Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Development of Ant Colony Optimization (ACO) Algorithms Based on Statistical Analysis and Hypothesis Testing for Variable Selection

C. M. Pessoa, C. Ranzan, L. F. Trierweiler, and J. O. Trierweiler

Group of Intensification, Modeling, Simulation, Control, and Optimization of Processes (GIMSCOP)
Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre, RS, Brazil (Tel: +55(51) 33084167,
e-mail: {carolpes, cassiano, luciane, jorge}@ enq.ufrgs.br)

Abstract: Obtaining reliable models from experimental data is a point of deep interest in all areas of research. Since the quality of the model depends on the number of selected variables, it is important to develop methods that identify the best ones. This work proposes a method of variable selection based on the Ant Colony Optimization (ACO) algorithm. Using data from a *Saccharomyces cerevisiae* fermentation, several criteria for trail update and model comparison were implemented and the obtained models were compared. The use of the length of the confidence interval produced the best results, finding the optimal model more frequently.

Keywords: Modelling, Variable Selection, Spectral Data, Ant Colony Optimization, Hypothesis Testing, Confidence Interval

1. INTRODUCTION

In order to achieve production increment, one of the alternatives is to invest in optimization techniques. This often means investment in new or better control methods, which is linked to two important tasks: obtaining a good model for the variable of interest and monitoring the process (Yamuna & Ramachandra, 1999).

Modelling consists on finding a causal relationship between variables. Regression analysis, combined with statistical techniques to quantify the confidence of the model, appears as the main tool used for this purpose, (Sykes, 1993). Among the different forms of regression, linear is certainly the most widely used. In this context, techniques like Partial Least Squares (PLS), Principal Component Analysis (PCA) and Principal Component Regression (PCR) appear as the primary regression methods, useful in the quantitative analysis of data (Geladi *et al.*, 2004).

The quality of the model depends not only on which variables are used in the regression, but on how many. A small subset of predictor variables is often preferable against using all available data, because it reduces costs and time spent in the measurements, tends to present a more simple physical interpretation, and, in the case of multiple linear regression (MLR), reduces the uncertainty of prediction, since this uncertainty increases with the ratio between the number of explanatory variables and the number of samples used in the calibration (Brown *et al.*, 2009). It is important to notice that, in general, each variable has different difficulty levels and cost involved in measuring them, and this aspect must be taken into account during variable selection.

Process monitoring, however, is highly dependent on the type and quality of sensors used. In recent years, optical sensors have become increasingly important in biotechnological applications. Optical sensors can be interfaced through glass window in reactors. Therefore, it is an in-situ, non-invasive method that gives real-time measurements (Hantelmann *et al.*, 2006, Scheper *et al.*, 1999). Several types of spectroscopy are possible through this technique, fact that makes models capable of dealing with spectral data so attractive. In this context, fluorescence sensors have being investigated for the determination of biomass and viable cells, bioreactor characterization, metabolic studies (transition aerobic/ anaerobic) and especially the monitoring of bioprocesses (Solle *et al*., 2003, Hitzmann *et al*., 1998). The development of a method capable of working with spectral data in order to identify spectral regions related to response variable can enable the development of optical sensors tailored to specific process, which would improve its control, making it more efficient and economic.

In the case of variable selection, a fairly common approach is the combination of suitable criteria that evaluate the quality of a subset of predictors combined with an algorithm that optimizes these criteria (Brown *et al.*, 2009). This approach is used in this work, applying the Ant Colony Optimization (ACO) algorithm as optimization method. Due to the advantages of spectral data, this work addresses the use of ACO for selecting components from spectroscopic analyses. When applied for different kind of data, the algorithm must take into account the difficulty level and cost involved in each variable measurement when selecting them.

Ant Colony Optimization algorithm is based on the hypotetical collective behavior of ants when searching for food sources. During this search, the ants secrete pheromones to mark their path, but they evaporate over time. In nature, ants that travel the shortest path return to the nest more quickly, so that the path traveled by these individuals has a

higher concentration of pheromone. This trail acts as a decoy for other ants and, in time, all individuals of the colony tend to go trough this optimal (shortest) way (Allegrini & Olivieri 2011).

