Original Research Paper

Volume - 14 | Issue - 02 | February - 2024 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Life Science

STATISTICAL EXPERIMENTAL DESIGNS FOR BIOPROCESS OPTIMIZATION: A FOUNDATION TO ARTIFICIAL INTELLIGENCE

Wahid Mohd	Research and Scientific Studies Unit, College of Nursing & Allied Health Sciences, Jazan University, Jazan-45142, Saudi Arabia.
Mehta Anamika	Department of Life Sciences, Sharda University, Greater Noida, U.P. (India)
Soni Nipunjot	Department of Biotechnology, General Shivdev Singh Diwan Gurbachan Singh Khalsa College, Patiala.
Chaudhary Payal	Department of Microbiology, University of Burdwan, Golapbag, Burdwan - 713104, West Bengal, India.
Jawed Arshad*	Research and Scientific Studies Unit, College of Nursing & Allied Health Sciences, Jazan University, Jazan-45142, Saudi Arabia. *Corresponding Author

ABSTRACT Amidst the growing market competition, technologies and processes are short-lived. Even the biotechnological processes that take years to develop are quickly superseded by more improved and efficient ones. Therefore, the rapid development of new technologies, Innovation and quicker optimization has become the need of the hour. Optimization of culture media is normally carried out by the classical One-Factor-At-a-Time (OFAT) approach. OFAT is generally used by biotechnologists at their initial research due to its simplicity and requires only basic knowledge of mathematics. On the other hand, it is uneconomical, needs a large number of runs to be carried out, leads to pseudo-optimal results, and takes a long time for optimization. Contrary to this approach, statistical and mathematical, e.g. Plackett–Burman, Response Surface Methodology (RSM), D-optimal Designs, Central Composite Designs (CCD), machine learning etc., prove to be quicker, economical, require less number of experiments and provide real-optima for the desired results, be it maximizing the protein expression, purification, or minimizing the by-products/ toxins, etc. Despite the huge potential, the adaptation of statistical methods for process optimizations is slow and sluggish. In this article, we compare and explain the basics of classical and modern process optimization methodologies and provide a simpler way to quickly adopt a machine learning methodology for scientists, who are averse to the complicated mathematics involved. Statistical Design (SDs) builds the foundation of Artificial Intelligence (AI). AI is generally used to fine tune the results of SDs with or without gaps or incompleteness in the data sets or nonlinear programs are very useful in the field of medium optimization. This study would serve the upcoming researchers to plan and devise statistical optimization strategies quickly and efficiently with high reproducibility.

KEYWORDS : Screening, Optimization, OFAT, RSM, CCD, ANOVA

1. INTRODUCTION

Screening and optimization of a biotechnological process are integral to picking the right drug candidate, selecting optimum conditions for higher metabolite/recombinant protein productivity, etc. Any fermentation process aims to maximize biomass or metabolite production per unit volume. In today's competitive world, where developments are rapid and the technology quickly gets obsolete, time and productivity are crucial to the success of any biotechnological operation. The productivity of intra-cellularly expressed recombinant protein depends on the cell population and per unit cellular protein expression (Jawed A, 2008). Cell growth and product formation are generally considered to have a stoichiometric relation, where the carbon or the primary energy source, the nitrogen source, trace metals and minerals along with oxygen from the cell culture medium are transformed into biomass, products/by-products, etc. Volumetric productivity, called q, refers to the amount of metabolite or desired product produced per unit volume per hour. It can be given by

$$q\left(\frac{g\ product}{l * h}\right) = \frac{dC_{product}}{dt}$$

where

q: Specific rate of product formation (g product liter-1 h-1)

dC = change in the product concentration

dt = change in time

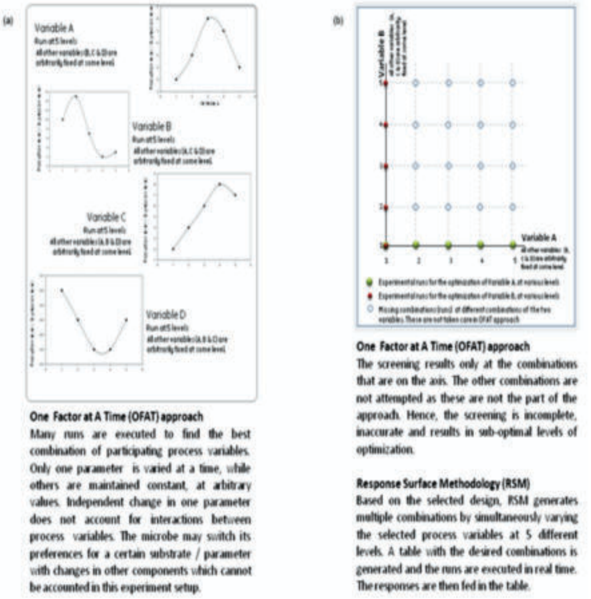
As the concentration of cells increases due to continuous growth, productivity can be best assessed after the process concludes.

Chemically defined culture media requires the microbe to synthesize every cellular component along with primary and secondary metabolites from simple, chemically defined substrates. Therefore, culture medium formulation initially focuses on boosting cell growth by providing balanced nutrients in sync with the cellular requirements, cell physiology and required cell density of the microorganism of interest. Once the components are optimized, complex components are added to fulfill the requirement of the culture for vitamins and other trace components. Previously published/used media for the related microbes can be used as a base to optimize the process under consideration. The inclusion of complex nutrients, such as tryptone, yeast extract, peptone, and casein hydrolysate, etc., help microbes to grow, improve the quality and amount of the protein expressed, by providing readily available substrates, biosynthetic precursors, vitamins, etc. (Kishimoto and Suzuki. 1995). There is a lot of information, in literature, advocating a significant increase in yields (approximately ten folds), of recombinant proteins just by the optimization of the medium composition. Some authors have attributed the increase in the expression level of recombinant protein with the addition of organic (undefined) nitrogen source(s) to reduce the burden on the cell due to the availability of biosynthetic precursors (Zabriskie et al. 1987; Kweon et al. 2001). Others have termed the increase in the expression of recombinant protein due to the suppression of protease activity by the amino acids present in these complex nutrients (Mizutani et al. 1986; Tsai et al. 1987; George et al. 1992). Besides this, complex nitrogen sources have also been reported to increase plasmid stability by Matsui et al. 1990 and improve the copy number of the plasmid (Shin et al. 1997). At the same time, the inhibition of protein synthesis beyond a certain concentration of complex nitrogen sources is also reported (Rinas et al. 1989; Matsui et al. 1990; Li et al. 1998). If cell growth along with the synthesis of the metabolite/required product is attained, the next step normally is the optimization of medium components for maximizing cell growth and desired metabolite production. There exist two main categories of culture medium/process optimization techniques, currently used in different parts of the world. The classical method or OFAT, that is One – Factor – At – a – Time and advanced machine learning optimization techniques. The successful formulation of an optimized medium is easy but its slow, time-consuming when optimization is approached using the OFAT approach. A step ahead of the classical method is employing the statistically designed experiments that can quicken the optimization process. Several methodologies have been employed in the past for the formulation of a complex medium. It enlists Plackett – Burman Design (1946), Response Surface Methodology (Box and Wilson, 1951), Expert System Approach (Kishimoto and Suzuki. 1995), etc. Recently, the application of machine learning techniques makes new dimensions in the field of medium optimization. These methodologies demonstrate great potential to be employed for the successful development of the fermentation medium. Traditional cell culture/fermentation medium optimization is a search for the optima by changing process parameters step by step and quantifying cell growth or metabolite production at every step. Modern medium optimization approaches break this routine and search for global optima by simultaneously varying multiple components. Although the machine learning approach (AI) is more effective than the classical or statistical one, yet the application of machine learning finds its foundation rooted in the data obtained through classical or statistical designs. Therefore, irrespective of the SDs or AI methods used, the optimization process should be executed systematically, which requires the framework of overall process design and requirements.

