technique for controlling the performance of a project by comparing the amount of work up to a certain moment to the estimations made before the project start. In this way, there is a measure of the amount of work performed and the amount of work remaining to achieve project completion.

The EVM technique has its roots in the Project Management Institute. It measures a project's progresses in an objectively way and provides an early warning of poor performance issues, if any [10]. It is widely accepted and well documented that implementing EVM would bring added value to a project monitoring scheme, especially in construction projects. Several authors highlight the use of this methodology as a predictive analysis tool [14]-[22], while other authors have used this method as a project management planning and control technique [14]-[22].

In practice, project progress is evaluated by comparing EV indexes and estimates against past values, against similar projects or against other several criteria proposed by literature [23]. Taking this into account, this article will determine the typical behavior of construction projects through a cluster analysis, to be used as a standard for comparison.

The concept of ES is an extension of the concept of EVM, and it refers to the amount of additional time needed to reach the established progress goals (in the case where construction is behind schedule). If construction is ahead of the expected schedule, it can be defined as the range of time in which construction can make no progress without producing a delay in the schedule [21].

EVM and ES require the setting of a series of concepts to determine the main indexes. Planned Value (PV) is defined as the total planned time until the end of the project (setting in the planning stage), i.e., PV refers to the contract schedule predicted/required before the construction stage. On the other hand, the Earned Value (EV) refers to the executed time during the construction stage, i.e., the performed schedule of an ongoing project. EVM uses these concepts to define indicators for schedule efficiency and time delay determination: schedule variation (SV) and schedule performance index (SPI) [24].

SV index determines if a project is delayed or ahead compared to its original schedule. A SV value below zero implies that a project is progressing as planned, while a positive value implies the contrary (1).

SPI is a measure of schedule performance. If SPI is less than 1, the index shows that the project is ahead of schedule, whereas if SPI values greater than 1 are obtained, it means that the project is delayed (2).

$$SV = EV - PV \tag{1}$$

$$SPI = \frac{EV}{PV} \tag{2}$$

Finally, the concept of ES is an extension of the EVM. This concept measures the additional time that a delayed project requires to reach completion or the time amount that a project that is ahead of schedule can be maintained without progress until it generates a schedule delay (3).

$$ES = \frac{EV-PV}{PV} \tag{3}$$

ES is generally used for construction schedules efficiency determination. As an example, [13] conducted a study with 16 American projects, while other authors have applied this methodology and proved their effectiveness [25]-[27].

Fig. 1 shows the performance indicators summary. Three curves of construction project works can be seen: planned schedule, early schedule and delayed schedule. Planned schedule refers to the contract schedule predicted before the construction stage. The early schedule curve shows the overall performance of a work with a construction rate higher than expected, i.e. is ahead of the planned schedule. Finally, the delayed schedule curve illustrates the typical behavior of a work that has not met the planning stage requirements, i.e. is behind planned schedule.

For the planned schedule curve, the planned value (PV) and the tenths of the planned value (used to standardize projects with different construction times) are indicated in Fig. 1. As regards to the early (E) and delayed (D) schedule curves, the representations of EV indicators are shown, (EV_E and EV_D, respectively). Also, the SV values that are obtained as the difference between the EV_E or EV_D and the PV, are shown (SV_E and SV_D).

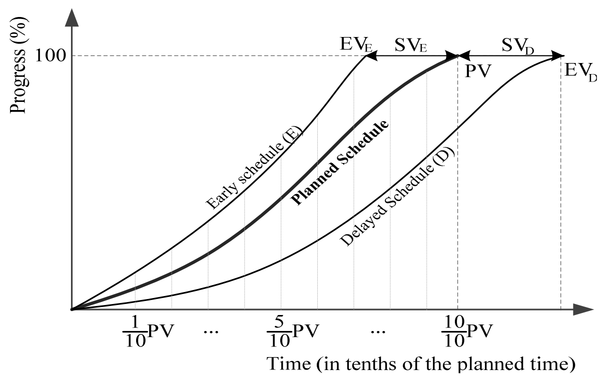

Fig. 1 EVM and ES indexes for construction project behaviour characterization

Authors like [7] and [28] have used EVM indicators to evaluate construction projects performance; [29] also validated the use of EVM indicators with several case studies from the civil industry. Given the importance of some EVM indicators for the control and management of construction, EVM and ES indicators will be used to characterize the construction projects database to determine the typical behavior.