Dorigo and Gambardela (1997) developed the first version of ACO seeking solution for the Traveling Salesman problem, a problem of combinatorial optimization search in the space of permutations (Ranzan, 2014). Currently, several studies have been published regarding the application of the ACO method for screening variables, among which can be mentioned the work of Ranzan (2014), Allegrini and Olivieri (2011), Hemmateenejad *et al.* (2011), Mullen *et al.* (2009) and Socha *et al.* (2008).

Ranzan *et al.* (2014) applied the Sum of Squares Errors (SSE) as a criterion for updating the pheromone trail and to compare models. The goal was to predict the content of protein in different brands of flour based on NIR spectral data. The results showed the use of ACO as a filtering tool made possible the selection of important spectral regions, increasing the coefficient of determination of generated models by 60% compared to other methods which used the full spectrum, such as PCA and PCR.

Other optimization algorithms have also been used in variable selection, and the two stochastic optimization algorithms most known and applied in the field of chemometrics are simulated annealing and genetic algorithm (Cerny, 1985, Kirkpatrick *et al.*, 1983). Moreover, other methods, such as tabu search, artificial colonies of bees, particles swarm and harmonic search can also be used for this application (Ghasemi *et al.*, 2012, Mello & Pinto, 2008).

2. STATISTIC METRICS

Establishing appropriate criteria to evaluate the generated models is also crucial in order to obtain the optimal result. Among the parameters useful in this evaluation, the most used are the root mean square error of calibration (RMSEC) and prediction (RMSEP), which examines the fit of the model to the set of calibration and testing data evaluating the reproducibility of the data, and the coefficient of determination R^2, which is a measure of the proportion of variability explained by the fitted model.

This coefficient is used quite frequently due to its simplicity, but there are some disadvantages in its interpretation, such as the increase of its value by the addition of terms in the model. The adjusted coefficient of determination (R_a^2) is a variation that can be used to solve these problems, since it takes into account the number of degrees of freedom associated with the sum of squared error (SSE) and the sum of total squares (SST) (Walpole *et al.*, 2012).

Also, the use of hypothesis tests is very useful when analyzing models. The t-Student test (or t-test), for example, allows to test hypotheses about the coefficients and build their confidence intervals (Wilcox, 2012). Basically, the hypotheses being tested are:

$$H_0: \beta_j = 0 \qquad H_1: \beta_j \neq 0 \qquad (1)$$

where β_j is a given model parameter j and j=0,1,…,k.

The rejection or not of the hypothesis H_0, called null hypothesis, depends on the level of significance chosen, on the parameter estimator and its variance and on the standard deviation of errors. If the null hypothesis is not rejected, the variable associated with β_j explains an insignificant amount of change in y in the presence of the others regressors, and therefore can be removed from the model.

Even considering the estimators as unbiased, they are unlikely to estimate the parameters β_j accurately. Thus, it is preferable to determine an interval where it is possible to assume, with a given confidence, that it contains the true value of parameter β_j, called the confidence interval. The smaller the confidence interval, there is less uncertainty in the model parameter (Walpole *et al.*, 2012).

The use of such tests has the advantage of evaluate the contribution of each parameter separately, rather the adequacy of the model as a whole. Optimization methods can, therefore, use this information in combination with statistical models to identify and select the most relevant variables.

Another way to assess the contribution of each predictor is making use of F-tests to compare subsets of variables against the full model. The higher the F value, the worse the submodel is when compared to the full model. This aspect will be better discussed in the next section.

3. ACO MODIFICATIONS

The version of ACO implemented in this study is a modification of the one used by Ranzan *et al.*, (2014), which is based on pheromone trail evolution during spectral group scanning. Initially, all spectral components are marked with the same pheromone concentration. The ACO routine selects random spectral components to compose a group that is evaluated using the objective function for process variable prediction. Based on objective function error, the pheromone concentration, associated with each spectral component at the evaluated spectral group, is updated. For the subsequent spectral group selection, the random selection chooses spectral components associating the same random trigger and a cumulative density of pheromone for the full range of spectral elements. This association brings into evidence significant elements inside the spectral range, and, after few iterative runs, a pheromone profile is established, and the regions with high pheromone density highlight the significant excitation/emission pairs for process variable prediction.

