# Hybrid Models Combining Mechanistic Models with Adaptive Regression Splines and Local Stepwise Regression

**ARTICLE** *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · NOVEMBER 2002

Impact Factor: 2.59 · DOI: 10.1021/ie0107744

Belmiro P. M. Duarte

Instituto Politécnico de Coimbra

# Hybrid Models Combining Mechanistic Models with Adaptive Regression Splines and Local Stepwise Regression

**Belmiro P. M. Duarte*,[†] and Pedro M. Saraiva[‡]**

*Instituto Superior de Engenharia de Coimbra, Quinta da Nora, 3030 Coimbra, Portugal, and Department of Chemical Engineering, University of Coimbra, Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal*

This paper introduces a hybrid modeling approach based on the combination of prior knowledge, under the form of mechanistic models, with tools devoted to the extraction of knowledge from operating data. The first module captures first-principles system behavior, whereas the second models output residuals between real data and mechanistic predictions. The empirical module comprises two sequential tools: one to partition the time domain into zones where residuals are well-fitted by time-dependent univariate piecewise linear polynomials and the other to fit local regression models of residuals within such zones, considering process inputs and mechanistic predictions as independent variables. Time domain partitioning is carried out by an adaptive regression splines (ARS) approach, whereas the construction of time-discrete regression models for each of the different zones thus identified is achieved by stepwise regression (SR). The quality of time domain partitioning and the performance of this hybrid modeling approach are evaluated with data obtained from a simulated fed-batch penicillin fermentation process. The ARS module adequately captures the zones where system behavior presents explicit time-dependent trends, and the SR module produces local models with good interpretability. The results obtained show that our hybrid approach outperforms other techniques applied to the same problem, particularly when local regression models include autoregressive terms.

## 1. Introduction

First-principle models, based on fundamental conservation laws, are widely used to address many process engineering tasks and goals.[1] Such mechanistic models represent process behavior through a set of equations that express existing scientific knowledge. They allow for both a certain degree of extrapolation, beyond the regions over which experimental data are available, and the estimation of unmeasured state variables. However, these advantages are sometimes counterbalanced by the inability of first-principles models to forecast real plant behavior with sufficient accuracy and their failure to take full advantage of all available empirical knowledge or data. These drawbacks primarily arise from the assumptions and simplifications introduced in mechanistic models to compensate for the lack of sufficient understanding of all of the relevant physical phenomena taking place in the system and/or the need to simplify the complexity of the resulting mathematical problem formulation.

Alternative modeling techniques that are based on the statistical treatment of operating data have also been explored. These include nonparametric estimators, such as artificial neural networks (ANNs), Fourier series, kernel estimators, and smoothing splines.[2] Another group of commonly used multivariate tools covers empirical parametric methods, including partial least squares (PLS), principal component analysis (PCA), regression analysis, and time-discrete models.[3−5]

Such empirical approaches, also known as “black-box” tools, extract knowledge directly from operating data, under a small number of assumptions about the true underlying process behavior. They are fitted to real data according to a minimization criterion that is usually related to a norm of the differences (“residuals”) between estimates and observations.[6] Purely empirical methods permit only limited extrapolation beyond the domain of the data from which they were derived; they also ignore any mechanistic knowledge that might be available about the process and its underlying physics, thereby potentially leading to unreasonable results (e.g., negative concentration predictions). Moreover, the modeling of multiple input/output systems with black-box tools usually requires large amounts of data for reliable predictions to be obtained.

In view of the above reasons, approaches that combine the properties of mechanistic (“white-box”) models with those of empirical (black-box) techniques, integrating the best of both paradigms, seem to be quite appealing. Such approaches, known as “hybrid” or “gray-box” models, aim to achieve good extrapolation properties, some degree of process behavior rationalization, facility of model development, and focus on relevant phenomenological parameter fitting.

The combination of mechanistic and empirical models can be performed in either serial or parallel arrangements. The serial approach involves a nonparametric estimator fed with operating data used to estimate parameters that are then provided to the mechanistic model. This approach was successfully used by Psichogios and Ungar[7] to model a fermentation process. In the case of dynamic models, outputs from the mechanistic model deriving from previous stages are also fed into the nonparametric module to achieve better-updated parameters through recurrent training procedures. Schubert et al.[8] used this approach to model yeast cultivation. Later, Oliveira[9] followed the same approach

* To whom correspondence should be addressed. Tel.: 351 239 790200. Fax: 351 239 790283. E-mail: nop61212@mail.telepac.pt.

[†] Instituto Superior de Engenharia de Coimbra.

[‡] University of Coimbra.

10.1021/ie0107744 CCC: $25.00
Published on Web 11/27/2002

to control biotechnological processes through a fuzzy expert system used to identify the current phase of operation and perform its hierarchical supervision. Acuña et al.[10] also used a serial structure to model bioprocesses kinetic rate expressions.

Parallel structures generally involve a mechanistic module and a nonparametric estimator, both being fed with the same data simultaneously. While the mechanistic module estimates the system behavior based on first principles, the nonparametric module aims to forecast the corrections (residuals) that have to be added to the mechanistic model predictions to obtain the "true" process behavior. Su et al.[11] proposed a parallel structure to model a complex chemical reactor system through the integration of mechanistic knowledge with a recursive ANN. Thompson and Kramer[12] also applied a parallel design combined in series with a parametric output model and a preprocessing module to a fed-batch penicillin fermentation plant. Shum and Myers[13] described a parallel structure applied to octane control in platforming units. Van Can et al.[14] used a parallel gray-box approach to model and control a laboratory pressure vessel. Duarte and Saraiva[15] have shown that a parallel structure based on mechanistic models and multivariate adaptive regression splines (MARS) can outperform other alternatives. In all of the above cases, the authors reported that hybrid structures provide better interpolation and range extrapolation properties than any of the individual mechanistic or empirical modules that comprise them used alone.

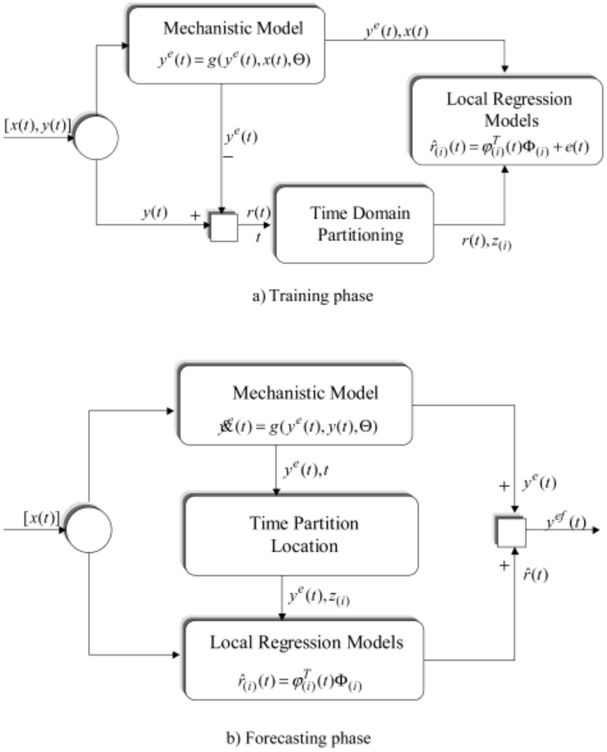
In recent years, chemical engineering researchers have devoted reasonable amounts of effort to model and control hybrid systems whose dynamics depend on discrete/logic and continuous variables.[16] Global models formulated as sets of local models can be viewed as the equivalent of hybrid systems modeling where discrete variables represent the domains of validation for each local relationship.

