

Research Article

A Critical Review of Algae Process Optimization Using Design of Experiment Methodologies

Bohnen FM^{1*} and Brück TB^{1,2}¹Bohnen & Brück Sustainable Innovations UG, Germany²Division of Industrial Biocatalysis, Department of Chemistry, Technical University of Munich (TUM), Germany

Corresponding author

Frank Bohnen, Bohnen & Brück Sustainable Innovations UG, Kattowitz Str.34, D-51065 Köln, Germany, Tel: +49-1771911792; E-mail: bohnen@bb-si.de

Submitted: 09 August 2013

Accepted: 11 September 2013

Published: 13 September 2013

Copyright

© 2013 Bohnen and Brück

OPEN ACCESS

Keywords

- Algae processes optimization
- Cultivation parameter
- Design of experiment
- Analysis tools
- Box-Cox transformation

Abstract

Climate change, limited fossil resources and steady population growth drive industrial development of sustainable biomass based processes for production of fuels, platform and specialty chemicals. In this context phototrophic microalgae cultivation processes are poised to play a central role for development of next generation bioprocesses due to high cellular growth rates and excellent yields of value adding products. Additionally, microalgae cultivation does not compete with food production and does not require agricultural land. However, industrial scale realization of microalgae cultivation has been hampered by technical issues such as optimal light and gas supply and the economic configuration of the bioreactor system. The resulting multi-parameter space complicates a targeted process optimization. To streamline experimental process optimization mathematical modeling using design of experiment (DoE) methodologies can be applied. Initial experimental data can be used to refine primary mathematical models. Therefore, DoE can be applied for consolidation of an iterative process optimization. However, DoE model choice and optimization in a dependent multi-parameter space is complex and requires diligent analysis of model design and parameter outcome. Testing data for normality is essential to derive a viable model and experimental data. If these prerequisites are not met misguided approaches to process optimization may result. In this study we reanalyze published data and demonstrate methods to streamline interpretation of DoE data. Further, methodologies and strategies for data transformation are presented that improve model evolution. To the best of our knowledge, this is the first time that a sequential strategy for DoE model evolution was applied to an algae cultivation process. Guiding iterative process optimization for algae cultivation by robust DoE models will significantly contribute to accelerate process scale-up and time to market scenarios. Ultimately, these factors determine success of an industrial process design.

INTRODUCTION

The eminent end of fossil fuel resources, climate change and a continuously growing world population drive the development of sustainable fuel and commodity processes [1]. However, current industrial processes for production renewable energy carriers are associated with the redirection of food to fuel feedstocks, destruction of agricultural land and excessive use of potable water [1-2]. Therefore, there is an urgent need to establish new technologies that do not compete with food production and do not result in land use change. A leading technology option is the photoautotrophic cultivation of microalgae species for production of biofuels, renewable chemicals, commodities and even pharmaceuticals [3-8].

A significant advantage of microalgae technology is efficient use of CO₂, rapid growth rates and the accumulation of high concentrations of value adding products [3-10]. Further, algae can be grown on marginal lands using salt instead of potable fresh water. Therefore, there is no competition with production of food resources required to maintain an ever growing population [3]. The efficiency of algae cultivation and hence the yield of biogenic feedstocks for renewable processes is highly dependent on microalgae species, the cultivation mode and cultivation conditions [3-11]. Numerous studies have examined the cultivation of various algae strains under different cultivation mode, particularly examining different photobioreactor configurations with open pond cultivation systems [3-9]. This

is of particular value to estimate the economic efficiency of specific cultivation methods [9]. However, these studies were predominantly conducted under different cultivation conditions, varying media composition, gas input, illumination and temperature regimes [4-8]. Therefore, data for these studies are mostly not directly comparable.

To derive robust and scalable algae cultivation however requires systematic studies of process parameters [9]. Of particular interest is the importance of a particular cultivation parameter as this will determine the strategy for downstream optimization processes. To achieve process parameter prioritization, mathematical models can rapidly guide the experimental validation procedures. To this end a few algae cultivation studies have employed the mathematical toolbox of design of experiment (DoE) methodologies [10-11].

However, the application of design of experiments methodology is complex with pitfalls that can lead to data misinterpretation. In this paper we re-analyze literature data [10] and discuss the design of experiments methodologies. Ultimately, this DoE based approach can be a basis for a systematic strategy targeted at the optimization of algae cultivation processes. To derive a scalable, robust process it is imperative to obtain an exact mathematical description of complex algae systems, which can guide process optimization to find the optimal growth conditions.

METHODS

DOE models and data sets for cultivation of the saline, lipid forming microalgae of *Dunaliella terticola butcher* (ATCC 30929), which involved media composition and illumination regimes were taken from the literature [10]. The study examined the effects of culture media components such as KNO_3 (N-source) and NaCl as well as the importance of the illumination on biomass and triglyceride yields. A key finding was that the concentration of the nitrogen source governed the biomass and lipid yield, while the illumination regime was a subordinate process parameter. To test the validity of these models we have reanalyzed the DoE data sets using the design of experiments software tools of the Statistica 9 software suite [12]. To examine the validity of the literature data we included tests on normality, analysis of variance (ANOVA) and used the statistically relevant terms at the $\geq 95\%$ probability level for mathematical model building and regression analysis [13].

