

# Statistical Quality Monitoring of Chemical Processes: A Stochastic Approach

ALICIA TINNIRELLO

Departamento Ciencias Básicas

Universidad Tecnológica Nacional Facultad Regional Rosario

Zeballos 1341 (2000) Rosario. Santa Fe

ARGENTINA

*Abstract* :-The aim of this work is the parameter estimation of correlated serial disturbance measurements of output data structure from continuous chemical processes. The chemical process is simulated with HYSYS simulator to obtain the dynamic response of outputs when random disturbances implicate the process inputs. Then, stochastic models are applied to take into account these unexplained disturbances, with the scope of monitoring the product quality. As an illustration, the approach is applied to the water concentration output of a Propylene Glycol plant composed by a CSTR (continuous stirred tank reactor) and a distillation unit, when stochastic perturbations are introduced into the feed molar flows. The proposed statistical approach proves to be satisfactory to predict the correlative data structure and in addition, it allows to forecast future variable values after identifying a given data pattern.

*Key-Words*: Quality process control, disturbance, stochastic models, time series, forecasting

## 1 Introduction

The typical industrial chemical process consists of an interconnected network of individual processing units. The inherent inertial components present in it, the transportation lag due to flow through connecting pipes and the fact that many process variables of importance are sampled and measured frequently, are some of the elements, among others, that promise that current observations are strongly related to past observations. Then, measurements of controlled process variables are seriously correlated.

The most important effect of serial correlation in data processing is that all the underlying assumptions (independence of observations; normally and independently distributed random data) behind traditional charting methods of statistical process control (SPC) break down. Specifically, these conventional control charts will give misleading results in the form of many false alarms. Therefore, statistical techniques used as tools to monitor the variability of a given process and to anticipate and/or identify process changes before adverse results occur, are generally inadequate for monitoring quality of continuous flow processes [1][5][6].

## 2 Case Study: Propylene Glycol Plant

The methodology is applied to a propylene glycol plant [7]. In this process, propylene oxide and water react to produce propylene glycol in a continuous stirred tank reactor (CSTR). The products of the reaction are then separated in a distillation tower. The process is closed loop stable by means of four control loops controlling the reactor temperature, reactor liquid level, column 9 tray temperature and condenser pressure.

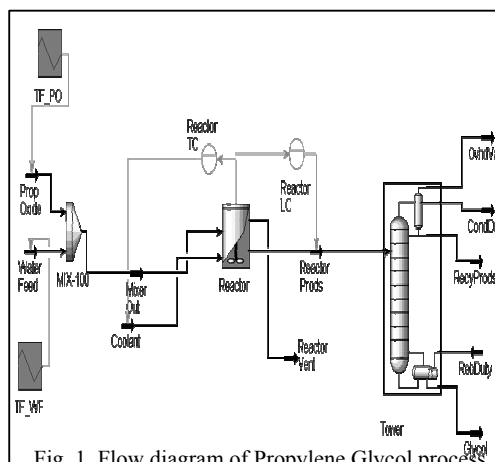


Fig. 1. Flow diagram of Propylene Glycol process

Dynamic responses of process variables are obtained by dynamic simulation, when random perturbations are introduced on the feed molar flows.

### 2.1 Preliminary Statistical Analysis

In this study, the time series analyzed, is the result of 200 observations per each minute of simulated water output concentration. The statistical information about this data is obtained by the measurements of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF).

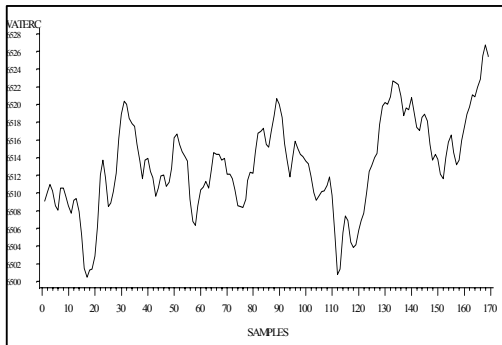


Fig.2 Distillation column outlet water concentration

The disturbances presence produces a tendency of many processes variables to drift away from their target values. These disturbances may be of deterministic nature, such as steps, exponential rises, etc., to a new level value, or they may be of a more random nature. Most often, we cannot partition the process variable into systematic component and random component. This is known as a state variable description. Instead, a realization of a process is observed. In such instances, we can mathematically model these process variables by difference equations or time series models [2] [6].

### 2.2. Dynamic Stochastic Models

The general mixed ARIMA( p, q ) process is:

$$\phi_p(B) (1-B)^d Y_t = \theta_q(B) a_t \quad (1)$$

where the stationary operator AR is

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2)$$

and the invertible operator MA is

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (3)$$

with the introduction of the backshift operator  $B^j y_t = y_{t-j}$ .

