Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Nonstationarity and cointegration tests for fault detection of dynamic processes

Gang Li *, **S. Joe Qin** **, **Tao Yuan** ***

* *Gang Li is with Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, California 90089, USA. (Email: gli816@usc.edu)*
** *The Mork Family Department of Chemical Engineering and Materials Science, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email: sqin@usc.edu*
*** *Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email: tyuan@usc.edu*

Abstract: As continuous industrial processes often operate around a desirable region of profitability, the measurement series for most process variables act as stationary series. However, there are inevitably some observed time series which are nonstationary caused by unexpected disturbances. Some series grow slowly for a long time with the equipment aging, and others appear to wander around as if they have no fixed population mean. For these series, traditional dynamic PCA or other statistical modeling methods are not applicable because the statistical properties of variables are time variant. In this paper, nonstationarity test is adopted to distinguish nonstationary series from stationary series. After that, cointegration analysis is used to describe the stochastic common trends and equilibrium error, which can be used to construct monitoring indices. Case study on Tennessee Eastman process shows that the proposed nonstationary process monitoring can efficiently detect faults in the nonstationary dynamic process.

Keywords: dynamic processes, nonstationary multivariate series, nonstationarity test, cointegration analysis, unit root test

1. INTRODUCTION

Principal component analysis and its variants are widely used in the area of statistical process monitoring for industrial processes(Qin, 2012). Generally speaking, when an industrial process operates under normal situation, the measured variables approximately follow a multivariate stationary stochastic process. If the sampling interval is long and the samples seem to be time independent, static PCA model performs well in describing cross correlations among measurements. Otherwise, if the auto-correlation between variables are evident, dynamic models are preferred to model the process. Different from static PCA models, dynamic models capture not only cross correlations between variables but also serial correlations among measurement series.

One of the most widely used dynamic model was proposed by Ku et al. with a lagged version of PCA to process multivariate variables (Ku et al., 1995), called dynamic PCA model (DPCA). DPCA conducts the singular value decomposition on an augmented matrix of time lagged process variables. Li and Qin investigated the relationship between dynamic PCA and a subspace identification method (SIM) and proposed a consistent dynamic PCA algorithm, called indirect dynamic PCA (IDPCA) (Li and Qin, 2001). Ding et al. combined SIM and model based fault detection technologies to propose another fault detection scheme (Ding et al., 2009). Using the linear dynamic state space description of processes, the residual generator of the parity space method after identifying a subspace model is equivalent to indirect dynamic PCA modeling for normal data. Negiz and Cinar used a canonical variate (CV) state space model to describe dynamic processes, which is equivalent to a vector autoregressive moving-average time-series model (VARMA) (Negiz and Çinar, 1997). In order to reveal the latent structure hidden in the multivariate time series, a dynamic latent variable model was proposed by Li et al. similar to the concept of dynamic factor analysis (Li et al., 2011). In their model, multivariate time series share several common autocorrelated

⋆ This work was supported by the members of Texas-Wisconsin-California Control Consortium and the IAPI Fundamental Research Funds (2013ZCX02-01).

stochastic trends, while the remaining subspace consists of independently identically distributed series and modeling residuals.

However, the aforementioned approaches all assume that the measured series possesses stationary (i.e. time-invariant) means and variances, rather than nonstationary cases. In practice, some observed series contain a slow-varying trend due to the equipment aging, and other observed series act like a random walk under the influence of various disturbances. In this case, traditional PCA or DPCA models are inadequate to describe relations among variables, and a new model framework is desired to depict the process structure. Nonstationarity test is a tool to test whether a time series is stationary or not, which is studied frequently in the econometrics (Phillips and Xiao, 1998). If two or more time series are nonstationary, but a linear combination of these time series is stationary, these series are cointegrated (Banerjee and Hendry, 1992). This paper will apply nonstationarity test technology to identify nonstationary time series and use cointegration representation to describe latent structures for nonstationary multivariate processes, and further propose the statistical indices for process monitoring.

The rest of the paper is organized as follows. In Section 2, several nonstationarity tests are reviewed and compared to identify whether a time series is stationary. Then, the cointegration analysis is adopted to describe the nonstationary multivariate time series in the Section 3. Following that, fault detection indices are proposed in Section 4 based on the proposed model. In Section 5, we use a case study on Tennessee Eastman process to illustrate the effectiveness of the method. Finally, conclusions are given in the last section.