RESULTS AND DISCUSSION

Re-Evaluation of *Dunaliella sp.* growth parameters according to

i. Dependence of biomass, lipid yield on light intensity and medium salt concentrations

The influence of light intensity on algae growth is one of the most critical variables for any algae cultivation scale up, since photosynthesis rate is a function of light intensity [2,4]. However, intensive illumination particularly at the beginning of the cultivation phase can be counterproductive as it leads to bleaching of photosynthetic pigments and cell death [4]. For industrial scale up of cultivation systems it is of critical importance to determine

the optimal operating window in terms of light intensity, since illumination for outdoor algae biomass production is dependent on the geographic latitude and the average climate conditions. For small laboratory applications the illumination can be tuned by lamps, but electricity costs need to be controlled and can be an economic bottleneck. In any economic scenario measures have to be taken to apply the optimum light intensity to the growth media [14].

Algae based lipids are regarded as the future oil source for biodiesel [3]. Massart and Hantson report on the influence of light intensity, potassium nitrate and sodium chloride on biomass productivity and lipid content [10]. The authors used *Dunaliella tertiolecta* (ATCC 30929) that has been cultured in a modified NORO medium. For optimal phototrophic growth carbon dioxide was applied as a sole carbon source by bubbling air (0.036% v/v CO_2) through the medium. Effects of light intensity, potassium nitrate and sodium chloride concentrations were investigated by DoE methodologies using a face centered response surface area design with three factors (independent variables) in the ranges 100, 200, 300 $[\mu\text{mol m}^{-2} \text{s}^{-1}]$ for light intensity, 1, 2, 3 $[\text{g l}^{-1}]$ for potassium nitrate and 10, 30, 50 $[\text{g l}^{-1}]$ for sodium chloride. The dependent variables (output variables) measured were biomass productivity $[\text{mg l}^{-1} \text{day}^{-1}]$ and lipid content [wt% of dry mass] after 6 days of culture. We have tested the validity of the applied model and resulting data sets using a test of normality and ANOVA.

ii. Normality Test of available data sets

For the proper analysis of designs of experiments a normal distribution of the dependent variables is required, since all statistical models are based upon a normality premise. In our re-examination of the data [10] we applied the Shapiro-Wilk test (Figure 1) on normality for biomass productivity $[\text{mg l}^{-1} \text{day}^{-1}]$, and lipid content [wt% of dry mass].

The normal probability plot (Figure 1) shows that for the algae lipid content the dependent variables biomass productivity and dry biomass follow a normal distribution, while lipid content did not pass the normality test. If the data set adheres to a normal distribution, the data points should follow a diagonal line, and the probability value should be greater than the probability level chosen. Sample distribution in Figure 1 indicates that data for biomass productivity have a much better fit to the diagonal line compared to the lipid content. This visual representation is confirmed by the outcome of the Shapiro-Wilk test: The probability level is set to $p=0.05$ for the Null-hypothesis, that data do not follow a normal distribution. For biomass productivity $p=0.20483$ for the Shapiro-Wilk test and greater than $p=0.05$. Therefore the Null-hypothesis has to be rejected. Consequently, the data can be regarded to follow a normal distribution. For lipid content the Shapiro-Wilk test calculates $p=0.00077$, that is smaller than the $p=0.05$ probability level. In this case the Null-hypothesis is accepted. The data for lipid content do not follow a normal distribution. To obtain normality for lipid content the raw data can be subjected to a mathematical transformation. Albeit ANOVA is regarded to be very robust even for non-normal distributed data [6], it potentially can lead to inconsistent models. In the course of this report we demonstrate that the author's modeling and interpretation of the light intensity influence on

