MODELING BIOPROCESS SCALE-UP UTILIZING REGULARIZED LINEAR AND LOGISTIC REGRESSION

Muhammad Farhan[1], *Antti Larjo*[1,2], *Olli Yli-Harja*[1], *Tommi Aho*[1*]

[1]Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
[2]Aalto University School of Science, 15400, FI-00076 Aalto, Finland
muhammad.farhan@tut.fi, antti.larjo@tut.fi, olli.yli-harja@tut.fi, tommi.aho@tut.fi

ABSTRACT

Bioprocess scale-up from optimized flask cultivations to large industrial fermentations carries technical challenges and economical risks. Essentially, the prediction of optimal process conditions in large fermentations based on small scale experiments is non-trivial. For example, common statistical methods encounter problems with the high-dimensional, small sample size and, on the other hand, the use of various scale-up criteria requires *a priori* knowledge that may be difficult to obtain. We propose a novel computational scale-up approach applicable to various bioprocesses. The method bases on regularized linear and logistic regression. With embedded feature selection, it automatically identifies the most influential parameters and predicts their values in large scale. In addition, the method predicts the resulting large scale yield. As a case study, we examined the production of a cytotoxic compound. We predicted scale-up from flask and 2L to 30L fermentations and found that, in both cases, the product yield predictions are close to experimentally observed yields.

Index Terms— Bioprocesses modeling, scale-up, regression model, yield prediction, regularized logistic regression.

1. INTRODUCTION

Today, fermentation bioprocesses produce many of the active pharmaceutical ingredients, enzymes and fine chemicals. Because of the high costs in large industrial scale, the process optimization is performed in laboratories at smaller scales, for example in flasks. An important step in bioprocess development is the process scale-up where the production obtained in laboratory size equipment (for example, flasks) is scaled up to industrial size (for example, 1000 liter fermenters).

During the last decades, extensive research has been performed [1, 2] for identifying applicable scale-up strategies

that would result at least the same product yield in large scale that has been obtained in the preceding small scale process optimization. However, it is generally acknowledged that there does not exist only one strategy that is applicable for all process types but the strategy depends on the process characteristics and the produced product [2]. Commonly, the search for a suitable scale-up strategy starts by characterizing the key stress factors and parameters influencing cell growth and product yield. Then, the process is optimized in small scale, and a so-called scale-up criterion is established. Scale-up criteria are different kinds of conversions of specific operational parameters into criterion values that are maintained constant across scales. The criteria suggested in the literature include, for example, volumetric mass transfer coefficient $k_L a$, power per unit volume, concentration of dissolved oxygen, impeller tip speed, pH change of the medium, and mixing time [3, 4, 5, 6, 7]. When applying scale-up criteria, the process developer assumes that cell growth and product yield remain constant if the selected criterion value is kept constant across scales. This makes it possible to determine the values of particular operational parameters in different scales. However, the use of scale-up criteria is able to provide only a partial solution as they can be used to determine the values of few parameters only while a typical bioprocess involves tens of parameters.

Apart from traditional approach of defining criteria, the scale-up task has also been approached by the methods of statistical modeling as conceived by the authors of [8, 9, 10]. Response surface methodology (RSM) is efficient in optimization of several variables as well as studying interactions between them. In [10], RSM is used to optimize the production within a single scale and to help in scale-up of extracellular protease from *Bacillus* sp. The authors of [8] suggest an RSM-based methodology for examining whether a single parameter appears as an issue in scale-up. However, neither of these methods can be used to predict the values of operational parameters in large scale based on the samples in small scale.