III. Research Method

A. Projects Database Characterization

For prediction model development some formal aspects must be taken into account. Since all the projects have different construction times, to standardize and compare data, the values of estimated completion time need to be discretized. The expected construction time (planned duration) will be considered over 10 equal parts, ten tenths (Fig. 1). By doing this, two projects with different construction times can be compared [30].

The database includes 90 Uruguayan housing construction projects, and for each one, the planned schedule and the ES are taken into consideration. This allows to determine EVM indexes through the use of (1)-(3). Also, for the cluster analysis verification phase, an analogous 15 Spanish housing projects database was used. Finally, a large number of authors mention planned and executed total times for different civil construction projects [34]-[37]. From these authors, only [35]-[37] will be taken into account for this study, as [31]-[34] do not publish data of the projects intermediate timelines.

B. Cluster Analysis

The cluster analysis is a multivariate technique used to classify a set of individuals or items into homogeneous groups. This type of analysis is often used for determining taxonomies or behaviors of similar groups [38]. Establishing construction works taxonomy is highly relevant, because it allows managers to make decisions based on accurate data.

For a cluster analysis, the selection of the grouping variables plays a fundamental role to obtain results. Conceptual and practical considerations should be taken into account and only variables that are specifically related to the object of analysis should be included in the study.

The complete process can be structured as follows: (1) there is a set of N individuals whose information is encrypted by a set of n variables (in this case, N matches the 90 construction projects mentioned, with 11 variables to be measured). (2) A similarity criterion is established. In this case, the EVM indicators described above will be taken into account. (3) A classification method is selected to determine the structure of the grouping. In this case, the maximum distance method (furthest neighbor or complete linkage) will be used, measuring the Euclidean distance. (4) Finally, the structure obtained by tree diagrams or dendrograms is specified.

In order to perform the classification, it must be determined how similar the elements are depending on how different their representations turn out to be in the space of the selected variables. Most of the similarity indexes are based on Euclidean distance. This is the result of measuring the spatial distance between two individuals *i* and *j*, indicated by $d\ (i,\ j)$. The value of $d\ (i,\ j)$ is always a positive value and the higher this value, the greater the difference.

If the study items are represented as vectors in the space of the variables, W_1 ... W_{i} ... W_{90}, then each vector will include the percentage work completion values in each tenth ($\frac{\mathrm{m}}{10}\mathrm{t}$, where m indicates the number of the tenth), and it will include the value of the ES index (delay measure), thus being able to fully characterize a construction work through the vector from (4). Then, $d\ (i,\ j)$ is determined by the vector difference between W_i - W_j, [39].

$$W_i = \left(\frac{1}{10}t, \quad \frac{2}{10}t, \quad \dots \quad \frac{9}{10}t, \quad \frac{10}{10}t, \quad ES\right), i = 1 \dots .90 \quad (4)$$

Regarding classification methods, there are different tendencies to be found on literature. The most commonly used are: hierarchical methods, optimization methods, density (or mode-seeking) methods and "clumping" (or partition) methods.

Hierarchical methods are highly worked and recognized [40], which is why they will be used in this research. In this technique, the items are not partitioned into clusters at one time, but successive partitions at different levels of aggregation are made. Additionally, the "complete linkage" classification method will be used. This method considers that the distance or similarity between two clusters must be measured according to their most disparate elements, i.e. the distance between two clusters is given by the maximum distance (or minimum similarity) between their components.

C. Validation

Once the optimal construction projects behavior grouping has been obtained, it will be validated through the use of the Spanish and the literature databases. The validation consists in the testing of these databases in the four obtained clusters.

A first analysis of the clusters centroids values shows a great variability in the data. That is why the linear regression that best fits will be determined. All tenths of the Uruguayan database will be analyzed using the best fits tool. The tenths that most influence the value of the ES indicator will be obtained and therefore will be taken into account for the adjustment study and validation phase.

The regression that best fits the predictive variables will be obtained using a best subsets model. The best subsets models have the highest values of R^2, thus guaranteeing high predictability. Regression of the best subsets is an efficient way to identify models that fit the data appropriately using as few predictors as possible. Models that contain a subset of the predictors can estimate the regression coefficients and predict future responses with less variance than the model that includes all the predictors. The best fits tool is used in multiple regression studies to find the predictor variables that assurance the best linear adjustment on the response variable; in this case it will be looked for which tenth or combination of tenths produce the best linear adjustment on the ES indicator.

Once the best-fitting predictors have been obtained (one or more of the W_i elements), their confidence intervals will be studied to verify the statistical dissimilarity between principal predictors of the clusters and to ensure that the validation method is reliable. In the case that the combination of predictors results in statistically similar confidence intervals, the next best combination that provides dissimilarity will be sought.

