

Model Discrimination and Selection in Evolutionary Optimization of Batch Processes with Tendency Models

Ernesto C. Martínez*

INGAR-Instituto de Desarrollo y Diseño (CONICET-UTN)
Avellaneda 3657, S3002 GJC, Santa Fe, Argentina

Abstract

Batch processes increasingly feature products with short market windows that make the development of a detailed kinetic model unattractive in terms of both time and economy. Tendency modeling is already established as a systematic methodology for timely optimization of batch processes using a gray-box model approach. In this work, the problem of effective discrimination among alternative tendency models is addressed using rank correlation methods and pair-wise bisection approach to model concordance. The Kendall tau statistics is used to measure model correlation and concordance with regards to alternative optimum predictions. Less promising tendency models are gradually eliminated using a *softmax* criterion which trades off exploitation with exploration for experiment planning.

Keywords: Tendency models, Rank correlation methods, Model concordance and rivalry, Batch processes, Evolutionary optimization,

1. Introduction

Considering the large uncertainty and variability present in batch environments, the development of an appropriate mathematical representation of the process is a tough challenge. One central concern is how process modeling can be best pursued considering poor knowledge about phenomena involved, sparse and biased measurements of key chemical variables and uncontrollable variations in process behavior from batch to batch. A model for a batch process cannot be entirely knowledge-driven or data-driven alone (Bonvin, 1998). Tendency modeling (Bonvin and Rippin, 1990; Fotopoulos *et al*, 1994,1996; Georgakis, 1995) is a systematic way of combining scarce data with first principles. In order to achieve the goal of optimal operation of batch processes in the face of uncertainty, a number of requirements are imposed on Tendency modeling for it to be successful (Martínez and Wilson, 2003). The first is how data gathering is best biased towards the most profitable section of the operating region bearing in mind that model adaptation (parameters and structure) is simultaneously occurring with evolutionary optimization. This will bring considerations about the influence of model uncertainty on the uncertainty of the resulting optimal

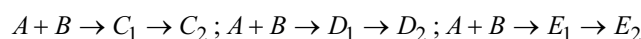
* Author email: ecmarti@ceride.gov.ar

operating condition. Another important issue is about iterating on model parameterization. As the process condition moves closer to the optimal, the local validity of the model can be improved by data selection and their relative weighting, thus enhancing the evolutionary search towards the optimum.

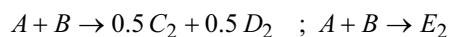
2. Tendency modeling

A “tendency model” is a low order, nonlinear, dynamic model that approximates the stoichiometric and kinetic relationships of a process using the available plant data along with fundamental knowledge of the process characteristics (Georgakis, 1995). The model structure and parameters are incrementally updated as more data becomes available. The main use of the tendency model is to determine a direction towards the optimum. For this to be feasible, the tendency model should be able to extrapolate to operating conditions quite different from those used in the initial development of the model.

As a case study, consider the following 6 reaction, 8 species mechanism, where D_2 is the desired product:



The intermediate species, C_1, D_1 and E_1 , were assumed to be non measurable and, in fact, had completely disappeared by the end of each of the batch runs. Process data was only assumed to be available at the end of the run. The performance of the process is assessed using a productivity index J that rewards the production of the desired product, penalizes the formation of by-products, penalizes the cost of the reactants and solvent and penalizes long batch times. As a result of using the Tendency modeling approach to experimental data obtained from the operating conditions listed in Table 1, the following stoichiometric model was identified and validated (Fotopoulos *et al.*, 1996):



Using Power-law kinetics and assuming the rate orders were equal to the corresponding stoichiometric coefficients three different models were obtained using different data sets. The estimated parameters were the pre-exponential factors and activation energies for the two reactions which made up the structure of the Tendency models are shown in Table 2. The model **M0** is obtained using process data from the first eight runs, whereas **M1** is the model obtained using the first nine runs. The model **S1** in turn is obtained using process data for operating conditions corresponding to runs # 1 through 8 plus run number ten. The operating conditions for run # 9 is the optimal solution predicted by **M0** whereas run # 10 is a “suboptimal” operating condition, yet in the gradient direction predicted by **M0**. Runs # 11 and #12 are the optimal operating conditions as predicted by models **M1** and **S1**, respectively. Run #13 is the actual optimum of the process.

3. Model discrimination

Once some tendency models are available, two crucial questions to decide the operating conditions for the next run are: i) which tendency models are *concordant* and which are *rivals* regarding the improvement direction, ii) how a model is selected accordingly.