Here, we aim at preparing a scale-up model that makes it possible to determine the values of operational parameters in large scale fermenters given the values of operational parame-

ters in small scale cultivations or fermentations. The model is required to scale-up the given small scale sample such that the product yield in the predicted large scale sample and the given small scale sample are roughly the same. The prediction task can be formulated into a statistical problem which has the following challenges: First, typical problems in bioprocess modeling have a small number of samples with respect to the number of parameters (that is, the problems are of type Large P, Small N). Second, many operational parameters obtain categorical values and their analysis is difficult using conventional methods. Third, different equipment have different parameter types, and their values are not directly comparable (e.g., flask shaking vs. fermenter mixing). Fourth, the processes exhibit nonlinear behavior that is difficult to capture. The proposed method copes with all these challenges using the state-of-the-art methods of statistical modeling. The advantage of using the proposed approach over the approaches based on constant criteria, like $k_L a$, is that the proposed approach identifies the relationships between various operational parameters, as well as their effects on the process outcome. Therefore, unlike when using one of the constant criteria, the proposed method does not need *a priori* knowledge in criterion selection but the method automatically selects the most important variables from a large set of variables. Consequently, the final model only contains variables that have effect on the process outcome.

The rest of the paper is organized as follows. Section 2 describes the proposed methodology for scale-up. Section 3 describes experimentation and discusses the obtained results. Finally, Section 4 concludes the paper.

2. SCALE-UP METHODOLOGY

In this section, we describe our scale-up methodology. It consists of four steps: encoding of categorical variables to incorporate them into modeling, product yield prediction using regularized linear regression, yield correspondence-based data rearrangement and scale-up modeling using regularized linear and logistic regression. Fig. 1 highlights the procedure for scale-up model development and its testing.
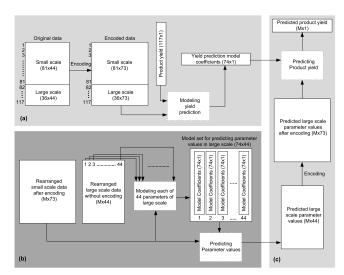
A multiple regression model [11] is a model with multiple, p, independent variables $x_{i1}, x_{i2}, \ldots, x_{ip}$ per sample i and involves $p + 1$ regression coefficients $b_0, b_1, \ldots, b_p$. The model is called multiple linear regression model when these coefficients are linearly related to each other to predict the dependent variable y_i given by the expression

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip}, \quad (1)$$

equivalently written in matrix form for $i = 1, 2, 3, \ldots, N$ as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} = [\mathbf{1X}]\boldsymbol{\theta}, \quad (2)$$

where $\mathbf{y} = [y_1 y_2 \ldots y_N]^T$, $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_N]^T$, $\mathbf{x}_i = [x_{i1} x_{i2} \ldots x_{ip}]^T$, and $\boldsymbol{\theta} = [b_0 b_1 b_2 \ldots b_p]^T$. Estimation of

Fig. 1. Methodology for scale-up modeling and testing. Procedure for (a) modeling yield prediction in all scales, (b) modeling operational parameter values in large scale using the data at a smaller scale and (c) testing the developed models.

the coefficients $\hat{\boldsymbol{\theta}}$ is performed such that the sum of squared error in the observed and predicted values of y_i is minimized.

In the light of above modeling scheme, the general problem statement that can be defined for the scale-up is to

$$\text{find } \hat{\boldsymbol{\theta}}_j : \mathbf{x}_{Lj} = [\mathbf{1X}_S]\boldsymbol{\theta}_j \text{ given that } |\mathbf{y}_L - \mathbf{y}_S| \leq \epsilon \quad (3)$$

where $j = 1, 2, \ldots, p$ and which results in

$$\hat{\mathbf{X}}_L = [\hat{\mathbf{x}}_{L1} \hat{\mathbf{x}}_{L2} \ldots \hat{\mathbf{x}}_{Lp}], \text{ such that}$$

$$\hat{\mathbf{y}}_L = [\mathbf{1}\hat{\mathbf{X}}_L]\hat{\boldsymbol{\theta}}^{yieldprediction} \approx \mathbf{y}_L, \quad (4)$$

where the subscripts L and S are used for large and small scale data respectively and all values of $\hat{\boldsymbol{\theta}}$ are determined by using regularized linear and logistic regression for numerical and categorical variables respectively. The value of allowed difference ϵ can be chosen. In our case study (described in Section 3), ϵ was set to 0.2, and the values of the product yields in the experiments were normalized to the range (0, 1] in order to anonymize the data.