Spanish and literature databases will be compared to the clusters, taking into account key predictors, in order to classify validation databases projects in to clusters. A construction project will belong to a cluster when all its main predictors are in the 95% confidence interval of that cluster. Since the dissimilarity between the data of the main predictors has been proven, when a project belongs to a cluster, statistically it can be said that it does not belong to any other.

After both validation databases have been classified into the different clusters, the absolute value of the difference between the average delay of the cluster (ES centroid for each cluster) and the delay of each one of the projects belonging to it will be measured. Average of all the differences represents the accuracy to predict the deviations of the schedules.

IV. Results and Discussion

A. Projects Database Characterization

The 90 Uruguayan construction projects were processed. Tables I and II show the average progress for each tenth and the ES and EVM indexes averages. These data allows for the construction of the projects' progress characteristic curves. The 95% Upper Confidence Interval (UCI) and Lower Confidence Interval (LCI) maximum (Max) and minimum (Min) database values are shown. From the database analysis it can be deduced that construction projects show an average delay of 34.68% with a 95% confidence interval that includes delays ranging from 28.95% to 40.41%.

Tables III and IV show the 15 Spanish construction projects characterizations that will be taken into account for the cluster analysis validation. Tables V and VI show the literature review processed data which will be used in the cross-validation. Data published by [35]-[37] was transformed using the work progress measurement system proposed by [30].

Data proposed by Uruguayan, Spanish and literature databases do not show great dissimilarities during the course of the construction works. However, deviation becomes evident when contrasting completion times, where Uruguayan projects have a tendency of showing larger delays than their counterparts, the literature database is the one that presents the shortest average completion time.

TABLE I
URUGUAYAN CONSTRUCTION PROJECTS DATABASE CHARACTERIZATION

Tenths of PV[a]	Mean	Deviation	UCI	LCI	Max	Min
1/10t%	5.50	4.17	6.36	4.64	17.62	0.00
2/10t%	7.69	3.48	8.41	6.97	20.00	0.72
3/10t%	8.10	3.46	8.82	7.39	20.42	2.08
4/10t%	8.34	3.15	9.00	7.69	16.34	0.00
5/10t%	8.72	3.48	9.44	8.00	18.00	0.88
6/10t%	9.49	3.80	10.27	8.70	30.13	3.01
7/10t%	9.50	3.53	10.23	8.77	20.13	2.51
8/10t%	9.23	3.39	9.93	8.53	21.23	3.32
9/10t%	8.98	3.67	9.74	8.22	24.17	0.01
10/10t%	8.12	3.67	8.88	7.36	16.27	0.00

[a]Values of estimated completion time considered over 10 equal parts [30]

B. Cluster Analysis

The cluster analysis of the Uruguayan database was carried out. Fig. 1 shows that with a combination of 4 conglomerates an acceptable level of similarity is reached, obtaining then conglomerates with significantly different behaviors, see Tables VII and VIII.

TABLE II
URUGUAYAN CONSTRUCTION PROJECTS DATABASE EVM/ES INDEXES

EVM/ES Indexes	Mean	Deviation	UCI	LCI	Max	Min
ES%	34.68	27.74	40.41	29.95	120.00	-6.25
SV%	8.21	6.41	9.54	6.89	26.00	-2.00
SPI	1.35	0.28	1.40	1.29	2.20	0.94

TABLE III
SPANISH CONSTRUCTION PROJECTS DATABASE CHARACTERIZATION

Tenths of PV	Mean	Deviation	UCI	LCI	Max	Min
1/10t%	5.69	2.06	6.73	4.64	10.51	2.85
2/10t%	7.15	1.96	8.15	6.16	10.41	2.89
3/10t%	8.45	1.89	9.40	7.49	10.98	4.98
4/10t%	8.55	1.64	9.38	7.72	11.21	5.44
5/10t%	8.82	1.94	9.80	7.84	11.83	4.08
6/10t%	9.60	1.61	10.42	8.79	13.09	6.89
7/10t%	9.79	1.75	10.68	8.90	13.13	6.36
8/10t%	9.29	1.58	10.09	8.49	13.11	6.93
9/10t%	8.96	1.83	9.89	8.04	10.94	4.22
10/10t%	7.37	2.36	8.56	6.17	10.36	3.01