In the present article, we discussed the commonly used techniques apply for establishing the appropriate combination or concentration of nutrients and that would support the desired cell growth and/or synthesis of a microbial product(s). However, optimization should be attempted within the context of the overall process requirements.

1.1 The classical method of medium optimization

Traditionally, medium optimization with defined components is performed in a series of shake-flask experiments termed as One-Factor at a Time (OFAT). In OFAT experiments, a single participating factor is varied sequentially, while all others are fixed arbitrarily at their respective levels. The success and validity of this design approach relies on the expertise and knowledge of the researcher. The experiment(s) therefore, are sequentially repeated and continued with selected variable(s) until the maximum cell growth or protein expression/production is achieved. This factor is now fixed at its 'best' level, and another factor is varied until all variables are screened. The resultant shift in response (cell concentration/cell growth rate and/or product yield) is contrasted to that of the earlier experiments and the response is plotted against each variable. Figure 1 presents a quick summary of the general optimization process.



One Factor at A Time (OFAT) approach

Many runs are executed to find the best combination of participating process variables. Only one parameter is varied at a time, while others are maintained constant, at arbitrary values. Independent change in one parameter does not account for interactions between process variables. The microbe may switch its preferences for a certain substrate / parameter with changes in other components which cannot be accounted in this experiment setup.

One Factor at A Time (OFAT) approach

The screening results only at the combinations that are on the axis. The other combinations are not attempted as these are not the part of the approach. Hence, the screening is incomplete, inaccurate and results in sub-optimal levels of optimization.

Response Surface Methodology (RSM)

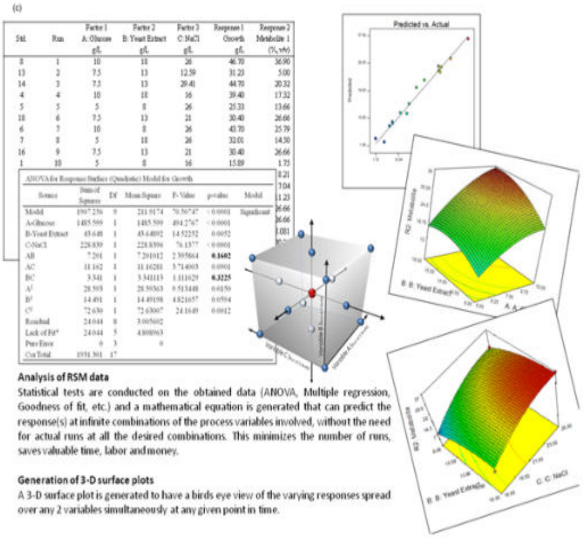
Based on the selected design, RSM generates multiple combinations by simultaneously varying the selected process variables at 5 different levels. A table with the desired combinations is generated and the runs are executed in real time. The responses are then fed in the table.

As it is evident from Figure 1(a,b), numerous combinations of variables are possible, but with the OFAT approach, not all the combinations are attempted. The optimum results predicted through this approach, are not mostly inaccurate as the data generation is fragmented with discrete experiments that are mutually exclusive and have provided no account of medium component interactions, whatsoever. This form of experimentation can be regarded as trial and error which requires luck, experience, skill and intuition for its success. By employing this methodology, Cocaign-Bousquet et al., (1995) revealed, the exact microbial/cellular nutritional requirements, simplifying the synthetic medium formulation for the prolonged growth of *L. lactis*. Several carbon and nitrogen sources along with amino acids were employed in step-by-step experimentation for producing a cytotoxic metabolite from *Stachybotrys chartarum called* verrucarcin (El Kady and Moubasher, 1982). Fewer successful single factor optimization methods have been reported in the past too (Chary et al. 1989; Monot et al. 1982).

The fermentation or the culture medium is not a simple formulation, where the effect of every component can be isolated from the other. There is a considerable interaction of each component with the rest. Microbes too behave in a very complicated way, concerning the uptake and utilization of medium components. A medium component may be preferentially utilized over the others; uptake of a nutrient can change depending on competing components, a constituent might suppress the effect of by-products released by the organism, diminishing concentration of one component may switch the preference of the culture for another component. Classical OFAT methods do not account for such kind of interaction of medium components and do not accurately show the absolute effect of one component on growth or other parameters under consideration. OFAT is, therefore, inefficient, unreliable, results in pseudo-optimal conditions and relies on intuition and experience for its success. Additionally, OFAT methodologies are simpler and obvious but these consume a lot of time and resources (Singh et al 2017).

1.2 Statistical method of medium optimization

An alternative optimization strategy that is slowly becoming popular in academia, as well as industry, is the application of statistical designs and optimization experiments that allows the researcher to estimate the effect of one or more than one independent variable simultaneously (Greasham and Inamine. 1986; Greasham and Herber. 1997). The strategies used for such optimization are collectively called the Design of Experiments (DOE). The approach is generally hierarchial and begins with primary screening to differentiate those variables (normally three to five or sometimes more depending on the culture requirement) that have a significant effect on the anticipated response from a pool of relevant variables (≥ five) with a minimum of testing (Figure 1©).



Std.	Run	Factor 1 A: Glucose g/L	Factor 2 B: Yeast Extract g/L	Factor 3 C: NaCl g/L	Response 1 Growth g/L	Response 2 Metabolite 1 (%, w/v)
8	1	10	18	26	46.70	36.90
13	2	7.5	13	12.59	31.23	5.00
14	3	7.5	13	29.41	44.70	20.32
4	4	10	18	16	39.40	17.32
5	5	5	8	26	25.33	13.66
18	6	7.5	13	21	30.40	26.66
6	7	10	8	26	43.70	25.79
7	8	5	18	26	32.01	14.50
16	9	7.5	13	21	30.40	26.66
1	10	5	8	16	15.89	1.75
						8.21
						7.04
						11.23
						26.66
						26.66
						1.081

ANOVA for Response Surface (Quadratic) Model for Growth

Source	Sum of Squares	Df	Mean Square	F-Value	p-value	Model
Model	1907.256	9	211.9174	70.50747	< 0.0001	Significant
A-Glucose	1485.599	1	1485.599	494.2767	< 0.0001	
B-Yeast Extract	43.648	1	43.64892	14.52252	0.0052	
C-NaCl	228.839	1	228.8394	76.1377	< 0.0001	
AB	7.201	1	7.201012	2.395864	**0.1602**	
AC	11.162	1	11.16281	3.714003	0.0901	
BC	3.341	1	3.341113	1.111629	**0.3225**	
A^2	28.593	1	28.59363	9.513448	0.0150	
B^2	14.491	1	14.49198	4.821657	0.0594	
C^2	72.630	1	72.63007	24.1649	0.0012	
Residual	24.044	8	3.005602			
Lack of Fit*	24.044	5	4.808863			
Pure Error	0	3	0			
Cor Total	1931.301	17				

Analysis of RSM data

Statistical tests are conducted on the obtained data (ANOVA, Multiple regression, Goodness of fit, etc.) and a mathematical equation is generated that can predict the response(s) at infinite combinations of the process variables involved, without the need for actual runs at all the desired combinations. This minimizes the number of runs, saves valuable time, labor and money.