In this article, we propose and test a parallel hybrid modeling structure comprising a mechanistic model and a nonparametric module, with the latter computing a correction to the predictions of the former. Techniques based on univariate adaptive regression splines (ARS) are used to partition the time domain and then combined with stepwise regression to construct time-discrete local models. The approaches tested and compared also comprise global and local time-discrete models, depending on exogenous and exogenous plus autoregressive terms. Our hybrid modeling structure was applied and used to forecast the concentration of penicillin produced by a simulated fed-batch penicillin fermentation process.

In the next section, we present the suggested hybrid structure, its different modules, and their connections, and we describe the theoretical basis underlying each of the procedures involved. Then, we evaluate and compare the predictive accuracy of the proposed hybrid model, varying the local character of the models and the types of terms involved, against those of purely mechanistic models and purely empirical models, leading to a number of relevant conclusions.

## 2. Proposed Hybrid Structure

The parallel hybrid structure proposed and tested in this paper combines a mechanistic module with a parametric empirical module (Figure 1). The mechanistic model tries to capture system behavior from first-principles relations predicting process outputs $y^e(t)$ from process inputs $x(t)$. To improve the adequacy of this mechanistic approach in modeling the operating data, some of its parameters are fitted by a least-squares nonlinear optimization algorithm. This step leads to the construction of an "optimal" mechanistic model that is adjusted to data but preserves entirely the structure of the assumed first-principles relationships. Then, an empirical parametric model attempts to estimate the difference (residual) $r(t)$, between the true process output (measurement) $y(t)$, and the mechanistic model prediction $y^e(t)$, as a function of $\varphi^T(t)$, which includes covariates $x(t)$ and $y^e(t)$.



a) Training phase

b) Forecasting phase

**Figure 1.** Hybrid model structure.

Consider a set of data acquired during process operation comprising observations of process inputs, $x(t)$, and outputs, $y(t)$. We want to be able to estimate $y(t)$ for a given $x(t)$. In the training phase, inputs $x(t)$ are fed into a mechanistic dynamic model $\dot{y}^e(t) = g(y^e(t),x(t),\Theta)$, where $\Theta$ is a vector of mechanistic model parameters (some of them previously fitted to data), thus leading to corresponding $y(t)$ estimates, $y^e(t)$. The set of residuals $r(t) = y(t) - y^e(t)$ is then fed into the empirical regression module, together with the corresponding $x(t)$ and $y^e(t)$ values. The space of residuals $R$ is first partitioned into time-dependent zones corresponding to intervals locally modeled by splines with the aim of minimizing the global variance, modeling $r(t) = h(t)$ with ARS, where $h(\cdot)$ represents a space of piecewise linear polynomials. The purpose of this task is not to find the time-dependent model of residuals, but rather to partition the time domain into intervals where process behavior is described by well-defined trends, designated here as zones, within which the variance of the residuals with respect to piecewise-spline-based models is minimized. The generic zone $z_{(i)}$ comprises the values of $t$ such that $z_{(i)} = \{t{:}t \geq t_{(i)} \cap t \leq t_{(i+1)}\}$, where $t_{(i)}$ is the $i$th split of time domain. Locally, the inputs of the residuals are expanded to accommodate the influence of independent variables from previous sampling times. That is, for each zone $z_{(i),}$ vectors of expanded inputs ($x_{\mathrm{exp},(i)}$) and expanded predicted outputs ($y^{\mathrm{e}}_{\mathrm{exp},(i)}$) are defined as $x_{\mathrm{exp},(i)} = [x_{(i)}(t), x_{(i)}(t - 1), ..., x_{(i)}(t - nd_{\mathrm{i},(i)})]$ and $y^{\mathrm{e}}_{\mathrm{exp},(i)} = [y^{\mathrm{e}}_{(i)}(t), y^{\mathrm{e}}_{(i)}(t - 1), ..., y^{\mathrm{e}}_{(i)}(t - nd_{\mathrm{o},(i)})]$, where $nd_{\mathrm{i},(i)}$ represents the maximum time delay of inputs in zone $z_{(i)}$ and $nd_{\mathrm{o},(i)}$ represents the maximum time delay of mechanistic predictions that are to be taken into account. Then, local expanded vectors form the final set of independent variables, $\varphi^{\mathrm{T}}_{(i)}(t)$, with $\varphi^{\mathrm{T}}_{(i)}(t) = [x_{\mathrm{exp},(i)}, x^2_{\mathrm{exp},(i)}, ..., x^{no_{\mathrm{i},(i)}}_{\mathrm{exp},(i)}, y^{\mathrm{e}}_{\mathrm{exp},(i)}, y^{\mathrm{e}^2}_{\mathrm{exp},(i)}, ..., y^{\mathrm{e}^{no_{\mathrm{o},(i)}}}_{\mathrm{exp},(i)}]^{\mathrm{T}}$, where $no_{\mathrm{i},(i)}$ is the maximum order of inputs assumed for $z_{(i)}$ and $no_{\mathrm{o},(i)}$ is the maximum order of mechanistic predictions considered. Finally, linear models of $r(t)$ in each of the zones $z_{(i)}$, designated by $r_{(i)}(t)$, are constructed following a stepwise regression procedure, fitting models of the form

$$\hat{\mathrm{r}}_{(i)}(t) = f(\varphi^{\mathrm{T}}_{(i)}(t),\Phi_{(i)}) \tag{1}$$

where $f$ is the space of linear functions, $\hat{r}_{(i)}(t)$ is the estimate of residuals in zone $z_{(i),}$ and $\Phi_{(i)}$ is a vector of local parameters fitted to minimize the square norm between $\hat{r}_{(i)}(t)$ and $r(t)$ for $t \in z_{(i)}$.

The global model obtained for forecasting thus comprises three types of information: (i) subset of $\varphi^{\mathrm{T}}_{(i)}(t)$ variables included in each local model; (ii) regression coefficients for each local model, $\Phi_{(i)}$, $\forall i$; and (iii) time domain knots, $t_{(i)}$, $\forall i$.

