

PCA-BASED METHOD OF IDENTIFICATION OF DOMINANT VARIABLES FOR PARTIAL CONTROL

Jobrun Nandong^{1*}, Yudi Samyudia¹ and Moses O. Tadé²

¹Department of Chemical Engineering, Curtin University of Technology, Sarawak Malaysia, CDT 250, 98009 Miri, Sarawak, MALAYSIA

²Department of Chemical Engineering, Curtin University of Technology, GPO Box U1987, WA 6845, AUSTRALIA

*Email: jobrun.n@curtin.edu.my

ABSTRACT

Since the early use of automatic control, the Partial Control strategy has frequently been adopted in complex chemical processes having more process variables than manipulated variables. The key idea of Partial Control is to find the dominant variables which can be controlled to constant setpoints and in turn leads to acceptable variations in the operating objectives in the face of external disturbances occurrence. Although the idea seems simple to understand, the identification of the dominant variables can be a daunting task where presently this is largely done based on extensive process knowledge and experience. In this paper, we present a novel methodology to identify the dominant variables based on Principal Component Analysis. The method can greatly facilitate the implementation of Partial Control strategy because it does not require extensive process experience and knowledge. The effectiveness of the methodology is demonstrated based on its application to a complex extractive fermentation process.

INTRODUCTION

The growing trend in process integrations for improving profit and reducing waste has led to increasing challenges in modern process control. Stephanopoulos and Ng (2000) highlighted two key issues to be resolved in modern process control today: (1) translation of implicit operating objectives to sets of feedback-controlled variables, and (2) explicit formalization of the *control philosophy* of the overall plant – i.e. control structure design problems. With respect to the second issue, the Feedback Optimizing Control Structure, and its variant Self-Optimizing Control Structure and Partial Control concept offer a formal framework, within which one can address the control structure problems in a systematic and theoretically-founded manner.

Partial Control Structure (PCS) concept is common in industry because it is straightforward to implement, robust, cost-effective, and easy to understand. The applications of PCS strategy date back to the early use of automatic process control when technology limitations and cost factors, made it necessary to find only a few simple measurements and actuators to control the process (Tyreus, 1999). The key challenge in the implementation of PCS rests on the difficulty in identifying the so-called dominant variables. Currently, this is done through the use of extensive process knowledge and experience i.e. lack of systematic tool to identify dominant variables. In connection to this limitation, recently Nandong et al (2009) proposed novel technique which can be used to identify the dominant variables without any need for extensive knowledge and experience. In this paper, the aim is to extent the concept described in

(Nandong et. al., 2009) by proposing a generalized problem formulation for partial control such that it could be solved easily via the PCA-technique. Additionally, in this paper, we provide a formal definition for dominant variable and criteria for finding the dominant variables set, which is referred to as *dominant variable criteria*.

PARTIAL CONTROL CONCEPT

Lets the plant to be controlled is given by:

$$\dot{X}(t) = F(X(t), U(t)) \quad (1)$$

Where the system state vector is $X \in R^{n_x}$, input variable vector (including the disturbances) is $U \in R^{n_u}$ and F is generally a nonlinear function of its arguments. The output vector $Y \in R^{n_y}$ can be expressed as:

$$Y(t) = G(X(t), U(t)) \quad (2)$$

Lets the measured output vector $Y_p \in R^{n_p}$ be a vector of process variables (excluding input variables), which define the process specifications (or operating objectives) and all state variables of interest where $Y_p \subseteq Y$ i.e. $n_p \leq n_y$. Note that, Y_p can be given by the nonlinear relationship as:

$$Y_p(t) = G_p(X(t), U(t)) \quad (3)$$

Because the variables in Y_p define the control objectives (e.g. stability, product specifications, etc), it is important to control them at specified setpoints Y_p^{ss} . Let the U_{mv} be the manipulated variable vector where $U_{mv} \subseteq U$, then depending on the number of available manipulated variables in U_{mv} , one can either use exact control or partial control strategy. According to Mayuresh et al. (2000) Partial Control can be defined as:

Definition 1: Partial Control

The system described by Eq. 1 and Eq. 2, without any constrains on U_{mv} is said to be partially controllable if the performance variable Y_p can be moved to and maintained within an acceptable range of an arbitrarily prescribed set point Y_p^{ss} , starting from an arbitrary initial point, by an appropriate (possibly non-unique) choice of the steady-state value of U_{mv} . Thus the control should be able to ensure that $Y_{p,min}^{ss} < Y_p < Y_{p,max}^{ss}$ in the face of external disturbances occurrence.