2. NONSTATIONARITY TEST FOR UNIVARIATE TIME SERIES

2.1 Unit root test

Unit root tests are a popular tool to test whether a time series is unit root nonstationary. A unit root process is a data-generating process whose first difference is stationary. In other words, the simplest unit root process x_k has the following form

$$x_k = x_{k-1} + e_k. \tag{1}$$

where e_k is a stationary series. A unit root test attempts to determine whether a given time series is consistent with a unit root process. Unit root test often uses the following regression model

$$x_k = \mu + \phi x_{k-1} + \sum_{i=1}^{p} \alpha_i \triangle x_{k-i} + \epsilon_k \tag{2}$$

where μ is the intercept, $\triangle$ is the difference operator such that $\triangle x_k = x_k - x_{k-1}$ and ϵ_k is the independent identical distribution (*iid*) innovation process with zero mean. The null hypothesis of unit root test is $\phi = 1$ in (2), which indicates the process is unit root nonstationary. The alternative hypothesis is $|\phi| < 1$, which indicates the process is trend stationary. Define $\hat{\phi}$ as the ordinary least squares estimator of ϕ, i.e.

$$[\hat{\mu}, \hat{\phi}, \hat{\alpha}_1, \cdots, \hat{\alpha}_p]^T = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{X} \tag{3}$$

where

$$\Psi = \begin{bmatrix} 1 & x_0 & \triangle x_0 & \ldots & \triangle x_{1-p} \\ 1 & x_1 & \triangle x_1 & \ldots & \triangle x_{2-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n-1} & \triangle x_{n-1} & \ldots & \triangle x_{n-p} \end{bmatrix} \tag{4}$$

and $\mathbf{X} = [x_1, x_2, \ldots, x_n]^T$, n is the sample length of the regression. Under the null hypothesis of $\phi = 1$, according to the functional central limit theorem, the limit distribution of $n(\hat{\phi} - 1)$ converges in distribution to the following function with $n \to \infty$ (Shumway and Stoffer, 2011):

$$n(\hat{\phi} - 1) \xrightarrow{d} \frac{\sigma[W(1)^2 - 1 - 2W(1)\int_0^1 W(t)dt]}{2\omega[\int_0^1 W^2(t)dt - [\int_0^1 W(t)dt]^2]} \tag{5}$$

where $W(t)$ is standard Brownian motion in $[0, 1]$. Further, σ, ω are two parameters and can be estimated consistently as $\hat{\sigma}^2 = \sum \hat{\epsilon}_k^2 / n$ and $\hat{\omega}^2 = \hat{\sigma}^2 / (1 - \sum \hat{\alpha}_j)^2$, respectively. Because the limiting distribution (5) does not have a closed form, quantities of the distribution must be computed by numerical approximation or by simulation. The lower tail critical value of the distribution can be found in a table with a given significance. If the test statistic is greater than the critical value, the null hypothesis is accepted.

If $p = 0$ is selected, $\sigma = \omega$ in distribution (5) and the test statistic $n(\hat{\phi} - 1)$ is well known as the unit root or Dickey-Fuller (DF) statistic. If $p > 0$, the DF test is extended to the so-called augmented Dickey-Fuller test (ADF). In ADF test, the choice of p is crucial and should be large enough to capture the essential correlation structure in the original series. According to the work by Said and Dickey (1984), the lag length p is enough to be $n^{1/3}$. An alternative test is the Phillips-Perron (PP) test, which differs from ADF mainly in how they deal with serial correlation and heteroskedasticity in the errors (Shumway and Stoffer, 2011). ADF and PP tests are two commonly used unit root tests, which test the coefficients estimation of the regression model (2).

2.2 Stationarity test

While the null hypothesis of unit root test is nonstationarity, stationarity tests with the Largrange multiplier (LM) principle can test the null hypothesis of trend stationarity. Kwiatkowski et al. proposed a well known test to identify a random walk from a stationary series, which is so called KPSS test (Kwiatkowski et al., 1992). It assumes that the target series can be decomposed into the sum of a deterministic trend, a random walk, and a stationary error:

$$\begin{cases} x_k = \mu + r_k + e_k. \\ r_k = r_{k-1} + u_k \end{cases} \tag{6}$$

where e_k is stationary series, r_k is a standard random walk,u_k is iid$(0, \sigma_u^2)$. The model is known as the component representation. Let e_k be the residuals from the regression of x_k on the deterministic trend as follows:

$$\hat{e}_k = x_k - \hat{\mu} \tag{7}$$

Define $\hat{\sigma}_e^2 = \sum \hat{e}_k^2/n$ as the estimate of the error variance from the regression, and $S_k = \sum_{i=1}^k \hat{e}_i$ as the partial sum process of the residuals, then the LM statistic can be constructed as follows (Phillips and Xiao, 1998)

$$LM = \frac{\sum S_k^2/n^2}{\hat{\sigma}_\epsilon^2} \tag{8}$$

Under the null hypothesis of stationarity and e_k is an iid process, the LM statistic converges to the following function in distribution

$$LM \xrightarrow{d} \int_0^1 V^2(t)dt \tag{9}$$

where $V(t) = W(t) - tW(1)$ is a standard Brownian bridge, and $W(t)$ is a Brownian motion. When there is a serial dependence in e_k, the statistic can be modified by replacing the variance estimate $\hat{\sigma}_\epsilon^2$ in (8) with the long run variance $s^2(l)$ (Kwiatkowski et al., 1992):

$$s^2(l) = \hat{\sigma}_e^2 + \frac{2}{n}\sum_{s=1}^{l}[(1 - \frac{s}{l+1})\sum_{k=s+1}^{n}\hat{e}_k\hat{e}_{k-s}] \tag{10}$$

The parameter l can be chosen as $n^{1/2}$. Given a significance level, the critical values can be calculated by numerical simulation or found from a table. The KPSS test is a upper tail test, thus when the test statistic (8) is below the critical value, the null hypothesis is accepted.

There are also some other efficient tests for nonstationarity, such as the point optimal test and the local best invariant test, which consider the case that ϕ in model (2) is close to 1.

3. CO-INTEGRATION ANALYSIS FOR MULTIVARIATE INTEGRATED SERIES

If the observed series are found to be unit root nonstationary, the population variance of series is not a constant, which makes the monitoring with traditional statistical methods inadequate. However, in the econometrics area, a phenomenon called co-integration has been observed frequently. When two or more observed series are nonstationary but a linear combination of these series is stationary, these time series are called cointegrated. Engle and Granger had initiated a discussion on the representation, estimation and testing for cointegrated series and introduced the following concepts (Engle and Granger, 1987).

Definition 1. If a series x_k has a stationary, invertible, autoregressive moving average (ARMA) representation after differencing d times with no deterministic component, it is called integrated of order d, denoted as $x_k \sim I(d)$.

For simplicity, only $d = 0, 1$ will be discussed in the paper. The results can be generalized to other cases. According to the definition, $x_k \sim I(1)$ is a unit root nonstationary, and $x_k \sim I(0)$ is stationary.

Definition 2. For a vector series $\mathbf{x}_k$, if all components of $\mathbf{x}_k$ are $I(d)$, and there exists a nonzero vector β so that $z_k = \beta^T \mathbf{x}_k \sim I(d-b), b > 0$, $\mathbf{x}_k$ is called cointegrated of order d, b, denoted as $\mathbf{x}_k \sim CI(d, b)$. and β is called the cointegrating vector.

For the bivariate case, there is at most one cointegration relation in $\mathbf{x}_k$, which was solved with a residual based procedures by Engle and Granger (1987). In the case that more than one relation may exist, say r cointegration relations, the cointegrated vector is extended to a cointegration matrix B with rank r, and $\mathbf{z}_k$ is a vector series with r dimension. There are two tasks in the cointegration analysis: i) testing r, and ii) identifying B. There are several representations for cointegrated processes, including vector error correction and common trends model.

3.1 vector error correction and Johansen test

Suppose that a multivariate series is generated by a vector autoregressive model of order p with Gaussian errors

$$\mathbf{x}_k = \mu + \sum_{i=1}^{p} A_i \mathbf{x}_{k-i} + \epsilon_k \tag{11}$$

where $\mathbf{x}_k \in \mathbf{R}^m$ is a m-dimensional series, $A_i \in \mathbf{R}^{m\times m}$ is the coefficient matrix, and ϵ_k is a m-dimensional white noise process distributed as $N(0, \Sigma_\epsilon)$. Here, we consider that $\mathbf{x}_k \sim I(1)$, and $\triangle \mathbf{x}_k$ are stationary. Define

$$\Pi = \sum_{i=1}^{p} A_i - I_m \tag{12}$$

then $\Pi \mathbf{x}_k \sim I(0)$ and $rank(\Pi) = r$ is the number of linearly independent cointegration relations (Reimers, 1992). Decompose Π into two full-ranked matrices $\Pi = AB^T$, where $A \in R^{m\times r}$ is the loading matrix and $B \in R^{m\times r}$ is the cointegration matrix. Note that the decomposition is not unique. Define $B_i = -\sum_{j=i+1}^{p} A_j$, then the model (11) can be reparameterized as a vector error-correction (VEC) model as follows (Reimers, 1992)

$$\triangle \mathbf{x}_k = \mu + AB^T \mathbf{x}_{k-1} + \sum_{i=1}^{p-1} B_i \triangle \mathbf{x}_{k-i} + \epsilon_k \tag{13}$$