2.1. Encoding of categorical variables

Bioprocesses often involve categorical variables which hold categorical labels rather than numerical values. Categorical variables with two labels are called dichotomous variables and can be directly entered as predictor or predicted variables in a multiple linear regression model [11]. In such a case using the expression in (1) only requires that the labels in a categorical variable be replaced with binary code like 0 and 1. However, in most cases the categorical predictors hold more than two labels which cannot be directly incorporated into the regression model. They require some other coding or transformation in order to be incorporated in the regression analysis. One way is to code a categorical variable with k labels

into $k-1$ dichotomous variables. For example, if a categorical variable has six labels then five dichotomous variables could be constructed that would contain the same information as the original variable, see the matrix c_1 at the bottom. This process of encoding a categorical variable into a number of separate, dichotomous variables is called dummy coding [11].

There exist many different ways or coding systems to dummy code a categorical variable. The coding system should be chosen so that it highlights the comparisons that are meant to be done among the different labels. Moreover, if there was high correlation or linear dependency among the variables, the regression modeling would become inaccurate. Therefore, the appropriate coding system and the regression methodology should minimize correlation and linear dependence. Moreover, often the processes data consists of samples for which the values of specific categorical variables for a particular set of measurements may not vary significantly (many experiments are performed with one or two changed parameters). In such cases, using binary coding is not a proper choice since it may cause singularity issues because of giving the same binary value to every label of all the categorical variables. Instead, unique codes and proper contrasts are required for each categorical variable.

One of the feasible coding systems is called contrast coding in which the labels are coded in such a manner that creates contrast among a set of labels [11]. The contrast is typically produced by giving the same variable positive and negative values for the labels between which the contrast is meant to be created. An example of contrast coding based dummy coding of a categorical variable with six labels is given in the matrix c_2 below. Here, the first new variable creates contrast between the groups of first two and the rest of the labels. The second new variable creates contrast between the first two labels. The third new variable creates contrast between groups of label 3, label 4 and of label 5, label 6 and so on. The advantage of defining such codes is that even if the labels for different variables remain unchanged for some samples, the values held by the subsequently created dummy coded variables would not only be different but also have contrast among the labels. Using contrast values helps in finding such coefficients of the model which can later be used to correctly distinguish between the categorical labels when the coded values are presented. Therefore, in the first step, fixed length dummy codes are defined in this way for each of the categorical variables while the numerical variables remain unchanged.

$$c_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, c_2 = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & -1 \end{bmatrix}$$

2.2. Product Yield Modeling

Once the original data is translated into a form that can be used for modeling, that is, after dummy coding, a generalized product yield prediction model is needed, like (2) where $\mathbf{x}_i$ takes the sample values and y_i takes the values of the product yield for experiment sample i. The resulting model coefficients can be used to predict the product yield for a given measurement sample at any scale, using the model presented in (2). The significance of developing such a model is that it is needed at a later stage when the large scale sample parameter values are predicted using the small scale samples, and it is desired to predict the product yield for that scaled-up sample to check how accurate the achieved scale-up is. Fig. 1(a) illustrates the concept of yield prediction modeling.

The problem with such process modeling is that they contain tens of different experimental variables and the number almost gets doubled after dummy coding. Not all of the variables are essential for modeling, and also the complexity of the model increases with the amount of variables used in the regression modeling. Moreover, incorporating all the variables in regression may lead to over-fitting. Therefore, the selection of the best subset of variables for model development, that is, feature selection, is required. Regularized regression with embedded feature selection has been found to be very effective in such situations [12, 13]. It is normally used when there are more variables than measurements, that is $P \gg N$, or the variables are linearly dependent on each other or overfitting is prohibiting the generalization of the solution. Here, we propose to use the regularized regression method of Least Absolute Shrinkage and Selection Operator (LASSO) that penalizes on the coefficients magnitude by adding a penalty term to the prediction error. That term includes a constant factor λ by which coefficients are translated to shrink them towards zero as well as towards each other [12]. Therefore, it always gives sparse solution, that is, many of the coefficients become zeros. This is how it automatically incorporates feature and model selection into optimization [14] as only the best variables, corresponding to non-zero or significant coefficients, are selected and the model is simplified.