The p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively, and  $\{ a_t \}$  is a white noise process with mean zero and variance  $\sigma_a^2$ . (We assume  $a_t$  is normally distributed, that is,  $a_t \sim N(0, \sigma_a^2)$ .)

$\{ \phi_i \}$  and  $\{ \theta_i \}$  are the coefficients of MA (p) and AR(q) polynomials.

More generally, the differenced series  $(1-B)^d$  follows the stationary ARMA(p,q) process. For some  $d > 0$  where  $\phi(B)$  is a stationary autoregressive operator.

Thus, an homogeneous nonstationary series can be reduced to a stationary series by taking a suitable difference of general series. Equation (1) is known as an Autoregressive Integrated Moving Average (ARIMA) model of order (p,d,q) [11].

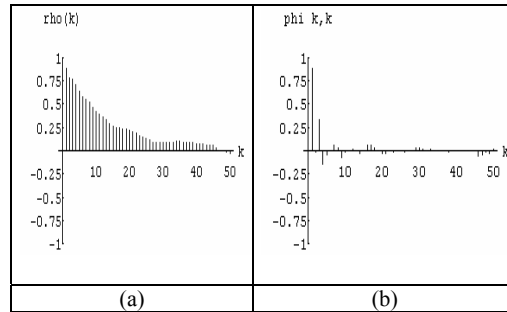


Fig.3: (a) Sample ACF. (b) Sample PACF

It can be observed that both sample ACF and sample PACF tail off and the observation pattern is characterized by correlative structure. The target is to find an adequate time series model, which represents the variability of simulated data referred to water concentration outputs. The methodology here proposed take into account the stochastic behavior of the process outputs which includes: (i) to model the correlative structure; (ii) to use the chosen model to remove the autocorrelation from data; (iii) to apply control charts to the residuals and eventually (iv) to forecast future variable values after identifying a data pattern [9].

### 3 Modeling the Data Correlative Structure

Water outputs are analysed in attempt to identify a data pattern and to determine whether data is correlated. The sample autocorrelation function (ACF) and the

sample partial autocorrelation function (PACF) of the original series (Fig. 3), gives an indication of the orders of AR or MA processes an also, an idea about a trend, nonconstant variances, or other nonnormal and nonstationary phenomena. This understanding provides a basis for postulating data transformation.

An appropriate transformation is applied, to lead them at the stationary form. Differencing of order  $d=1$  was a suitable difference and the proposed models were estimated by SAS System [10] and checked with the help of various diagnostic statistics that indicate how well the models fit data.

Possible candidate models:

Model A:

$$\{\phi_i\} \{0.748525, -0.472845\} \text{ and} \\ \{\theta_i\} \{-0.0706612\}, \text{ where } \sigma_a^2 3.043 \cdot 10^{-10}$$

Model B:

$$\{\phi_i\} \{0.66263, -0.43047\}, \text{ and} \\ \{\theta_i\} \{-0.042946\}, \text{ where } \sigma_a^2 2.961 \cdot 10^{-10}$$

### 3.1 Model fitting procedure

The adequacy of each model is checked by: looking at the behavior of the residuals of a fitted ARIMA( $p, d, q$ ) process (if them appears to be a stationary random sequence), through the Portmanteau test and the chi-square statistics used in the test for lack of fit, computed using the Ljung-Box formula .

The  $\chi^2$  test statistics for the residuals series indicates whether the residuals are uncorrelated (white noise) or contain additional information that might be used in a more complex model. The hypothesis that the residuals are uncorrelated shown by the  $\chi^2$  tests, can or cannot be rejected.

### 3.2 Diagnostic checking

The conditional least square estimation method used allows us to know the model parameter estimates (the estimated value for each parameter and the standard error and  $t$ -value for the estimate). As  $t$ -values obtained for autoregressive parameters are highly significant ( $t \gg 2$ ) for both proposed models and standard error estimates are also satisfactory, both models are adequate to describe the variability of water concentration in the output product. Besides, parameter estimates are not highly correlated,

so co-linearity may not have influenced the results.

As in general, when comparing candidate models, smaller AIC (Akaike's Information Criterion) and SBC (Schwarz's Bayesian Criterion) statistics indicate the better fitting model, model A, ARIMA (2,1,1) resulting to be better according to its lower values for these statistics.

In addition, autocorrelations of residuals allows us to check the goodness of fitting models. The chi-square statistics for series residuals, which is computed using the Ljung-Box formula, indicates whether the residuals are uncorrelated (white noise) or not. In this study case, the  $\chi^2$  test values do not reject the no-autocorrelation hypothesis of the residuals [12].