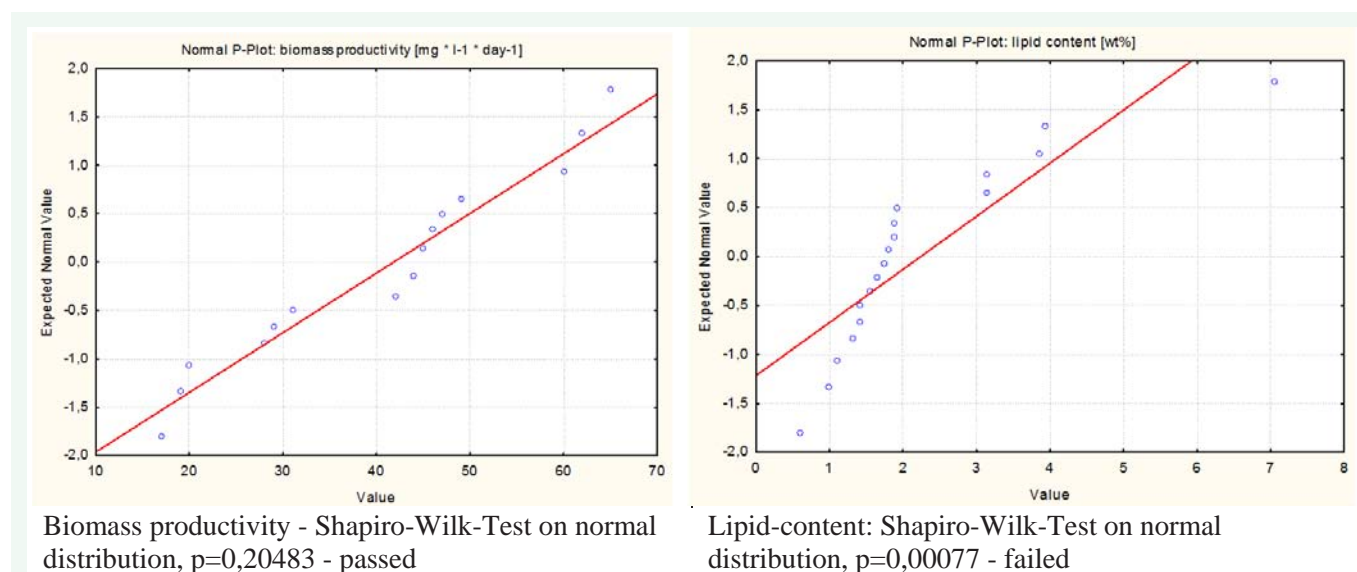


Figure 1 Normal probability plot and Shapiro-Wilk test on normality for biomass productivity [mg l⁻¹ day⁻¹], and lipid content [wt% of dry mass] for data generated by [10].

the lipid content was limited. We will further show how the Box-Cox transformation of lipid content data leads to more robust models and data interpretation. Our data re-analysis indicates that the original authors' data interpretation can be improved to a valuable outcome.

iii. ANOVA and modeling for biomass productivity

For the determination of the statistically relevant terms an analysis of variance was performed. The ANOVA bases on an equation describing the linear, quadratic and quadratic interaction terms for the biomass productivity response on light intensity, potassium nitrate and sodium chloride (see Equation 1).

Equation 1:

$$\text{Biomass productivity} = k_0 + k_1 [\text{KNO}_3] + k_2 [\text{NaCl}] + k_3 [\text{Light}] + k_{11} [\text{KNO}_3]^2 + k_{22} [\text{NaCl}]^2 + k_{33} [\text{Light}]^2 + k_{12} [\text{KNO}_3] [\text{NaCl}] + k_{13} [\text{KNO}_3] [\text{Light}] + k_{23} [\text{NaCl}] [\text{Light}]; [\text{Eq. 1}]$$

Our re-calculation of the statistically relevant equation terms on biomass productivity confirm the author's findings. In statistical terms only k_2 [NaCl], k_3 [Light], k_{33} [Light]² and k_0 (intercept) contribute at a 95% confidence level to the biomass productivity response. The influence of the nitrogen source in the chosen range is statistically insignificant. Consequently, the model should only contain the coefficients k_0 , k_2 , k_3 and k_{33} and therefore is reduced to the statistically relevant terms resulting in (See Equation 2).

Equation 2:

$$\text{Biomass productivity} = k_0 + k_2 [\text{NaCl}] + k_3 [\text{Light}] + k_{33} [\text{Light}]^2; [\text{Eq. 2}]$$

In the original paper the authors used all equation terms for modeling. Comparing the results of the modeling for the all factor equation (Equation 1) with our statistically refined equation (Equation 2) the differences of the predicted values are

negligible. However, the confidence and prediction intervals are significantly smaller in case of the statistically refined model and therefore the predictability of the model is more accurate.

Table 1 shows the calculation for the factor levels: 200 [μmol m⁻² s⁻¹] for light intensity, 2 [g l⁻¹] for potassium nitrate and 30 [g l⁻¹] for sodium chloride.

The improved predictability of Equation 2 can be explained by statistic descriptors. The mean square residual or mean square error is lower for the refined model (Equation 2) since the residual degrees of freedom are higher for the refined model compared to the all factor model (Equation 1). This outcome

Table 1: Comparison of the two models expressed in Equation 1 and 2.

Regression	Eq. 1	Eq.2
Constant	-6,90536	-1,45000
(1)KNO ₃ [g · l ⁻¹](L)	-0,39107	
KNO ₃ [g · l ⁻¹](Q)	0,10714	
(2)NaCl [g · l ⁻¹](L)	-0,18107	-0,44000
NaCl [g · l ⁻¹](Q)	-0,00098	
(3)light intensity [μmol · m ⁻² · s ⁻¹](L)	0,45234	0,44300
light intensity [μmol · m ⁻² · s ⁻¹](Q)	-0,00069	-0,00071
1L by 2L	-0,00625	
1L by 3L	0,00625	
2L by 3L	-0,00094	
MS Residual	Eq. 1	Eq. 2
	23,15045	16,36429
Degrees of Freedom (df) for Error Term Calculation	Eq. 1	Eq. 2
	8	14
Prediction	Eq. 1	Eq. 2
Predicted	45,8	45,8
-95,% Conf.	41,5	42,7
+95,% Conf.	50,2	48,8
-95,% Pred.	33,9	36,5
+95,% Pred.	57,7	55,0

exemplifies that the essence of any effective DoE model building is the iterative elimination of statistically insignificant terms, which leads to a more simplified and statistically more accurate model. Therefore, the simplified process model in (Equation 2) provides for an accelerated and exact model interpretation.