Generalized Problem Statement

We propose a more generalized and direct representation of a performance measure ϕ_p as following:

$$\phi_p = F_p(U, Y, B) \quad (4)$$

Here, $B \in R^{n_p}$ is a vector of process parameter and F_p is a function of its arguments. Note that in this generalized form, the performance measure can be directly one of the elements in the measured output set Y_p (i.e. $\phi_p = y_p$) where $y_p \in Y_p$ e.g., y_p could be the specified purity in the distillate. Additionally, ϕ_p can be an implicit function of the process variables and parameters e.g. optimum profit.

Now, Eq. 4 can be explicitly written in terms of dominant and minor variables as:

$$\phi_p = F_p(U_{D,p}, Y_{D,p}, B_{D,p}, U_{M,p}, Y_{M,p}, B_{M,p}) \quad (5)$$

Here, $U_{D,p}, Y_{D,p}$ and $B_{D,p}$ correspond to the sets of inputs, outputs and parameters that have dominant effects on ϕ_p and thus, referred to as *dominant variables*. Meanwhile the $U_{M,p}, Y_{M,p}$ and $B_{M,p}$ are sets of inputs, outputs and parameters that have only minor contributions to ϕ_p and referred to as *minor variables*.

Let $\Omega_p = \{U_{D,p}, Y_{D,p}, B_{D,p}\}$ and $\Psi_p = \{U_{M,p}, Y_{M,p}, B_{M,p}\}$ are the sets corresponding to dominant and minor variables respectively. Furthermore, assuming that the contributions of the dominant and minor variables to the performance measure ϕ_p can be combined linearly so, Eq. 5 can be further written as:

$$\phi_p = F_{D,p}(\Omega_p) + F_{M,p}(\Psi_p) \quad (6)$$

Where $F_{D,p}$ and $F_{M,p}$ are functions describing the contributions of dominant and minor variables sets respectively to ϕ_p . The variation in the performance measure ϕ_p from the steady-state value as a result of external disturbance occurrence can now be written as:

$$\Delta\phi_p = \Delta F_{D,p} - \Delta F_{M,p} \quad (7)$$

$$\Delta F_{D,p} = F_{D,p}(U_{D,p}, Y_{D,p}, B_{D,p}) - F_{D,p}(U_{D,p}^{ss}, Y_{D,p}^{ss}, B_{D,p}^{ss}) \quad (8)$$

$$\Delta F_{M,p} = F_{M,p}(U_{M,p}, Y_{M,p}, B_{M,p}) - F_{M,p}(U_{M,p}^{ss}, Y_{M,p}^{ss}, B_{M,p}^{ss}) \quad (9)$$

The notations $U_{D,p}^{ss}, U_{M,p}^{ss}, Y_{D,p}^{ss}, Y_{M,p}^{ss}, B_{D,p}^{ss}$ and $B_{M,p}^{ss}$ correspond to steady-state values.

Now, a generalized representation involving multiple performance measures can be expressed as:

$$\Phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} = \begin{bmatrix} F_{D,1}(\Omega_1) \\ F_{D,2}(\Omega_2) \\ \vdots \\ F_{D,n}(\Omega_n) \end{bmatrix} + \begin{bmatrix} F_{M,1}(\Psi_1) \\ F_{M,2}(\Psi_2) \\ \vdots \\ F_{M,n}(\Psi_n) \end{bmatrix} \quad (10)$$

Given the Eq. 10, the key problem in partial control is to identify the set of dominant variables (Ω_p) corresponding to the performance measure ϕ_p where $p = 1, 2, \dots, n$. In this paper, we demonstrate the use of novel technique (Nandong et al., 2009) which can be conveniently applied to solve this complex problem.

Dominant Variable

From Eq. 6 and Eq. 10, the dominant variable can be clearly defined for a particular performance measure ϕ_p for $p = 1, 2, \dots, n$.

