Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 64 (2015) 586 – 593

Procedia Computer Science

Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2015 October 7-9, 2015

A residual Grey prediction model for predicting S-curves in projects

José Ramón San Cristóbal*, Francisco Correa, María Antonia González, Emma Diaz Ruiz de Navamuel, Ernesto Madariaga, Andrés Ortega, Sergio López, Manuel Trueba

ETS de Náutica, University of Cantabria, Santander 39004, Spain

Abstract

S-curves are usually taken as expression of project progress and have become a requisite tool for project managers through the execution phase. The common methodology for predicting S-curve forecasting models is based on classifying projects into groups and producing a standard S-curve for each group using multiple linear regression techniques. Traditional regression models taken to fit individual projects require a large amount of data and make many strict assumptions regarding statistical distribution of the data. The grey system theory, however, is well suited to study the behavior of a system with incomplete information or limited amount of discrete data. Easy of use and accuracy, two significant criteria for project managers when choosing a forecasting model, are considered two additional attributes of the grey system theory. This paper proposes a residual Grey prediction model to forecast the actual cost and the cost at completion of a project based on the grey system theory. Results show that the accuracy of the forecasting model is highly efficient.

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge

Keywords: S-curves; Forecasting; Grey model.

1. Introduction

It is assumed that the profile of the cumulative cost versus elapsed time on projects takes the shape of an S-curve[1]. The common characteristics of all systems that demonstrate this behaviour (S-curve) include slow growth, followed

* Corresponding author. Tel.: +0 34 942 202267; fax: +0 34 942 201303.
E-mail address: jose.sancristobal@unican.es

1877-0509
Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge
doi:10.1016/j.procs.2015.08.570

by rapid growth, which in turn is followed by slow growth again to an asymptotic maximum. The reason is that projects start slowly when the resources necessarily need to set up, and then projects start to accelerate once all resources have been acquired[2,3]. Even the largest projects start with an initially small number of tasks, but soon begin to tackle multiple activities simultaneously. These parallel interconnected activities increase the spending greatly compared to the work at the beginning which in comparison is usually more limited. A cap on the total cost limits the budget and forces this large spending rate to decline. Because the tasks depend on each other, they cannot all end simultaneously. Activities dwindle to a small number again and eventually the entire project ends. Thus, in projects, the S-curve is driven by the multiple interconnected activities that occur in the middle of the project life[4].

S-curves are usually taken as expression of project progress and have become a requisite tool for project managers through their execution phase[5]. The guide to the Project Management Body of knowledge (the PMBOK) defines the S-curve based on its appearance as: "graphic display of cumulative costs, labour hours, percentage of work, or other quantities, plotted against time." The name derives from the S-like shape of curve (flatter at the beginning and end, steeper in the middle[4].

The literature review suggests that S-curves can be used for several purposes, as a target against which the actual progress of the project can be evaluated at any point in time to monitor whether the project is on schedule[6], to forecast the likely duration of the project once the contract price and cumulative expenditure are known, and even to manage cash flow, current performance status, future necessary cost/duration, etc., for running projects[5,7-9]. Jepson[10] argued that S-curves representing labour man hours or labour costs can only act as indices for financial control.

Traditional regression models taken to fit individual projects could not be well fitted. These methods require a large amount of data and make many strict assumptions regarding statistical distribution of data. Few data, extreme values, emerging changes, classifications of projects, uncertainties and uniqueness always exist in the project engineering environment. This paper proposes a residual Grey prediction model to forecast the actual cost and the cost at completion of a project based on the grey system theory, which is well suited to study the behaviour of a system with incomplete information or limited amount of discrete data. The paper begins by reviewing the literature on methods for predicting S-curves. Next, the residual grey forecasting method is presented and applied to a road building project. Finally, there is a concluding section with the main findings of the paper.