The architecture of our hybrid approach for forecasting purposes begins with the feeding of inputs into the mechanistic model, which uses them to estimate $y^{\mathrm{e}}(t)$ values. These estimates, together with the inputs $x(t)$ and time $t$, are fed into the empirical regression module, which locates the zone $z_{(i)}$ to which $t$ belongs and estimates $\hat{r}(t)$ using the corresponding local regression model. Finally, $\hat{r}(t)$ is added to $y^{\mathrm{e}}(t)$ to produce the combined prediction estimate $y^{\mathrm{ef}}(t) = y^{\mathrm{e}}(t) + \hat{r}(t)$ for the process outputs. One should notice that such an approach is directed toward measured output variables, as it is not possible to estimate or model residual components through an empirical module if no measurements are available. Estimator/observer structures to forecast nonmeasured outputs, as a replacement for direct measurements, can be employed, but in such cases, an additional component needs to be modeled by the empirical module, namely, the difference between the estimates and the true values.

Our approach is particularly powerful when the structure of the residuals is time-dependent, resulting in local time inadequacies for the mechanistic model that are different from one time zone to another. Thus, we consider here a fixed mechanistic model structure and complement it with local-time empirical models aimed at capturing the behavior of the residuals in each time zone. Through a detailed analysis of the empirical components added in the several time zones, one can also obtain relevant insights about how to adjust the underlying structure of the mechanistic model so that its prediction performance improves over time.

MARS is a nonparametric kernel estimator technique developed by Friedman[17] to achieve acceptable data representation accuracy through localized adjustments. The model structure has the form of an expansion of product spline basis functions, with the number of basis functions and local parameters automatically derived from the data. The algorithm involves (i) a forward procedure to select certain spline functions, (ii) a backward procedure to prune unnecessary splits until the most parsimonious set is found, and (iii) a smoothing procedure to provide the final model with a certain degree of continuity at the knots. The general form for this model is

$$\hat{Y} = \sum_{m=1}^{M} a_m B_m(X) \tag{2}$$

where $\hat{Y}$ is the estimate; $M$ is the maximum number of basis functions allowed; $a_m$, $m = 1, .., M$, represents constants; $X$ is the vector of independent variables, here also called covariates; and $B_m(\cdot)$ are basis functions defined by

$$B_m(X) = \prod_{k=1}^{K_m} H[s_{km}(X_{\mathrm{v}(k,m)} - T_{\mathrm{k}m})] \tag{3}$$

where $K_m$ is the number of splits that give rise to $B_m(X)$, $H[\cdot]$ represents the step function, $s_{km}$ is $\pm 1$, $X_{\mathrm{v}(k,m)}$ is the vector of independent variables of the $m$th basis function for the $k$th split, and $T_{\mathrm{k}m}$ represents the knots of the splits for the $m$th basis function.

Despite the global character of the nonparametric model originated by MARS, locally, it is built of parametric basis functions and based on recursive partitioning or tree-based regression.[18] The basic principle underpinning this group of adaptive techniques is to split the data iteratively into two regions with respect to the independent variables to minimize a squared-error-based criterion.

The forward procedure starts with the initialization of a maximum number of basis functions to be used, $M$. Then, $B_1(X)$ is set equal to 1, and the counter $m$ is set equal to 2. In each of the loops of $m$, two basis functions, $B_m(\cdot)$ and $B_{m+1}(\cdot)$, are added to the model, and the space containing the $m - 1$ basis functions already included is searched via an internal loop to find the best split of the data. To perform such a search, an enumerative procedure that varies the index $v(k,m)$ with respect to the covariates not already included in the model for all knots and basis functions, represented by $k$ and $m$ values, respectively, is carried out. The space of covariates $X_{\mathrm{v}(k,m)}$ is searched, with univariate split functions $b^+(x_v - l) = \max\{0, +1 \times (x_v - l)\}$ and $b^-(x_v - l) = \max\{0, -1 \times (x_v - l)\}$ built for every eligible knot (data value) $l \in N$, where $x_v$ is the $v$th covariate and $N$ is the number of data values. New interaction basis functions are created by multiplying previous basis functions with truncated linear functions including a new covariate $v$. Internally, a third loop searches the knots at which the model, including new basis functions, leads to the minimum "lack-of-fit", a criterion based on squared error loss. The iterative procedure chooses the knots $T_{\mathrm{k}m}$, the covariates $v$, and the basis functions $m$. The criterion to be minimized is the sum of squared residuals weighted by a factor that penalizes model complexity, resulting in the final overall metric called the general cross-validation (GCV) index, given by

$$\mathrm{GCV}(m) = \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}[Y_i - \hat{Y}(X_i)]^2}{[1 - C(m)/N]^2} \tag{4}$$

where $Y_i$ are the observed output values and $1/[1 - C(M)/N]^2$ represents a model complexity penalty, with

$C(m) = C1(m) + d$, $C1(m) = \text{trace}[B(B^TB)^{-1}B^T] + 1$, and $d = 3$. Additional information on the MARS procedure can be found in the open literature,[17,19−21] with the backward and smoothing algorithms conveniently discussed by Friedman.[17]

Our adaptive regression splines module guarantees the continuity of the model $h(t)$ at the knots $t_{(i)}$. On the other hand, local regression models are trained with the values at the extremes of each zone taken into account, so that continuity of the residuals is achieved throughout the domain. Therefore, we will be using MARS here only for partitioning of the time domain, thus resulting in an application of adaptive regression splines (a particular univariate version of MARS). This is achieved by assuming that $t$ is the unique covariate of a MARS model for residuals $r(t)$. The number of basis functions $m$ is iteratively increased by 2, with the corresponding model and the statistic GCV($m$) computed for each value of $m$. This procedure is stopped when the relative variation in two successive iterations is lower than a user defined tolerance, tol, that is

$$\Delta\text{GCV}(m) = \frac{\text{GCV}(m) - \text{GCV}(m-2)}{\text{GCV}(m)} < \text{tol} \qquad (5)$$

where $\Delta$GCV($m$) is the relative variation due to the increase in the number of basis functions $m$. The model thus generated is made up of $m$ basis functions, $k_m$ knots $T_{km}$, and local regression coefficients $a_m$. Together with the extremes of the time domain, the knots $T_{km}$ give rise to the set of cutting planes, designated as $t_{(i)}$ with $i \in [1, k_m + 2]$, that bound each zone $z_{(i)}$ with $i \in [1, k_m + 1]$.

Our ARS procedure thus intends to achieve a representation of discrete/logic components, establishing the time domain zones described mathematically by different discrete finite values. Then, a different set of algebraic relations, with inputs and mechanistic predictions as independent variables, is associated with each of the discrete time zones obtained earlier. The discrete component captures the time dependency of the residuals, whereas the continuous component captures the interactions between the spaces of time-dependent variables and the space of residuals. The representation of the features of residuals by such a model aims to obtain an acceptable degree of precision without compromising model complexity and interpretability.

Because the construction of local regression models at each time interval is supported by sound statistical criteria, noise present in the measured data is addressed properly. Only significant data structures and relationships are captured in our empirical components, so that no noise or data overfitting occur.

