

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

Advanced Statistical Model for the Prediction of Bio-Calcareous Minerals

Ramya, R.¹, Sarayu, K.^{2*}, Yuvaraj, P.³, Annaselvi, M.², Ramachandra Murthy, A.², Nagesh R. Iyer⁴,
Kalimuthu, P.¹

Paavai Engineering College, Pachal, Namakkal, India¹

CSIR-Structural Engineering Research Centre, Taramani, Chennai, India²

Engineering Consultant, L&T Infra Engineering Ltd., Chennai, India³

Former Director, CSIR-Structural Engineering Research Centre, Taramani, Chennai, India⁴

*Corresponding Author

ABSTRACT: Quantification of crystals produced by bacteria and the conditions required for it is time consuming and is a tedious process, which can be effectively handled by the statistical method. This paper presents the Multivariate Adaptive Regression Splines (MARS) model to predict the time (hrs) required for the calcium carbonate crystals obtained from various bacteria with optimization of various intrinsic and extrinsic factors. Multivariate Adaptive Regression Splines (MARS) is an implementation of solving regression-type problems. It is also used to predict the values of continuous dependent variables from a set of independent variables. The input variables of MARS are the intrinsic and extrinsic factors such as temperature (°C), pH, bacterial concentration (CFU/ml), calcium concentration (g/l), organic N₂ concentration (g/l) and inorganic N₂ source. The time required for maximum quantity of crystal production is the estimated outcome of MARS. MARS is a non-parametric regression algorithm. MARS model have been developed using MATLAB software. The total dataset have been divided into 70% for training and 30% for testing in the MARS. The predicated results are in good agreement with the experimental values.

KEYWORDS: Multivariate Adaptive Regression Splines (MARS); bio-calcareous minerals; bacteria; intrinsic factors; extrinsic factors.

I. INTRODUCTION

Concrete is one of the most commonly used building materials around the world. Usually concrete consists of cement, coarse aggregate, fine aggregate and water. Many structures have suffered severe degradation since their construction due to the combined effects of deicing salts, freeze-thaw cycles, aggressive environments, and drastically increased live loads. One of the major problems that the civil engineers of today face is to preserve, maintain, and retrofit the structures. Maintaining of cracked structures involves applying a concrete mortar onto the damaged surface. This process can be time consuming and expensive because it is difficult to gain access to damaged regions especially located underground and at greater heights.

Crack in concrete reduces the strength of structures. The cost required for repair of crack is high and time consuming; the alternate approach for repairing is self-healing. Microorganisms play an important role in improving sand properties, repair of limestone monuments and sealing of concrete cracks to highly durable material and finally enhance the durability of building materials. Self-healing concrete is a product that will biologically produce limestone to heal cracks that appear on the surface of concrete structures. The self-healing concrete is one that senses its crack formation and reacts to cure itself without human intervention. Self-healing of cracks in concrete would contribute to a longer service life of concrete structures and would make the material not only more durable but also more sustainable. Various healing mechanisms are autogenous healing, healing of cracks by superabsorbent polymer (SAP), calcium

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

carbonate precipitating micro-organisms, encapsulated polymer, etc. In this study calcium carbonate based healing is used to increase the strength and durability.

Bio-calcareous minerals are used to overcome this problem. The biological mode of synthesis of the minerals is known as the bio-mineralization process. Calcium carbonate and silica are examples of bio-minerals. It is quite widespread among the prokaryotic and the eukaryotic organisms ^[1]. Calcium carbonate is one such bio-mineral. Many bacterial species originated from various natural habitats have been reported to precipitate calcium carbonate ^[5] in invitro and natural conditions which are influenced by various abiotic factors that determine the route and environment for the calcium carbonate precipitation ^[4]. These are the composite materials that are made up of both mineral and organic components. There are various factors that influencing the calcium carbonate crystal formation. They are temperature, pH, calcium concentration, bacterial concentration, nitrogen sources.

Predicting the calcium crystal production in experimental method is tedious and time consuming, so we go for statistical method. There are various advanced statistical model for prediction of calcium carbonate crystal. In this paper quantity of calcium carbonate crystal produced and time required for precipitation by various bacteria are incorporated in Multivariate Adaptive Regression Splines (MARSplines) method. Multivariate Adaptive Regression Splines (MARSplines) is an implementation of techniques popularized ^[2] for solving regression-type problems. There are a large number of methods available for fitting models to continuous variables, such as a linear regression, nonlinear regression and regression trees. MARS is a flexible, more accurate, and faster simulation method for both regression and classification problems ^[2]. MARS gives the relationship in non linear form between the dependent and the predictor variables besides identifies the interaction among predictor variables. MARS is based on a divide and conquers strategy, partitioning the training data sets into separate regions; each gets its own regression line. MARS model is developed by using MATLAB software. In MARS model total data set is divided into 70% for training and 30% for testing.