2. Methods for predicting S-curves

Various mathematical formula forms for S-curves have been developed[5,7,11]. Hardy[12] analysed 25 different types of projects and found that there was no close correlation between the values considered even when separating them into different categories. Bromilow and Henderson[13] used four general building projects to develop their value S-curve. Drake[14] collected projects from regional health authorities and classified them into different cost categories, the author fitted and S-curve into each of these categories but no figures were published of the number of projects analysed or of the level of accuracy of the fitted function. Kenley and Wilson[3] proposed an ideographic methodology to build individual construction project cash flows model based on the logit transformation approach. Skitmore[15] utilized three approaches, analytic, synthetic, and hybrid, in combination with six alternative models to determine the best approach/model combination for the available data and forecasts for future expenditure flows. Kaka[2] used a stochastic model based on historical data with logit transformation technique to incorporate variability and inaccuracy in their forecasts and decision-making. Barraza et al.,[8] developed stochastic S-curves to provide probability distributions of budgeted cost and planned elapsed time for a given percentage of progress in order to evaluate cost and time variations. Hwee and Tiong[16] developed an S-curve profile model from cost-schedule integration equipped with progressive construction-data feedback mechanisms. Mavrotas et al.,[1] modelled cash flows based on a bottom-up approach from a single contract to the entire organization with an S-curve based on a conventional non-linear regression model. Blyth and Kaka[9] proposed a model that standardized activities to produce an individual S-curve for an individual project using a multiple linear regression model. Chao and Chien[17] proposed an empirical method for estimating project S-curves that combined a succinct cubic polynomial function and a neural network model based on existing S-curve formulas and attributes of the project. Cheng and Roy[18] proposed an evolutionary fuzzy decision model for cash flow prediction using time-dependent support vector machines and S-curves. Cheng et al.,[6] proposed a progress payment forecasting approach using S-curves for the construction phase. The authors improve the traditional grey prediction model by applying the golden section and bisection method to build a short-interval cost-forecasting model. Maravas and Pantouvakis[19] developed an S-surface cash flow model based on fuzzy set theory to predict the working capital requirements of projects. Lin et al.,[5] proposed a construction project progress forecasting approach which combines the grey dynamic prediction model and the residual modified model to forecast the current project progress during the construction phase. Chen et al.,[20] estimated project's profitability at completion using a multivariate robust regression model to test how well the key variables in project initiation and planning phases predict project profitability.

There are limitations that existed in the previous studies at developing models to forecast S-curves such as poor classification of projects, statistical assumptions, unique of construction projects, etc. Most developed formulas have been based on different classified groups according to criteria such as type of project, duration of contract, type of procurement, and size of company[21]. The characteristic of grouping distribution could display different S-curves due to the fact that each construction project is unique, should be modelled separately and greatly different from one project to another. In addition, the groups of the same classified projects still could have a variety of shapes of S-curves due to uncertain factors, while the different type of projects may have almost the same shape of curves[5]. Thus, poor choice of project grouping could be done in practice[3,9]. Limitations of traditional linear regression is that extreme values could unduly influence the logistic transformation analysis when the data approaches either 0% or 100% and that it should be to gather a lot of history sample data and make many strict assumptions to distribution of samples[5]. Different parameter values of the fitting model should be used for different types of construction projects rather than the general practice of applying its parameter values to all types of work[15]. In addition, it is different for traditional statistical models to reflect real growth trends among different stages because parameters such as cost/period can grow at different speeds during the whole project[5].

In forecasting there is a common assumption that accuracy is the primary criterion in selection among forecasting techniques. However, new techniques and criteria are desirable in the selection and evaluation of forecasting techniques from practical perspective[22]. Easy of use, easy of using available data and interpretation, flexibility, etc., are significant criteria to project managers who are concerned about what technique will provide high-speed and works well over various work-related situations. The grey prediction model gets rid of any aforementioned strict assumptions and the sample is suited for limited data to construct a forecasting model. Few data, emerging changes, uncertainties and uniqueness always exist in the construction project engineering context[5].