Definition 2: Dominant Variable

The dominant variable set Ψ_p with respect to a performance measure ϕ_p (Eq. 6) is defined as the smallest sub-set of variables that can be formed from set $\Sigma = U \cup Y \cup B$ (i.e. $\Psi_p \subset \Sigma$) such that when they are controlled to constant values in the presence of disturbances occurrence, the magnitude of variation in the performance measure (i.e. $|\Delta\phi_p|$) will be within an acceptable limit $\Delta\phi_{p,max}$ i.e. $|\Delta\phi_p| \leq \Delta\phi_{p,max}$. Or from Eq. 7, when the dominant variables are kept constant, then $\Delta F_{D,p} = 0$ and

thus, $\Delta\phi_p = \Delta F_{M,p} \leq \Delta\phi_{p,max}$ otherwise Ψ_p is not the dominant variable set for the given performance measure ϕ_p .

Remark 1:

Note that, the dominant variable set with respect to the performance measure ϕ_i is not necessarily the same as the dominant variable set with respect to another performance measure ϕ_j , i.e. $\Omega_i \neq \Omega_j$ for $i \neq j$. Moreover, some of the elements in Ω_i might be in Ω_j as well, so $\Omega_i \cap \Omega_j \neq \emptyset$ and in this case the performance measures are correlated with each other. On the other hand, in case that $\Omega_i \cap \Omega_j = \emptyset$ then the performance measures are completely uncorrelated with each other.

METHODOLOGY

Principal Component Analysis

The dataset reduction can be effectively performed using a well-established multiprojection analysis technique called *Principal Component Analysis* (PCA). Assuming that the data matrix X has m rows (observations) and n columns (variables), then application of PCA will decompose X into the sum of outer product of vectors t_i and p_i plus a residual matrix E :

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_k p_k^T + E \tag{11}$$

Where $k \leq \min(m,n)$, the vector t_i is known as scores and the vector p_i is called loadings. While the scores contain the information on how the samples or observations relate to each other, the loadings contain the information on how variables are interrelated.

The key objective of applying PCA to a matrix of dataset X is to reduce the large set of interactive variables into much smaller set of factors, and yet these factors should be able to adequately describe the variability in the original dataset. For more details about the PCA, interested readers could refer to Wise and Gallagher (1996).

Proposed Concept of Dominant Variable Identification

Fig. 1 illustrates the successive applications of PCA on the original dataset X – i.e. successive dataset reductions. The first application of PCA (i.e. 1st level of dataset reduction) on the original dataset X (Fig. 1a) generates two uncorrelated sub-groups (orthogonal groups) of smaller datasets X_1 and X_2 . The subscript 1 is to indicate the subset of variables and performance measures that occupy the 1st and 3rd quadrants. And the subscript 2 is to indicate those variables and performance measures that occupy 2nd and 4th quadrants. Here, X_1 and X_2 are called the first level of reduced datasets from the original dataset X .

First application of PCA on X (Fig. 1a) divides the dataset into two sub-datasets:

$$X = [X_1 \ X_2] \tag{12}$$

From Fig. 1a, the sub-dataset X_1 contains the performance measure of interest. But, at this level of dataset reduction probably it remains unclear which of the variables (out of 7) that strongly correlate with the performance measure i.e. dominant variables. So, another PCA is applied to X_1 (Fig. 1b) and generates even smaller two sub-datasets:

$$X_1 = [X_{11} \ X_{12}] \tag{13}$$

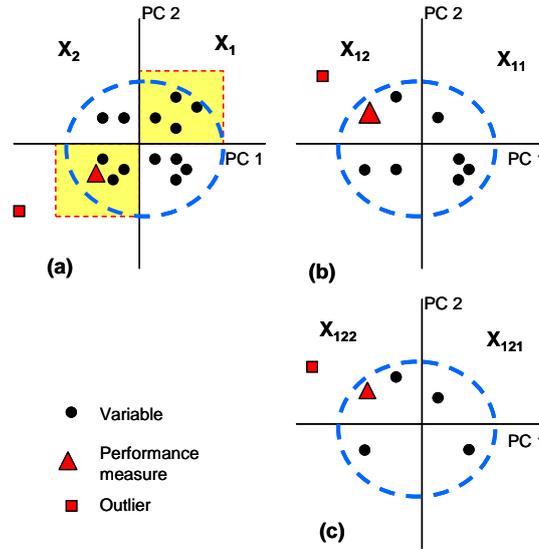


Fig. 1: Generalized concept of dataset reduction using PCA to identify dominant variables

Now, from Fig. 1b it can be clearly seen that the performance measure of interest is contained within X_{12} . Notice that, the number of variables which are potentially the dominant variables for the performance measure is now reduced from 7 to 4. Another PCA can be applied to X_{12} to further reduce this sub-dataset as:

$$X_{12} = [X_{121} \ X_{122}] \tag{14}$$

Finally, from Fig. 1c one can identify that the performance measure is now within X_{122} and only two variables that are correlated with the performance measure. Thus, these two variables are deemed to be the dominant variables for the performance measure.