Generation of 3-D surface plots

A 3-D surface plot is generated to have a birds eye view of the varying responses spread over any 2 variables simultaneously at any given point in time.

Once the significant components are deciphered, as the next step, Full-factorial, Fractional-factorial, Central Composite Design (CCD), Response Surface Methodology (RSM), Plackett-Burman method, Taguchi design, etc., are employed with the selected components (Roseiro et al. 1992; Plackett and Burman 1946), for detailed parameter selection, process optimization and robustness testing. Optimization of cell culture medium with defined components by Plackett-Burman design has been reported by many workers (Metzger et al. 1984; Singh & Tripathi 2008; McIntyre et al. 1996; Kisaalita et al. 1993). Plackett-Burman method is too simplistic to be executed in the current scenario for complete optimization. On the other hand, a full factorial search generates a huge number of experiments, that defeats the benefits of statistical optimization. RSM is productive only when a few crucial variables (between two or three) are to be examined (Garcia-Ochoa et al. 1992). The target is to usually find the combinatorial levels of these process factors that support the best suitable response in a time-dependent manner, employing RSM. There are many publications available that employ the successful use of RSM have been described (Zhang et al. 1996). The fermentation process involving *C. bombicola*, for the production of an important surfactant sophorolipid, Casas et al. (1997) initially used a 4-factor, 2-level factorial design, subsequently a 3-factor, 3-level response-surface design to improve and formulate an optimized synthetic medium. The curvature-effect analysis inferred that Mg^{2+}in the culture medium had no significant effect while nitrogen and phosphorus sources showed their presence was crucial for the culture growth. Other workers have used RSM for biosorption of chromium (Margarita et al. 2005), ethyl butyrate production by lipases (Jose et al., 2005) Another easy-to-use and quicker optimization methodology is the Sequential Simplex methodology (Leggett et al., 1983; Spendley et al., 1962), the optimal combination is approached, via series of sequential steps toward better and optimal results, where the maximum value can be obtained with a nominal number of steps. The methodology uses Pattern-Seeking Approach (PSA). PSA is non-statistical but it gives a fair idea for identification of the process optima rapidly. DOE is increasingly being used in the industry, but still, many scientists and researchers resort to the use of OFAT techniques for process or product development. DOE employs simple/complex statistical methods to arrive at the real optima, taking all the parameters into considerations. The results generated are sequentially predicted via a mathematical model, making the process robust and reproducible. In a DOE setup, a large number of process variables can be optimized using fewer experiments which reduces the experimental waste generation, saves time and money. DOE uses mathematical equations to select the factors that significantly affect the results. The process variables are altered simultaneously so that the effect of one variable on other variables can be accounted and studied. Once identified, their effect on results (growth/protein production, etc.) is quantified and tested for significance. Using ANOVA and multiple regression analysis DOE is capable of predicting the most important variable for increasing/ decreasing the metabolite production or toxic by-products generation respectively. The analysis can easily be done using various software, such as Statistica®, Design-Expert®, MATLAB, etc. Table 1 lists in brief the various methodologies and their details. The objective of this article is to encourage young researchers to use statistical techniques for their experimental work. We found Design-Expert™ Software from Statease Inc. easy to use and user-friendly for researchers with little knowledge of Mathematics and statistics, compared to the other software packages.

1.3 Machine learning approach to medium optimization

Artificial Intelligence (AI) techniques, such as Expectation-maximization, Deep Learning/Artificial Neural Network, Genetic Algorithms, Simulated Annealing are some of the machine learning tools applied in the various optimization studies. An artificial neural network is one of the most used tools of the machine learning approach. It is a self-learning approach which is generally used for the prediction, and to find out the optimal solution quickly. Zhang et al., (2020) used the hybrid of artificial neural networks and a genetic algorithm for the optimization of the four basic conditions (agar concentration, light time, culture temperature, and humidity) of plant tissue culture using a three-layer neural network to predict the differentiation rate of melon. Badhwar et al., (2020) compared the non-linear hybrid mathematical tools GA–ANN and GA–ANFIS for improved pullulan production from Bapat PM, Wangikar *Aureobasidium pullulans* and process optimization (substrate concentration, incubation period, temperature, pH and agitation speed). Bapat and Wangikar (2004) used machine.learning.based approach such as Genetic algorithm (GA), Neighborhood analysis (NA) and Decision Tree technique (DT) for the optimization of rifamycin B from *Amycolatopsis mediterranei* S699.

2. Methodology

Statistical optimization methods are approached in a well-defined way, moving from simple to complex models. The methodology begins with simple designs meant for screening and moves further to optimization and robustness testing/ Model Validation. To understand the design method more let's take an example. For screening design, let's consider 5 factors, e.g., Glucose concentration (A), Yeast extract concentration (B), NaCl (C), Temperature (D) and pH (E) as independent parameters that affect cell growth (Response 1) and metabolite production (Response 2). Starting with Plackett-Burman design, all the parameters are varied at 2-levels (a higher and a lower level). The design mainly estimates the main effects of participating components. This design can be used to screen a few significant components from a larger array. It is assumed that in the real scenario methodology, only a few factors are significant that affect the response. The general 2-level factorial screening design is shown in the form of Table 2a. Each parameter is represented for its units, type and the lower and higher experimental boundaries. The Experimental table generated by Design-Expert software is shown in Table 2b. Each column indicates the experimental parameter with its different levels, usually +1 and -1. The notation "+" indicates as 'High' level and "-" indicates a 'Low' level of the selected experimental range. The total number of runs is calculated as level [no. of factors (n) – 1]. For 5 factors, varied at 2 levels, the total number of runs will be $2^{5-1} = 16$. Looking into the experimental details in Table 2b, only 16 runs are enough to perform the complete set. The total number of runs increases as the parameters are varied at 3 or 5 levels or we can say as the number of variables increases. Each run is executed and the results; i.e. cell growth in this example is fed in the corresponding row. Based on these results, the Analysis of Variance (ANOVA) is carried out and the significance of each variable against the response is calculated. Based on the significance (F-value and p-values), process variables showing the maximum effect on the response are selected for further optimization steps. Screening designs are beneficial as these balances the cost of experimentation and the information obtained from the experiments. Screening runs can be considered as a prelude to further experimentation, mainly for detailed optimization (RSM) studies. Screening designs consider fewer experimental points and therefore can be performed quickly. Screening designs can be upgraded to multi-level factorial designs for deeper insights and analysis of the obtained responses.