TABLE IV
SPANISH CONSTRUCTION PROJECTS DATABASE EVM/ES INDEXES

EVM/ES Indexes	Mean	Deviation	UCI	LCI	Max	Min
ES%	30,24	23,65	42,21	18,27	30,24	23,7
SV%	6,60	4,97	9,11	4,09	6,60	4,97
SPI	1,30	0,24	1,42	1,18	1,30	0,24

TABLE V
LITERATURE CONSTRUCTION PROJECTS DATABASE CHARACTERIZATION

Tenths of PV	Mean	Deviation	UCI	LCI	Max	Min
1/10t%	6.06	1.12	7.15	4.97	7.23	4.73
2/10t%	7.65	0.58	8.22	7.08	8.32	6.90
3/10t%	8.53	0.51	9.02	8.03	8.96	7.90
4/10t%	8.75	0.50	9.24	8.25	9.23	8.09
5/10t%	8.47	1.52	9.96	6.98	9.96	6.45
6/10t%	9.20	0.33	9.52	8.88	9.63	8.93
7/10t%	8.89	0.62	9.49	8.28	9.37	8.03
8/10t%	9.04	1.14	10.15	7.93	10.57	8.02
9/10t%	9.53	1.63	11.13	7.93	11.20	7.36
10/10t%	8.56	0.64	9.18	7.93	9.16	7.98

TABLE VI
LITERATURE CONSTRUCTION PROJECTS DATABASE EVM/ES INDEXES

EVM/ES Indexes	Mean	Deviation	UCI	LCI	Max	Min
ES%	21.88	9.24	30.93	12.82	21.88	9.24
SV%	5.00	2.71	7.65	2.35	7.00	1.00
SPI	1.22	0.09	1.31	1.13	1.29	1.08

To assess the dissimilarity between the four clusters formed, an ANOVA analysis was performed for the SV and SPI control variables. The significant differences detected for these indicators were analyzed by Tukey and Fisher method. It was proven that for the selected control variables the clusters are statistically dissimilar [41]. Moreover, the data in Tables VII and VIII are graphically represented in Fig. 3, where the four typical behaviors of the construction projects can be identified.

Cluster C1 and cluster C2 define the intermediate stages of the construction projects behaviors, where C2 contains projects that have a shorter completion time than the projects contained in C1. Cluster C3 and cluster C4 define the extreme behaviors for construction projects, where cluster C3 contains projects that end on time, while cluster C4 contains projects with the greatest delays.

Construction projects from C1 usually start with a performance below the planned rate and end with an average delay of 64%, while construction projects from C2 start with an adequate construction rate, suffer minor delays, culminating then with a total delay of almost 27% of the planned time. Cluster C3 defines the construction projects that maintain the expected behavior. These types of projects perform a 5% of work advance in the first two tenths and maintain a construction rhythm greater than 10% in the remaining tenths. Thus being able to maintain a curve that is between the expected schedule and early schedule, and finish on the initially estimated time. Finally, C4 defines the construction projects that suffer the greatest delays and therefore have completion times furthest from those originally planned. This type of project begins with a poor performance; suffer great delays and setbacks; and finally end up with an average delay of 120% of the planned completion time.

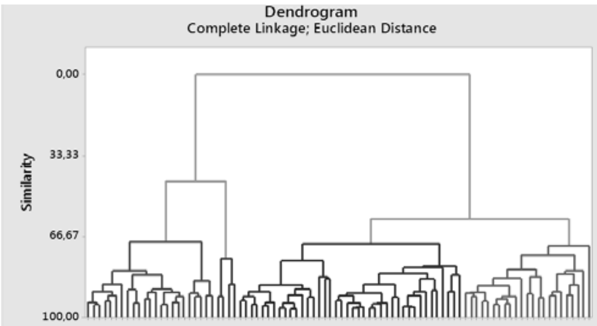

Fig. 2 Clustering dendrogram using complete linkage method

TABLE VII
CENTROIDS FOR 4 CONGLOMERATES COMBINATION

Variables	C1 [b]	C1 UCI	C1 LCI	C2	C1 UCI	C1 LCI
1/10t%	.	5.90	2.75	5.57	6.93	4.21
2/10t%	5.84	6.70	4.97	7.83	8.84	6.82
3/10t%	5.25	5.97	4.54	7.53	9.36	5.69
4/10t%	6.03	6.81	4.24	8.57	9.25	6.81
5/10t%	7.09	8.03	6.15	8.42	9.43	7.40
6/10t%	7.72	8.53	6.42	9.68	10.77	8.51
7/10t%	8.76	10.09	7.43	8.90	9.83	7.98
8/10t%	8.29	9.43	7.15	8.79	9.59	7.99
9/10t%	8.73	10.09	7.38	10.03	11.22	8.84
10/10t%	9.17	10.36	7.98	8.96	10.00	7.92
ES %	64.57	68.31	60.74	27.08	29.44	24.55

[b] C1, …, C4 represents Cluster N°1, …., Cluster N°4 centroid.