Further, the mean square (MS) residual is a qualitative measure how good the fit corresponds to the data points. In this context, a low value for MS residual indicates an improved mathematical description of the process. If the model leads to a higher degrees of freedom value (df-value) more data points can be used for the error term calculation and the confidence and prediction intervals become smaller. Therefore, the refined model in Equation 2 is more powerful for process prediction purposes. In fact, narrowing of the confidence and prediction intervals indicates that the refined model in (Equation 2) is more suitable for establishing industrial process scale-up since the derived risk analysis will be more accurate.

iv. ANOVA and modeling for lipid content

In the original report [10] data for the lipid content has not been subjected to a normality test. Subsequently, ANOVA data analysis and model building was performed with the raw, non-normalized data.

By contrast, in our study application of the Shapiro-Wilk test for the lipid content confirmed a non-normal data distribution. In consequence, the normality prerequisite for variance analysis is not given for the lipid content.

For the determination of the statistically relevant factors the original publication describes an analysis of variance. The ANOVA bases on an equation describing the linear, quadratic and quadratic interaction terms for the biomass productivity response on light intensity, potassium nitrate and sodium chloride (see Equation 3).

Equation 3:

$$\text{Lipid Content} = k_0 + k_1[\text{KNO}_3] + k_2 [\text{NaCl}] + k_3 [\text{Light}] + k_{11} [\text{KNO}_3]^2 + k_{22} [\text{NaCl}]^2 + k_{33} [\text{Light}]^2 + k_{12} [\text{KNO}_3] [\text{NaCl}] + k_{13} [\text{KNO}_3] [\text{Light}] + k_{23} [\text{NaCl}] [\text{Light}]; [\text{Eq. 3}]$$

In the original paper the authors inferred based on ANOVA analysis that only the terms k_0 (intercept), $k_2 [\text{NaCl}]$ and $k_{22} [\text{NaCl}]^2$ are statistically significant. Therefore it was concluded that light intensity has no statistical relevance. Consequently, the proposed growth model for *Dunaliella* sp. was summarized as in (See Equation 4)

Equation 4:

$$\text{Lipid Content} = k_0 + k_2 [\text{NaCl}] + k_{22} [\text{NaCl}]^2; [\text{Eq. 4}]$$

In the original paper the authors used all coefficients for the modeling (Table 2). Inconsequently, it was concluded based on Equation 3 that medium light intensity at highest NaCl concentration gives the best results regarding lipid content and no photo inhibition takes place.

ANOVA analysis in the original paper shows, that the quadratic term for light intensity has a p-value of 0.08. This is close to the chosen probability level having a p-value of 0.05. This indicates, that in a refined model the term $[\text{Light}]^2$ could

be of statistical relevance. For clarification it has to be stated that a p-value of 0.05 equals 95% probability that the term has statistical relevance. The null-hypothesis assumes that the term is not statistically relevant.

However, the deviation of a more accurate model based upon statistically relevant data should follow a procedure of consecutive elimination of the most irrelevant terms followed by an ANOVA for the model without the excluded terms. Since the exclusion of terms changes the model, the ANOVA results and regression coefficients change as well. Following the procedure of step-by-step elimination of the statistically least relevant terms the p-values of the remaining terms usually show an evolution towards lower numbers, since the model gets more accurate in statistical means.

v. Modeling of the raw lipid content data

For the evolution of a statistical model for the non-transformed data we followed the above described consecutive term elimination followed by ANOVA. In this context, (Figure 2) shows the sequential evolution of the model. Examination of (Figure 2) indicates, that the term $[\text{light}]^2$ in iteration step 4 moves below the 0.05 p-threshold and therefore is statistically relevant within the 95% probability level. Contrary to the original report ANOVA analysis indicates that the quadratic term of light intensity is statistically significant.

This is a major difference regarding the mathematical model of the experimental data. The claim on independence of the system to light intensity should be revised since the refined model indicates that photo-inhibition takes place [15,16].

The re-evaluation of the raw data leads to a different model (refined raw data model):

Equation 5:

$$\text{Lipid Content} = k_0 + k_2 [\text{NaCl}] + k_{22} [\text{NaCl}]^2 + k_{33} [\text{Light}]^2$$

The outcome of the iterative modeling approach results in the statistically relevant terms: linear and quadratic sodium chloride concentration and quadratic light illumination. Again, potassium nitrate concentration variation has no significant impact on the lipid content. The influence of sodium chloride is predominant and high light intensities lead to photo-inhibition.

vi. Modeling of the Box-Cox transformed lipid content data

If dependent variables do not follow normal distribution the raw data must be transformed. In this respect, a Box-Cox transformation of the data is a straightforward methodology that can be applied to identify a suitable transformation algorithm. In principle, the Box-Cox transformation is a power transformation

Table 2: General Box-Cox transformation rules according to [13].