Other statistical designs which can be used in place of PBD are Taguchi orthogonal array methods, Taguchi designs are generally known as orthogonal arrays, which consist of a set of fractional factorial designs ignoring the interaction terms and concentrating only on the estimation of main effects. Taguchi uses the orthogonal arrays for arriving to optima, i.e., La(b^c),
where,
La = number of experimental runs,
b = number of levels of each factor, and
c = number of variables.

For example, L4 2^3 design consists of up to 3 factors at 2 levels each. There are 4 rows.

Statistical designs can be developed with process variables with several levels. In a biological research setup, two or three-level designs are the most common. The L18 Taguchi design widely used for its simplicity and ease. When a Taguchi design is created, the experimental levels of each variable replace the old data while storing the current one.

While 2-level factorial designs account for the main effects, multi-level factorial designs analyze the main effects as well as the interaction effects, using 3 to 5 level factorial designs. Employing Central Composite Design (CCD), under the umbrella of factorial design, fractional factorial generates a table with runs equal to $2^5 = 32$, while the full factorial generates even more runs (~50; mostly center point replicates/duplicates for higher efficiency and to assess lack of model fit due to curvature). Normally a fractional factorial set-up is sufficient for analyzing and interpreting the results in the case of any biotechnological process, though full factorial experiments may be required for Neural networks (a technique based on attached nodes similar to neurons in the living system. The technique requires training the system by repeated input and analysis. Alternatively, simple multilevel factorial designs permit us to develop predictive mathematical or statistical models that incorporate the main effects of each process variable/factor as well as their interactions. Table 3a lists the experimental design with four components for Central Composite Design (CCD) considered under RSM. Table 3b generates the experimental design with variations at 5 levels to be performed and response to be inserted against each trial run.

2.1 Designs and execution of optimization experiments

Once the significant factors are selected after screening, the next step is to optimize them simultaneously. To do this, a multi-level fractional factorial experiment is carried out. This statistical methodology is collectively called the Response Surface Methodology (RSM). Central Composite Design (CCD) is normally used for the optimization of microbiological/biotechnological unit operations under RSM. For a 3 factor CCD, varied at 5 levels; Table 3b demonstrates the number of runs required to optimize the participating factors. Once the CCD table is generated, all experiments are performed at random to calculate the response achieved per run. If the ratio of maximum to the minimum response obtained is less than 10, the data can be utilized without any transformation. In the example shown in Table 3a, the ratio of maximum to minimum responses is <10, therefore, no transformation is required. The value of the response of each run is fed in the row corresponding to each run in the design table. After all the runs are completed and responses fed, the obtained data is analyzed for the best fit model equation. The obtained responses are fitted into different regression models, e.g. Mean, Linear, 2 Factor Interaction (2FI), 3 Factor Interaction (3FI), Quadratic, etc., and the corresponding F-value / p-value are calculated. The statistical model with the highest F-value shows better response fitting and p-values below 0.05 are considered significant.

Response surface methodology (RSM)

Input and output of the experimental designs are used for the development of the model. In RSM, ANOVA is generally performed for the selected model shows Fischer's coefficient and p-values for all the participating factors and the interaction terms (Table 3 b, c). The individual and interaction terms with p-values higher than 0.05 and 0.10 respectively are removed from the model, as these don't support hierarchy (Table 3 d, e). In biological systems, we have a prior idea about the crucial components, hence the inclusion or exclusion of factors may be considered if these don't alter the 'significant' status of the overall model.

Second most important result interpretation through RSM is the curve fitting which generates a model equation that can be used to predict responses. For cell growth (Y), the constructed equation with the obtained responses can be depicted as:

$Y = 2.19 + 0.35*A + 0.30*B + 0.06*C + 0.27*D + 0.07*A*B - 0.01*A*C + 0.12*A*D + 0.14*B*D + 0.01*C*D - 0.02*A^2 - 0.10*B^2 - 0.02*C^2 - 0.13*D^2$

Similarly, metabolite production (Y') can be depicted by

$Y' = 78.9 - 3.02*A + 0.06*B + 1.40*C + 3.28*D - 11.23*A*B - 0.89*A*C + 5.97*A*D - 3.86*B*D - 0.97*C*D - 11.48*A^2 - 11.48*B^2 - 11.48*C^2 - 9.87*D^2$

where A, B, C and D represent experimental factors such as Glucose (%-w/v), Yeast extract (%-w/v), NaCl (%-w/v) and Phosphates (%-w/v) and the terms AB, AC, AD, BD and CD are interaction terms, while A^2, B^2, C^2, $D^2$are quadratic terms in the relevant equation. As the p-value for interaction term BC was found to be high, showing that its components B and C don't interact, and the interaction term BC has very less effect on the product formation. Removing it from the model makes the equation a better fit and increases the accuracy of the predicted conditions. If the ANOVA shows a model to be significant with non-significant 'lack of fit', the model can be used to generate and predict responses accurately. Third important result interpretation RSM is the predicted responses that are generated in the form of a 2D-contour plot or 3D surface plot (Figure 2a-e). These graphs can be used, to predict the responses at an infinite number of the desired combination of factors, without performing the experiment in real time. This ability of RSM helps researchers to find a real optimum, even if they did not perform the experiment at the exact same combination of factors at the desired levels. Depending on the aim of the experiment, one either attempts to maximize (cell growth, protein expression, metabolite production, recovery, etc.) or minimize a response (toxicity, inhibition, etc.). Using the model equation developed after ANOVA, one can predict the experimental conditions which lead to the desired results (maximum or minimum response). This can also be achieved by response surface graphs plotted in between any two factors while keeping the remaining others fixed at their selected levels. Experimental runs are finally executed at the predicted conditions to verify the authenticity and accuracy of the model prediction.

The solution of second-order polynomial equations

The model generated in terms of second-order polynomial equations can be solved by finding out the solution of this equation. it can be done either by mathematical analysis by doing a double regression equation or by using some statistical software like Minitab, Matlab etc.