It should be noted that the initial behaviors of the four clusters show similarities and it is only after the third tenth that the behavior begins to differ. This is the reason why a study of the best subsets was carried out. Through this statistical analysis, it will be determined which one or more of the tenths are determinant to the projects delay, i.e. ES index; and only these key predictors will be taken into account during the validation phase. This will ensure that the classification of new projects is accurate, avoiding comparing tenths that are not related to the ES value.

TABLE VIII
CENTROIDS FOR FOUR CONGLOMERATES COMBINATION

Variables	C3 [b]	C3 UCI	C3 LCI	C4	C4 UCI	C4 LCI
1/10t%	6.75	8.40	5.11	5.57	7.86	0.97
2/10t%	9.77	11.33	8.20	7.83	8.81	1.43
3/10t%	10.50	12.12	9.87	4.53	4.73	3.30
4/10t%	10.44	12.03	9.85	3.57	4.47	3.35
5/10t%	10.95	12.59	9.31	8.42	9.87	6.64
6/10t%	11.09	13.19	11.00	5.68	6.48	4.50
7/10t%	11.40	13.12	9.68	8.90	11.73	5.80
8/10t%	11.34	13.14	9.54	8.79	9.20	4.61
9/10t%	8.21	9.51	6.91	4.37	6.42	2.31
10/10t%	6.49	8.09	4.88	3.02	5.29	0.74
ES %	5.70	7.98	3.32	106.74	118.45	95.02

[B] C1, …, C4 represents Cluster N°1, …., Cluster N°4.

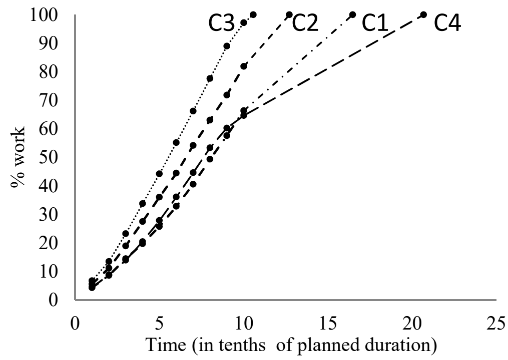

Fig. 3 Cluster graphical representation

C. Validation

The validation phase consists in determining how appropriate the obtained clusters are. The best method to classify new projects in the proposed clusters will be determined and then the degree in which the clusters conform to the validation data bases will be calculated.

After an exhaustive analysis of the Uruguayan construction projects database, it is concluded that not all tenths have the same ES predictive capability. This is reaffirmed in Fig. 3, where it can be seen that the centroids have a similar behavior at the beginning of the projects, separating its behavior only once 30% of the planned time has passed. It is for this reason that determining which one or more of the tenths has a better predictive capability in the ES response is sought.

The regression that best fits the predictive variables will be obtained using a best subsets model. Table IX shows all combinations for W_i predictors and the coefficient of determination (R^2) of the predictive model that is obtained by using each combination. The best combination of three predictors was selected for the following reasons: (1) it contains the smallest number of predictors that obtain the greater R^2 and (2) all predictors are at the initial phases of the project, which will allow predictions at earlier stages of the projects. In the selected combination (row 5, Table IX), the percentage of the ES data variation that is explained by W_i predictors value is 75.19% (R^2).

After selecting the best predictor variables combination, a means analysis was performed. Means analysis are commonly used to determine whether the mean of each group differs from the overall mean. Using this statistical analysis, it was possible to demonstrate that the best subsets that were selected are dissimilar. This can be seen represented in Figs. 4-6.

Afterward, the statistical dissimilarity between the main predictors was demonstrated, the new projects of the validation databases were classified among the proposed clusters. It can be said that a new construction project belongs to a cluster if all its elements belonging to the centroids of the main predictors (its value is within the confidence interval for each one of the key predictors). Table X can be used as a reference to classify a new project within the four clusters obtained. For the construction project to be owned, it must belong to the three key predictors between ICU and LCI.