The stepwise regression algorithm uses sweeps of the covariance matrices resulting from $\varphi_{(i)}^T(t) \times \varphi_{(i)}(t)$ to move variables into (forward) and out of (backward) the model. The procedure is iterative, and in each step, $t$-statistics and $F$-statistics are computed for each variable. The incorporation or removal of a particular variable is thus judged by comparing these values with assumed $F$-to-enter ($F_{\text{in}}$) and $F$-to-remove ($F_{\text{out}}$) scores.[22] At each stage, this iterative procedure demands an update operation over the covariance matrix, called sweeping. The sweeping algorithm used here is due to Hemmerle,[23] and the corresponding sweeping operator was introduced by Goodnight.[24]

## 3. Application to a Simulated Penicillin Fermentation Process

We now consider the application of the time partitioning algorithm and hybrid modeling approach introduced in the previous section. The data that we will be considering for that purpose results from simulating a fed-batch penicillin fermentation process. A mechanistic model describing its behavior was proposed by Thompson[25] and used to test a parallel hybrid approach combining mechanistic knowledge with a radial basis neural network,[12] as well as a parallel hybrid approach using mechanistic knowledge and MARS.[15]

**3.1. Simulated Process Measurement Data.** We assume that the true behavior of the process is described by the following model

$$\frac{dB}{dt} = B(\mu - D - c_L) \qquad (6)$$

$$\frac{dS}{dt} = -\sigma B + (S_f - S)D \qquad (7)$$

$$\frac{dP}{dt} = q_p B - P(D + k) \qquad (8)$$

$$\frac{dV}{dt} = F \qquad (9)$$

Here, $B$ is the biomass concentration, $D$ is the dilution factor (given by $F/V$, where $F$ is the feed flow and $V$ is the volume of the reactor contents), $S_f$ is the initial substrate concentration, $S$ is the substrate concentration, and $P$ is the penicillin concentration.

The kinetics of the fermentation is assumed to be described by the following equations

$$\mu = \mu_m \frac{S}{k_x B + 10} \qquad (10)$$

$$c_L = c_{L\max} \frac{B \exp(-S/100)}{k_L + B + 1} \qquad (11)$$

$$\sigma_l = \frac{\mu}{Y_{x/s}} + \frac{q_p}{Y_{p/s}} + m_x \qquad (12)$$

$$m_x = m_{xm} \frac{B}{B + 10} \qquad (13)$$

$$q_p = 1.5 q_{pm} \frac{SB}{4k_p + SB[1 + S/(3k_I)]} \qquad (14)$$

where $\mu$ is the specific growth rate, $c_L$ is the specific cell lysis rate, $\sigma_l$ is the specific substrate consumption rate, $q_p$ is the specific product formation rate, and $m_x$ is the maintenance energy. The values of the parameters in the above expressions are given in Table 1.

**Table 1. Parameter Values for True Fed-Batch Penicillin Fermentation Model[12]**

| parameter | value | parameter | value |
|---|---|---|---|
| $c_{L\max}$ | 0.0084 h$^{-1}$ | $k_x$ | 0.3 |
| $k_L$ | 0.05 | $k_I$ | 0.1 g/L |
| $m_{xm}$ | 0.029 h$^{-1}$ | $q_{pm}$ | 0.004 h$^{-1}$ |
| $k$ | 0.01 h$^{-1}$ | $Y_{p/s}$ | 1.2 |
| $k_p$ | 0.0001 g/L | $Y_{x/s}$ | 0.47 |
| $\mu_m$ | 0.11 h$^{-1}$ | | |

**Table 2. Batches Simulated for Generation of True Process Data**

| batch | conditions |
|---|---|
| 1 | standard ($F = 0.11$ L/h, $S_f = 525$ g/L, $B_i = 5.0$ g/L) |
| 2 | low substrate feed concentration ($S_f = 480$ g/L) |
| 3 | random substrate feed concentration ($S_f$ is varied as a piecewise constant function of time, with its value being changed every 3 h; the value in each interval is chosen from a normal distribution between 475 and 575 g/L with a mean of 525 g/L) |
| 4 | ramp substrate feed concentration ($S_f$ increases linearly from 475 to 575 g/L over the batch) |
| 5 | random feed flow rate ($F$ is varied as a piecewise constant function of time, with its value being changed every 3 h; the value in each interval is chosen from a normal distribution between 0.13 and 0.09 L/h with a mean of 0.11 L/h) |
| 6 | high substrate feed concentration ($S_f = 575$ g/L) |
| 7 | high initial biomass concentration ($B_i = 10$ g/L) |
| 8 | medium initial biomass concentration ($B_i = 7.5$ g/L) |
| 9 | low feed flow rate ($F = 0.09$ L/h) |

The true process measurements were obtained by using eqs 6−14 to simulate a number of batches, each lasting for approximately 200 h. In each batch, the concentrations of biomass, substrate, and penicillin and the volume of liquid inside the reactor were sampled every 3 h, and the corresponding simulated "measurements" were created by adding to them normally distributed noise with zero mean and standard deviations of ±0.52 g/L, ±2.6 g/L, ±0.16 g/L, and ±0.24 L, respectively, for each of the above four output variables. These values correspond to about 1% of the range of variation for each of these variables over a typical batch. Nine different batch scenarios were simulated under the conditions listed in Table 2, thus generating a data set with 612 records. The initial conditions of the reactor, assumed for all batches were $S(0) = S_f$, $B(0) = B_i$, $P(0) = 0.0$ g/L, and $V(0) = 2.0$ L, where $B_i$ is the initial concentration of biomass in the reactant bulk.

**3.2. Mechanistic Process Model.** Because, in practice, we will not be able to derive the mechanistic model that exactly matches process operating data, to illustrate our approach, we will now assume that only partial mechanistic knowledge is available, leading to a mechanistic model, assumed to be known and available for this fed-batch penicillin fermentation process, that has a structure similar to that of the true process model (unknown except for the sake of generating simulated process data), but somewhat simpler. The dynamic relations are the same as before (eqs 6−9), but we consider now for our approximate mechanistic model specific rate kinetics, with eq 10 replaced by

$$\mu = 1.2\mu_m \frac{S}{5k_x + S} \tag{15}$$

thereby rendering the specific growth rate independent of the biomass concentration. Furthermore, the specific substrate consumption rate uses different amounts of substrate and maintenance energy, leading to the replacement of eq 12 by

$$\sigma_l = \frac{\mu}{1.1Y_{x/s}} + \frac{q_p}{0.9Y_{p/s}} + 0.9m_{xm} \tag{16}$$