Remark 2:

Notice that, the PCA plots for the successive dataset reductions only involve two principal components i.e. PC1 and PC2. Plotting of PC1 and PC2 is considered adequate if the sum of variances of PC1 and PC2 is at least 70% of the total variances in the dataset.

Dominant Variable Criteria

In order to ensure that the dominant variables are contained within the reduced dataset (regardless of the level of dataset reduction), the following 3 criteria must be observed:

1. *At least one performance measure in the dataset.*
2. *At least one variable in the dataset.*
3. *At least one outlier exists within the dataset.*

If one or more of these criteria are not fulfilled, then the set of variables obtained cannot be guaranteed as the dominant variable set. Thus following the previous illustration in Fig. 1, notice that all of the 3 criteria are fulfilled through all of the stages in the dataset reductions.

Algorithm

The detailed procedure for obtaining the dominant variables by the PCA-based technique is described in Nandong et al (2009). Here only important steps are outlined as following:

Step 1: Performance Measures Specifications

Determine all the required operating objectives or performance measures e.g. product yield, productivity and optimum profit.

Step 2: Design of Experiment

The key aim is to generate data of the process plant of interest. This can be done by applying the concept of design of experiment (e.g. factorial design). At each run, the process plant is perturbed by applying inputs perturbations. The required values (e.g. outputs, inputs, models, etc) are calculated for each inputs perturbation set at both of the nominal and the steady-state perturbed operating levels.

Step 3: Identification of Dominant Variable

Successive dataset reductions are applied on the dataset until the dominant variables are identified.

Step 4: Control Structure Design Decisions

Here, the tasks involve the selections of controlled variables from the available dominant variables, manipulated variables for the available valves or streams and control connections. Note that, for cases where the dominant variables are closely correlated with among other, only subset of the dominant variables are necessarily selected as controlled variables.

CASE STUDY: CONTINUOUS EXTRACTIVE ALCOHOLIC FERMENTATION

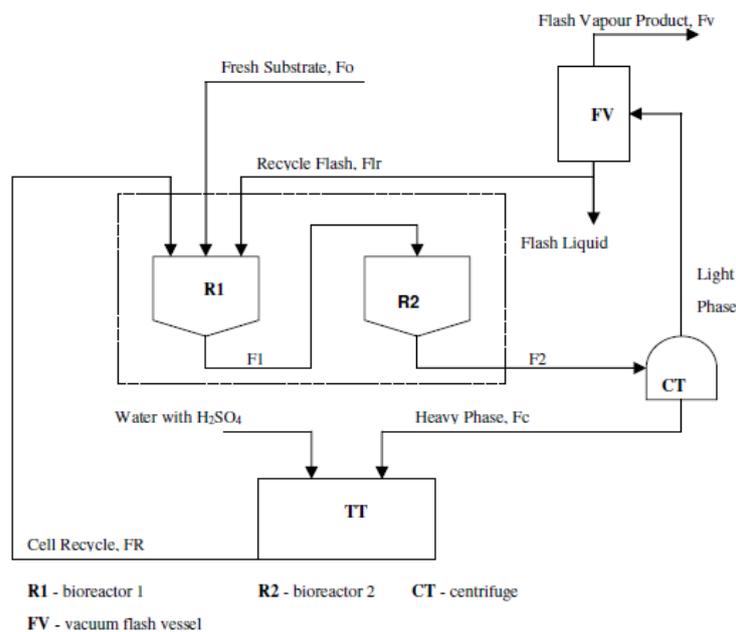


Fig. 2: Two-stage continuous extractive alcoholic fermentation process

Figure 2 shows the two-stage continuous extractive alcoholic fermentation process, which consists of two bioreactors in series (R1 and R2), a centrifuge (CT) to separate cells from broth liquid, a vacuum flash vessel (FV) to partially remove ethanol from the broth and a treatment tank (TT) in which the cells are treated before being recycled back to R1. There are two recycle loops which are (1) flash recycle, and (2) cell recycle.