3. A residual Grey forecasting method

The grey system theory, originally presented by Deng[23], focuses on model uncertainty and information insufficiency in analyzing and understanding systems seeking mathematical relations and movement rules. The grey system puts each stochastic variable as a grey quantity that changes within a given range. It does not rely on statistical method to deal with the grey quantity. It deals directly with original data, and searches the intrinsic governing laws from the available data[24]. In the grey system theory there are three systems classified by the degree of information completed. A white system is defined as the case where information in it is fully known; while a black system is defined as the case where information is unknown or nothing in the system is clear. A system with partial information known and partial information unknown is defined as a grey system. Among the various forecasting models that have been developed, the Grey prediction model requires fewer data and less complicated mathematical calculation. This characteristic is the core of the Grey system theory[6], which has been successfully applied to many fields including wafer fabrication, opto-electronics, electricity costs, integrated circuits, and meteorology[25].

The most commonly used grey forecasting model is GM (1,1)[26], which indicates one variable is employed in the model and the first differential equation is adopted to match the data generated by the accumulation generating operation (AGO). The AGO reveals the hidden regular pattern in the system development and converts a series lacking obvious regularity into a monotonously increasing series to reduce the randomness of the series, and increase the smoothness of the series.

The grey dynamic prediction model should be operated in accordance with the principle of keeping the same dimension of data series. The minimum number of data must be four in consecutive order without bypassing any data[28]. That is to say, a new data is attached on tail end of the original data series and the first data in the original data series should be removed before the next forecasting operation. This operation could be performed step-by-step according to the above equation to get a new predicted value for each subsequent period.

In order to establish the GM (1,1) model, the raw data series of number $X^{(0)}$ is assumed to be

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\} \quad (1)$$

where n is the total number of modelling data. These data are fluctuating in a definite range. In order to find out the regular patterns, the series of data is treated by 1-AGO:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \quad (2)$$

So

$$\begin{aligned} x^{(1)}(1) &= x^{(0)}(1) \\ x^{(1)}(2) &= x^{(0)}(1) + x^{(0)}(2) \\ &\ldots \\ x^{(1)}(n) &= x^{(0)}(1) + x^{(0)}(2) + \ldots + x^{(0)}(n) \end{aligned} \quad (3)$$

then, a series of number $X^{(1)}$ is formed

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\} \quad (4)$$

The GM (1,1) model can be constructed by establishing a first order differential equation for $X^{(1)}(k)$ as:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (5)$$

where parameters *a* and *u* are called the developing coefficient and grey input, respectively. In order to obtain the solution to this differential equation, we can get the parameters *a* and *u* by using the least square method[26]. Then

$$\hat{a} = [a, u]^{(T)} = (B^T B)^{-1} B^T Y \quad (6)$$

where

$$B = \begin{bmatrix} -\frac{1}{2}\left(x^{(1)}(1) + x^{(1)}(2)\right) & 1 \\ -\frac{1}{2}\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ & \vdots \\ -\frac{1}{2}\left(x^{(1)}(n-1) + x^{(1)}(n)\right) & 1 \end{bmatrix} \quad (7)$$

and

$$Y_N = \left(x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\right)^T \quad (8)$$

The solution of the differential equation (5) is

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{u}{a}\right)\exp(-ak) + \frac{u}{a} \quad (9)$$

Then, according to the inverse accumulated generating operation (IAGO), we can get the modeling calculated values $\hat{X}^{(0)}$

$$\hat{X}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n)\} \quad (10)$$

where

$$\hat{x}^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1), (k = 2, 3, \ldots, n) \quad (11)$$

and

$$\hat{x}^{(1)}(1) = x^{(0)}(1) \tag{12}$$

The residual series is a sequence that could be derived from the operations by grey dynamic prediction model as follows:

$$e(k) = \hat{x}^{(0)}(k) - x^{(0)}(k), (k = 2, 3, \ldots, n) \tag{13}$$

where $\hat{x}^{(0)}(k)$ is the prediction value, $x^{(0)}(k)$ is the real value and $e(k)$ is the residual value. Thus, the residual series is transformed as a nonnegative sequence as follows. Given the series $Y = (Y_1, Y_2, \ldots, Y_n)$, the nonnegative sequence of this series is:

$$LY = Lg10(Y + 1 - \min(Y)) \tag{14}$$

Then, the grey dynamic prediction model GM(1,1) is again introduced as:

$$\hat{e}^{(1)}(k+1) = \left(e^{(0)}(1) - \frac{u_e}{a_e} \right) \exp(-a_e k_e) + \frac{u_e}{a_e} \tag{15}$$

where

$$\frac{a_e}{\mu} = (B_e^T B_e)^{-1} B_e^T Y_{ne} \tag{16}$$

$$B_e = \begin{bmatrix} -\frac{1}{2}\left(e^{(1)}(1) + e^{(1)}(2)\right) & 1 \\ -\frac{1}{2}\left(e^{(1)}(2) + e^{(1)}(3)\right) & 1 \\ & \vdots \\ -\frac{1}{2}\left(e^{(1)}(n-1) + e^{(1)}(n)\right) & 1 \end{bmatrix} \tag{17}$$

and

$$Y_{Ne} = \left(x_e^{(0)}(2), x_e^{(0)}(3), \ldots, x_e^{(0)}(n)\right)^T \tag{18}$$

Through inverse AGO to get the predictive error’s value as:

$$\hat{e}^{(0)}(k) = \hat{e}^{(1)}(k) - \hat{e}^{(1)}(k-1), (k = 2, 3, \ldots, n) \tag{19}$$

An error measure is used to asses the accuracy in terms of closeness of fit as well as to provide a basis for model performance evaluation. The evaluation criterion to measure the percent of prediction accuracy is the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n} \sum_{k=1}^{n} \frac{\left| x^{(0)}(k) - \hat{x}^{(0)}(k) \right|}{x^{(0)}(k)} \tag{20}$$

Lower MAPE values are better because they indicate that smaller percentage errors are produced by a forecasting model. Following Lewis[27] (1982), less than 10 percent is highly accurate forecasting. Values between 10 percent and 20 percent, between 20 percent and 50 percent, and higher than 50 percent are considered indicators of high, average, and low prediction accuracy, respectively.

4. Application

The purpose of this section is to illustrate the use of the proposed model, which can be a useful tool for project managers in controlling and revising the gap between the estimated and actual S-curve during the course of a project. With this aim, the GM (1,1) model is applied to the road building project shown in Table 1. The project was undertaken in the mountain pass La Braguía in the North of Spain from June 2012 to July 2013. Table 2 shows the actual cost, the forecasted value and the corresponding MAPE values.

Table 1. Data of the project.

Task	Description	Predecessors
A	Demolitions	-
B	Transport of Earths	-
C	Walls	A
D	Longitudinal drainage	A
E	Transversal drainage	B
F	Granular and asphalt capes	C, D
G	System of road signs	F, E
H	Markings on the road	F, E
I	Landscape integration	B

Table 2. Actual cost, forecasted value, error and MAPE values.

Date	Real	Forecasted	Error	MAPE
Jun-12	25,567			
Jul-12	66,293	59,985		
Ago-12	78,293	84,338		
Sep-12	124,073	118,577		
Oct-12	191,367	166,717	-24,650	
Nov-12	259,845	295,589	26,842	1.5
Dic-12	285,612	338,051	80,578	2.7
Jan-13	290,843	365,047	60,337	1.3
Feb-13	303,489	406,482	7,246	3.8
Mar-13	316,431	313,012	-4,728	1.1
Apr-13	320,690	305,523	9,348	3.0
May-13	336,756	312,135	-5,725	2.2
Jun-13	349,379	346,961	-3,794	1.6