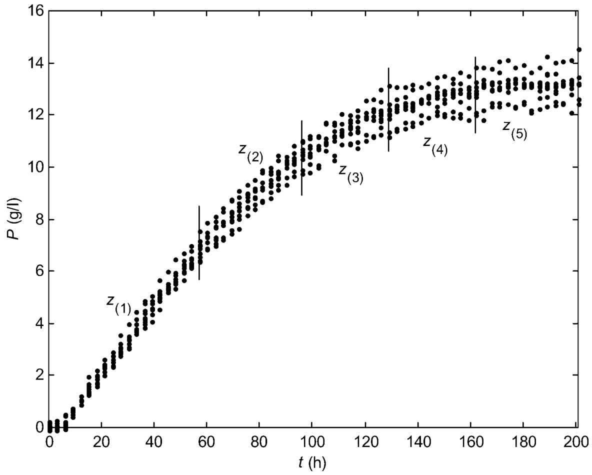


**Figure 2.** Time partitioning of variable $P$.

**Table 3. Optimal Parameter Fitting for Mechanistic Fed-Batch Penicillin Production Model**

| parameter | initial guess ($h^{-1}$) | final adjusted value ($h^{-1}$) |
|---|---|---|
| $\mu_m$ | 0.11 | 1.095 |
| $q_{pm}$ | 0.004 | $1.360 \times 10^{-3}$ |
| $m_{xm}$ | 0.029 | $4.787 \times 10^{-2}$ |
| $k$ | 0.01 | $-2.999 \times 10^{-3}$ |

The assumption that the specific production rate includes a reduced maximum product formation rate and an increased inhibitory effect leads to

$$q_p = 0.9q_{pm}\frac{S}{1.1k_p + S[1 + S/(0.5k_I)]} \tag{17}$$

Finally, cell lysis is completely ignored, thus supporting the assumption that $c_L = 0$. Parameters $\mu_m$, $q_{pm}$, $m_{xm}$, and $k$ in the model comprising eqs 6−9 and 15−17 were fitted to simulated process measurement data generated as described in section 3.1 by using a multiresponse nonlinear optimization procedure, leading to the values listed in Table 3.

Estimates for the several output variables obtained from this model with fitted parameters, for the same scenarios and sampling intervals used to generate the simulated data, correspond to predictions of the substrate, biomass, and penicillin concentrations and the volume content, represented by $S^e$, $B^e$, $P^e$, and $V^e$, respectively.

**3.3. Time Partitioning Algorithm.** To illustrate the behavior of our time splitting algorithm, we now describe its application to the penicillin concentration measurement data ($P$) obtained from the nine simulated training scenarios. The tolerance value, tol, was set to 0.05, and the number of splits obtained was 5, corresponding to $t_{(1)} = 0.0$ h, $t_{(2)} = 57.0$ h, $t_{(3)} = 96.0$ h, $t_{(4)} = 129.0$ h, $t_{(5)} = 162.0$ h, and $t_{(6)} = 201.0$ h, with the first and the last cuts corresponding to the beginning and the end of the batches, respectively. Figure 2 presents the time domain partitioning that corresponds to the zones $z_{(i)}$, $i = 1, ..., 5$, and shows qualitatively the adequacy of the algorithm in the identification of time zones with specific features.

The partition of time with respect to residuals $P - P^e$ used in our hybrid structure is presented in Figure 3. The time splits $t_{(i)}$, $i = 1, ..., 6$, with characteristic trends of residual features are $t_{(1)} = 0.0$ h, $t_{(2)} = 30.0$ h, $t_{(3)} = 96.0$ h, $t_{(4)} = 129.0$ h, $t_{(5)} = 165.0$ h, and $t_{(6)} = 201.0$ h. As expected, time intervals are different from one case to another, because in the former they are based

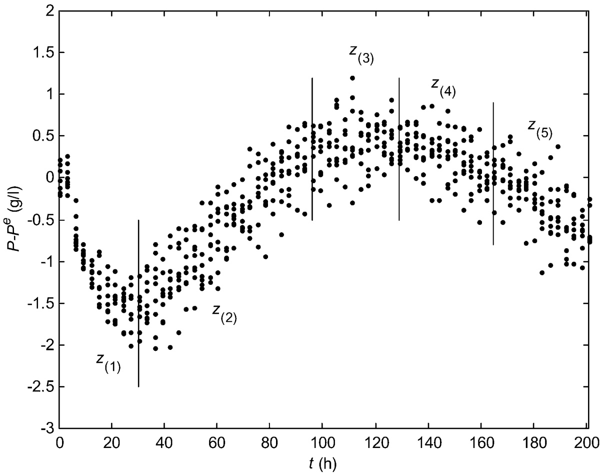


**Figure 3.** Time partitioning of residuals $P - P^e$.

**Table 4. List of Independent Variables Included in the Local Regression Models Used for Predicting $P - P^e$**

| zone | independent variables |
|---|---|
| $z_{(1)}$ | $t$, $P^e(t)$, $V^e(t-2)$ |
| $z_{(2)}$ | $F(t)$, $B^e(t)$, $B^e(t-2)$, $B^{e^2}(t)$ |
| $z_{(3)}$ | $t$, $S_f(t-1)$, $F^2(t-1)$, $S^e(t-1)$ |
| $z_{(4)}$ | $F(t)$, $S^{e^2}(t-2)$, $P^e(t)$ |
| $z_{(5)}$ | $t$, $S_f(t)$, $F(t)$, $S^e(t-2)$ |

on the $P$ profile curves, whereas in the latter, they are associated with the time evolution profiles for the $P - P^e$ residuals, i.e., the assumed mechanistic model's lack of fit across time. Although we concentrate here on the prediction of $P$, the same approach might also be adopted to predict any of the other outputs. For instance, if one were interested in predicting the substrate concentration, $S$, then the time domain would be partitioned with respect to the residuals $S - S^e$, leading to a different set of time intervals, related this time to the $S - S^e$ residual time profiles.

The adjustment of local empirical models within each time zone was achieved by forward stepwise regression, assuming $F_{in} = 3.857$ and $F_{out} = 2.713$, corresponding to probabilities for type II errors of 0.05 and 0.10, respectively.[26] The independent variables used to fit local models for the penicillin concentration ($P$) were time $t$; inputs $S_f$, $B_i$, and $F$; and mechanistic predictions $S^e$, $B^e$, $P^e$, and $V^e$, with a maximum time delay allowed for each independent variables of 2 and a maximum order also of 2. Table 4 lists the independent variables included in the best local models found for each zone of the time space, using a hybrid structure where the empirical modules do not comprise autoregressive terms. Once again, a similar approach can be adopted to build local empirical models associated with the prediction of any of the other plant outputs, aside from $P$.

Therefore, univariate ARS is used to identify suitable time intervals but not to locally model the residuals or variables themselves, because this task is achieved through a stepwise regression analysis. The SR procedure is used to find local models of low complexity, thus building parsimonious structures with good interpretability that differ from one time zone to the next (according to the underlying assumed mechanistic model lack-of-fit structure within each of the time intervals). Time-dependent cubic splines are sufficiently accurate to build local models when small partitions with nonsignificant autocorrelation and/or cross-correlation functions are found. However, in many situations, the available mechanistic model residual structure in a given time interval reflects the absence of additional relationships between the predicted outputs and process inputs, mechanistic estimates or autoregressive terms, which are therefore fed into our SR modules and included in them whenever such terms are found to provide statistically significant added prediction capabilities. On the other hand, variables that do not contribute significantly to this goal are not included in the final local empirical components built for each particular time zone. As a particular case, one might end up getting cubic splines models for a given time zone in case the SR module includes no variables other than time in its structure.