Approximate Lambda	Suggested transformation of the dependent variable
- 1	Reciprocal
- 0.5	Reciprocal Square Root
0	Natural Logarithm
0.5	Square Root
1	None

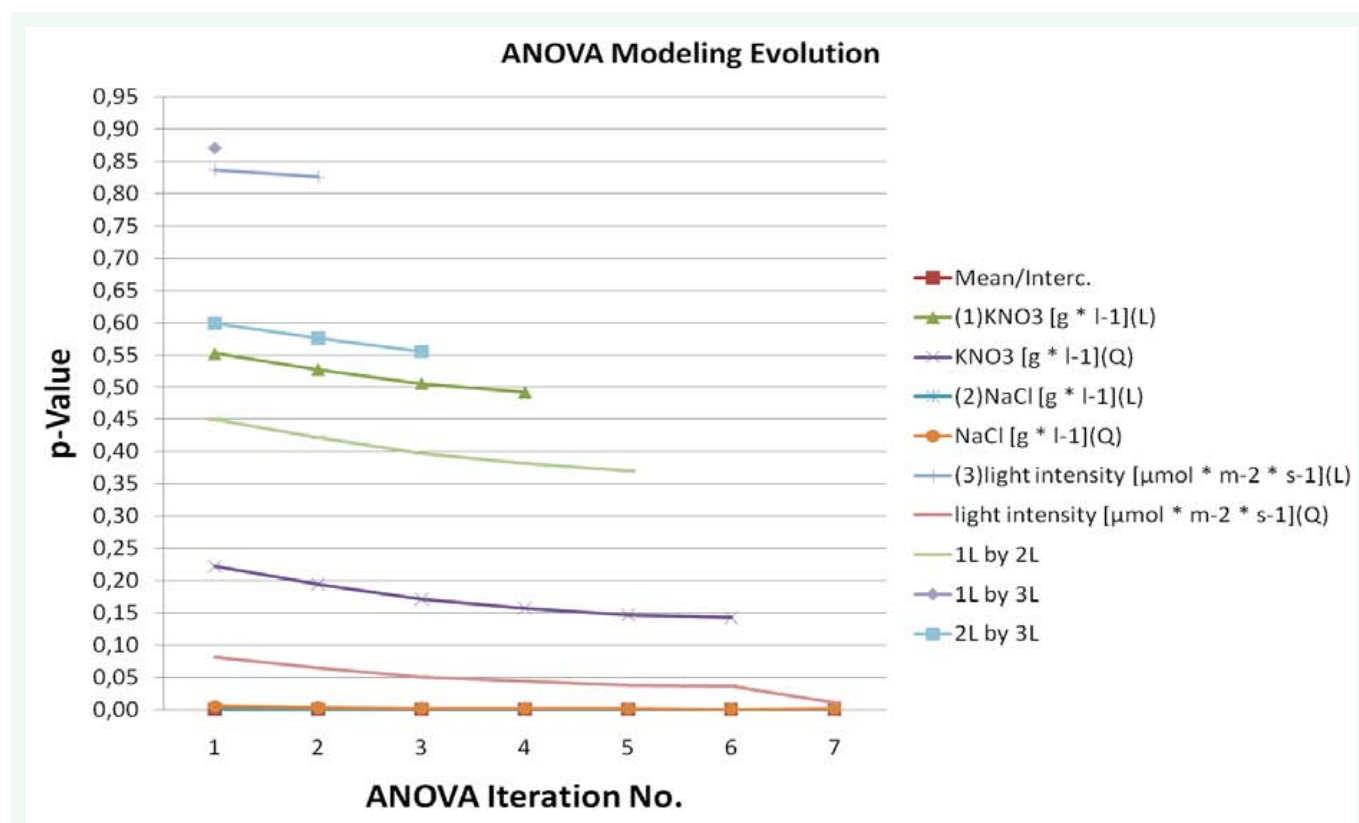


Figure 2 Consecutive evolution of the mathematical model based on stepwise ANOVA.

that can shift the raw data towards normality [13]. The most important parameter resulting from this procedure is the lambda (λ) value. If $\lambda \neq 0$, transformation λ equals the exponent applied in the power transformation for each dependent data point, if $\lambda = 0$, each data point of the dependent variable should be transformed to its natural logarithm value. As indicated in various literature reports it must not be the exact λ value [13,17]. In this respect, (Table 2) summarizes λ values and corresponding transformation procedures that can be applied to normalize experimental data [3,18,17].

We ran a Box-Cox transformation on the raw data for lipid content. The iteration led to $\lambda = -0.0183$ which is close to Zero. Adhering to rules in (Table 2) the lipid concentration data were transformed to its natural logarithm. To validate if the transformed data $\ln(\text{lipid content})$ follow a normal distribution, the Shapiro-Wilk test was applied.

Figure 3 shows that the natural logarithm-transformed data passed the test on normality and therefore was used for the ANOVA and model evolution. The initial formula for the modeling is shown in equation

Equation 6:

$$\ln(\text{Lipid Content}) = k_0 + k_1 [\text{KNO}_3] + k_2 [\text{NaCl}] + k_3 [\text{Light}] + k_{11} [\text{KNO}_3]^2 + k_{22} [\text{NaCl}]^2 + k_{33} [\text{Light}]^2 + k_{12} [\text{KNO}_3] [\text{NaCl}] + k_{13} [\text{KNO}_3] [\text{Light}] + k_{23} [\text{NaCl}] [\text{Light}];$$

Using the consecutive elimination of the least significant term followed by ANOVA of the term reduced model the evaluation

converged to a model with the statistical significant terms [NaCl], $[\text{NaCl}]^2$ and $[\text{Light}]^2$ leading to equation 7.