A schematic summary of steps within ACO implementation used in this work can be seen in Fig. **1**. See Ranzan *et al.* (2014) for more detail about this algorithm

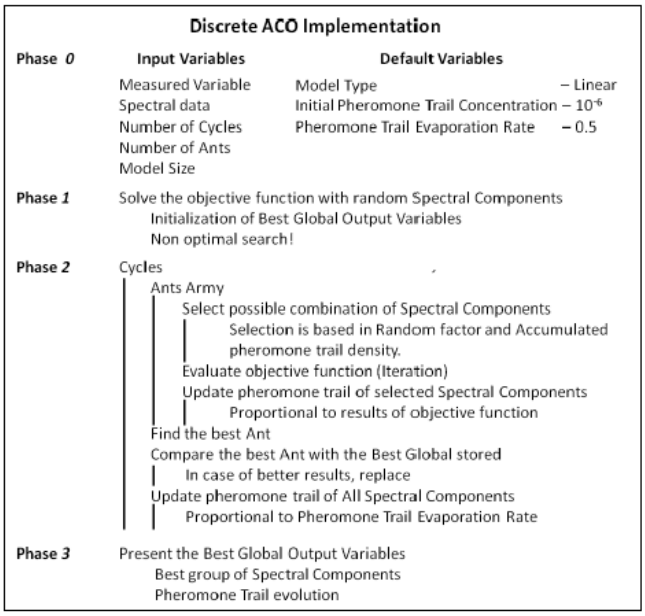

Fig. 1. Schematic summary of steps within ACO implementation.

3.1 Pheromone trail update

An important question is the update of pheromone trail of each spectrum component based on its performance within the same model. The original algorithm, suggested by Ranzan *et al.* (2014), uses SSE both as pheromone trail update and model selection criterion. As SSE is a property of the model as a whole, its use may erroneously increase the pheromone concentration of a spectrum component present in a good model that has not contributed significantly to the response.

Therefore, besides the use of an indicator of the model quality, such as SSE, it is important to introduce in the algorithm metrics capable of assessing the importance of each variable within the model. In order to study this aspect, this work proposes the implementation of 3 metrics for the model (besides SSE) and 3 metrics for model component analysis as criterion for trail update.

The global metrics chosen are: the adjusted coefficient of determination (R_a^2), due to reasons discussed in section 2; the logarithm of R_a^2, in order to enhance the region of interest, i.e., values close to one; the absolute value of a modified R^2, called RR (Silveira, 2012), as a way of enhancing the difference in the correlation, given by (2):

$$\boldsymbol{RR = -\log\left(\frac{1}{(1-R^2)}\right)} \qquad (2)$$

For all three above, the increment on the pheromone trail at each iteration is exactly the same for all components in the selected model. The pheromone concentration is directly proportional to the R_a^2 and the RR, and inversely proportional to the SSE and the logarithm of R_a^2.

The individual metrics chosen were: the t-test of hypothesis, directly associating the absolute t-value to the spectral component; the length of the confidence interval of the coefficient associated with each variable, which can be used as criterion once the variables are staggered; the F-test of hypothesis, associating the F-value of a submodel to the spectral component absent in it.

In other words, the F –test method works as follows: first, the algorithm chooses k variables among the 150 available and generates a model. Then, one of the components is withdrawn from the original group and another model, with size $k_{sub} = k - 1$, is constructed. The submodel is then compared to the full model through the F-test. The F value is associated with the variable not included in the subgroup. Therefore, variables that are important for the final model will have a higher F statistic, since not using it results in a model worse than the full one. These variables with higher F-value will then receive a higher increment in the pheromone trail.

At each iteration, the 3 metrics above lead to different increment in the pheromone trail for each spectral component. This increment is directly proportional to the t-value and the F-value, and inversely proportional to the length of the confidence interval.

3.2 Criteria for model selection

The second modification is regarding the comparison of generated models. In this aspect, besides the SSE, were also considered different metrics in order to find the criterion which emphasizes the best model the most.

The criteria chosen are: the ratio between SSE and R_a^2; the product of the error and the logarithm of R_a^2; the RR coefficient.