II. MATERIAL AND METHODS

A. Collection of Bacteria:

Bacterial samples collected from Rameshwaram are used as a source for the present study. 30 different bacterial samples were taken and their growth pattern along with maximum crystal production is monitor. Bacteria can grow in culture medium designed with basic essential nutrients such as carbon and nitrogen source. In the present study calcium is supplemented along with the basic nutrients to have maximum crystal production at the minimal time without affecting the bacterial growth.

B. Growth in specific medium:

The bacterial species are grown in specific media for maximum calcium carbonate crystal precipitation. Media has organic and inorganic nitrogen source. The growth and crystal development are observed.

C. Parameters Monitored:

Calcium carbonate precipitation mediated by bacteria is basically decided by five intrinsic and extrinsic factors.

1. Temperature
2. pH
3. Bacterial concentration
4. Calcium concentration
5. Nitrogen source

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

Each of these components were varied individually by keeping others constant.

C.i. Temperature:

Bacteria may grow at a variety of temperature from close to freezing to near the boiling point of water. In present study, mesophilic bacteria is used that grows at (16 and 40°C). So a temperature range of 15, 20, 25, 30, 35, 40 and 45°C is maintained in this study.

C.ii. pH:

The pH in the microenvironment of the organism is an important factor that controls the survival and the metabolic activity, which in turn, controls the secretion of the products. So, a pH range of 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0 is maintained in the present study. High pH conditions are found to favour the formation of CO_3^{2-} from HCO_3^- that leads to calcification of the bicarbonate generated^[3].

C.iii. Bacterial concentration:

Minimal amount of bacteria is required for the growth in the given media. Optimal amount of inoculum is added to make sure, not excess or limited quantities are added for maximum crystal production. Bacterial concentration ranging from 100, 250, 500, 750, 1000, 1250 and 1500 µl is used in the present study.

C.iv. Calcium concentration:

Bicarbonate synthesized by the bacterial cell is generally excreted out of the cell, where it combines with the calcium available in the environment to precipitate (microbially induced calcium carbonate). So, the calcium involved in this mechanism is supplied either by the medium or may result from the support material to which the bacterium is attached. This ensures the fixation of the excess toxic calcium in the environment, enabling the bacteria survival in the unfavorable condition. The calcium concentration range of 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 and 5.5g/l is used in the present study.

C.v. Nitrogen concentration:

Bacteria needs nitrogen source for their growth along with other supplements. This nitrogen source may be organic or inorganic. The usage of organic nitrogen source is easy for bacteria. The organic nitrogen source varies from 4, 6, 8, 10, 12, 14 and 16g/l and inorganic nitrogen source varies from 20, 22, 24, 26, 28, 30 and 32g/l in the present study.

D. MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS):

Multivariate Adaptive Regression Splines is a flexible nonparametric regression modeling. The main aim of MARS model is to predict relationship between dependent variables and independent variables. MARS develops the multivariate piecewise linear regression. MARS is widely accepted for following reason (i) MARS is capable of modeling complex non-linear relationship among variables without strong model assumptions, (ii) MARS can capture the relative importance of independent variables to the dependent variable when many potential independent variables are considered and (iii) MARS does not need long training process and hence can save lots of model building time, especially when the dataset is huge^[6]. For predicting MARS model required training data, including input variable, targeted output variables. In this method total data set is divided into several splines in equal intervals. Each spline is further divided into many subgroups and several knots. The knots are located between several input variables or same input variables at different intervals. The MARS model approximates a regression function, called a basis function (BF), using smoothing splines to generally represent the data in each subgroup^[2]. Between any two knots, the model can characterize the data either globally or by using linear regression. The BF is unique between any two knots, and is shifted to another BF at each knot^[2]. The two BFs in two adjacent domains of data intersect at the knot to make model outputs continuous^[2]. The aim of the MARS model is to develop relationship between dependent variables and

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

independent variables. The MARS form the bended regression line from spline to spline and from subgroup to subgroup to fit for data. In general, MARS function can be represented by using the following equation^[2].