Neural Network Model

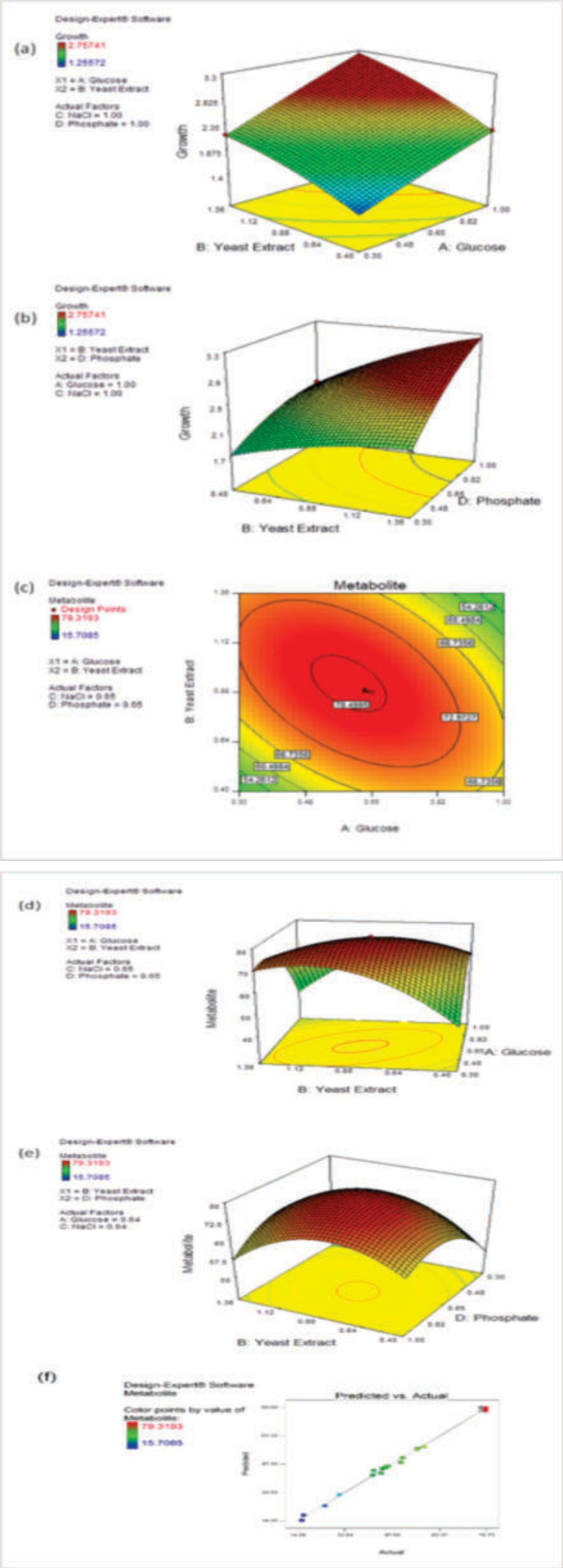
ANN model contains three layers of neurons: An initial input layer that corresponds to the input variables; a sandwiched hidden layer and a final output layer. Training and test set were prepared randomly, and a feed-forward back-propagation network is generally used to train the input variables of considered design such as CCD. Medium component concentrations were normalized using log-sigmoidal function

: m= 1/(1+ exp (-n))

n: medium component concentrations as considered in design

m: output which act as input for the network.

Performance of the generated network during the training program till the final adaptation can also be predicted beforehand. The developed model can be used to obtain the optimum concentration of the input variables. In MATLAB, there are several methods to solve the equation using Genetic Algorithm, Quadratic Programming etc.



3. Quick tips

Every protein / microbial species is unique. Therefore, it's highly recommended that designing and selection of factors should be done on a case-to-case basis.

- The selection of experimental factors should be logical. There is a plethora of literature available that discusses the effect of a large number of culture media and other components on cell growth, metabolite or protein production etc.
- Existing literature can be considered as a base for designing the screening experiments that will quicken the optimization and validation process.
- Process variables should be selected based on experimentation by OFAT or PB design. Random selection of process variables may lead to exhaustive experimentation without reaching the optimal conditions.
- Runs should be randomized to avoid any bias. Runs should be divided into 2 – 3 blocks (sets) if the number of runs exceeds the handling capacity. The main factors most of the time affect the process, but it is not necessary that all the factors interact with all the other factors. There can be strong, weak or no interaction between one or many components present in the experimental design space.
- The factors in the model, with a 'not significant' *p-value*, should be removed from the model to increase the overall accuracy of the model and prediction of response. If the p-values for the main factors are not significant, one should reconsider the selection of that factor or alter the experimental range.
- If the responses at the centre points (replicates) vary widely, either whole setup should be repeated, or analysis should be carried out after normalization of centre points. Bias arising from instruments, procedures should be minimized, if not removed. The difference between predicted vs. actual can be checked in the form of a table or in the form of a graph for comparing the variation (Figure 2f).
- Each set of runs should be performed in triplicate and the obtained responses should be averaged out before feeding into the design table for analysis.

4. CONCLUSION

The optimization of biotechnological processes requires considerable time and labour. Researchers generally adhere to simple, easy, time consuming and almost obsolete methods for process and media optimizations, due to its straightforward approach.

While these methods lead to optimization with pseudo – optimum results, the biggest shortcoming is the lack of insight into the behaviour of target microorganism with the process variables and the mutual interactions among them. Statistical methods, on the other hand, are quick, require less time to ~~be~~ execute~~d~~, result in real optimum, provide a deep insight into the interactions among different medium components, as well as the relevant physical parameters, such as temperature, pH, oxygen saturation, viscosity, etc.

Statistical methods like RSM and CCD employ multivariate regression analysis, to predict the experimental responses, at the infinitesimal combination of the experimental variables. A researcher can use the statistical model or model equation to predict the desirable responses (either maximum or minimum or at a specified target level).

The benefits of statistical methods for any process optimization outweigh the ease of the classical on-factor-at-a-time approach. The authors hope that this article would introduce the researchers to statistical methods of process optimization and encourage them to attempt the same; without getting into many technical jargons. We would also be glad to assist in designing any studies based on CCD, RSM, etc.

5. Steps to optimization methodology

1. Select process variables based on the previously conducted in-house experiments. Components should be selected based on the intended use of the final product. The purity of the components should be high if the product is to be considered as a drug or therapeutic candidate, enzymes for diagnostic use or medical use. If the product is to be used as a detergent component, animal feed or in paper and pulp processing, crude substrates can also be used.
2. Once the components are selected, select a 2-level factorial experiment using PB design.
3. Execute the runs as directed **(Table 2)** and feed the results obtained in the corresponding rows.
4. Perform Analysis of Variance (ANOVA) to calculate the F-value of the model and individual components.
5. If the model statistics show it to be significant, select the top 4 – 5 highly significant process parameters ($p< 0.1$ or 0.05) and a higher F-value. Selecting more than 5 significant variables is possible but that would result in a large number of runs. A large number of runs (~10 variables or more) may not provide significant improvements. This would extend the time to optimize and may compromise the economy of the optimization process.
6. Select Central composite design from the RSM section using Design Expert® software.
7. Select the number of factors and range as selected by the previous PB experiment.
8. Once selected, proceed to the next step for the generation of the Response Surface Design matrix.
9. After the design matrix is generated, perform the experiment at the given set of conditions.
10. Feed the results in the corresponding runs and proceed for ANOVA.
11. Check the ANOVA table and find the p-value for the individual (A, B, C…etc.) and interaction (AB, AC, BC, AD, etc.) components.
12. Find the individual components that have p-value > 0.5 and deselect them to make the model predictability more accurate.
13. Proceed to the next step for 3-D response surface graph generation to have a bird's eye view of the experimental output. The graph gives a visual idea of the experimental conditions that can be attempted.
14. As the next step, proceed to the optimization section, where the desirable conditions can be selected graphically based on the optimization needs.
15. After the conditions are fixed, one-click output can be generated with the specific value of each participating component within the design space.
16. The suggested conditions should be replicated, and the production/expression should be evaluated against the predicted values by the software.

Acknowledgements

The authors sincerely acknowledge the support provided from the Deanship of Scientific Research, Jazan University, Saudi Arabia, in the form of access to Saudi digital library for this work.

Financial statement

No financial support was available for this study.