Table 2 summarizes all the criteria implemented and presents the legend used for each combination of criteria. Each pair of criteria (C1, C2), being C1 the trail update criterion and C2 the model selection criterion, is shown as a number. This will favor the interpretation of the results along this paper.

Table 1. Summary of all criteria implemented in ACO algorithm for updating the pheromone trail (C1) and comparing/selecting models (C2).

		Model selection (C2)			
		SSE	SSE/R^2_a	SSE * log R2a	RR
Trail Update (C1)	SSE	1	2	3	4
	R^2_a	5	6	7	8
	log R^2_a	9	10	11	12
	RR	13	14	15	16
	Length of the confidence interval (LCI)	17	18	19	20
	t-test	21	22	23	24
	F-test	25	26	27	28

The different criteria introduced in Table 1 will be tested using the case study discussed in the next section. It is important to notice that case number 1 (using SSE both as C1 and C2) is the one implemented by Ranzan *et al.* (2014), being used only as reference for the results produced in this work.

4. CASE STUDY

4.1 Two-dimensional fluorescence spectroscopy

Fluorescence spectroscopy is based on emission of fluorescence of fluorophores present in a sample because of emission / remission of low energy light. The reemitted light is proportional to the concentration of the fluorophore in the sample and has a wavelength equal to or greater than the excitation (Hitzmann et al., 1998, Solle et al., 2003). Thereby, it is based on the remission of light with spectral shift presented by some chemical species. Radiation with certain wavelengths is sent to the medium, electronically exciting the chemical bonds of certain molecules. When returning to its initial state, these components emit fluorescence in wavelengths of different lengths that are filtered by lenses or monochromators and processed by a data acquisition system.

4.2 Experimental Data

The fluorescence experimental data used in this work consists of two fermentative batch cultivations of glucose by Saccharomyces cerevisiae H620 growing in a 1.5L bioreactor at constant temperature and pH, 30ºC and 5.5, respectively, with Schatzmann medium supplementation. During cultivation, fluorescence spectra were collected every 6 minutes, using a BioView fluorometer (Delta Light & Optics, Denmark), as described by Stärk *et al.* (2002). Each spectrum contained 150 fluorescence pairs with excitation/emission wavelengths: 15 filters in the region of 270 to 550 nm for excitation and 15 filters in the region of 310 to 590 nm for emission, both with a bandwidth of 20 nm, collected equidistantly. The spectral data is then formed by 150 pairs of excitation-emission wavelength, and its summary can be seen in Fig.2. The experimental data was segmented so that first batch is used only in the models calibration phase, and batch two is used only in the model test phase.

There is a distinction between fluorescence pairs due to the fluorescence scanning method performed by the equipment. Since fluorescence measurements are made by varying the wavelengths of excitation and emission, each cell of the data matrix is composed by an excitation and an emission wavelength. However, as the phenomenon of fluorescence is due to the absorption and emission of energy, the emitted wavelength cannot have a higher energy than that used to excite it, thereby pairs whose fluorescence emission wavelength is lower than the excitation show no real information, only measurement noise.

Due to this fact, the fluorescence data matrix, in the form as presented in Fig.2, has zero for values of fluorescence intensity for pairs above the main diagonal of the matrix. Thus, valid values of fluorescence intensity are located below the diagonal of the matrix (fluorescence pairs located in the matrix diagonal are those whose excitation wavelength is equal to the emission).

Excitation wavelength (nm) \ Emission wavelength (nm)	310	330	350	370	390	410	430	450	470	490	510	530	550	570	590	Open
Open	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	0
550	0	0	0	0	0	0	0	0	0	0	0	0	0	0	134	135
530	0	0	0	0	0	0	0	0	0	0	0	0	0	131	132	133
510	0	0	0	0	0	0	0	0	0	0	0	0	127	128	129	130
490	0	0	0	0	0	0	0	0	0	0	0	122	123	124	125	126
470	0	0	0	0	0	0	0	0	0	0	116	117	118	119	120	121
450	0	0	0	0	0	0	0	0	0	109	110	111	112	113	114	115
430	0	0	0	0	0	0	0	0	101	102	103	104	105	106	107	108
410	0	0	0	0	0	0	0	92	93	94	95	96	97	98	99	100
390	0	0	0	0	0	0	82	83	84	85	86	87	88	89	90	91
370	0	0	0	0	0	71	72	73	74	75	76	77	78	79	80	81
350	0	0	0	0	59	60	61	62	63	64	65	66	67	68	69	70
330	0	0	0	46	47	48	49	50	51	52	53	54	55	56	57	58
310	0	0	32	33	34	35	36	37	38	39	40	41	42	43	44	45
290	0	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
270	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Fig.2. Diagram showing the fluorescence pairs used to acquire the spectral data, as well as the number assigned to each pair.