Here we have N measurement samples forming the predictor variables vector $\mathbf{X} \in \mathbb{R}^{N \times P}$ and predicted or response variable vector $\mathbf{y} \in \mathbb{R}^+$. Assuming that inputs x_{ij} is standardized, that is, it has zero mean and unit norm, if the linear regression model is similar to (1), then the estimate of the model coefficients provided by the shrinkage method of LASSO is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta} \sum_{i=1}^{N} (y_i - b_0 - \sum_{j=1}^{p} x_{ij} b_j)^2 \qquad (5)$$

$$\text{subject to } \sum_{j=1}^{p} |b_j| \leq t,$$

which by using (1) and (2) is equivalent to minimizing the prediction error-based Lagrange function given by

$$\|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1, \qquad (6)$$

where $\lambda > 0$ is the Lagrange multiplier also called regularization parameter which controls the amount of shrinkage of the

coefficients (at some large value of λ, all coefficients are zero) and $\|\boldsymbol{\theta}\|_1$ is the l_1-norm of the coefficient vector by which the error function is penalized. The solution in (5) is non-linear in y_i, and therefore there is no closed form expression of the problem but it may be solved using quadratic programming or the *Least Angle Regression* algorithm [15, 12]. It gives solutions for different values of λ and choosing the optimal solution from them is non-trivial. A graph called regularization path visualizes the values of the coefficients for all the values of $\lambda > 0$ and may help in finding the optimal solution. However, in practice, cross-validation is usually performed over a set of values of λ to estimate the prediction error and to pick the optimal solution corresponding to the minimum prediction error. Here, we use 10-fold cross-validation to select the optimal model coefficients and also to ensure that the model is general enough to give a very low prediction error for the product yield even for an unseen sample.

2.3. Yield Correspondence-based Data Rearrangement

Since the aim in the process development is to optimize the process in smaller scale and to preserve the product yield in the large scales, the product yield can be used as a reference for correspondence-based data rearrangement. The idea is that if a pair of samples at both the scales have the product yield within a specific range then there exist a one-to-one correspondence between the sample pair. Since exactly matching product yield values is highly unlikely, so there has to be some tolerance band for the product yield correspondence. As stated in (3), here we use ± 0.2 units tolerance for the difference in product yield. Each sample of large scale is used to find the corresponding sample(s) from the small scale such that the difference in the product yield is within ± 0.2 units. One large scale sample can have more than one corresponding small scale samples. In that case, the large scale sample is replicated as many times as there are corresponding small scale samples. Therefore, after rearrangement, the new data contains equal number of samples for both the scales.

2.4. Scale-up Modeling

Once the data rearrangement is performed, the aim is to develop optimal linear models to predict the variable values in the large scale based on the variable values in the small scale (see (3)). In the previous section, we discussed how regularized linear regression can produce sparse models, however, the response variable in that case was numerical whereas the operational parameters are categorical as well. If the response variable is categorical, that regression modeling technique cannot be used alone to develop a model for prediction or classification of its labels. Here we exploit sparse logistic regression, a framework of LASSO [16, 17] to solve this problem.

The essence of logistic regression is the logistic function which is used to model the posterior probability density func-

tion (PDF) for each class or label. These class probability densities are then used to define the classifier. The PDF for the class $k = 1, 2, ..., K$ is modeled as

$$p_k(\mathbf{x}) = \exp(\boldsymbol{\theta}_k^T \mathbf{x})/(1 + \sum_{j=1}^{K} \exp(\boldsymbol{\theta}_j^T \mathbf{x})), for k \neq K, \quad (7)$$

$$\text{and } p_K(\mathbf{x}) = 1/(1 + \sum_{j=1}^{K} \exp(\boldsymbol{\theta}_j^T \mathbf{x})), \quad (8)$$

where $\mathbf{x} = [1 x_1 x_2 ... x_p]^T$ denotes the augmented predictor vector and $\boldsymbol{\theta}_k = [b_{k0} b_{k1} b_{k2} ... b_{kp}]^T$ are k set of coefficients of the models, one for each of the k categorical label, and are obtained by maximizing the penalized log-likelihood given by