Table 1. Data set for the case study

Run #	Batch time [min]	Temperature [K]	Initial Amount A [ml]	Initial Amount B [ml]	$J, \$/\text{day/liter}$
1	400	385	360	200	+3.46
2	400	385	360	300	+12.30
3	400	385	600	200	-4.6
4	400	385	600	300	+6.53
5	400	410	360	200	-4.23
6	400	410	360	300	+1.92
7	400	410	600	200	-12.70
8	400	410	600	300	-5.70
9	81	368	500	411	+6.34
10	240	383	490	331	+23.17
11	92	370	510	406	+16.15
12	132	380	506	420	+31.40
13	172	377.1	491.2	437	+38.10

Table 2. Parameterizations of alternative Tendency models

Parameter	Model M0	Model M1	Model S1
A_1	1.43	40.91	2097.52
E_1	3798.0	6393.8	9900.3
A_2	1.105×10^9	5.462×10^9	3.530×10^{12}
E_2	20124.3	21317.3	26921.4

As a robust guideline for answering the above questions on statistical grounds the following *monotonicity* assumption is proposed here: “the values of the chosen objective function $y(\mathbf{x})$ should exhibit evidence, in a statistical sense, of a definitive improvement as the operating conditions \mathbf{x} move closer to the predicted optimum \mathbf{x}^* .” In mathematical terms this assumption states the existence of a monotonic relationship ϕ defined as follows:

$$y(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{x}^*\|), \quad \phi \text{ is monotonic} \quad (1)$$

The observed data will fit this guideline of a local optimum to varying degrees, so one would like to quantify the degree of fit. A typical nonparametric measure used for this purpose is *Kendall’s correlation coefficient* τ , otherwise known as *Kendall’s tau* (Gibbons, 1993).

3.1. The bisection approach

The core idea is graphically depicted in Fig. 1: if a perpendicular bisector is created between a pair of experimental points \mathbf{x}_1 and \mathbf{x}_2 in the dataset, two tendency models are *concordant* if their predicted optima are both on the same side of the bisector as the point providing the greater $y(\mathbf{x}_i)$. It is easy to imagine creating bisectors for all possible $(n-1)$ pairs in the data set and counting the number of bisectors that satisfy the model concordance criterion. This test of concordance can be carried out for all possible pairs

of points in the data set. The number of satisfied bisectors (concordances) minus the number of unsatisfied bisectors (discordances), divided by the total number of bisectors will exhibit a Kendall's tau distribution. The Kendall statistic for a sample dataset is calculated as

$$\tau = (n_c - n_d) / \frac{1}{2} n(n-1) \quad (2)$$

The underlying idea of the bisection approach is that two concordant tendency models will exhibit highly correlated ranks regarding the increasing distance from the predicted optima to each data point. The concordance testing procedure for two models is summarized in a systematic procedure as follows.

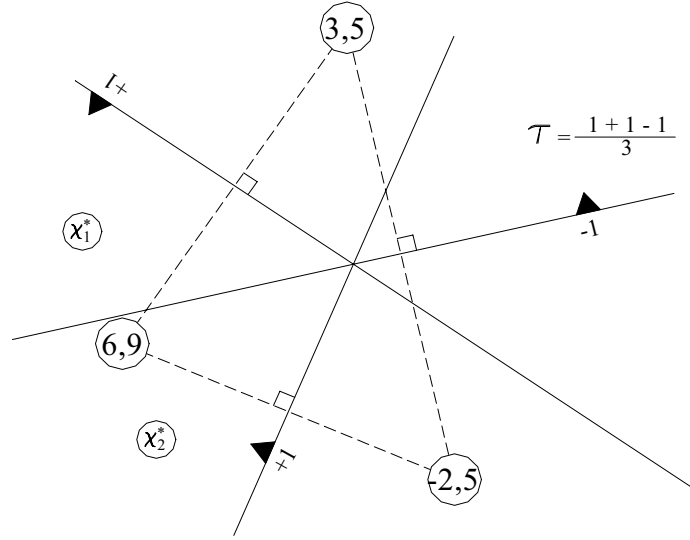


Figure 1. Pairwise bisection approach to model concordance and rivalry

Step 1. For model “1” use its predicted optimum x_1^* along with the data set to rank the values of the objective function $y(x_i)$ in ascending order of $\|x_i - x_1^*\|$. Let's denote θ_i^1 the set of ranks for the objective function thus obtained.

Step 2. Repeat Step 1 for model “2” using its predicted optimum x_2^* ; denote θ_i^2 the resulting set of ranks for the objective function.

Step 3. Apply the Kendall's tau correlation approach to test for independence between θ_i^1 -ranks and θ_i^2 -ranks, against the alternative hypothesis for positive or negative correlation between ranks for the chosen degree of confidence α .

If the hypothesis for positive correlation between θ_i^1 -ranks and θ_i^2 -ranks can be accepted, the tendency models can be considered as *concordant* to the chosen level of significance. Conversely, if the hypothesis for negative correlation between ranks is the one to be accepted with a level of confidence, tendency models are considered as *rival* to this extent.