Equation 7:

$$\ln(\text{Lipid Content}) = k_0 + k_2 [\text{NaCl}] + k_{22} [\text{NaCl}]^2 + k_{33} [\text{Light}]^2;$$

The analysis of the natural logarithm transformed data shows the same statistically relevant terms compared with our revised raw data analysis. On first sight it might be not necessary to apply a Box-Cox transformation. However, comparing the models will show significant differences.

vii. Comparison of the different models for lipid content

Table 3 shows the regression results, the mean square residuals, degrees of freedom for error term calculation, the prediction results for the center point of the design of experiments at $[\text{KNO}_3] = 2$, $[\text{NaCl}] = 30$ and $[\text{Light}] = 200$ for model comparison and the adjusted regression coefficients for the ANOVA effect estimates. The regression equations for the refined models (Equation 5 and Equation 7) contain less terms compared with the all term Equation 3. This can be deduced from the degrees of freedom. Therefore, the error term calculation is more accurate for Equation 5 and Equation 7 and results in broader confidence and prediction limits for the all term model (Eq3).

Albeit the refined models (Equation 5 and Equation 7) contain identical statistically relevant terms, the model using Box-Cox transformed data (Equation 7) is more accurate. In contrast to the other models the Box-Cox transformed data meet the normality premise. ANOVA calculations indicate, that the MS residual for

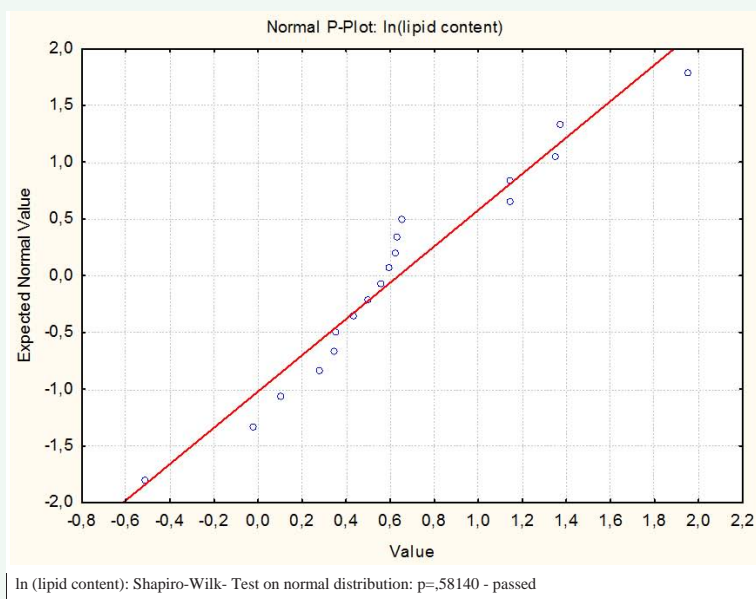


Figure 3 Normal probability plot and Shapiro-Wilk-Test results for natural logarithm Box-Cox transformed lipid content data.

Table 3: Comparison of the regression models 3, 5 and 7 regarding lipid content.

Regression	Equation 3	Equation 5	Equation 7
	Raw data published results	Raw data Refined model	Box-Cox (ln) transformed data Refined model
Dependent Variable	Lipid Content	Lipid Content	Ln(Lipid Content)
Regression equation			
Constant	-1,68880	1,817063	0,189049
(1)KNO ₃ [g * l ⁻¹](L)	2,49021		
KNO ₃ [g * l ⁻¹](Q)	-0,64393		
(2)NaCl [g * l ⁻¹](L)	-0,24326	-0,075170	-0,003764
NaCl [g * l ⁻¹](Q)	0,00470	0,002535	0,000606
(3)light intensity [μmol * m ⁻² * s ⁻¹](L)	0,03684		
light intensity [μmol * m ⁻² * s ⁻¹](Q)	-0,00010	-0,000004	-0,000003
1L by 2L	0,01125		
1L by 3L	-0,00048		
2L by 3L	0,00008		
MS Residual	Equation 3	Equation 5	Equation 7
	0,640	0,509	0,079
Degrees of Freedom (df) for Error Term Calculation	Equation 3	Equation 5	Equation 7
	8	14	14
Prediction	Equation 3	Equation 5	Equation 7 Back transformed
Predicted	2,09	1,70	1,68
-95,% Conf.	1,36	1,01	1,42
+95,% Conf.	2,82	2,38	2,08
-95,% Pred.	0,11	-0,35	0,89
+95,% Pred.	4,08	3,75	3,18
Effect Estimates (ANOVA)	Equation 3	Equation 5	Equation 7
Regression coefficient R ² adjusted	0,73	0,78	0,85

the transformed data is much lower. Consequently, error term calculation is more accurate. For the reason of comparability the prediction data for natural logarithm transformed data were re-transformed by an e^x function. It can be deduced from table 3, that the confidence and prediction intervals are more constraint for the model using Box-Cox transformed data and most relaxed for the all term raw data model. In addition, the regression coefficients for the ANOVA effect estimates is optimized for our recalculation (Equation 7).

