Chemometrics and Intelligent Laboratory Systems 157 (2016) 133–139

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Model development and surface analysis of a bio-chemical process

Dazhi Jiang [a], Wan-Huan Zhou [b], Ankit Garg [c], Akhil Garg [d,*]

[a] Department of Computer Science, Shantou University, Shantou 515063, China
[b] Department of Civil and Environmental Engineering, Faculty of Science and Technology, University of Macau, Macau, China
[c] Department of Civil and Environmental Engineering, Indian Institute of Technology, Guwahati 781001, India
[d] Department of Mechatronic Engineering, Shantou University, Shantou 515063, China

ARTICLE INFO

Article history:
Received 29 April 2016
Received in revised form 28 June 2016
Accepted 18 July 2016
Available online 20 July 2016

Keywords:
Phytoremediation
Lead removal
Statistical modelling
Biochemical
Genetic programming
Cross-validation

ABSTRACT

Phytoremediation, is a promising biochemical process which has gained wide acceptance in remediating the contaminants from the soil. Phytoremediation process comprises of biochemical mechanisms such as adsorption, transport, accumulation and translocation. State-of-the-art modelling methods used for studying this process in soil are limited to the traditional ones. These methods rely on the assumptions of the model structure and induce ambiguity in its predictive ability. In this context, the Artificial Intelligence approach of Genetic programming (GP) can be applied. However, its performance depends heavily on the architect (objective functions, parameter settings and complexity measures) chosen. Therefore, this present work proposes a comprehensive study comprising of the experimental and numerical one. Firstly, the lead removal efficiency (%) from the phytoremediation process based on the number of planted spinach, sampling time, root and shoot accumulation of the soil is measured. The numerical modelling procedure comprising of the two architects of GP investigates the role of the two objective functions (SRM and AIC) having two complexity measures: number of nodes and order of polynomial in modelling this process. The performance comparison analysis of the proposed models is conducted based on the three error metrics (RMSE, MAPE and R) and cross-validation. The findings reported that the models formed from GP architect using SRM objective function and order of polynomial as complexity measure performs better with lower size and higher generalization ability than those of AIC based GP models. 2-D and 3-D surface analysis on the selected GP architect suggests that the shoot accumulation influences (non-linearly) the lead removal efficiency the most followed by the number of planted spinach, the root accumulation and the sampling time. The present work will be useful for the experts to accurately determine lead removal efficiency based on the explicit GP model, thus saving the waste of input resources.

1. Introduction

Soil contamination is a problem of national concern that is responsible for degradation of human health and environment [1]. The contamination has rapidly increased in last few decades as a result of waste and wastewater discharged from anthropogenic sources [2]. The methods such as the ion exchange, precipitation, reverse osmosis and evaporation (physio-chemical methods) can be applied for decontaminating however, these methods require a lot of resources and thus expensive to implement. However, phytoremediation, a technique involving biochemical mechanisms (adsorption, transport, accumulation and translocation; Fig. 1 [3]) has gained wide acceptance in remediating the contaminants from the soil using vegetation.

It has advantage of being economical and also environment friendly. Phytoremediation works on principle of natural processes occurring in plant and therefore it is not dependent on any external resources and is easy and reasonably inexpensive to implement. For the phytoremediation on metal contaminated soils, the quantification of the relationship between the metal-tolerant plant species and chemical properties of soil is vital.

The use of spinach for phytoremediation of metal-contaminated soils has been reported in previous studies [4–6]. Experimentally it has been found that, the removal efficiency of heavy metals varies with the metal ion concentration and plant density. There is a risk in this method that, with this removal, there is a certain amount of accumulation of heavy metals in root as well as shoot of plants [7,8]. Any relationship between removal efficiency and accumulation in roots, shoots along with plant density and concentration will be of prime interest in evaluating the risk and design of bio-remediation measures for contaminated soils [8]. In this context, the quantification of the relationships based on the statistical response surface methodology (RSM) can be applied. However, the mechanism of the formulation of models using RSM methodology is based on the prior assumptions such as the form of the model, residual and data correlation assumptions, etc. These models are built on the training data and not tested for the testing data beyond the input range in a realistic condition. Therefore, this induces the ambiguity in prediction ability of the model on the testing

http://dx.doi.org/10.1016/j.chemolab.2016.07.010
0169-7439/

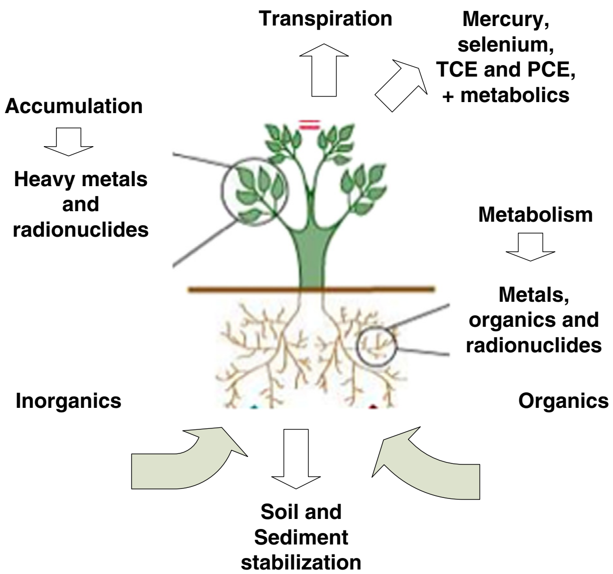

Fig. 1. Bio-chemical processes in during phytoremediation.

samples. Alternatively, the artificial intelligence (AI) approach of genetic programming (GP) in formulating the decision support models can be applied. This AI approach works on the optimization and genetic algorithm principle and its mechanism supports to evolve the mathematical models [9–13] explicitly from the given data. The decision support model can also suggest the precise selection of shoot and root properties for the maximizing the lead removal efficiency. Past quantitative studies [14–16] involving the applications of GP in modelling of systems have reported that the performance of the GP models depends the architect (objective function, parameter settings and complexity measure) selected.