$$\hat{\boldsymbol{\theta}}_{1,2,...,K} = \arg\max_{\boldsymbol{\theta}_{1,2,...,K}} [\frac{1}{N} \sum_{i=1}^{N} \log p(x_i) - \lambda \sum_{j=1}^{K} \|\boldsymbol{\theta}_j\|_1], \quad (9)$$
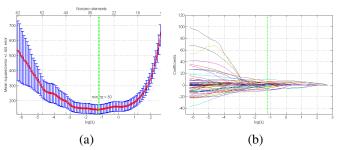
where $\boldsymbol{\theta}_{1,2,...,K} \in \mathbb{R}^{(p+1) \times K}$ and whose quadratic approximation gives rise to an equivalent penalized iteratively reweighted least squares problem that can easily be solved by coordinate descent algorithm [16]. Again cross-validation governs the selection of the optimal model coefficients such that the prediction error is minimal. When a measurement sample is presented and its corresponding class label is to be predicted, the model coefficients and the predictor values are used to compute the probability densities for every class labels using (7) and (8). The class with the highest probability is the predicted class label for the categorical variable.

Hence regularized linear regression is used to develop models for numerical variables whereas logistic regression framework of LASSO is used for modeling categorical variables. Scale-up is realized by using these models in predicting the value of every individual variable at large scale. This concept is illustrated in Fig. 1(b).

3. EXPERIMENTAL RESULTS AND DISCUSSION

Materials: Our case study contains data about 117 samples from a bioprocess that produces a cytotoxic compound called anthracycline. Experiments were performed in flasks (81 samples), 2L fermenters (24 samples), and 30L fermenters (12 samples). Since the experiments at 30L fermenters are expensive to perform, the process optimization has been performed in flasks and 2L fermenters and the number of samples in 30L is much smaller. This typical situation highlights the need of developing efficient scale-up modeling approaches. Researchers should be able to complete the process optimization in small scale, and have a model that predicts the values of operational parameters in the large scale. However, in order to develop a general model, the required minimum number of samples in each scale is difficult to determine and usually remains unknown.

Testing the Product Yield Modeling: Each sample is composed of around 40 typical bioprocess variables such as strain,

Fig. 2. Product yield prediction model selection (Green dashed vertical line) using the (a) error plot and (b) regularization path plotted as a function of $log(\lambda)$.

broth medium, broth adsorbent, fermentation time, temperature, agitation, aeration, pH, etc. After dummy coding the categorical variables, the number of variables increased to more than 70. All the samples were exploited when developing the yield prediction model using regularized linear regression. Here, 10-fold cross-validation was performed to determine the optimal model coefficients based on the minimum prediction error so that the model is general enough to predict the product yield for an unseen sample. Fig. 2(a) plots the mean square error (MSE) with the standard error obtained from cross-validation as a function of logarithm of the regularization parameter λ. The vertical green dashed line is at the point with the minimum error. Fig. 2(b) plots the regularization path as a function of $\log(\lambda)$. The value of λ yielding minimum prediction error gives the optimal solution (coefficients at $\log(\lambda)$ = -1.2127 in regularization path) with 43 out of 73 model coefficients as zero. Half of the 30 non-zero coefficients (corresponding to dummy-coded variables) appear significant, and only around 15 real variables remain with a true impact on the product yield. This highlights the capability of our regularized regression-based approach for automatically selecting the important variables from a large set of variables. A benefit of our approach is that no biological *a priori* knowledge is needed about the variables or their relationship with the product yield. Fig. 3 compares the experimentally observed and the predicted values of the product yield. It is evident that despite of small sample size, categorical variables and high-dimensionality of the data, the methodology identifies a model that is both general and accurate.
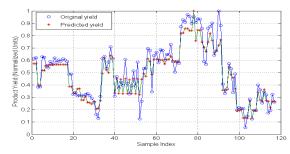
Testing the Scale-up Modeling: In order to test the scale-up modeling, samples with specific product yields were selected from the small scale data to predict the variable values of their corresponding large scale samples. As mentioned earlier, the value for every individual variable of a large scale sample was predicted using the respective model. That is, the values of numerical variables were predicted using (2) whereas the categorical variables were predicted using the PDF models of (7) and (8). Then the predicted variable values were given to product yield prediction model that predicted the product yield in large scale. Since the aim was that product yield remains constant in scale-up, this predicted value should be

within $\pm$ 0.2 units tolerance range of the product yield at the small scale. Fig. 1, in particular Fig. 1(c), shows how the testing of scale-up is performed.