TABLE IX
BEST SUBSETS REGRESSION - POSSIBLE COMBINATIONS

R^2	1/10t	2/10t	3/10t	4/10t	5/10t	6/10t	7/10t	8/10t	9/10t	10/10t
58.78			X							
54.19				X						
65.88				X		X				
62.54			X					X		
75.19			**X**	**X**		**X**				
74.77			X	X				X		
81.12			X	X		X		X		
77.89			X	X				X	X	
86.74			X	X		X		X		X
82.43		X	X	X				X	X	
90.21			X	X		X		X	X	X
88.13		X	X	X		X		X		X
93.13		X	X	X		X		X	X	X
91.46	X		X	X		X		X	X	X
94.38		X	X	X	X	X		X	X	X
94.10	X	X	X	X		X		X	X	X
95.63	X	X	X	X	X	X		X	X	X
94.52	X	X	X	X		X	X	X	X	X
95.77	X	X	X	X	X	X	X	X	X	X

[B] C1, …, C4 represents Cluster N°1, …., Cluster N°4.

TABLE X
CLUSTERING CLASSIFICATION TABLE

Variables	3/10t (%)	4/10t(%)	6/10t(%)
C1 - UCI	5.97	6.81	8.53
C1 - LCI	4.54	4.24	6.42
C2 - UCI	9.36	9.25	10.77
C2 - LCI	5.69	6.81	8.51
C3 - UCI	12.12	12.03	13.19
C3 - LCI	9.87	9.85	11.00
C4 - UCI	4.73	4.41	6.48
C4 - LCI	3.35	3.35	4.50

A practical example will be described below. The Spanish database W_{E3} that belongs to "El Cañaveral" construction project will be used for demonstration purposes, see Table XI. Taking into account the key predictors (3/10t, 4/10t, and 6/10t) and using Table X, it can be said that this project belongs to C2 cluster. The predictor 3/10 is between 5.69% and 9.36%, the predictor 4/10 belongs to the confidence interval that goes from 6.81% up to 9.25%, and the predictor 6/10 is between 8.51% and 10.77%, satisfying the proposed classification criteria. This same study has been carried out for each of the projects of both validation databases, obtaining a classification of the validation projects in to the proposed clusters.

TABLE XI
W_{E3}– EL CAÑAVERAL CONSTRUCTION PROJECT

1/10t	2/10t	**3/10t**	**4/10t**	5/10t	**6/10t**	7/10t	8/10t	9/10t	10/10t	ES
4.21	7.41	**9.32**	**8.23**	4.08	**10.09**	9.13	9.55	9.55	8.48	25,00

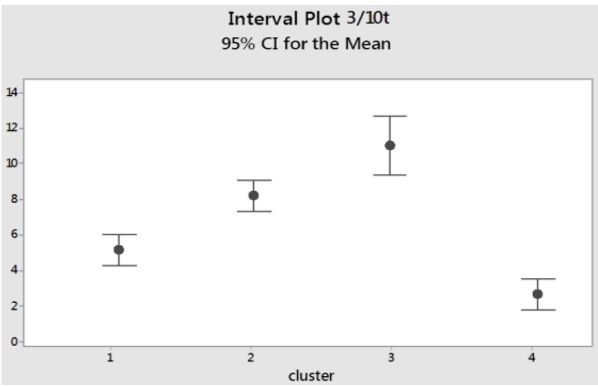

Fig. 4 Interval plot for 3/10 t – Clusters for Uruguayan database

Finally, to obtain the method prediction capability, the absolute value of the difference between the average delay of the cluster (ES centroid for each cluster) and the delay of each one of the projects belonging to it is calculated. For the example case this value is calculated as the absolute value of the difference between 27.08% and 25.00%, i.e. 2.08%. For all the projects belonging to the two validation databases, this difference is calculated. The predictive capacity of the model is given by the average of all the calculated differences. Using this criterion, a difference of 2.36 % was obtained for the Spanish database, and 4.72% for the literature database, averaging a value of accuracy in the prediction of 3.54%.