Nandong et al. (2006) conducted optimization and controllability study on this process where the kinetic data was obtained from Costa et al. (2001). Note that, this system is highly nonlinear. To identify the dominant variables for developing partial control of this process, the previously proposed algorithm is applied.

Step 1: Performance Measures Specifications

Three performance measures are chosen: (1) ethanol yield (Yield), (2) substrate conversion (Conv), (3) and ethanol productivity (Prod). These performance measures are calculated at steady-state levels i.e. nominal and perturbed operating levels. It is interesting to note that Yield and Productivity have opposite trends, i.e. increase in Yield would reduce the Productivity. So, our objective is maintain their optimal trade-off values despite the disturbance occurrence.

Step 2: Design of Experiment

The inputs selected for the design of experiment are (1) fresh substrate flowrate (F_0), (2) fresh substrate concentration (S_0), (3) cell recycle ratio, and (4) flash liquid recycle ratio. Nominal values of F_0 , S_0 , R and r are $100\text{m}^3/\text{hr}$, $120\text{kg}/\text{m}^3$, 0.225 and 0.27 respectively. These nominal values correspond to the optimal trade-off between yield and productivity (Nandong et al., 2006). The magnitude of input perturbation is $\pm 20\%$ of these nominal values. Based on factorial design there are $2^4 = 16$ number of experimental runs i.e. this corresponds to 16 perturbed operating levels. In total there are 17 runs including the one at the nominal operating level.

Step 3: Identification of Dominant Variables

Successive dataset reductions are performed using PCA on the original dataset which consists of 24 elements (4 inputs, 10 outputs, 6 process parameters and 4 performance measures) and 17 observations (1 nominal operating level and 16 perturbed operating levels). The output variables are viable cell (X_v), substrate (S), and ethanol (Et) concentrations plus bioreactor temperature (T) and bioreactor liquid level (L). The parameters are (1) growth, (2) substrate consumption, and (3) ethanol formation rates i.e. rx , rs , and rp respectively.

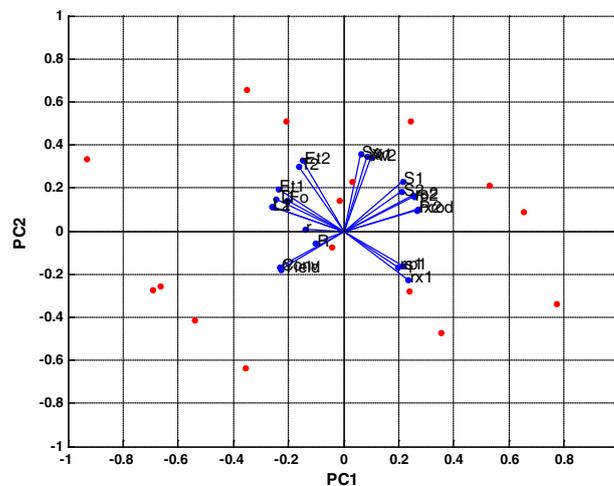


Fig. 3: PCA plot for dataset X: variances of PC1 + PC2 = 80%

Fig. 3 shows the PCA plot for X dataset for the first two principal components (PC1 and PC2). From the plot (enlarged version not included), it can be seen that all of the performance measures Yield, Conv and Prod in X_1 i.e. none in X_1 . There are 9 variables that may correlate with the performance measures are S_o , Xv_1 , Xv_2 , S_1 , S_2 , rs_2 , rp_2 , rx_2 and R . Notice that, we can see a few outliers in 3rd quadrant, thus X_1 fulfils all of the 3 dominant variable criteria – so dominant variables are among the 9 variables. Now, we apply PCA on X_1 in order to further reduce the number of variables.