The objective in our case study is to use the flask and 2L experiments as two alternative small scale data and to develop models that are able to determine the values of operational parameter at 30L. Moreover, we aimed to determine whether the scale-up to a 30L fermenter is possible directly from the flask scale, or is it the only option to base the scale-up on fermentations at 2L scale. If direct scale-up from flask to 30L is possible, that would improve the cost-efficiency in process development. Therefore, flask experiment data was first used as the small scale data and 30L data as the large scale data. This setting produced 330 measurement sample pairs after data rearrangement. Then, the values of operational parameters in 30L experiments were predicted using the small scale samples. The derived large scale samples were then provided to the product yield prediction model to predict the product yields. The performance of the scale-up strategy was evaluated by comparing the predicted product yields with the experimental product yields at the large scale. Fig. 4(a) shows the comparison of the experimental and the predicted product yields at the large scale. In general, the values of the product yield using the derived large scale samples are very much in accordance with the $\pm$ 0.2 units tolerance range of the experimental large scale product yield. This is also ascertained by the root-mean-square error (RMSE) of the product yield which is 0.172 and is within the tolerance range of 0.2.

The total prediction error of the product yield is composed of two different errors sources: the errors in predicting the parameter values at the large scale, and the error caused by the yield prediction model. The small total prediction error of our approach suggests that the scale-up is not only good in terms of predicted product yield but also in the sense that the derived large scale samples are quite similar to the experimentally tested large scale samples. That is, there is only little difference in the values of operational parameters in the predicted large scale samples and the experimental large scale samples. In particular, this is true for the 15 operational parameters that have significant effect to the product yield.

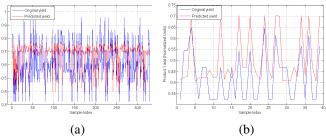In the second step in our case study, samples from the ex-

Fig. 3. Result of product yield (normalized to the range (0, 1]) prediction model for the samples from flask experiments compared to the experimentally observed product yields.

periments at 2L and 30L fermenter were considered as the small scale and large scale data, respectively. After applying the same procedure described in the previous paragraph we derived large scale (30L) samples and predicted their product yield. The comparison of the experimental product yield at large scale with the predicted product yield is presented in Fig. 4(b). It shows that the scale-up from 2L to 30L is satisfactory since in most of the cases the product yield from the scaled-up samples is within the $\pm$ 0.2 units tolerance range. Again, this is confirmed by the RMSE value of the product yield which is 0.14 and is within the tolerance range of 0.2. Thus, in this case study, our modeling methodology was able to achieve the scale-up with very similar accuracy from flask to 30L and from 2L to 30L. Finally, with the results of Fig. 4 we can infer that it is appropriate to perform 30L fermentations directly after flask experiments rather than to perform experiments at 2L vessels too. Therefore, in future we may consider omitting fermentations at the 2L scale.

4. CONCLUSION

A novel statistical approach to model the scale-up of bioprocesses was presented. Regularized regression with the embedded feature selection property of LASSO and its logistic regression classification framework provided effective tools for modeling the scale-up. The approach contained two modeling tasks. First, a model was developed to predict the product yield scale-independently. The model was found to be general enough and providing satisfactory results in different scales. Second, the scale-up modeling (i.e., the prediction of operational parameters) was realized by developing a separate model for each parameter in the large scale such that the value of the parameter is predicted based on the operational parameters in a small scale. Illustrations revealed that the scale-up was successfully achieved from flask to 30L fermenter, as well as from 2L to 30L fermenter. This was ascertained by the RMSE values that were well within the specified range of 0.2. Because of similar performance in these two cases, the idea of omitting the 2L fermentations in future is supported. Instead of these lab-size fermentations, the values of operational parameters in 30L fermenters could be determined based on flask experiments. The future work contains

(a)   (b)

Fig. 4. The predicted product yield (normalized to the range (0, 1]) after scale-up to 30L experiment from (a) flask experiment and (b) 2L experiment, compared to the experimentally observed product yield of 30L experiment samples.

more detailed characterization of the presented methodology, for example, by testing with different experimental data with various production organisms and products.