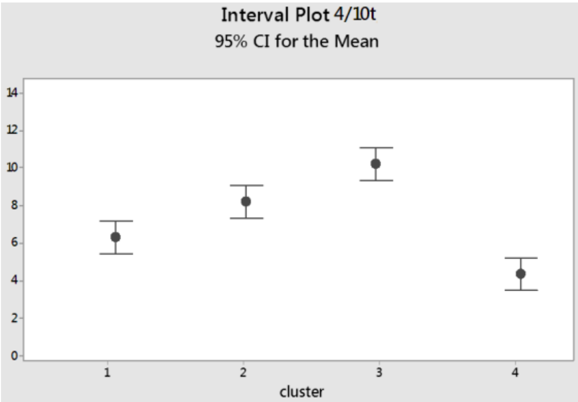

Fig. 5 Interval plot for 4/10 t – Clusters for Uruguayan database

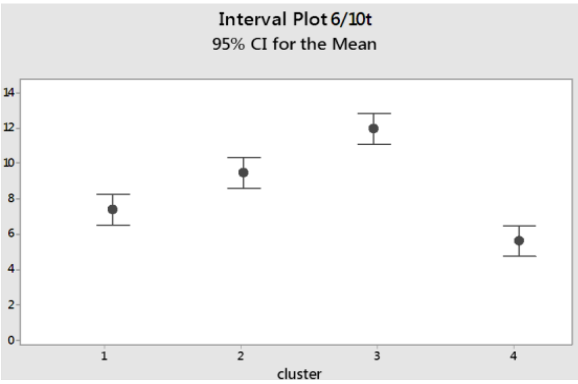

Fig. 6 Interval plot for 6/10 t– Clusters for Uruguayan database

V. Conclusion

Four typical behaviors of construction projects were detected by the use of a cluster analysis. The study was successfully validated through an ANOVA using Tukey and Fisher methods. A method to determine the belonging of new projects to the proposed clusters was developed using an Uruguayan database and cross-validated with two additional databases: Spanish and Literature.

It could be concluded that it is possible to have an acceptable predictive capacity at 30% of the planned time, but it is only at 60% that more accurate data about delays behavior is obtained. Moreover, in the validation phase an error of 3.54% was obtained; transforming the classification criterion into a powerful tool for the prognosis and management of construction projects.

Acknowledgment

We are grateful to National Housing Agency (Uruguay) for giving us permission to use State data, to architect Ariel Beltrand, Housing Programmes Manager, to architect Gabriel Borderre, Construction Division Manager, and to architect Jorge Risso, Housing Programmes Chief Officer, for their support and keen interest in this project.

References

[1] M. Sambasivan and Y. Soon, “Causes and effects of delays in Malaysian construction industry.” *International Journal of Project Management, Volumen* 25, pp. 517-526, 2007.

[2] M. Alkhathami, “Examination of the correlation of critical success and delay factors in construction projects in the kingdom of Saudi Arabia”, *Pittsburgh: Doctor of Philosophy Thesis: University of Pittsburgh*, 2004.

[3] M. Mohamad, “The factors and effect of delay in government construction project: Case study in juantan.” *Malaysia Pahang: Bachelor´s Degree thesis*, 2010.

[4] S. Assaf and S. Al-Hejji, “Causes of delay in large construction projects.” *International journal of project managment*, 4(24), pp. 349-357, 2006.

[5] M. Mulenga, A. Clinton. and T. Wellington, “Effects of construction projects schedule overruns: A Case of the Gauteng Province, South Africa”, *Procedia Manufacturing* , pp. 1-6, 2015.

[6] W. Chan and M. Kumaraswamy, “A comparative study of causes of time overruns.” *International Journal of Project Management*, 1(15), pp. 55-63, 1997.

[7] A. De Marco, T. Narbaev. “Earned value-based.” *Journal of Facilities Management*, 11(1), pp. 69-80, 2013.

[8] G. Aristondo, “Cooperativismo, ayuda y autogestión.” *Revista electrónica geográfica y ciencias sociales*, 2(146), pp. 41-98, 2003.

[9] E.C, “Strengthening project internal, How to enhance the role of. s.l.” *Office for Official Publications of the European Communities*, 2007.

[10] PMBOK, “Project Management Body of Knowledge” (PMBOK®). 3rd Edition ed. s.l, *Project Management Institute*, 2004.

[11] S. Brlecic, M. Dimitric and M. Dalsaso, “Effective Project Management Tools for Modern.” *Pomorski zbornik*, Volumen 51, pp. 131-145, 2016.

[12] E. Norman, S. Brotherton and R. Fried, “Work Breakdown Structures: The. New Jersey”, *John Wiley & Sons*. 2008.

[13] W. Lipke, “Achieving normality for cost.” *The Measurable News, Volumen (Fall/Winter)*, pp. 6-11, 2003.

[14] W. Abba and F Niel, F., “Integrating technical performance measurement.” *The Measurable News*, Volumen 4, pp. 6-8, 2010.

[15] F. Anbari, “Earned value project management method and extensions.” *International Journal of Project Management*, 34(4), p. 12–23, 2003.

[16] V. Blanco, “Earned value management: a predictive analysis tool.” *Navy Supply corps newsletter*, 66(2), pp. 24-27, 2003.