Fig. 4 shows the PCA plot for X_1 sub-dataset. From the plot, one can identify that all of the performance measures are in X_{12} sub-dataset i.e. Yield and Conv in 2nd quadrant and Prod in 4th quadrant. The result indicates that while Yield and Conv are positively correlated with each other, the Prod is deemed to be negatively correlated with Yield and Conv. Thus, this is consistent with previous report in (Nandong et al., 2006; Costa et al., 2001). The variables that are in X_{12} are R , rx_2 , S_1 and S_2 . So we have reduced the number of variables from 9 to 4 where these 4 variables are most likely the dominant variables for the Yield, Conv and Prod. In other words, if these variables are kept constant, the steady-state variations or offsets of these performance measures will be guaranteed to be small in the face of disturbance occurrence. As these performance measures relate to steady-state performance and so also the dominant variables corresponded to them. Thus, no guarantee however that the dynamic performance is good when the dominant variables are controlled. So, to ensure good dynamic performance, it suggested to define the dynamic performance and similar procedure can be used to find the dominant variable/s for the dynamic performance. In this way, the dynamic performance degradation can be guaranteed to be small when its dominant variable/s is controlled. Note that, the search for the dominant variable/s for dynamic performance is not conducted in this study. Notice that, there is an outlier (observation 6) in 4th quadrant. Thus, X_{12} fulfils all of the 3 dominant variable criteria – so the obtained dominant variable sets is consistent.

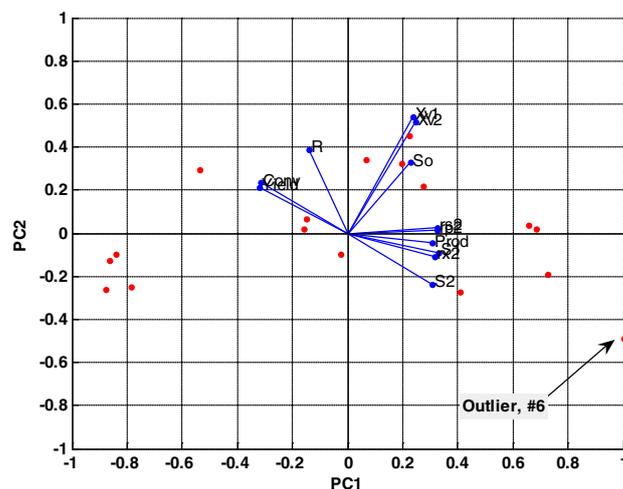


Fig. 4: PCA plot for X_1 sub-dataset: variances of PC1 + PC2 = 84%

Step 4: Control Structure Design Decisions

From the 4 dominant variables, only two are selected as controlled variables which are S_2 and rx_2 . We do not need to control S_1 because this is tightly correlated with S_2 , thus

controlling S_2 will lead to indirect control of S_1 . For the cell recycle ratio (R), since this is an input, we can choose it as one manipulated variables owing to its strong influence on both selected controlled variables. Another, manipulated variable is the fresh substrate flowrate (F_0).

Using the RGA analysis, the pairings are R - S_2 and F_0 - rx_2 . Linear PI controllers are used for both control loops. In this study, the PI parameters tuning are not optimized. In this case, we adopt simple tuning procedure based on Ziegler-Nichols for each loop followed by detuning to reduce controllers aggressiveness. The performance of the decentralized 2×2 MIMO control is tested against the step change in fresh substrate concentration (S_0) with magnitude of 30 kg/m^3 i.e. 25% of nominal value.

Dynamics Simulation

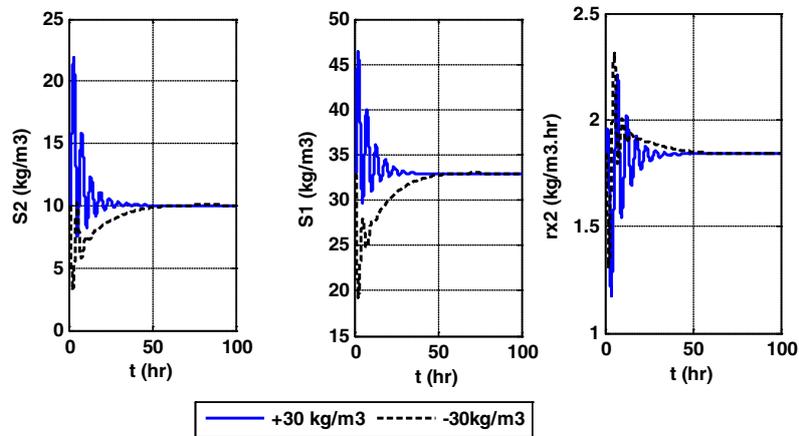


Fig. 5: Closed-loop responses to step change in S_0 : $\pm 30 \text{ kg/m}^3$

Fig. 5 shows the closed-loop responses in S_1 , S_2 and rx_2 when subject to step disturbance change in S_0 . Notice that the responses in S_1 and S_2 exhibit similar shape implying closed correlation between them i.e. as anticipated from the previous analysis. Although the disturbance magnitude is severe, the system remains stable. But, the response to step down in S_0 is significantly slower than that to step up in S_0 . This shows the system is highly nonlinear.