In summary, since the model diagnostic tests show that all the parameter estimates are significant and the residual series is white noise, the estimation and diagnostic-checking stages are complete. Thus, the resulting ARMA (2,1,1) model is:

$$(1 - 0.748525 B) + (1 + 0.472845B) \\ (1-B) Y_t = (1 + 0.0706612 B) a_t \quad (4)$$

This processes can be expressed:

$$(1 - 0.27568 B - 0.353936 B^2) (1-B) Y_t = \\ (1 + 0.0706612 B) a_t \quad (5)$$

Then

$$Y_t = 1.27568 Y_{t-1} + 0.078256 Y_{t-2} - 0.353936 \\ Y_{t-3} + a_t + 0.0706612 a_{t-1} \quad (6)$$

The Eq.(6) expressed the present process value  $Y_t$  through the previous values and the present and previous values of  $a_t$  .

The Box-Jenkins methodology requires that the model to be used in describing and forecasting a time series be both stationary and invertible [3]. These properties were analyzed with satisfactory results, so we proceed to forecasting values of the water output concentration.

## 4 Forecasting

The forecasting technique produce the forecast points, the forecast errors and the prediction interval forecast that are established in our case at 95% of confidence. This information is showed through the plots, (Fig. 4) where the point lines represent the adjusted series and the confidence intervals. The series values and the points forecasts are

showed at 200 in observations and continue

ten values ahead. The accuracy of these predictions by the mean absolute deviation was measured, which is simply the average of the absolute deviations for all forecasts, with the simulated techniques in which the historical dates were forecasted choosing models A y B to see which of them provides the most accurate simulated prediction. Also we control the residuals values that are obtained comparing the date series and the adjusted series.

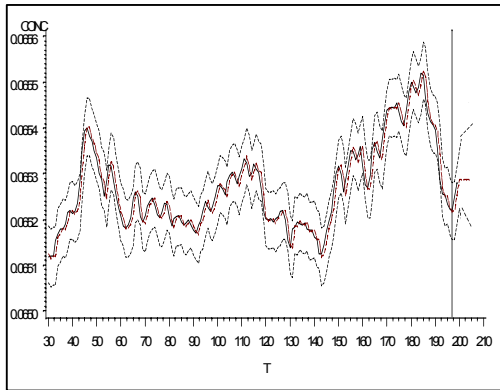


Fig. 4: Water concentration forecasts

Finally we must point out that the forecast points should be considered at the work confidence interval to take right decisions [4].

## 5 Conclusion

Having tentatively identified a model describing the water concentration outputs at the distillation unit; it can expressed another time series generated by the model in order to develop a control scheme for the propylene glycol production to detect special events.

Thus the forecasting model will give forecasts points and a prediction interval forecasts which follow the trend of the sample values; these results allow us to identify randomly different occurring steps changes simulated in the process, in order to design a control algorithm to satisfy a desired objective. In the case of product quality control, a reasonable objective is to try to minimize the variance of the water output concentrations from the target.

### References:

[1] Alwan, L. and H. Roberts. "Time Series Modeling for Statistical Process Control", *J. Econ. Bus. Stat.*, **6**, 87-95 (1988).

[2] Babatunde A. Ogumlaike, W. Hammon Ray "Process Dynamics, Modeling and Control" Oxford University Press (1994)

[3] Box, G. and G. Jenkins, "Time Series Analysis: Forecasting and Control", Holden-Day, San Francisco (1976).

[4] Bowerman Bruce – O'Connell Richard *Forecasting and Time Series: An applied approach* – Duxbury Press (1993).

[5] English, J., Krishnamurthi, M. and T. Sastri, "Quality Monitoring of Continuous Flow Processes", *Computers Ind Eng.*, **20**, 2, 251-260 (1991).

[6] Harris, T. and W. Ross. "Statistical Process Control Procedures for Correlated Observations", *The Canadian Journal of Chem. Engng.*, **69**, 2, 48-57 (1991).

[7] Hysys , Hyprotech Ltd. (1998).

[8] Mac Gregor, J. "On Line Statistical Process Control", *Chem. Eng. Prog.*, **84** (10), 21-31 (1988).

[9] Montgomery D., "Introduction to Statistical Quality Control", Wiley, Toronto (1985).

[10] SAS Institute Inc. SAS/ETS User's Guide Version 6 Second Edition (1993).

[11] Wei, W. "Time Series Analysis: Univariate and Multivariate Methods", Addison Wesley, Toronto (1990).

[12] Wolfram Research *Time Series Pack Reference and User's Guide* (1995)