A total of 190 spectra were collected from each cultivation. The data obtained by BioView Spectrum Fluorometer was processed with MATLAB software (Ver. 5.3.0.10183 R11, The Mathworks, Inc., Natick, USA). Given that the efficiency of regression methodologies is highly associated with spectral data quality, it is useful to normalize the spectral signals prior to data analysis. This process helps in eliminating arbitrary offsets and multiplication factors. This was achieved by applying Standard Normal Variate (SNV) scaling to spectral data. This method essentially autoscales the samples, obtaining zero mean and standard deviation equal to 1 for each spectrum (Gemperline 2006, Wehrens 2011). Fig 4 exemplifies one of the 190 normalized 2D fluorescence spectra obtained for the first and second fermentations.

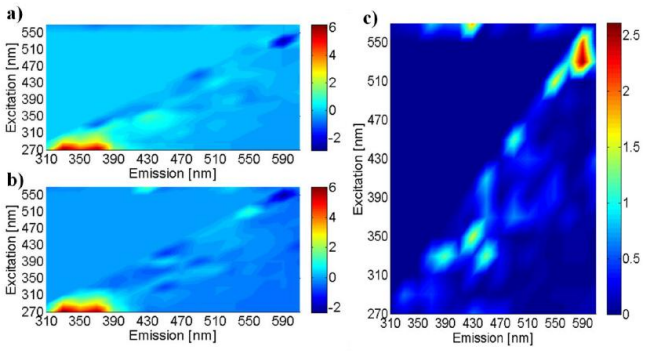

Fig. 3. Fluorescence spectra at t = 0, with standard SNV method of (a) fermentation 1 and (b) fermentation 2. (c) absolute difference in fluorescence intensity, pair-to-pair, between the normalized spectra.

The comparison of the absolute difference between the normalized initial spectral data of both fermentation (Fig. 3c) confirms the similarity between the spectroscopic data of the reaction medium. Also, after applying PCA to verify the similarity of spectroscopy fluorescence data obtained from two cultivations, is possible to notice that both spectral data are similar and no pre-processing is needed to perform a comparison between fermentations, as presented in Fig. 4.

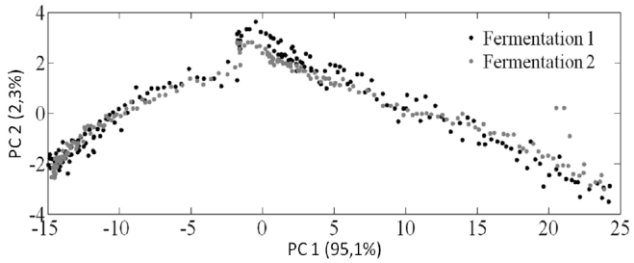

Fig. 4. Principal component 1 versus principal component 2 for both fermentations tests.

In order to effectively evaluate chemometric models based on fluorescence data, information about state variables must be available at the same sample range of fluorescence data. Since, originally, the total amount of off-line data is considerably smaller than fluorescence sampling data, it was used a dynamic model of the fermentative system, suggested by Ranzan (2014), to interpolate the state variables, obtaining off-line data at the same sampling range as fluorescence data.

3. RESULTS AND DISCUSSION

In order to assess the contribution of each modification to the final result, 100 (one hundred) evaluations for each combination were conducted. Each evaluation searched for the best linear model with 3 independent variables (k = 3), since a previous PCA analysis indicated the first 3 principal components describe about 98% of variability. Each evaluation used 50 cycles of 100 ants in the algorithm and all models involved were obtained by least squares method. Since there are 7 possible trail update criteria and 4 model selection criteria, there are 28 possible combinations to be compared in order to find the best method.