In order to obtain the values shown in Table 2, we follow the following process: with the data from June 12 to September 12, a series of number $X^{(0)}$ is formed:

$$X^{(0)} = \{25{,}567; 66{,}293; 78{,}293; 124{,,}073\}$$

This series of number is treated by 1-AGO (Eqs. 2 and 3). Then, we get a new series of number:

$$X^{(1)} = \{25{,}567; 91{,}938; 170{,}231; 294{,}304\}$$

Applying Eqs. (6), (7), and (8) we can get $\hat{a}$

$$\hat{a} = [a, u]^T = [-0.341; 41{,}607]$$

where

$$B = \begin{bmatrix} -58{,}791 & 1 \\ -131{,}085 & 1 \\ -232{,}268 & 1 \end{bmatrix}$$

and

$$\bar{Y} = [66{,}293;\ 78{,}293;\ 124{,}073]^T$$

The solution of the differential equation (5) is therefore

$$\hat{x}^{(1)}(k+1) = 147{,}756\exp(0.341k) - 122{,}111$$

Then, a series of calculated data $\hat{X}^{(0)}$ is given by inverse accumulated generating operation

$$\hat{X}^{(0)} = [25{,}645;\ 59{,}985;\ 84{,}338;\ 118{,}577;\ 166{,}717]$$

Operating in accordance with principle of keeping the same dimension of data series, that is, attaching a new data on tail end of the original data series and removing the first data, a new predicted value is obtained for each subsequent period. Performing this operation step-by-step, the residual series is obtained applying equation (13). Repeating the same process as above for this residual series, the forecasted values and their corresponding MAPE values shown in Table 2, are obtained.

All MAPE values are lower than 5 percent and only one value is higher than 3 percent. Both the initial and final MAPE values are lower than 2 percent, which indicates the small percentage error produced by the forecasting model at the initial and final stages of the project, as can be seen in Figures 1. According to Lewis' interpretation[32], these results show that the accuracy of the residual modified Grey prediction model to forecast the cost and the cost at completion of the project is highly efficient.

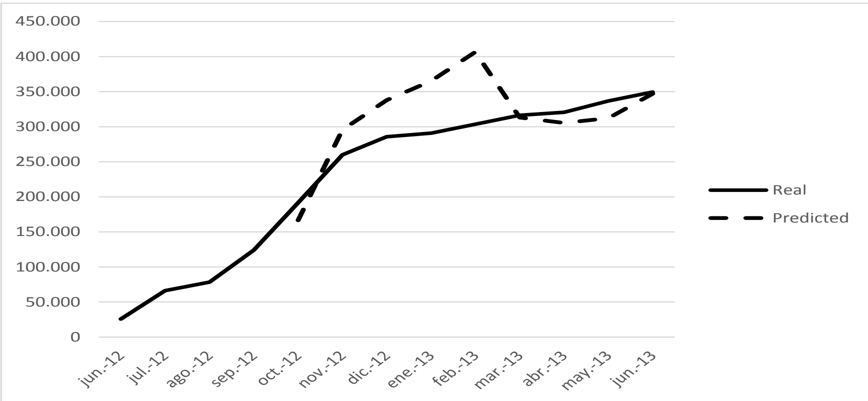

Fig. 1. Real and Predicted cost.

5. Conclusions

Accuracy has traditionally been the primary criterion in selection among forecasting techniques. However, new techniques and criteria are desirable in the selection and evaluation of these techniques from practical perspective. Easy of use and available data, interpretation, flexibility, etc., are significant criteria to project managers who are concerned about what technique will provide high-speed and works well over various work-related situations. The objective of this paper is to develop a reliable and easy cost forecasting method to assist project managers to control and revise the gap between the estimated and actual S-curves. In an environment featured by scarcity of data and uncontrollable factors, project managers always need a practicable and effectual tool to monitor real progress against the expected schedule at any time during the course of a project. This paper shows that the residual Grey prediction model can be utilized for prediction of the actual cost and cost at completion of a project. The advantages of the model are remarkable. When the number of data is not enough for mathematics statistics, or probabilistic analysis, the model provides an accurate, easy and stable method for predicting project progress in comparison with the traditional forecasting methods. Results show that the proposed approach is able to get accurate cost forecast. The low MAPE values obtained reflect that the accuracy of the residual Grey prediction model forecasting model is highly efficient.