[17] R. Bruke, “Project Management Planning and Control Techniques” 4th Edition ed. s.l.:s.n, 2003.

[18] D. Cioffi, “Designing project management: A scientific notation and an improved formalism for earned value calculations“, *International Journal of Project Management*, Volumen 24, pp. 136-144, 2006.

[19] D. Jacob, “Forecasting project schedule completion with earned value.” *The Measurable News*, pp. 7-9, 2006.

[20] E. Kim, W. Wells and M. Duffey,” A model for effective implementation of Earned Value Management methodology.” *International Journal of Project Management*, 21(5), pp. 375-382, 2003.

[21] W. Lipke. “A study of the normality of earned value indicators. *The Measurable News*, pp. 1-16, 2002.

[22] R. McKim,T. Hagazy, and M. Attalla, “Project performance control in in reconstruction projects”, *Journal of Construction Engineering and Management*, 126(2), pp. 137-141, 2000.

[23] R. Aliverdi, L. Moslemi, and S. Amir, S., “Monitoring project duration and cost in a construction project by applying.” *International Journal of Project Management*, Volumen 31, p. 411–423, 2013.

[24] W. Lipke, O. Zwikael, K. Henderson, and F. Anbari, “Prediction of project outcome, the application of statistical methods to earned value

management and earned schedule performance indexes." *International Journal of Project Management,* Volumen 27, pp. 400-407, 2009.
[25] Q. Fleming, and J. Koppleman, "Earned Value Project Management". 4th Edition ed. Pensilvania: *Project Managment Institute*, 2010.
[26] M. Vanhoucke, "Mesuring time-improving project perfomance using earned value managment." *Reserch and managment science*, Volumen 136, 2010.
[27] M. Vanhoucke, "Integrated project managment and control: first comes the theory, then the practice." *Managment for proffesionals*, Volumen 23, 2014.
[28] M. Liberatone, B. Pollack-Johnson, and C. Smith, "Project *Managment in construction: software use and research directions." Journal of construcion engineering and managment*, 127(2), pp. 101-107, 2001.
[29] L. Forbes, and S. Ahmed, "Modern construction: lean project delivery and integrated practices." CRC Press, 2010.
[30] N. Rudeli, A. Santilli, I. Puente, and E. Viles, "A statistical model for schedule prediction: Validation in a housing cooperative construction database." *Journal of Construction Engineering and Management*, 2017.
[31] A. Senoucia, A. Ismailb, and N. Eldina, "Time Delay and Cost Overrun in Qatari Public Construction Projects." *Procedia Engineering*, Volumen 164, p. 368 – 375, 2016.
[32] S. Vandevoorde and M. Vanhoucke, M., "A comparison of different project duration forecasting methods." *International Journal of Project Management*, Volumen 24, p. 289–302, 2006.
[33] S. Guo, J. Chen and C. Chiu, "Fuzzy duration forecast model for wind turbine construction project." *Automation in Construction*, 2016.
[34] A. Araszkiewicz, "Application of Critical Chain Management in Construction Projects Schedules in a Multi-Project Environment: a Case Study." *Procedia Engineering*, Volumen 182, p. 33 – 41, 2017.
[35] C. Sabaté, C. "De la Preservación del Patrimonio a la Ordenación del Paisaje. Intervenciones en Paisajes Culturales en Latinoamérica." *Gerencia Española de Cooperación Internacional para el Desarrollo*, Volumen 502-11-004-2, pp. 11-23, 2012.
[36] M. Callahan, D. Quackenbush, and J. Rowings, "Construction Project Scheduling." New York, NY: McGraw-Hill, Incorporated, 1992.
[37] A. Hernandez, and J. Loayza, "Planeamiento integral, control, construcción y análisis técnico ejecutado en un centro comercial en la ciudad de Arequipa", Perú: Pontificia Universidad Católica de Perú, 2013.
[38] J. Hair, R. Anderson, R. Tatham and W. Black., "Análisis multivariate 5ta edición", 799 pag, 2004.
[39] A. Scott, and M. Knott, "A cluster analysis method for grouping means in the analysis of variance." *Biometrics*, 507-512, 1974.
[40] W. Dillon, and M. Goldstein, "Multivariate analysis: Methods and applications", 1984.
[41] A. Santilli, N. Rudeli, E. Viles, M. Tanco, D. Jurburg, "Comportamiento tipo del desarrollo de obras cooperativas de viviendas Uruguayas.", *Memoria Investigaciones en Ingeniería*, 15 (2017) ISSN 2301-1092.