Remark 1. On one hand, if the goal of a VAR analysis is to determine relationships among the original variables, differencing loses information and may lead to a model misspecification, since long-term information is reflected in the levels. On the other hand, if the goal is to simulate an underlying data-generating process, integrated levels data can cause a number of problems. Fortunately, the VEC model provides intermediate options, between differences and levels, by mixing them together with the cointegrating relations. Since all terms of the VEC model are stationary, problems with unit roots are eliminated.

Johansen proposed a maximum likelihood procedure to estimate the parameters in (13), and formulated a likelihood ratio test of the hypothesis $H_0 : r = r_0$ against $H_1 : r > r_0$ based on the statistic (Johansen, 1995):

$$-2\ln Q = -n \sum_{i=r_0+1}^{m} \ln(1 - \hat{\lambda}_i) \tag{14}$$

where r_0 is the given value of r to be tested, $\hat{\lambda}$ is the $m - r_0$ smallest eigenvalues obtained by solving an eigenvalue equation:

$$|\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p}| = 0 \tag{15}$$

where

$$\begin{aligned} S_{ij} &= \frac{1}{n} R_i R_j, (i,j = 0, p) \\ R_0 &= \triangle X - \triangle X Y^T (YY^T)^{-1} Y \\ R_p &= X_{-p} - X_{-p} Y^T (YY^T)^{-1} Y \\ \triangle X &= (\triangle \mathbf{x}_1, \ldots, \triangle \mathbf{x}_n) \\ X_p &= (x_{-p+1}, \ldots, x_{n-p}) \\ Y &= (y_0, \ldots, y_{n-1}) \\ y_k &= (1, \triangle \mathbf{x}_{k-1}^T, \ldots, \triangle \mathbf{x}_{k-p+1}^T)^T \end{aligned}$$

The asymptotic distribution of the statistic is nonstandard and depends on the restriction of the deterministic term if existing, which can be found in Turner (2009). The lag parameters p can be jointly estimated with r_0 using an order criterion (Reimers, 1992). Once the rank r and lags p are determined, the cointegration matrix B can be estimated by an ML procedure or an LS method from (13).

3.2 common trend representation and Stock-Watson test

Another representation for cointegrated variables is known as common-trend representation or dynamic factor model:

$$\begin{cases} \mathbf{x}_k = \Gamma \tau_k + \mathbf{e}_k \\ \tau_k = \tau_{k-1} + \mathbf{u}_k \end{cases} \tag{16}$$

where $\mathbf{e}_k$ is a stationary multivariate series, $\Gamma \in R^{m\times(m-r)}$ is the loading matrix, $\tau_k \in \mathbf{R}^{m-r}$ is a $(m-r)$-dimensional random walk process, $\mathbf{u}_k$ is an *iid* process. Stock and Watson (1988) proposed an estimation procedure of the common trend representation (16) and test for the number of the random walks. First of all, they used principal component analysis to estimate the τ_k with $m-r$ principal components. Then the autoregressive coefficient matrix Φ in the form of $\tau_k = \Phi\tau_{k-1} + \epsilon_k$ was estimated by least squares regressor. At last, the following statistic is used to test the hypothesis of $H_0 : r = r_0$ against $H_1 : r > r_0$:

$$q_{m-r} = n(real(\hat{\lambda}_{min}) - 1) \tag{17}$$

where $\hat{\lambda}_{min}$ is the smallest eigenvalue of Φ, and $real(\cdot)$ means the real part of a number. Under the H_0 hypothesis, the estimated minimum eigenvalue should be insignificantly different from one. The asymptotic distribution of q_{m-r} is nonstandard and depends on r (Stock and Watson, 1988).