The calculation of the lipid content using the all term model (Equation 3) and the refined raw data model (Equation 5) is insignificant in statistical terms since the confidence limits are very broad. Only the model using the transformed data (Equation 7) allows an accurate mathematical description and could be used for industrial scale-up purposes.

CONCLUSIONS

Limited fossil resources, the climate change and steady population growth drive industrial development of sustainable biomass based processes for production of fuels, platform and specialty chemicals [1]. Due to excellent biomass and product yields phototrophic microalgae cultivation processes are poised to play a central role in the development of next generation renewable processes [2-4]. Microalgae cultivation does not compete with food production and does not require agricultural land. However, industrial scale realization of microalgae cultivation is complicated by technical issues such as optimal light and gas supply and the economic configuration of the bioreactor system. The resulting multi-parameter space complicates experimental process optimization. To streamline experimental process optimization mathematical modeling using design of experiment (DoE) methodologies can be applied [10]. However, devising meaningful models that can guide experimental process optimization requires exact analysis of process parameters and needs to adhere to statistical prerequisites for model building. Incoherent models and data analysis may prohibit accurate process optimization. In this study, we have re-analyzed literature data and DoE models for optimization of *Dunaliella tertiolecta* cultivation a closed bioreactor system [10]. We demonstrate that the literature applied model can be further improved and present strategies and models that lead to a more accurate interpretation of resulting data. In contrast to the reported mathematical model, our refined model clearly indicates that production of algae biomass and associated triglycerides is primarily dependent in light supply to the culture. If initial data sets, i.e. the light dependence of triglyceride formation [10], does not pass a normality test, methods such as the Box-Cox transformation should be applied to obtain normality. The transformed and normal distributed data can then be utilized for sequential model building and parameter examination. For rapid technical scale-up and commercialization of algae cultivation procedures high accuracy process modeling is required. Strict application of DoE design strategies presented here and statistical prerequisites will significantly improve algae cultivation processes, thereby accelerating technical scale-up and commercialization of algae cultivation processes. Good DoE models will not only accelerate commercialization but also enable an accurate risk analysis due to consideration of confidence and prediction limits.

In summary, design of experiment is a very powerful tool for the optimization of biotechnological processes and the mathematical modeling of the statistical relevant terms at a given probability level [18]. However, to avoid misinterpretation and to obtain accurate results, data analysis must follow several rules [19].

The DoE methodology requires normal distributed data [17,19,20]. After completion of the experimental series a test on normal distributed data is a prerequisite. In this context, the Shapiro-Wilk test is robust method. Failing the test on normality usually is caused by skewed distribution pattern or due to the experimental set-up.

To transform skewed distributions the Box-Cox transformation is the method of choice. If Box-Cox transformed data passes the test on normality the evaluation should be continued with transformed data. In case of Box-Cox transformation failure the whole data set and the experimental set-up has to be revised regarding e.g. inaccurate measurements, analytical precision, changes in raw materials or reaction conditions [21].

Accurate mathematical modeling should follow a consecutive elimination of the least statistical relevant terms based on the analysis of variance (ANOVA) until all terms of the mathematical model show statistical relevance at a given probability level. As a result simplified models with enhanced statistical relevance are obtained [22].

In case of Box-Cox transformation the mathematical model has to be back-transformed by the inverse function applied to the data [13].

The resulting mathematical description only contains statistical relevant terms and allows a solid scientific interpretation of the microorganism's response on the selected cultivation parameter [23].

In this study we analyzed published data applying the rules and prerequisites of DoE model building. We demonstrated that accurate model evolution results in improved predictability.

Our analysis adopting the sequential model evolution clearly indicates that this methodology leads to robust model building and simplified data interpretation.

To the best of our knowledge, this is the first time that a sequential model evolution approach [22] was adopted for algae process optimization [10,11,15,14,23-31]. Guiding iterative process optimization for algae cultivation by robust DoE models will significantly contribute to accelerate process scale-up and time to market scenarios. Ultimately, these factors determine success of an industrial process design.

REFERENCES

1. Demirbas A. Political, economic and environmental impacts of biofuels: A review. *Applied Energy*. 2009; 86: 108–S117.
2. Delucchi MA. Impacts of biofuels on climate change, water use, and land use. *Ann N Y Acad Sci*. 2010; 1195: 28-45.
3. Chisti Y. Biodiesel from microalgae beats bioethanol. *Trends Biotechnol*. 2008; 26: 126-131.
4. Brennan L, Owende P. Biofuels from microalgae—A review of