**Table 5. Batches Used for Model Testing and Comparison**

| batch | conditions |
|---|---|
| 1 | standard ($F = 0.11$ L/h, $S_f = 525$ g/L, $B_i = 5.0$ g/L) |
| 2 | low substrate feed concentration, medium initial biomass concentration and medium feed flow rate ($S_f = 475$ g/L, $B_i = 7.5$ g/L, $F = 0.10$ L/h) |
| 3 | high random substrate feed concentration ($S_f$ is varied as a piecewise constant function of time, with its value being changed every 3 h; the value in each interval is chosen from a normal distribution between 525 and 575 g/L with a mean of 550 g/L) |
| 4 | high substrate feed concentration ($S_f = 550$ g/L) |
| 5 | extrapolation: low feed flow rate ($F = 0.08$ L/h) |
| 6 | extrapolation: high substrate feed concentration, initial high biomass concentration and low flow ($S_f = 615$ g/L, $B_i = 12.5$ g/L) |

**3.4. Hybrid Modeling Test and Comparison.** The capability of the hybrid modeling structure presented in section 2 to predict the transient behavior of penicillin concentration was tested over six new fermentation batches, described in Table 5, that were not used before for training purposes. The first four involve conditions within the ranges employed to train the models (cf. Table 2), and the last two involve some degree of extrapolation: batch 5 has values of $F$ below the lower limit of the simulated training data, whereas the values of $S_f$ and $B_i$ in batch 6 exceed the corresponding upper limits.

The predictive performances of hybrid approaches based on four different empirical modeling strategies were then compared with those of purely mechanistic or empirical alternatives: Approach 1 consists of the mechanistic model with adjusted parameters. Approach 2 corresponds to time domain partitioning combined with local regression models, as a purely empirical modeling approach; and the remaining approaches are based on a hybrid parallel structure similar to the one exploited in section 2. Approach 3 uses the mechanistic model combined with global MARS, without time partitioning or the inclusion of past data, and inputs and mechanistic predictions as covariates for the MARS module. Approach 4 is based on the mechanistic model combined with a global time-discrete model of the ARX type[5] for the residuals. Approach 5 corresponds to the one introduced here, and thus combines the mechanistic model with time domain partitioning and local regression models for the residuals. Finally, approach 6 is similar to approach 5 but involves autoregressive effects of error on local regression models of residuals, assuming that the expanded vector of past residuals in zone $z_{(i)}$, $r_{\exp,(i)}$, is defined as $r_{\exp,(i)} = [r_{(i)}(t-1), r_{(i)}(t-2), ..., r_{(i)}(t - nd_{e,(i)})]$, where $nd_{e,(i)}$ represents the maximum time delay of autoregressive terms, so that $\varphi_{(i)}^{T}(t)$ com-

**Table 6. Characteristics of the Approaches Tested and Compared**

| approach | characteristics |
|---|---|
| 1 | mechanistic model with adjusted parameters |
| 2 | empirical model<br>time domain partitioning performed with MARS<br>local regression models of the variable to predict trained with SR<br>local models comprising autoregressive and exogenous effects of past sampling times |
| 3 | hybrid model<br>mechanistic model + global MARS model<br>past sampling times are ignored |
| 4 | hybrid model<br>mechanistic model + global ARX model<br>error model comprising autoregressive and exogenous effects of past sampling times |
| 5 | hybrid model<br>time domain partitioning performed with MARS<br>local regression models of error trained with SR<br>local models of error comprising exogenous effects of past sampling times |
| 6 | hybrid model<br>time domain partitioning performed with MARS<br>local regression models of error trained with SR<br>local models of error comprising autoregressive and exogenous effects of past sampling times |

prises additional rows that correspond to $[r_{\exp,(i)}, r_{\exp,(i)}{}^2, ..., r_{\exp,(i)}{}^{no_{e,(i)}}]^{T}$, where $no_{e,(i)}$ is the maximum order considered. Because the maximum time delay and order of autoregressive terms were also set at 2, $P(t) - P^e(t)$ was modeled in every time zone by taking into account the following additional terms: $P(t-1) - P^e(t-1)$, $P(t-2) - P^e(t-2)$, $[P(t-1) - P^e(t-1)]^2$, and $[P(t-2) - P^e(t-2)]^2$. The main characteristics of all of the above approaches considered and tested are listed in Table 6.

The first approach has four inputs ($t$, $F$, $B_i$, and $S_f$) and one output ($P$). In approach 2, the past values of $P$ are additional inputs for the empirical component. In approaches 3 and 5, the empirical component has four additional inputs ($V^e$, $S^e$, $B^e$, and $P^e$) and the output is $P - P^e$. Approaches 4 and 6 consider residuals $P - P^e$ as the output of the empirical model, with the error between the data and the mechanistic prediction in past sampling times as an additional input. Approach 2 comprises time domain partitioning with respect to the penicillin concentration, $P$, whereas, for approaches 5 and 6, time domain partitioning is performed with respect to residuals $P - P^e$. The optimal architecture for the MARS model used in approach 3 was found using a cross-validation algorithm. The maximum number of basis function in the model of approach 3 was set to 44, and the maximum number of interactions between independent variables was set equal to 8. The number of time partition zones found for approaches 5 and 6 is equal to 5.

For each testing scenario, we evaluated the predictions of each model against the true simulated process data, computing the corresponding prediction error with $n = 68$ points. Tables 7 and 8 and Figure 4 present performance statistics for each approach, including the mean, $\bar{r}$, and standard deviation, $\sigma_r$, for the prediction error, defined as

$$\bar{r} \equiv \frac{1}{n}\sum_{k=1}^{n}(P_k^{\text{true}} - P_k^{\text{predicted}}) \tag{18}$$

and

$$\sigma_r \equiv \sqrt{\frac{\sum_{k=1}^{n}(P_k^{\text{true}} - P_k^{\text{predicted}} - \bar{r})^2}{n-1}} \tag{19}$$

The absolute average values of these performance indexes appear in the seventh row of each table, and the last row shows the ratio of this average to the best (i.e., smallest) average obtained among all of the modeling approaches considered.

Our results illustrate the superior performance of hybrid strategies, in terms of both the mean and standard deviation of the prediction error, over that of purely empirical or mechanistic strategies. The approaches based on mechanistic models together with local regression and global MARS present better forecasting performances. The quality of prediction for this

**Table 7. Prediction Error Mean Values**

| batch | approach 1 | approach 2 | approach 3 | approach 4 | approach 5 | approach 6 |
|---|---|---|---|---|---|---|
| 1 | −0.3502 | −0.0159 | 0.0387 | 0.0424 | −0.0190 | −0.0133 |
| 2 | 0.1432 | 0.0880 | 0.0280 | 0.0320 | 0.0125 | 0.0135 |
| 3 | −0.2905 | 0.0337 | 0.0290 | 0.0624 | 0.0416 | 0.0376 |
| 4 | −0.4713 | 0.0431 | 0.0258 | 0.0805 | −0.0017 | −0.0065 |
| 5 | 0.1722 | −0.2337 | 0.0235 | −0.0335 | −0.0014 | 0.0076 |
| 6 | −0.9387 | −0.1955 | −0.0593 | 0.0127 | −0.1145 | −0.0815 |
| abs. average | 0.3994 | 0.1017 | 0.0340 | 0.0439 | 0.0325 | 0.0267 |
| SAMV[a] | 14.789 | 3.812 | 1.277 | 1.647 | 1.192 | 1.000 |

[a] SAMV = scaled average mean value.