Competing Interest Statement

The authors declare no competing interests exist.

Table 1. Table listing various Classical, Statistical Optimization and Artificial Intelligence techniques

Classical Optimization Techniques				
Experimental Design	Methodology	Benefits	Limitations	Suitability
Borrowing	Addition of components from similar processes.	Simple and easy	Too many options, too large experimental space	Small scale screening
Biological Mimicry	Adoption of similar processes	Simple to execute	Restricted flexibility	Small scale screening
Component Swapping	Removal / Swapping of components, one by one.	Easy to perform, no expertise required	Interactions unaccounte, erroneous, non-specific results	Small-scale screening experiments,
One-Factor-at-A-Time (OFAT)	Change the level of one component while keeping others at a fixed level.	Easy to apply, Simple, 2-D graphs easy to interpret	Lengthy, Erroneous, Interactions between components ignored	Small scale experiments, Screening studies
Statistical Design/ Methodology				
Experimental Design	Methodology	Benefits	Limitations	Suitability

Placket Burman Design	Screening of the components at two levels	Easy and simple to perform	A smaller number of runs, only two levels tested	Good to initiate the experimentation. Gives an idea of crucial / non-crucial factors.
Full Factorial Design	All combinations tested,	Every possible combination tested	Futile repetitions, redundant number of runs performed	Not a practical approach for biological experiments
Partial Factorial Design	Runs performed with one less than the total number of factors	Quick, Short, Easy, Good for screening,	One 2 level considered, Component interactions not accounted	Good for screening, Easy for screening experiments
Central Composite Design (CCD)	Fewer combinations (at 5 levels) executed than full factorial design,	Estimates curvature and directions, Fewer runs, Easy to perform	Lower number of runs may miss some critical combination, Lower coverage	Perfect for biological systems and optimization experiments,
Box-Behnken Design	Minimum required number of runs	Estimated the curvature with lower number of runs than CCD	Not as accurate as CCD, Not all combinations tested	Best for optimization with fewer factors, good for stepping up from screening design
RSM - Multiple experimental designs	Polynomial fit, Steepest accent, Peak determination, Steepest trail estimation, Multiple regression,	Wide application, Peak, plateau and trough differentiation,	Limited to Plotting in the form of 2-D / 3-D graphs. Only 2-Factors can be visualized	RSM contains a number of designs suited to various experimental needs
Nelder Mead (NM) simplex method	NM simplex method is based on a real-parameter black-box optimization method (n + 1 dimension) and works well with irregular objective functions.	The NM simplex method generally gives significant improvements in the primary experiments and provide quick and satisfactory outputs	When the function values, are uncertain the estimation of the process parameters and process controls are problematic,	Where a high level of accuracy in solution is not necessary,
Evolutionary Operation	Step by step training similar to crude Artificial Neural Networks (ANN)	Sophisticated than OFAT, analyses interaction between the factors.	Requires Mathematical skills	Can serve as starting point for moving from Statistical methods to ANN
Artificial intelligence				
Experimental type	Methodology	Benefits	Limitations	Suitability
Artificial Neural Networks (ANN)	Mimics the learning like the human brain	Good at pattern recognition, accurate predictions, good for a large amount of data	Gaps in the data disturb the pattern generation. The duplicity of results hinders pattern generation. Can be inappropriately applied	Refined results with accurate results, comprehensive data
Support Vector Regression (SVR)	support vectors to define a hyperplane which maps the relationship between the determinant variables and the response variables.	minimizes the structural risk rather than the empirical risk in the conventional neural networks is very effective in high dimensional spaces		recommended when the number of dimensions is larger than the number of samples
Gaussian Process (GP) regression algorithm	The GP is a non-parametric approach which uses a distribution over functions which is consistent with the observed data set	It can capture a wide range of relations between inputs and outputs. This uncertainty is not directly captured in neural networks		If the number of hidden layers in neural networks are increased to infinite number a large number of neural networks will converge to Gaussian process over functions
Fuzzy logic	It utilizes a series of rules using a fuzzy membership function. At first, fuzzy memberships are defined which explains the level of the components (low or high) in a fermentation medium	When a new medium composition is entered in its program, it predicts the output of the fermentation	High mathematical skill is required.	Highly variable data is required to analyze
Genetic Algorithms	Uses the theory of natural selection	Much improved results, the process improves over trials	Previous data abandoned with every new iteration,evolutionary process	Complex for biologists to apply mathematical expertise needed

Table 2(a). Table showing general2- level factorial screening design

S. No.	Factor	Units	Type	Low Level	High Level
1	Glucose	%-w/v	Numeric	-1	1
2	Yeast Extract	%-w/v	Numeric	-1	1
3	NaCl	%-w/v	Numeric	-1	1
4	Temperature	DegC	Numeric	-1	1
5	pH		Numeric	-1	1

Table 2(b). Table showing the experimental design matrix for screening withthe 2-levelsmall factorial design. (The responses for growth and metabolite produced are to be added to the designated rows against each experiment)

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Response 1	Response 2
Std.	Run	A: Glucose	B: Yeast Extract	C: NaCl	D: Temperature	E: pH	(Growth) R1	(Metabolite) R2

		(%-w/v)	(%-w/v)	(%-w/v)	(Deg C)		OD600 nm	(%, v/v)
12	1	+1	+1	-1	+1	-1		
3	2	-1	+1	-1	-1	-1		
6	3	+1	-1	+1	-1	+1		
4	4	+1	+1	-1	-1	+1		
5	5	-1	-1	+1	-1	-1		
15	6	-1	+1	+1	+1	-1		
9	7	-1	-1	-1	+1	-1		
14	8	+1	-1	+1	+1	-1		
8	9	+1	+1	+1	-1	-1		
2	10	+1	-1	-1	-1	-1		
7	11	-1	+1	+1	-1	+1		
10	12	+1	-1	-1	+1	+1		
1	13	-1	-1	-1	-1	+1		
11	14	-1	+1	-1	+1	+1		
13	15	-1	-1	+1	+1	+1		
16	16	+1	+1	+1	+1	=1		

Note: R1, R2: Response

Table 3(a). Table showing the factors selected and their low and high levels used in designing CCD experiment.