- technologies for production, processing, and extractions of biofuels and co-products. *Renewable and Sustainable Energy Reviews*. 2010; 14: 557-577.
5. Batan L, Quinn J, Willson B, Bradley T. Net energy and greenhouse gas emission evaluation of biodiesel derived from microalgae. *Environ Sci Technol*. 2010; 44: 7975-7980.
 6. Davis R, Aden A, Pienkos PT. Techno-economic analysis of autotrophic microalgae for fuel production. *Applied Energy*. 2011; 88: 3524-3531.
 7. Chisti Y, Yan J. Energy from algae: Current status and future trends: Algal biofuels – A status report. *Applied Energy*. 2011; 88: 3277-3279.
 8. Bahadar A, Khan MB. Progress in energy from microalgae: A review. *Renewable and Sustainable Energy Reviews*. 2013; 27: 128-148.
 9. Rodolfi L, Chini Zittelli G, Bassi N, Padovani G, Biondi N, Bonini G, et al. Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol Bioeng*. 2009; 102: 100-112.
 10. Massart A, Hantson AL. Optimization of the medium composition of the microalgae “*Dunaliella tertiolecta* butcher” in order to combine high cell density and accumulation of lipids for biodiesel production, Proceedings of the Third International Symposium on Energy from Biomass and Waste. Venice Italy. 2010.
 11. Shinde SD, Lele SS. Statistical Media Optimization for Lutein Production from microalgae *Auxenochlorella protothecoides* [SAG 211-7A]. *Int J Adv Biotechnology and Research*. 2010; 1: 104-114.
 12. StatSoft. Inc. STATISTICA (data analysis software system), version 9.0. 2009.
 13. Osborne JW. Improving your data transformations: applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*. 2010; 15: 1-9.
 14. Jeon YC, Cho CW, Yun YS. Combined effects of light intensity and acetate concentration on the growth of unicellular microalga *Haematococcus pluvialis*. *Enzyme and Microbial Technology*. 2006; 39: 490-495.
 15. Harker M, Tsavalos AJ, Young AJ. Use of response surface methodology to optimise carotenogenesis in the microalga, *Haematococcus pluvialis*. *J Appl Phycol*. 1995; 7: 399-406.
 16. Park KH, Lee CG. Optimization of Algal Photobioreactors Using Flashing Lights. *Biotechnol Bioprocess Eng*. 2000; 5: 186-190
 17. Lee KM, Gilmore DF. Statistical experimental design for bioprocess modeling and optimization analysis: repeated-measures method for dynamic biotechnology process. *Appl Biochem Biotechnol*. 2006; 135: 101-116.
 18. Mandenius CF, Brundin A. Bioprocess optimization using design-of-experiments methodology. *Biotechnol Prog*. 2008; 24: 1191-1203.
 19. Gilmour SG. Response surface designs for experiments in bioprocessing. *Biometrics*. 2006; 62: 323-331.
 20. Ilzarbe L, Alvarez MJ, Viles E, Tanco M. Practical Applications of Design of Experiments in the Field of Engineering: A Bibliographical Review. *Qual Reliab Engng Int*. 2008; 24: 417-428.
 21. Singh B, Kumar R, Ahuja N. Optimizing Drug Delivery Systems Using Systematic “Design of Experiments.” Part I: Fundamental Aspects. *Crit Rev Ther Drug Carrier Syst*. 2005; 22: 27-105.
 22. Noordin MY, Venkatesh VC, Sharif S, Elting S, Abdullah A. Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *J Mat Proc Technol*. 2004; 145: 46-58.
 23. Fang H, Zhao C, Song XY. Optimization of enzymatic hydrolysis of steam-exploded corn stover by two approaches: response surface methodology or using cellulase from mixed cultures of *Trichoderma reesei* RUT-C30 and *Aspergillus niger* NL02. *Bioresour Technol*. 2010; 101: 4111-4119.
 24. Harker M, Tsavalos AJ, Young AJ. Use of response surface methodology to optimise carotenogenesis in the microalga, *Haematococcus pluvialis*. *J Appl Phycol*. 1995; 7: 399-406.
 25. Silveira ST, Burkert JF, Costa JA, Burkert CA, Kalil SJ. Optimization of phycocyanin extraction from *Spirulina platensis* using factorial design. *Bioresour Technol*. 2007; 98: 1629-1634.
 26. Sun L, Wang C, Ma C, Shi L. Optimization of renewal regime for improvement of polysaccharides production from *Porphyridium cruentum* by uniform design. *Bioprocess Biosyst Eng*. 2010; 33: 309-315.
 27. Saelao S, Kanjana-Opas A, Kaewsuwan S. Optimization of Biomass and Arachidonic Acid Production by *Aureispira maritima* Using Response Surface Methodology. *J Am Oil Chem Soc*. 2011; 88:619-629.
 28. Tran HL, Kwon JS, Kim ZH, Oh Y, Lee CG. Statistical Optimization of Culture Media for Growth and Lipid Production of *Botryococcus braunii* LB572. *Biotechnol Bioproc Eng*. 2010; 15: 277-284.
 29. Wang B, Lan CQ. Optimizing the Lipid Production of the Green Alga *Neochloris Oleoabundans* using Box-Behnken Experimental Design. *Can J Chem Eng*. 2011; 89: 932-939.
 30. Wen ZY, Chen F. Application of statistically-based experimental designs for the optimization of eicosapentaenoic acid production by the diatom *Nitzschia laevis*. *Biotechnol Bioeng*. 2001; 75: 159-169.
 31. Kim W, Park JM, Gim GH, Jeong SH, Kang CM, Kim DJ, et al. Optimization of culture conditions and comparison of biomass productivity of three green algae. *Bioprocess Biosyst Eng*. 2012; 35: 19-27.

Cite this article

Bohnen FM, Brück TB (2013) A Critical Review of Algae Process Optimization Using Design of Experiment Methodologies. *JSM Biotechnol Bioeng* 1(3): 1015.