$$y = \hat{f}(x) = a_0 + \sum_{m=1}^M a_m B_m^{(q)}(x) \quad (1)$$

where,

a_0 = coefficient of the constant basis function, or the constant term

$\{a_m\}_1^M$ = vector of coefficients of the non-constant basis functions, $m = 1, 2, \dots, M$

$B_m^{(q)}$ are the basis functions that are selected for inclusion in the model of q^{th} order

$$B_m^{(q)}(x) = \prod_{k=1}^{k_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+^q \quad (2)$$

Where,

$B_m^{(q)}(x)$ = vector of non-constant (truncated) basis functions, or the tensor product spline basis

m = number of non-constant basis functions ($1, 2 \dots M$)

q = the power to which the spline is raised in order to control the degree of smoothness of the resultant function estimate, which in this case is equal to 1

'+' = denotes that only positive results of the right-hand side of the equation are considered; otherwise, the functions evaluate to 0.

s_{km} = indicates the (left/right) sense of truncation, which assumes only 2 values (± 1), representing the standard basis function and its mirror image. For s_{km} equal to +1, the basis function will have a value $x-t$ if $x>t$ and 0 if $x \leq t$. If it is -1, the basis function will have a value $t-x$ when $x<t$, while 0 if $x \geq t$

$x_{v(k,m)}$ = value of the predictor

$v(k,m)$ = label of the predictor ($1 \leq v(k,m) \leq n$)

n = number of predictors

t_{km} = "knot" location on the corresponding predictor space or region, or value that defines an inflection point along the range of the predictor

K = maximum level or order of interaction, or the number of factors, in the m^{th} basis function ($1, 2, \dots, K_m$).

Basis function is a function that represents information contained in one or more variables. The relationship between predictor variables and dependent variables are re-expressed by basis function. MARS forward and backward algorithm. In forward algorithm many basic function are required for better performance which gives over-fitting of basis function. In backward algorithm over-fitting basis function are deleted by generalized cross-validation (GCV). In present study backward algorithm is used in which over-fitting of function is eliminated. The Craven and Wahba proposed the Generalized Cross Validation as

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N \left[y_i - \hat{f}_M(x_i) \right]^2}{\left[1 - \frac{C(M)}{N} \right]^2} \quad (3)$$

Where,

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

N = Number of data

C(M) = Complexity cost function that varies based on basis function

M = Basis function

Friedman and Silverman suggested using equation as a lack-of-fit criterion but with an increased cost complexity function $\tilde{C}(M)$ to reflect the additional (basis Function) parameters that, along with the expansion coefficients (a_0, \dots, a_M), are being fit to the data. Such a cost complexity function can be expressed as:

$$\tilde{C}(M) = C(M) + d.M \quad (4)$$

Where,

C(M) = number of parameters being fitted

M = number of non-constant basis functions in the model, which is proportional to the number of (non-linear) basis function parameters

d = cost for each basis function optimization.

IV. RESULT AND DISCUSSION

MARS model is developed to predict the time required for calcium carbonate crystal production by various bacteria. The MARS model is developed using MATLAB software. The input vector are in different unit so normalization is done to have input data in same units^[6]. Below equation is used for the linear normalization of the data values between 0 and 1.

$$x_i^n = \frac{x_i^a - x_i^{\min}}{x_i^{\max} - x_i^{\min}} \quad (5)$$

where,

x_i^a and x_i^n - the i^{th} components of the input vector before and after normalization

x_i^{\max} and x_i^{\min} - the maximum and minimum values of all the components of the input vector before the normalization.

The value of coefficient of correlation (R) is determined by using the following formula

$$R = \frac{\sum_{i=1}^n (E_{ai} - \bar{E}_a)(E_{pi} - \bar{E}_p)}{\sqrt{\sum_{i=1}^n (E_{ai} - \bar{E}_a)^2} \sqrt{\sum_{i=1}^n (E_{pi} - \bar{E}_p)^2}} \quad (6)$$

where,

E_{ai} and E_{pi} - the actual and predicted values

\bar{E}_a and \bar{E}_p - mean of actual and predicted E values corresponding to n patterns.

The output vector obtained from the MARS model is a normalized data and hence, the normalized data is reverted to its actual value by using equation

$$x_i^a = x_i^n (x_i^{\max} - x_i^{\min}) + x_i^{\min} \quad (7)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

where,

x_i^n - the normalized result obtained after the test for the i th component.

x_i^a - the actual result obtained for i^{th} component

x_i^{max} and x_i^{min} - the maximum and minimum values of all the components of the corresponding input vector before the normalization.