Factor	Name	Units	Low Actual	High Actual	Low Coded	High Coded	Mean	Std. Dev.
A	Glucose	%-w/v	0.3	1.0	-1	1	0.65	0.276
B	Yeast Extract	%-w/v	0.4	1.36	-1	1	0.88	0.378
C	NaCl	%-w/v	0.3	1.0	-1	1	0.65	0.276
D	Phosphate	%-w/v	0.3	1.0	-1	1	0.65	0.276
	Response	Name	Analysis	Minimum	Maximum	Mean	Std. Dev.	Ratio
	Y1	Growth	Polynomial	1.256	2.757	2.032	0.366	2.196
	Y2	Metabolite	Polynomial	15.708	79.319	51.481	20.282	5.049

Table 3(b). Table showing the experimental design matrix for optimization with CCD. (The responses for growth and metabolite produced are to be added to the designated rows against each experiment)

		Factor 1	Factor 2	Factor 3	Factor 4	Response 1	Response 2
Std.	Run	A: Glucose	B: Yeast Extract	C: NaCl	D: Phosphate	Growth	Metabolite 1
		%-w/v	%-w/v	%-w/v	%-w/v	%-w/v	(%, v/v)
2	1	1.00	1.36	0.30	0.30	2.06	15.71
11	2	0.65	0.88	0.65	0.65	2.19	79.32
7	3	0.30	1.36	1.00	1.00	2.16	43.44
4	4	0.30	1.36	0.30	1.00	1.98	43.24
6	5	0.30	0.40	1.00	0.30	1.42	28.60
3	6	1.00	0.40	1.00	1.00	2.25	55.59
12	7	0.65	0.88	0.65	0.65	2.19	79.32
5	8	1.00	0.40	0.30	1.00	2.06	58.08
1	9	1.00	1.36	1.00	0.30	2.16	16.25
9	10	0.65	0.88	0.65	0.65	2.19	79.32
8	11	0.30	0.40	0.30	0.30	1.26	23.63
10	12	0.65	0.88	0.65	0.65	2.19	79.32
17	13	0.65	0.88	0.06	0.65	2.08	40.53
21	14	0.65	0.88	0.65	0.65	2.19	79.32
22	15	0.65	0.88	0.65	0.65	2.19	79.32
13	16	0.06	0.88	0.65	0.65	1.58	50.49
18	17	0.65	0.88	1.24	0.65	2.23	50.06
15	18	0.65	0.07	0.65	0.65	1.44	45.30
14	19	1.24	0.88	0.65	0.65	2.76	40.31
20	20	0.65	0.88	0.65	1.24	2.30	55.48
16	21	0.65	1.69	0.65	0.65	2.43	45.52
19	22	0.65	0.88	0.65	0.06	1.41	44.43

Table 3(c). ANOVA table for CCD (with all terms) ANOVA for Response Surface (Quadratic) Model for Growth Analysis of variance table [Partial sum of squares - Type III]

	Sum of		Mean	F	p-value	
Source	Squares	df	Square	Value	Prob > F	
Block	0.014871	1	0.0148			
Model	2.929469	13	0.2253	228.7704	< 0.0001	significant
A-Glucose	0.6939	1	0.6939	704.4522	< 0.0001	
B-Yeast Extract	0.496816	1	0.49681	504.3711	< 0.0001	
C-NaCl	0.05758	1	0.05758	58.45554	0.0001	
D -Phosphate	0.398945	1	0.39894	405.0111	< 0.0001	
AB	0.016884	1	0.01688	17.14031	0.0043	
AC	0.000335	1	0.00033	0.340274	0.5780	
AD	0.051631	1	0.05163	52.41657	0.0002	
BD	0.061863	1	0.06186	62.8037	< 0.0001	
CD	0.001341	1	0.00134	1.361096	0.2816	
A2	0.003651	1	0.00365	3.706845	0.0956	
B2	0.147308	1	0.14730	149.548	< 0.0001	
C2	0.006145	1	0.00614	6.238718	0.0411	
D2	0.241671	1	0.24167	245.3463	< 0.0001	
Residual	0.006895	7	0.00098			
Lack of Fit*	0.006895	3	0.00229			
Pure Error	0	4	0			
Cor Total	2.951235	21				

Note: * = Not significant; df = Degrees of freedom,

Table 3(d).ANOVA table for CCD (with only significant terms)

ANOVA for Response Surface Reduced Quadratic Model						
Analysis of variance table [Partial sum of squares - Type III]						
	Sum of		Mean	F	p-value	
Source	Squares	df	Square	Value	Prob > F	
Block	0.014871	1	0.014871			
Model	2.927793	11	0.266163	279.484	< 0.0001	significant
A-Glucose	0.6939	1	0.6939	728.6289	< 0.0001	
B-Yeast Extract	0.496816	1	0.496816	521.681	< 0.0001	
C-NaCl	0.05758	1	0.05758	60.46173	< 0.0001	
D-Phosphate	0.398945	1	0.398945	418.911	< 0.0001	
AB	0.016884	1	0.016884	17.72856	0.0023	
AD	0.051631	1	0.051631	54.21551	< 0.0001	
BD	0.061863	1	0.061863	64.95911	< 0.0001	
A2	0.003651	1	0.003651	3.834064	0.0819	
B2	0.147308	1	0.147308	154.6804	< 0.0001	
C2	0.006145	1	0.006145	6.45283	0.0317	
D2	0.241671	1	0.241671	253.7666	< 0.0001	
Residual	0.008571	9	0.000952			
Lack of Fit*	0.008571	5	0.001714			
Pure Error	0	4	0			
Cor Total	2.951235	21				

Note: * = Not significant; Df = Degrees of freedom

Table 3(e). ANOVA table for CCD for metabolite production

ANOVA for Response Surface Reduced Quadratic Model						
Analysis of variance table [Partial sum of squares - Type III]						
	Sum of		Mean	F	p-value	
Source	Squares	df	Square	value	Prob > F	
Block	46.58776	1	46.58776			
Model	8979.028	13	690.6945	199.6146	< 0.0001	significant
A-Glucose	51.82493	1	51.82493	14.9777	0.0061	

B-Yeast Extract	0.023821	1	0.023821	0.006885	0.9362	
C-NaCl	27.12857	1	27.12857	7.840311	0.0265	
D-Phosphate	61.09488	1	61.09488	17.65677	0.0040	
AB	418.245	1	418.245	120.8752	< 0.0001	
AC	6.345777	1	6.345777	1.833965	0.2177	
AD	118.1785	1	118.1785	34.15427	0.0006	
BD	49.54568	1	49.54568	14.31898	0.0069	
CD	7.626545	1	7.626545	2.204115	0.1812	
A2	2031.769	1	2031.769	587.1928	< 0.0001	
B2	2030.794	1	2030.794	586.9109	< 0.0001	
C2	2045.446	1	2045.446	591.1454	< 0.0001	
D2	1502.123	1	1502.123	434.1221	< 0.0001	
Residual	24.22098	7	3.46014			
Lack of Fit	24.22098	3	8.07366			
Pure Error	0	4	0			
Cor Total	9049.837	21				