5. REFERENCES

[1] Junker HB, "Scale-up methodologies for eschericia coli and yeast fermentation processes," *Bioscience and Bioengineering*, vol. 97, pp. 347–364, 2004.

[2] Schmidt FR, "Optimization and scale up of industrial fermentation processes," *Applied Microbiology and Biotechnology*, vol. 68, pp. 425–435, 2005.

[3] Garcia-Ochoa F and Gomez E, "Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview," *Biotechnology Advances*, vol. 27, pp. 153–176, 2009.

[4] Katzer W, Blackburn M, Charman K, Martin S, Penn J, and Wrigley S, "Scale-up of filamentous organisms from tubes and shake-flasks into stirred vessels," *Biochemical Engineering*, vol. 7, pp. 127–134, 2001.

[5] Marques MPC, Cabral JMS, and Fernandes P, "Bioprocess scale-up: quest for the parameters to be used as criterion to move from microreactors to lab-scale," *Chemical Technology and Biotechnology*, vol. 85, pp. 1184–1198, 2010.

[6] Rocha-Valadez JA, Estrada M, Galindo E, and Serrano-Carreon L, "From shake flasks to stirred fermenters: Scale-up of an extractive fermentation process for 6-pentyl-a-pyrone production by trichoderma harzianum using volumetric power input," *Process Biochemistry*, vol. 41, pp. 1347–1352, 2006.

[7] Seletzky JM, Noak U, Fricke J, Welk E, Eberhard W, and Knocke C et al, "Scale-up from shake flasks to fermenters in batch and continuous mode with corynebacterium glutamicum on lactic acid based on oxygen transfer and ph," *Biotechnology and Bioengineering*, vol. 98, pp. 800–811, 2007.

[8] Hsu Y-L and Wu W-T, "A novel approach for scaling-up a fermentation system," *Biochemical Engineering*, vol. 11, pp. 123–130, 2002.

[9] Ogawa S, Kamijima T, Miyamotao Y, Miyajima M, Sato H, and Takayama K et al, "A new attempt to solve the scale-up problem for granulation using response surface methodology," *Pharmaceutical Sciences*, vol. 83, pp. 439–443, 1994.

[10] Saran S, Isar J, and Saxena RK, "Statistical optimization of conditions for protease production from bacillus sp. and its scale-up in a bioreactor," *Applied Biochemistry and Biotechnology*, vol. 141, pp. 229–240, 2007.

[11] Stockburger DW, *Multivariate statistics: concepts, models, and applications*, Missouri State University, 2001.

[12] Hastie T, Tibshirani R, and Friedman JH, *The elements of statistical learning: Data mining, inference, and prediction*, Springer Series in Statistics, Springer, 2009.

[13] Kauppi J-P, Huttunen H, Korkala H, Jääskeläinen IP, Sams M, and Tohka J, "Face prediction from fmri data during movie stimulus: Strategies for feature selection," in *Proc. International Conference on Artificial Neural Networks ICANN*, 2011, pp. 189–196.

[14] Tibshirani R, "Regression shrinkage and selection via the lasso," *Royal Statistical Society Series B Methodological*, vol. 58, pp. 267–288, 1996.

[15] Efron B, Hastie T, Johnstone I, and Tibshirani R, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

[16] Friedman JH, Hastie T, and Tibshirani R, "Regularization paths for generalized linear models via coordinate descent," *Statistical Software*, vol. 33, pp. 1–22, 2010.

[17] Huttunen H, Kauppi J-P, and Tohka J, "Regularized logistic regression for mind reading with parallel validation," in *Winning submission to Mind reading from MEG, PASCAL Challenge in ICANN*, 2011, pp. 20–24.