**Table 8. Prediction Error Standard Deviations**

| batch | approach 1 | approach 2 | approach 3 | approach 4 | approach 5 | approach 6 |
|---|---|---|---|---|---|---|
| 1 | 0.6454 | 0.1672 | 0.1643 | 0.1737 | 0.1706 | 0.1630 |
| 2 | 0.7067 | 0.2060 | 0.1573 | 0.1894 | 0.1753 | 0.1505 |
| 3 | 0.8355 | 0.1586 | 0.2550 | 0.1837 | 0.2603 | 0.1899 |
| 4 | 0.6559 | 0.1766 | 0.1848 | 0.1761 | 0.1841 | 0.1842 |
| 5 | 0.6617 | 0.3264 | 0.2459 | 0.2192 | 0.1824 | 0.1866 |
| 6 | 0.6309 | 0.2815 | 0.2051 | 0.1710 | 0.1919 | 0.1668 |
| average | 0.6894 | 0.2194 | 0.2021 | 0.1855 | 0.1941 | 0.1735 |
| SASD[a] | 3.973 | 1.264 | 1.165 | 1.069 | 1.119 | 1.000 |

[a] SASD = scaled average standard deviation.

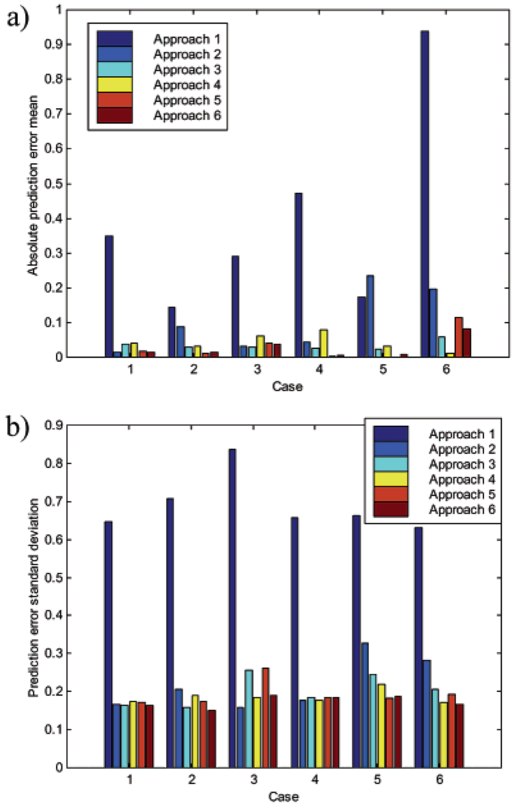


**Figure 4.** Comparison of performance for the approaches tested.

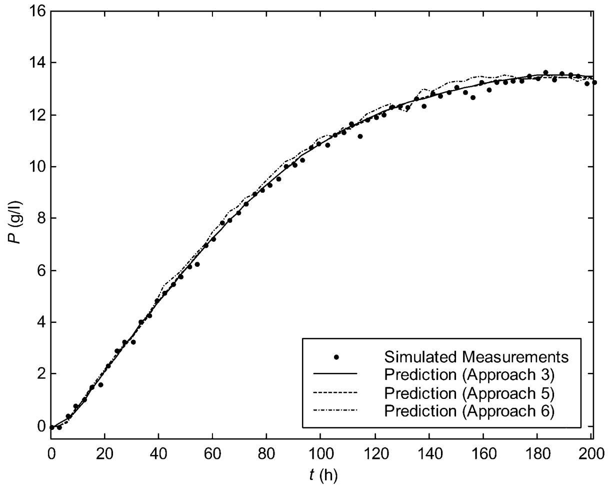


**Figure 5.** Comparison of model predictions for batch 3.

type of strategy is further improved by including the autoregressive terms in local error models, as shown by the performance associated with approach 6. Approach 4, which also involves a global ARX model, presents good performance regarding the standard deviation but lower quality relative to the absolute mean value. This is due to instability caused by autoregressive terms included in the model, with similar attenuated behavior found for approach 6. The better approaches capture all of the features of process dynamics, as the standard deviation of error has the same magnitude as the white noise added to generate simulated output measurements (±0.16 g/L). Furthermore, hybrid approaches present better performance in extrapolation (cases 5 and 6). Figure 5 compares the predictive performance of approaches 3 (mechanistic model combined with a global MARS model), 5 (mechanistic model combined with local regression models neglecting autoregressive terms in error models), and 6 (mechanistic model combined with local regression models involving autoregressive terms in error models) for the third batch of testing data. One can see that, in this example, all three of the top approaches provide quite good predictions for the simulated data sets, with the zone where larger lack-of-fit performance is obtained corresponding to the end of the penicillin production batch (zones 4 and 5 of time partitioning), as illustrated in Figure 6.

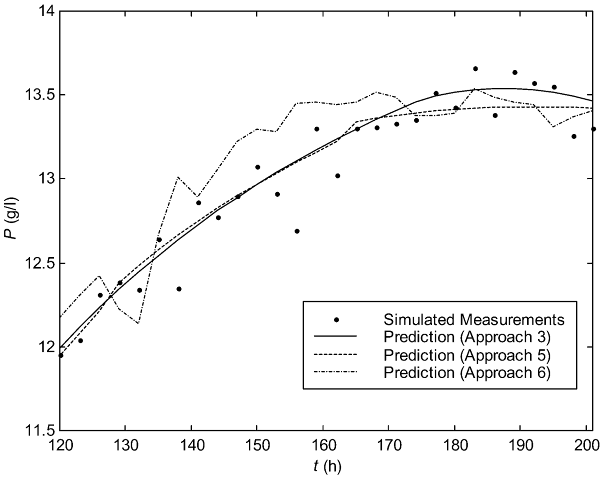


**Figure 6.** Comparison of model predictions for batch 3 ($t \in [120, 201]$ h).

The CPU time required by approach 5 in the training phase was 152 s, with 99.79% of this time being allocated to the nonlinear procedure used to fit the mechanistic model parameters. The efficiency of the training algorithm is thus highly dependent on the initial estimates assumed for such parameters, with the CPU time needed to partition the time domain and build local regression models being marginal (less than 1 s of CPU time). All of these values were obtained with a 733-MHz CPU processor with 256 MB of RAM.

## 4. Conclusions

This paper presents a hybrid modeling approach for chemical processes that combines a mechanistic modeling component, based on first-principles relations, together with an empirical component. The latter attempts to capture the discrepancy between the predictions of the mechanistic component and process data. The two modules are combined in parallel, with the mechanistic module being responsible for describing the dynamic features of the process from first principles, while the empirical module tries to overcome discrepancies between measured output values and mechanistic predictions. The empirical module comprises a time domain partitioning procedure to delimit zones where residuals are described by piecewise linear polynomials, performed by univariate ARS, and the construction of local regression models considering only exogenous or exogenous and autoregressive terms through a forward stepwise regression procedure.