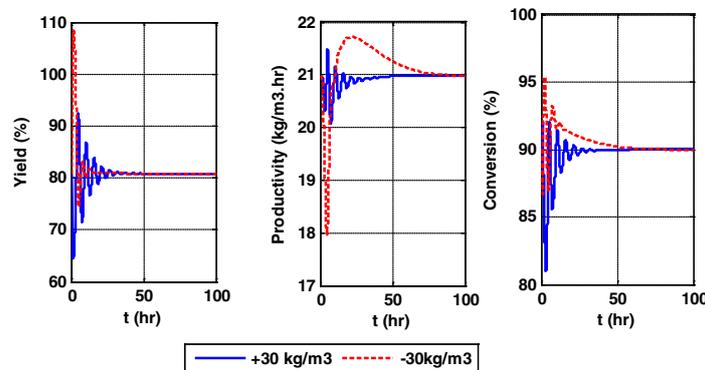


Fig. 6: Yield, Conv and Prod responses to step change in S_0 : $\pm 30 \text{ kg/m}^3$

Fig. 6 shows the closed-loop responses of Yield, Conv and Prod under the S_0 disturbance. To evaluate the effectiveness of the partial control structure, one can look

at the steady-state offsets in Yield, Conv and Prod. The steady-state offsets in Yield, Conv and Prod in this case is less than 0.05%, thus showing the effectiveness the partial control design in term of reducing the steady-sate offset. The dynamic response can be further improved by retuning of the PI controllers used, which are not optimized in this study.

It is important to note that the partial control developed in this case is only a basic design. Thus, its performance can be further improved by applying some of the PID enhancement techniques such as, cascade and ratio control. Furthermore, to make partial control design fully functional for large-scale plant, we generally need to add inventory control e.g. level control. Overall, although the partial control design adopted in this case study is very simple, it shows acceptable performance.

CONCLUSION

In this paper, we present the general formulation for partial control problem and demonstrate the use of the PCA-based technique to identify the dominant variables. This new technique can help engineers in the implementation of partial control strategy without any need for rigorous experience and knowledge. The effectiveness of the technique is demonstrated based on its application to complex highly nonlinear extractive fermentation process. From the set of dominant variables identified, only two of them are controlled using simple linear PI controllers. Despite the nonlinearity of the system, the partial control design shows acceptable performance even without optimizing the controller tunings. Interestingly, the steady-state offsets in the Yield, Conversion and Productivity are less than 0.05% when subject to large disturbance (i.e. 25% of nominal value) in fresh substrate concentration (S_0). In conclusion, the PCA-based technique can be used as a tool to overcome the current limitation in the application of partial control, which arises from the heavy reliance on process experience and knowledge.

ACKNOWLEDGEMENT

This work is part of the eScienceFund Project (no: 02-02-07SF0001).

REFERENCES

- Costa, A.C., Atala, D.I.P., Maugeri, F., Maciel, R., (2001), *Process Biochem.*, 37, 125
- Mayuresh, V.K., Shinnar, R., Rinard, I., Morari, M., (2000). *AIChE*, 46(12), 2456
- Nandong, J., Samyudia, Y., Tadé, M.O., (2006), *Chemical Product Process Modeling*, 1, Article 8
- Nandong, J., Samyudia, Y., and Tadé, M.O., (2009), *Ind. Eng. Chem. Res.* (under review)
- Stephanopoulos, G., Ng, C., (2000), *J. Proc. Cont.*, 10, 97
- Tyreus, B.D., (1999), *Ind. Eng. Chem. Res.* 38, 1432
- Vinnicombe, G. (2000) *Uncertainty and Feedback*. Imperial College Press: Singapore
- Wise, B.M., Gallagher, N.B. (1996), *J. Proc. Cont.*, 6, 329