For each evaluation, the RMSEC and the selected variables were computed. Also, an exhaustive search was conducted and the 3 components that generated the best model were determinate, as well as the minimum error possible, equal to 0.19 g/L. This information is valuable when quantifying the quality of a model and allows the evaluation of each criterion.

Once the minimum error possible was found through exhaustive search, a good way of comparing the 28 methods is to analyze how many times among 100 evaluations each one found the optimal model. The result is shown in Fig. 5:

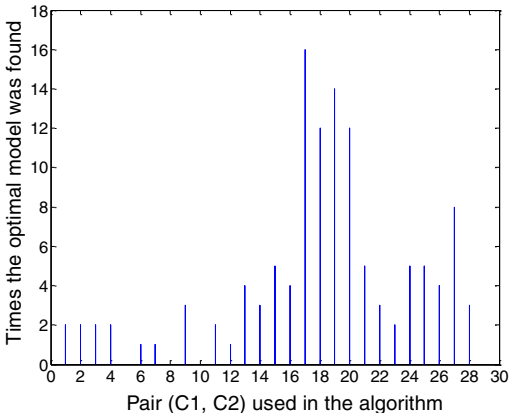

Fig. 5. Number of times each criteria combination (cf. Table 1) found the optimal result, among 100 evaluations.

As can be seen in Fig. 5, the criteria pairs 17, 18, 19, 20 showed the best results, finding the optimal model more frequently than the other cases. Of all 93 evaluations that found the optimal solution (among the 28 conditions), 45% belong to this group. This four pairs have the length of the confidence interval (LCI) as the criterion for trail update. Comparing the pair 17 with the pair 1 taken from literature (both use SSE as C2) it is possible to see the contribution of the modification proposed in this work, once the optimal model was found 8 times more frequently.

The use of different criteria for model selection (C2), however, does not seem to be significant for the final result. This is more evident when considering the group using SSE as C1 (numbers 1, 2, 3, 4): the use of 4 different C2 led to the same result. Since the parameters SSE, R2, Ra2 and RR are all mathematical related and evaluate, in different ways, the aspect aspects of the model, it makes sense the 4 possible criteria for model selections led to similar results.

Th LCI, t-test and F-test parameters has the advantage of providing information about the each component of the model, which results in a better trail update. Therefore, the methods using this kind of analyses as C1 found the optimal model more frequently. The LCI group was responsible for 45% of all evaluations that found the minimal error, followed by the F-test and the t-test group, responsible for 13% and 12%, respectively.

The methods 13-16 also must be highlighted, because although they use the RR parameter as C1, they found the optimal the same amount of times than the F-test methods. In this case, the enhancement this parameter provides in differences between values of R^2 seem to be great enough to compensate the lack of information about the components, resulting in a good trail update as well.

In order to evaluate the quality of each method, it is also convenient to define how far from the minimal error was the majority of evaluations. Fig. 6 shows the error threshold for 90% of the evaluations of each method. For example, using the criteria pair number 1, 90% of all evaluations found an error lower than 0.215 g/L. The goal is to make this threshold the smallest possible, in a way that 90% or more of the evaluations find a model much closer to the optimal one.

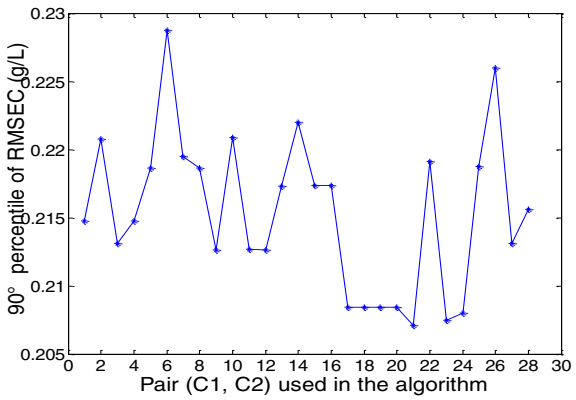

Fig. 6. Value of RMSEC capable of comprising 90% of errors found in all 100 evaluations (i.e., the 90º percentile), for each criteria combination.