In Table 3a the ranks obtained for models $M1$ and $S1$ after applying the three steps above are given, whereas in Table 3b depicts the ranks corresponding to $M1$ and the perfect model, i.e. the actual process. Using the ranks of Table 3a, the Kendall's tau is $\tau = -1$ which indicates that, according to the information available, models $M1$ and $S1$ are not concordant but rivals. Conversely, when $S1$ is compared to a "perfect" model (i.e., the one that predict the optimal solution) the resulting Kendall's tau is $\tau = +0.8$, a strong indication of concordance to a high level of significance ($\alpha < 0.001$). It can be concluded from the sample data available that model $S1$ is the model with better chances of providing a reliable estimation of the process optimum.

Table 3(a). $M1$ -ranks against $S1$ -ranks

M1 -ranks	2	1	4	5	3
S1 -ranks	4	5	2	1	3

Table 3(b). $S1$ -ranks against optimal ranks

S1 -ranks	4	5	2	1	3
Optimal ranks	5	4	2	1	3

3.2. Concordance among several models

The bisection approach can be extended to more than two models quite easily. Suppose we have obtained the set of ranks $\theta_i^1, \theta_i^2, \dots, \theta_i^k$ ($k > 2$) for alternative tendency models as in **Step 1** in Fig. 2. A measure of agreement in the prediction of the improvement can be developed using the *Kendall's coefficient of concordance* (Gibbons, 1993). Let's denote the sum of ranks given to the i th data point by the k models as $R_i, i=1,2,\dots,n$.

The sum of ranks for each model is $1+2+\dots+n=n(n+1)/2$, and hence the average rank for each of the n data points is $(n+1)/2$. If there is no agreement among the models and the model assign ranks to data points almost randomly, each rank for each model would be the average rank $(n+1)/2$ and the rank sum for each model would be equal to $(k(n+1)/2)$ because each is the sum of k ranks. The sum of squares of deviations of the actual rank sums around $k(n+1)/2$ is denoted by S and defined as

$$S = \sum_{i=1}^n \left[R_i - \frac{k(n+1)}{2} \right]^2 \quad (3)$$

On the other hand, if there is perfect agreement among the models, each of them would have ranks that are all the same and the rank sums would be some permutation of the numbers $1k, 2k, 3k, \dots, nk$. The sum of squares of deviations around $k(n+1)/2$ in the case of perfect concordance is

$$\sum_{i=1}^n \left[ik - \frac{k(n+1)}{2} \right]^2 \quad (4)$$

The Kendall's coefficient of concordance W for the group of k models is thus defined as the ratio between equations 3 and 4 which, after some algebraic manipulations, can be written as

$$W = \frac{12S}{k^2 n(n^2 - 1)} \quad (5)$$

The sum of squares of deviations under perfect model agreement is the maximum possible value of S and the therefore the value of W ranges between 0 and 1, with increasing values reflecting an increasing degree of concordance among the k models. In Appendix D of Gibbons (1993) a table for different values of n and k of the probabilities for the null hypothesis in the case of perfect agreement are provided. As an example, the hypothesis of concordance between $M1$, $S1$ and optimal ranks is tested. The resulting $W=0.015$ is too small a value that allows concluding that these three sets of ranks are not concordant with a high level of significance ($\alpha < 0.001$).

4. Model Selection

The main advantage of model discrimination is that the number of alternatives for model selection is significantly reduced. Instead of choosing among several models, it is possible to select among groups of concordant models. If all tendency models are concordant, model selection is trivial since all models provide quite similar directions for improvement. When there are M different groups of concordant models, a good exploration strategy is to carry out one experiment per each group choosing randomly within each group which model is used. A more elaborated *softmax* strategy is to calculate the average Q_j of the predicted performance value for group j , and choose the group for experimental planning using the probabilities:

$$e^{Q_j/T} / \sum_{m=1}^M e^{Q_m/T} \quad (5)$$

where T is an exploration parameter which decreases gradually towards zero. High values of T makes equiprobable selecting a model within any of the groups. As exploration is decreased the concordant group with the greatest Q_j is given more preference. As T tends to zero model selection becomes greedy towards the best group.

5. Final remarks

Model discrimination is a very important step so that the most appropriate model prediction is used while evolutionary optimization is being performed. A nonparametric approach based on rank correlation has been proposed to characterize concordance and rivalry between competing tendency models. The use of the Kendall's coefficient of concordance allows resorting to a precise measure for model discrimination and selection among several competing models. A Softmax selection criterion among groups of concordant models is proposed.

References

- Bonvin, D. and D. Rippin.. 1990. Chem. Engng. Sci 45, 3417.
- Bonvin, D., J., 1998. J. Proc. Cont. 8, 355.
- Fotopoulos, J., C. Georgakis and H. G. Stenger Jr., 1994.. Chem. Engng. Sci. 49, 5533.
- Fotopoulos, J., C. Georgakis and H. G. Stenger Jr., 1996. Chem. Engng. Sci. 51, 1899.
- Georgakis, G., 1995. Entropie 194, 34.
- Gibbons J. D., 1993. *Nonparametric statistical methods*, SAGE Publications, London.
- Martínez, E. C. and J. A. Wilson, 2003. Proc. Indian Natn. Sci. Academy 69, 403.