Therefore, the present work will explore the ability of the artificial intelligence approach of genetic programming (GP) based on the two new architects in formulating the decision support models for % removal efficiency of lead. The two new architects of GP are defined by the two new objective functions. One objective function to be investigated is the structural risk minimization principle (SRM) while the other is Akaike information criterion (AIC). The complexity measures based on the minimum order of the polynomial and the number of nodes will be used in the penalty term of these two objective functions. The procedure involving the experimentation planning and the modelling procedure of the % removal efficiency with respect to the four inputs is shown in Fig. 2. The % removal efficiency of lead is summarized statistically from laboratory experiments [8]. The rest of the procedure involves the settings of the architects of GP, models formulation, models analysis and validation and the surface analysis to find any physical interpretation from it.

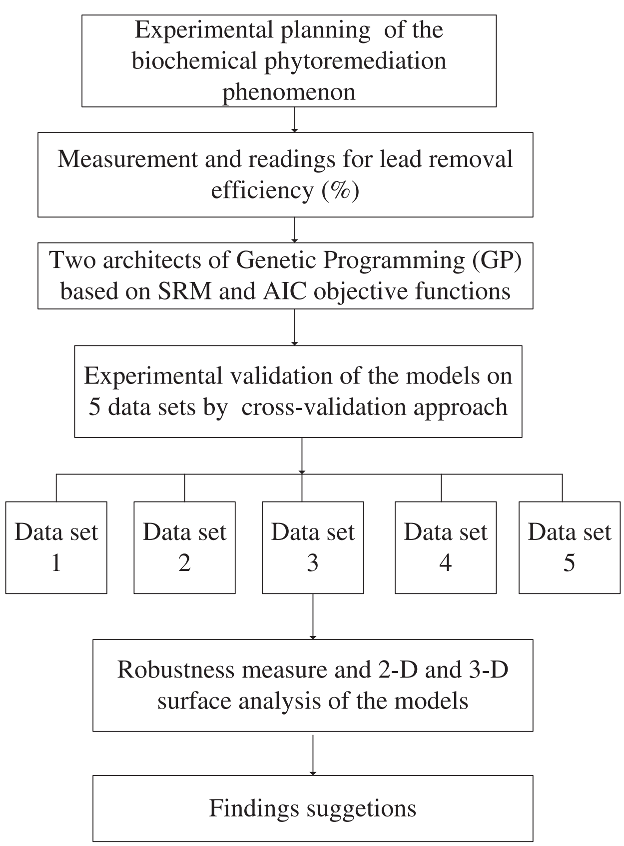

Fig. 2. Procedure of quantification of % removal efficiency of lead as function of % root accumulation, % shoot accumulation, no. of plants and sampling time.

Table 1
Inputs considered for the phytoremediation process for lead removal.

Inputs for the phytoremediation	Minimum	Maximum	Mean
Sampling time (days)	7	14	21
Number of planted spinach	2	6	4
Root accumulation (%)	1.28	2.55	1.81
Shoot accumulation (%)	1.26	2.64	1.84

2. Phytoremediation chemical process for measuring % removal efficiency of lead

The complete set-up including the experimental procedure for evaluating % removal efficiency of lead and other plant properties are kept the same as those mentioned in the work by [8]. Soil selected in their study was mainly sand (59%) and considerable clay content (11%). It has organic content of 0.72%. Spinach (*Spinacia oleracea*) was selected as plant for conducting phytoremediation in contaminated soil with lead. Spinach is native to central and southwestern Asia and can grow to a height of up to 30 cm. Table 1 shows the set of inputs used in the Phytoremediation experiment. The plants were grown in pots that contain 7 kg of lead-contaminated soil. Soils in pots were subjected to

Table 2
Experimental data comprising of % removal efficiency of lead and four inputs (x1, x2, x3 and x4) for soils vegetated with spinach.

No.	Y (% removal efficiency)	x1 (sampling time; days)	x2 (no. of planted spinach)	x3 (% root accumulation)	x4 (% shoot accumulation)
1	39.28	7	2	2.476	2.252
2	49.18	14	4	1.764	1.792
3	47.28	7	6	2.096	2.648
4	29.32	21	2	1.614667	1.668
5	49.16	14	4	2.54	2.54
6	44.92	21	6	1.289333	1.265333
7	53.18	21	4	1.678	1.714
8	49.7	14	4	1.728	1.782
.	57.3	14	6	1.778	1.786
.	48.16	14	4	1.642	1.626
.	59.6	21	6	2.276	2.292
.	32.8	14	4	1.601333	1.594667
.	48.62	14	4	1.596	1.628
.	45.2	21	2	2.556	2.496
.	42.28	7	6	1.457333	1.494667
36	25.06	7	2	1.457333	1.489333
37	48.62	14	4	1.73	1.748
38	38.16	14	2	1.674	1.688
39	49.12	14	4	1.744	1.718
40	36.22	7	4	1.632	1.648