According to Fig. 6, the best case in terms of reproducibility is pair 21, which found in 90% of evaluations equal or lower than approximately 0.207 g/L, an error 10% higher than the minimal RMSEC found by exhaustive search. However, the value found for methods 17-20 is very close to the best case, showing great results also in terms of reproducibility.

It is also important to evaluate the prediction of the models obtained. Since the cases using the length of confidence interval as criterion for trail update found the optimal more frequently and presented good reproducibility, this will be the ones used as reference. For this purpose, 2 models were plotted against the experimental data of biomass concentration in fermentation 2 (test group):

- Model containing the spectral pairs [33, 50, 57], which gave the optimal model in calibration phase;
- Model containing the spectral pairs [33, 49, 50], which were the most frequent spectral components chosen by the ACO algorithm using the LCI (criteria 17 to 20) in calibration phase.

The numbers inside brackets refer to one pair emission-excitation used in the fluorescence spectroscopy. To see which wavelengths are used in these pairs go to Fig.4.

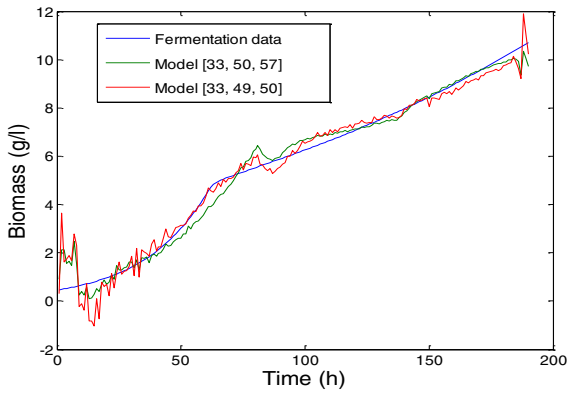

Fig. 7. Comparison between measured data (blue), best model found in calibration phase (green) and most frequent model chosen by ACO using LCI in calibration phase (red).

As shown in Fig. 7, the most frequent model found is very similar to the optimal one. Both optimal and most frequent models can predict biomass concentration with a small RMSEP: 0.42 and 0.54 $g.l^{-1}$, respectively. Although the model is not exactly the same, two of them (33 and 50) are identical. The difference between the models prediction is, therefore, due to the third components (57 for the first model and 49 for the second one), which are in different spectral regions, as can be verified in Fig.2. The algorithm implemented used only linear models, but the use of nonlinear models could improve significantly the application of this method in variable inference.

4. CONCLUSIONS

The Ant Colony Optimization, as others optimization methods, is an important tool for variable selection. However, many variations of this method can be found in literature. This work proposed the alteration of two steps within ACO implementation: the trail update and the model selection. The modifications of the second step, however, produced similar results.

In general, the use of LCI for trail update group showed the best optimization results, being responsible for 45% of all the times the optimum was found. Almost all models obtained by this method had a RMSEC smaller than 0.207 g/L, a value only 10% higher than the minimal error. Also, the most frequent model found by this method produced very good inference result, similar to the optimal one, especially when looking at the components composing both models: 2 of the 3 selected components are the same.

This result is in some way expected, once the length of the confidence interval gives information about each variable composing the model, instead of the full model. The t-test and the F-test, which also provides this kind of information, produced good results as well, although worse than those from LCI group. This fact is likely specific to the experimental data used and could be different if data of different spectroscopy or complexity are used.

REFERENCES

Allegrini, F. and Olivieri, A. C. (2011). A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis. *Analytica chimica acta*, 699, 18-25.

Bohm, G. and Zech, G. (2010). *Introduction to statistics and data analysis for physicists*. Deutsches elektronen-synchrotron.

Brown, S. D., Tauler, R., and Beata, W. (2009). *Comprehensive chemometrics: chemical and biochemical data analysis*. Elsevier.

Cerny, V. (1985). Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *Journal of optimization theory and applications*, 45, 41-52.

Dorigo, M. and Gambardela, L. M. (1997). Ant colony system: a cooperative learning approach to the traveling salesman problem. *Ieee transactions on evolutionary computation*, 1, 53-66.

Geladi, P., Sethson, B., Nyström, J., Lillhonga, T., Lestander, T., and Burger, J. (2004). Chemometrics in spectroscopy: part 2. Examples. *Spectrochimica acta part b: atomic spectroscopy*, 59, 1347-1357.