In transient processes, dynamic mechanistic models account for temporal aspects of behavior. However, most of the traditional hybrid modeling structures describe the discrepancies of data relative to mechanistic predictions regardless of time. On the other hand, the approach introduced here considers the effect of past time instants in building the local empirical component associated with each particular time interval. Our time domain is thus split and local regression models fitted to describe residuals for each of such particular time zones (with different complexities and empirical model structures associated with different time intervals, as is statistically required). Providing the expansion of the vector of independent variables to account for autoregressive and exogenous terms, a stepwise regression algorithm is used to generate local models with good interpretability, not achieved with classical ARX time-discrete models.

The hybrid approach presented here is independent of the precise form of its empirical component, and thus we examined a total of six different alternatives applied to the same problem and sets of data, covering empirical components such as a global time-independent MARS approach, a global time-discrete ARX model, a set of local regression models accounting for exogenous terms established by stepwise regression, and a set of local regression models accounting for exogenous and autoregressive terms also obtained through stepwise regression. The results thus obtained indicate that the latter approach leads to the best predictive capabilities that were found.

The empirical module composed of the parallel hybrid structure used in this paper can be based on different predictive tools (e.g., artificial neural networks and multivariate statistical models), and the comparison of performance might reveal other optimal combinations. Nevertheless, the battery of cases and approaches that we tested is rich enough to illustrate the advantages of hybrid tools, particularly those using local regression models and convenient time domain partitioning. In general, hybrid modeling tools seem to outperform other approaches, with their empirical component resulting in increased forecasting performance over that of purely mechanistic models with adjusted parameters. On the other hand, the superiority of a hybrid model over a purely empirical model cannot always be guaranteed: if the mechanistic component represents a particularly bad approximation to the true physical behavior of the process, one might be better off with a purely empirical model. However, the different scenarios and approaches covered here do show the potential and advantages that can, in general, derive from the adoption of hybrid models in the prediction of the behavior of many interesting chemical processes, both for interpolation and extrapolation purposes.

## Acknowledgment

The authors acknowledge financial support provided by FCT through research project POCTI/1999/EQU/32647.

## Literature Cited

(1) Bequette, B. W. *Process Dynamics Modeling, Analysis and Simulation*; Prentice Hall: Upper Saddle River, NJ, 1996.

(2) Eubank, R. *Spline Smoothing and Non-Parametric Regression*; Marcel Dekker: New York, 1988.

(3) Van Overschee, P.; DeMoor, B. *Subspace Identification of Linear Systems: Theory, Implementation, Applications*; Kluwer Academic Publishers: New York, 1996.

(4) Pearson, R. K. *Discrete-Time Dynamic Models*; Oxford University Press: New York, 1999.

(5) Ljung, L. *System Identification—Theory for the User*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 1998.

(6) Haykin, S. *Neural Networks—A Comprehensive Foundation*; Prentice Hall: Upper Saddle River, NJ, 1994.

(7) Psichogios, D. C.; Ungar, L. H. A Hybrid Neural Network—First Principles Approach to Process Modeling. *AIChE J.* **1992**, *38*, 1499.

(8) Schubert, J.; Simutis, R.; Dors, M.; Havlík, I.; Lubbert, A. Hybrid Modeling of Yeast Production Processes—A Combination of a Priori Knowledge on Different Levels of Sophistication. *Chem. Eng. Technol.* **1994**, *17*, 10.

(9) Oliveira, R. F. Supervision, Control and Optimization of Biotechnological Processes Based on Hybrid Models. Ph.D. Dissertation, Martin-Luther Universität, Halle-Wittenberg, Germany, 1998.

(10) Acuña, G.; Cubillos, F.; Thibault, J.; Latrille, E. Comparison of Methods for Training Grey-Box Neural Network Models. *Comput. Chem. Eng.* **1999**, *23*, S561.

(11) Su, H.-T.; Bhat, N.; Minderman, P. A.; McAvoy, T. J. Integrating Neural Networks with First Principles Models for Dynamic Modeling. In *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes (DYCORD+92)*; Balchen, J. G., Ed.; Pergamon Press: 1992.

(12) Thompson, M. L.; Kramer, M. A. Modeling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE J.* **1994**, *40*, 1328.

(13) Shum, S. K.; Myers, D. R. Intelligent System Applications: A Technology Licensor's Perspective. In *Intelligent Systems in Process Engineering: Part II: Paradigms from Process Operations*; Stephanopoulos, G., Han, C., Anderson, J. L., Wei, J., Eds.; Academic Press: New York, 1997.

(14) Van Can, H. J. L.; Hellinga, C.; Luyben, K. C. A. M.; Heijnen, J. Strategy for Dynamic Process Modeling Based on Neural Networks and Macroscopic Balances. *AIChE J.* **1996**, *42*, 3403.

(15) Duarte, B. P. M.; Saraiva, P. M. Hybrid Modeling through the Combination of Mechanistic Models and MARS. In *Proceedings of the 6th IFAC Symposium on Dynamics and Control of Process Systems*; Elsevier Science: New York, 2001.

(16) Branicky, M. S. Studies in Hybrid Systems, Modeling, Analysis and Control. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1995.

(17) Friedman, J. H. Multivariate Adaptive Regression Splines. *Ann. Stat.* **1991**, *19*, 1.

(18) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth: Belmont, CA, 1984.

(19) De Veaux, R. D.; Psichogios, D. C.; Ungar, L. H. A Comparison of Two Nonparametric Estimation Schemes: MARS and Neural Networks. *Comput. Chem. Eng.* **1993**, *17*, 810.

(20) Sekulic, S.; Kowalski, B. R. MARS: A Tutorial. *J. Chemom.* **1992**, *6*, 199.

(21) Chen, V. C. P. Applications of Orthogonal Arrays and MARS to Inventory Forecasting Stochastic Dynamic Programs. *Comput. Stat. Data Anal.* **1999**, *30*, 317.

(22) Kennedy, W. J.; Gentle, J. E. *Statistical Computing*; Marcel Dekker: New York, 1980.

(23) Hemmerle, W. J. *Statistical Computations on a Digital Computer*; Blansdell Publishing Company: Waltham, MA, 1967.

(24) Goodnight, J. H. A tutorial on the SWEEP operator. *Am. Stat.* **1979**, *33*, 149.

(25) Thompson, M. L. System Analysis, Control and Optimization of the Fed-Batch Penicillin Fermentation. M.Sc. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1984.

(26) Montgomery, D. C.; Runger, G. C. *Applied Statistics and Probability for Engineers*; John Wiley & Sons: New York, 1994.

*Received for review* September 17, 2001
*Revised manuscript received* July 22, 2002
*Accepted* October 18, 2002

IE0107744