References

1. Mavrotas, G., Caloghirou, Y., Koune, J. A model on cash flow forecasting and early warning for multi-project programmes: application to the operational program for the information society in Greece. *Int J Proj Manag* 2005;**23** (2):121-3.
2. Kaka, A.P The development of a benchmark model that uses historical data for monitoring the progress of current construction projects. Engineering, Construction and Architectural Management 1999;**6** (3):256-66.
3. Kenley, R., Wilson, O.D. A construction project cash flow model- An ideographic approach. *Constr Manage Econ*, ASCE 1986;**4**:213-32.
4. Cioffi, D.F. A tool for managing projects: an analytic parameterization of the S-curve. *Int J Proj Manag* 2005;**23**:215-22.
5. Lin, M.C., Tserng, H.P., Ho, S.P., Young, D.L. A novel dynamic progress forecasting approach for construction projects. *Expert Syst Appl* 2012;**39**;2247-55.
6. Cheng, Y.M., Yu, C.H., Wang, H.T. Short-Interval Dynamic Forecasting for actual S-curve in the construction phase. *J Constr Eng M ASCE* 2011;**137**:933-41.
7. Tuker, S.N A single alternative formula for Department of Health and Social Security S-curves. *Constr Manage Econ*, ASCE, 1988;**6**:13-23.
8. Barraza, G.A., Back, W.E., Mata, F. Probabilistic monitoring of project performance using S-curves. *J Constr Eng M ASCE* 2000;142-8.
9. Blyth, K., Kaka, A.. A novel multiple linear regression model for forecasting S-curves. Engineering, Construction and Architectural Management 2006;**13**(1):82-95.
10. Jepson, W.B. Financial control of construction and reducing the element of risk. Contract Journal 1969;**24**:862-4.
11. Miskawi, Z. An S-curve equation for project control. *Constr Manage Econ* 1989.
12. Hardy, J.V.Cash flow forecasting for the construction industry. MSc report. Dept. of Civil Engineering, Loughborough University of Technology; 1970.
13. Bromilow, F.J. and Henderson, J.A. Procedures for reckoning the performance of building contracts, 2nd ed., CSIRO. Division of Building Research. Highett Australia; 1970.
14. Drake, B.E. A mathematical model for expenditure forecasting post contract. Technician Israel Institute of Technology, Haifa; 1978.
15. Skitmore, M. A method for forecasting owner monthly construction project expenditure flow. *J Forecasting* 1998;**14**:17-34.
16. Hwee, N.G., Tiong, R.L. Model of cashflow forecasting and risk analysis for contracting. *Int J Proj Manag* 2002;**20**(5);351-63.
17. Chao, L.C., Chien, C.F. Estimating project S-curves using polynomial function and neural networks. *J Constr Eng M ASCE* 2009;169-77.
18. Cheng, M.Y., Roy, A.F.V. Evolutionary fuzzy decision model for cash-flow prediction using time-dependent support machines. *Int J Proj Manag* 2009;**29**(1):56-65.
19. Maravas, A., Pantouvakis, J.P. Project cash flow analysis in the presence of uncertainty in activity duration and cost. *J Project Manag* 2012;**30**(3):374-84.
20. Chen, H.L., Chen, C.I., Liu, C.H., Wei, N.C. Estimating a project's profitability: A longitudinal approach. *Int J Proj Manag* 2013;**31**:400-10.
21. Kaka, A.P., Price, A.D.F. Modeling standard cost commitment curves for contractors' cashflow forecasting. *Constr Manage Econ* 1993;**11**:271-83.
22. Yokum, J.T., Amstrong, J.S. Beyond accuracy: Comparison of criteria used to select forecasting methods. *J Forecasting* 1995:**11**;591-97.
23. Deng, J.L. Introduction to grey system theory. *J Grey Syst* 1989;**1**:1-24.
24. Mao, M., Chirwa, E.C. Application of grey model GM(1,1) to vehicle fatality risk estimation. Technological Forecasting and Social Change 2006;**73**:588-605.
25. Wu, L., Liu, S., Wang, Y. Grey Lotka-Volterra model and its applications. Technological Forecasting and Social Change 2012;**79**:1720-30.
26. Deng, J.L.Control problem of grey system. Systems and Control Letter 1982;**1**(5):288-94.
27. Lewis, C.D.. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. London: Butterworth Scientific; 1982.