Gemperline, P. (2006). *Practical guide to chemometrics*. Boca Raton, CRC/Taylor & Francis.

Ghasemi-Varnamkhasti, M., Mohtasebi, S. S., Rodriguez-Mendez, M. L., Gomes, A. A., Araújo, M. C. U., and Galvão, R. K. H. (2012). Screening analysis of beer ageing using near infrared spectroscopy and the successive projections algorithm for variable selection. *Talanta*, 89, 286-291.

Goss, S., Aron, S., Deneubourg, J. L., and Pasteels, J. M. (1989). Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, 76, 579 – 581.

Haltelmann, K., Kollecker, M., Hüll, D., Hitzmann, B., and Scheper, T. (2006). Two-dimensional fluorescence spectroscopy: a novel approach for controlling fed-batch cultivations. *Journal of biotechnology*, 121, 410-417.

Hemmateenejad, B., Shamsipur, M., Zare-shahabadi, V., and Akhond, M. (2011). Building optimal regression tree by ant colony system–genetic algorithm: application to modeling of melting points. *Analytica chimica acta*, 704, 57-62.

Hitzmann, B., T. Pekeler, C. Lindermann, S. Marose and T. Scheper (1998). Chemometric models for the on-line estimation of bioprocess variables from 2-D fluorescencespectra. *7th International Conference on Computer Applications in Biotechnology*. Osaka.

James, G., Witten, D., Hastie and T., Tibshirani, R (2013). *An introduction to statistical learning with applications in R*. Springer, New York .

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671-680.

Mello, P. A. and Pinto, J. C. C. S. (2008). Introdução à modelagem matemática e dinâmica não-linear de processos químicos. *Escola piloto virtual giuliano massaran*. Rio de janeiro.

Mullen, R. J., Monekosso, D., Barman, S., and Remagnino, P. (2009). A review of ant algorithms. *Expert systems with applications*, 36, 9608-9617.

Örkcü, H. and Bal, H. (2011). Comparing performances of backpropagation and genetic algorithms in the data classification. *Expert systems with applications*, 38, 3703–3709.

Ranzan, C., Strohm, A, Ranzan, L., Trierweiler, L. F., Hitzmann, B., and Trierweiler, J. O. (2014). Wheat flour characterization using nir and spectral filter based on ant colony optimization. *Chemometrics and intelligent laboratory systems*, 132, 133-140.

Scheper, T., Hitzman, B., Stärk, E., Ulber, R., Faurie, R., Sosnitza, P., and Reardon, K. F. (1999). Bioanalytics: detailed insight into bioprocesses. *Analytica Chimica acta*, 400, 121-134.

Settle, F. (1997). *Handbook of instrumental techniques for analytical chemistry*. Upper Saddle River, Prentice-Hall.

Silveira, L. (2012). *Development of an advanced control strategy for a Propane Recovery Unit*. Master's dissertation, UFRGS, Porto Alegre, RS, Brasil.

Socha, K. and Dorigo, M. (2008). Ant colony optimization for continuous domains. *European Journal of Operational Research*, 185, 1155-1173.

Solle, D., Geissler, D., Stärk, E., Scheper, E., and Hitzmann, B. (2003). Chemometric modelling based on 2D fluorescence spectra without a calibration measurement. *Bioinformatics,* 19, 173-177.

Stärk, E., Hitzmann B., Schügerl, K., Scheper, T., Fuchs, C., Köster, D., and Märkl, H. (2002). In-Situ-fluorescence-probes: A useful tool for non-invasive bioprocess monitoring. *Advances in biochemical engineering Biotechnology*, 74, 21 – 38.

Sykes, A. O. (1993). An Introduction to Regression Analysis. *Law and Economics Working Paper*, 20, 1-34.

Walpole, R. E., Myers, R., Myers, S. and Ye, K. (2012). *Probability & Statistics for Engineers & Scientists*. 9th ed., Pearson, Boston.

Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. 3rd ed. Elsevier.

Yamuna, R. K. and Ramachandra R. V. S. (1999). Control of fermenters – a review. *Bioprocess Engineering*, 21, 77-88.