The calcium carbonate crystals production by various bacteria are checked by varying the different parameters. From the experimental result, the optimized parameter values for crystal production are evaluated. The optimized values are shown in table 1.

Parameters	Optimum value
Temperature	30°C
pH	8
Bacterial concentration	1000 μ l
Calcium concentration	5.5g/l
Organic nitrogen source	4g/l
Inorganic nitrogen source	20g/l

Table 1. Optimized parameters

The time required for calcium carbonate crystal production for various bacteria are experimentally evaluated. By using MARS model the time required for calcium carbonate crystal production is predicated by MATLAB. The developed MARS model gives the coefficient of determination(R^2) value as 0.94, the predicted result are in good agreement with actual value. This is verified by using 30% of test data which had R^2 value is 0.95. This type of MARS model can be used for any type of work in future. Fig 1 shows the experimental value VS predicated MARS output for time required for crystal production.

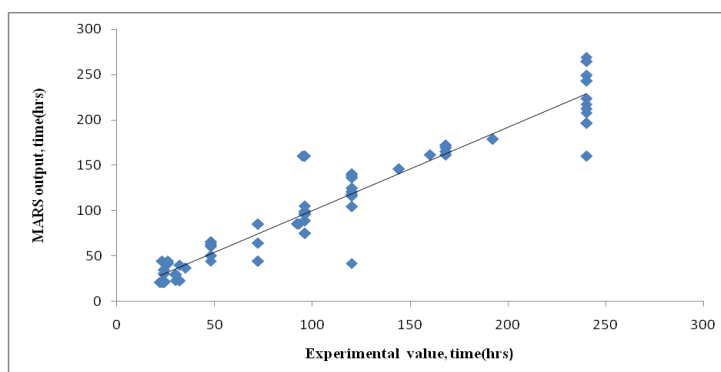


Fig 1: Experimental VS Predicated time value (hrs)

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

V. CONCLUSION

Biom mineralization based MARS model are developed to evaluate the time required for maximum quantity of calcium carbonate crystal production, by using various bacteria varied with five intrinsic and extrinsic factors. An overview of temperature, pH, bacterial concentration, nitrogen source, calcium concentration are described. MARS gives the relationship in non linear form between the dependent and the independent variables besides identifies the interaction among independent variables. MARS model is developed using MATLAB software for prediction of time required for calcium carbonate crystals formation. MARS is trained with about 70% of total data set and tested with about 30% of total data set. The evaluated result from the MARS model are in good agreement with the experimental value. The developed model can be used for predicting the time required for calcium carbonate crystal quantity in bio-mineralization process.

ACKNOWLEDGMENT

The Authors are thankful to the Director, CSIR – Structural Engineering Research Centre, Chennai for allowing us to carry out this work. The help and cooperation received from scientists and technical staffs of Computational structural mechanic group – Biotechnology laboratory are deeply acknowledged. This paper is published with the kind permission of the of the director, CSIR-SERC, Chennai.

REFERENCES

- [1] C. Barabesi, A. Galizzi, G. Mastromei, M. Rossi, E. Tamburini and B. Perito, Journal of Bacteriology, vol.189, pp.228–235, 2007.
- [2] J H. Friedman, " Multivariate adaptive regression splines", The Annals of Statistics, vol.19, no.1, pp.1-67, 1991.
- [3] A. Knoll, "Reviews in Mineralogy and Geochemistry", vol.54, pp.329–356, 2003.
- [4] M A. Rivadeneyra, J. Parraga, R. Delgado, A. Ramos-Cormenzana and G. Delgado, FEMS Microbiology Ecology, vol.48, pp.39–46, 2004.
- [5] C. Rodriguez-Navarro, M. Rodriguez-Gallego, K B. Chekroun and M T. Gonzalez-Munoz, Applied and Environmental Microbiology, vol. 69, pp.2182–2193, 2003.
- [6] K. Sarayu, N R. Iyer and A. Ramachandra Murthy, "Exploration on the Biotechnological Aspect of the Ureolytic Bacteria for the Production of the Cementitious Materials - a Review", December 2013.
- [7] P. Yuvaraj, A. Ramachandra Murthy, N R Iyer and Pijush Samui & S K Sekar, " Multivariate Adaptive Regression Splines to predict fracture characteristics of high strength and ultra high strength concrete beams", CMC , vol.3, no.1, pp.73-97, 2013.