REFERENCES

1. A El-Kady I., S, El-Maraghy, S., 1982. Screening of zearalenone producing Fusarium species in Egypt and chemically defined medium for production of the toxin. Mycopathologia. 78, 25-9.
2. Badhwar P, Kumar A, Yadav A, Kumar P, Siwach R, Chhabra D, Dubey KK. Improved Pullulan Production and Process Optimization Using Novel GA–ANN and GA–ANFIS Hybrid Statistical Tools. Biomolecules. 2020 Jan;10(1):124.
3. Bapat PM, Wangikar PP. Optimization of rifamycin B fermentation in shake flasks via a machine.learning.based approach. Biotechnology and bioengineering. 2004 Apr 20;86(2):201-8.
4. Box, G.E.P, Wilson, K.B., 1951. On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society Series B. 13, 1–45.
5. Casas, J.A., Delara, S.G., Garciaochoa, F., 1997. Optimization of a synthetic medium for Candida bombicola growth using factorial design of experiments. Enz. Microbial Technol. 21, 221 – 229.
6. Chary, C.V.K., Rambhav, S., Venkateswerlu, G., Ramachandran, L.K., 1989. Possible precursor for thiostrepton in Streptomyces azureus - culture medium production optimization. Ind. J. Microbiol. 29, 191 – 198.
7. Cocaign-Bousquet, M., Garrigues, C., Novak, L., Lindley, N.D., Loubiere, P., (1995) Rational development of a simple synthetic medium for the sustained growth of Lactococcus lactis. J. Appl. Bacteriol. 79, 108 – 116.
8. Garcia-Ochoa, F., Santos, V.E., Fritsch, A.P., 1992. Nutritional study of Xanthomonas campestris in xanthan gum production by factorial design of experiments. Enzyme Microbiol. Technol. 14. 991 – 996.
9. George,H.A., Powell, A.L., Dahlgren, M.E., Herber, W.K., Maigetter, R.Z., Burgess, B.W., Stirdivant, S.M., and Greasham, R.L., 1992. Physiological effects of TGFα-PE40 expression in recombinant Escherichia coli JM109. Biotechnol. Bioeng. 40. 437 – 445.
10. Greasham, R.L., Herber, W.K., 1997. Design and optimization of growth media. In: Rhodes PM Stanbury PF (Eds.) Applied microbial physiology - a practical approach. 53 – 74. Oxford University Press Oxford.
11. Greasham, R.L., Inamine, E., 1986. Nutritional improvement of processes, In: Demain AL Solomon NA (Eds.) Manual for industrial microbiology and biotechnology. 41 – 48. ASM Washington.
12. Haque, S., Khan, S., Wahid, M., Mandal, R.K., Tiwari. D., Dar, S.A., Paul, D., Areeshi, M.Y., Jawed, A., 2016. Modeling and optimization of a continuous bead milling process for bacterial cell lysis using response surface methodology. RSC Advances. 6, 16348-16357.
13. Jawed, A., 2008. Studies on process development for the production of recombinant staphytokinase. PhD Thesis, CSIR –IMTECH, Jawaharlaal Nehru University, New Delhi, India.
14. Jose, M., Rodriguez. N., Elena. R., Elizabeth, C., 2005. Biosynthesis of ethyl butyrate using immobilized lipase: a statistical approach. Process Biochem. 40. 63 – 68.
15. Kisaalita, W.S., Slininger, P.J., Bothast, R.J., 1993. Defined media for optimal pyoverdine production by Pseudomonas fluorescens. Appl. Microbiol. Biotechnol. 39. 750 – 755.
16. Kishimoto, M., Suzuki, H., 1995. Application of an expert system to high cell density cultivation of Escherichia coli. J. Ferment. Bioeng. 80. 58 – 62.
17. Kweon, D.H., Han, N.S., Park,K.M., Seo, J.H., 2001. Overproduction of *Phytolacca insularis* protein in batch and fed-batch culture of recombinant *Escherichia coli.* Process Biochem. 36. 537 – 542.
18. Leggett, D.J., 1983. Instrumental simplex optimization – experimental illustrations for an undergraduate laboratory course. J. Chem. Educ. 60. 707 – 710.
19. Li, Y., Chen, J., Mao, Y –Y., Lun, S –Y, Koo, Y –M., 1998. Effect of additives and fed-batch culture strategies on the production of glutathione by recombinant Escherichia coli. Process Biochem. 33. 709 – 714.
20. Margarita, E.R.C., Monica, A.P. DeSilva., Selma, G.F.L., 2005. Biosorption of chromium using factorial experimental design. Process Biochem. 40. 779 – 788.
21. Matsui, T., Sato, H., Sato, S., Mukataka, S., Takahashi, J., 1990. Effects of nutritional conditions on plasmid stability and production of tryptophan synthase by a recombinant Escherichia coli. Agric. Biol. Chem. 54. 619 – 624.
22. McIntyre, J.J., Bull, A.T., Bunch, A.W., 1996. Vancomycin production in batch and continuous culture - Amycolatopsis orientalis culture medium optimization. Biotechnol. Bioeng. 49. 412 – 420.
23. Metzger, L.S., Dotzlaf, J.E., Foglesong, M.A., 1984. Development of a defined medium for tyosine producing strains of Streptomyces fradiae. Abstr. Annu. Meet Am. Soc. Microbiol. 199.
24. Mizutani, S., Mori, H., Shimizu, S., Sakaguchi, K., Kobayashi, T., 1986. Effect of amino acid supplement on cell yield and gene product in Escherichia coli harbouring plasmid. Biotecnol. Bioeng. 28. 204 – 209.
25. Monot, F., Martin, J–R., Petitdemange, H., Gay, R., 1982. Acetone and butanol production by Clostridium acetobutylicum in a synthetic medium. Appl. Environ. Microbiol. 44. 1318 – 1324.
26. Plackett, R.L., Burman, J.P., 1946. The design of optimum multi-factorial experiments. Biometrika. 33. 305 – 325.
27. Rinas, U., Kracke-Helm, H.A., Schurgerl, K., 1989. Glucose as a substrate in recombinant strain fermentation technology. Appl. Microbiol. Biotechnol. 31. 163 – 167.
28. Roseiro, J.C., Esgalhado, M.E., Amaral Collaco, M.T., Emery, A.N., 1992. Medium development for xanthan production. Process Biochem. 27. 167 – 175.
29. Shin, C.S., Hong, M.S., Bae, C.S., Lee, J., 1997. Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2]. Biotechnol. Prog. 13. 249 – 257.
30. Singh, V., Haque, S., Niwas, R., Srivastava, A., Pasupuleti, M., Tripathi, C.K., 2017. Strategies for Fermentation Medium Optimization: An In-Depth Review. Frontiers in Microbiology. 7:2087.
31. Singh, V., Tripathi, C.K., 2008. Production and statistical optimization of a novel olivanic acid by Streptomyces olivaceus MTCC 6820. Process Biochemistry. 43. 1313-7.
32. Spendley, W., Hext, G.R., Himsworth, F.R., 1962. Sequential application of simplex designs in optimization and evolutionary operation. Technometrics. 4. 441 – 461.
33. Tsai, L.B., Mann, M., Morris, F., Rotgers, C., Fenton, D., 1987. The effects of organic nitrogen and glucose on the production of recombinant human insulin-like growth factor in high cell density *Escherichia coli* fermentations. J. Ind. Microbiol. 2. 181 – 187.
34. Zabriskie, D.W., Wareheim, D.A., Polansky, M.J., 1987. Effect of fermentation feeding strategies prior to induction of expression of a recombinant malaria antigen Escherichia coli. J. Ind. Microbiol. 2. 87 – 95.
35. Zhang Q, Deng D, Dai W, Li J, Jin X. Optimization of culture conditions for differentiation of melon based on artificial neural network and genetic algorithm. Scientific Reports. 2020 Feb 26;10(1):1-8.
36. Zhang, J., Marcin, C., Shifflet, M.A., Salmon, P., Brix, T., Greasham, R., Buckland, B., Chartrain, M., 1996. Development of a defined medium fermentation process for physostigmine production by Streptomyces griseofuscus. Appl. Microbiol. Biotechnol. 44. 568 – 575.