4. FAULT DETECTION BASED ON COINTEGRATION REPRESENTATION

In the study on economics, nonstationary test and cointegration analysis are directly performed on economical data collected in the long term. Here, these kinds of modeling are used in the monitoring of industrial processes, which should be adapted for the special use. As is mentioned, cointegration model can describe nonstationary variables with strong correlations, which are called cointegrated variables. However, it is desired to construct monitoring indices to indicate whether there is a disturbance or break in the normal structure. Different from traditional PCA, only equilibrium error process z_k can be acquired as a stationary multivariate series, which is proper for monitoring.

However, if $x_k \sim I(1)$, $\triangle\tau_k \sim I(0)$ is stationary, and the difference information of common trends can also be used for monitoring purpose. The remaining directions which are orthogonal to $col(B)$ contain the common stochastic trends in $\mathbf{x}_k$, which can be extracted as

$$\tau_k = B^{\perp T}\mathbf{x}_k \tag{18}$$

where $col(B)$ means the space spanned by the columns of B, and $B^{\perp}$ means the full-column-rank matrix which is orthogonal to $\hat{B}$. For a new observation $\mathbf{x}_k$, we calculate the $\mathbf{z}_k$ and $\triangle\tau_k$ as

$$\begin{aligned} \mathbf{z}_k &= \hat{B}^T\mathbf{x}_k + \hat{\mu} \\ \triangle\tau_k &= B^{\perp T}(\mathbf{x}_k - \mathbf{x}_{k-1}) \end{aligned} \tag{19}$$

The T^2 statistic for z_k and $\triangle\tau_k$ are listed in Table 1.

Table 1. Fault detection indices

Statistics	Calculation	Control limit
T_z^2	$\mathbf{z}^T\Lambda_z^{-1}\mathbf{z}$	$\frac{r(n^2-1)}{n(n-r)}F_{r,n-r,\alpha}$
T_τ^2	$\triangle\tau^T\Lambda_\tau^{-1}\triangle\tau$	$\frac{(m-r)(n^2-1)}{n(n-m+r)}F_{m-r,n-m+r,\alpha}$

Λ_z, Λ_τ are covariance of z_k and $\triangle\tau_k$, respectively. $F_{a,b,\alpha}$ is the critical value of F-distribution with degree a, b at the significance α

The statistic T_z^2 monitors the equilibrium relations among nonstationary series, which is similar to SPE index in a PCA model, while the statistic T_τ^2 monitors the variations in the nonstationary components, which is similar to T^2 index in a PCA model.

5. CASE STUDY ON TE PROCESS

In this section, the proposed method is investigated for nonstationary variables in the Tennessee Eastman Process (TEP). TEP was created to evaluate process control and monitoring methods, such as PCA, PLS, and Fisher discriminant analysis (FDA) (Chiang et al., 2000). The process contains 12 manipulated variables and 41 measured variables. Process measurements are sampled with an interval of 3 minutes. 19 composition measurements are sampled with time delays that vary from six minutes to fifteen minutes, which are not used in this study. Therefore, 22 process measurements and 11 manipulated variables, i.e. XMEAS(1-22) and XMV(1-11), are chosen as $\mathbf{X}$.

Firstly, 480 original normal samples are chosen to test the nonstationarity of all variables and find out the nonstationary variables. We use ADF test, PP test and KPSS test with lags $p = 2$ and without linear time trend to identify nonstationary series in TE process. The test results are listed in Table 2, where '0' represents to accept the hypothesis of unit root nonstationarity and '1' represents to accept the hypothesis of trend stationarity. From the table, 7 variables seem to be nonstationary. However, there are only four nonstationary series, which are plotted in Figure

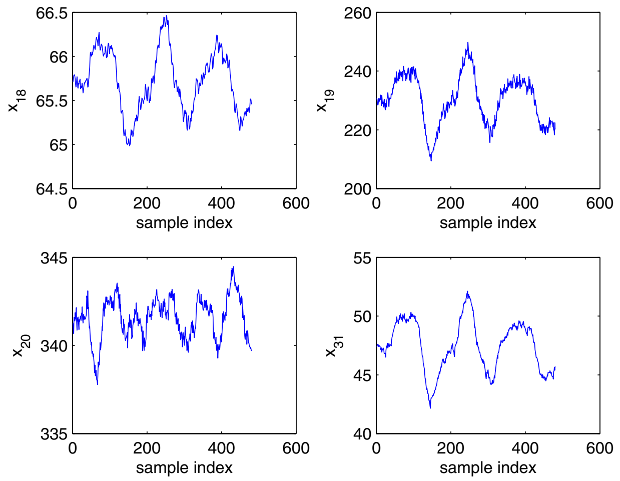

Fig. 1. nonstationary series in TE process

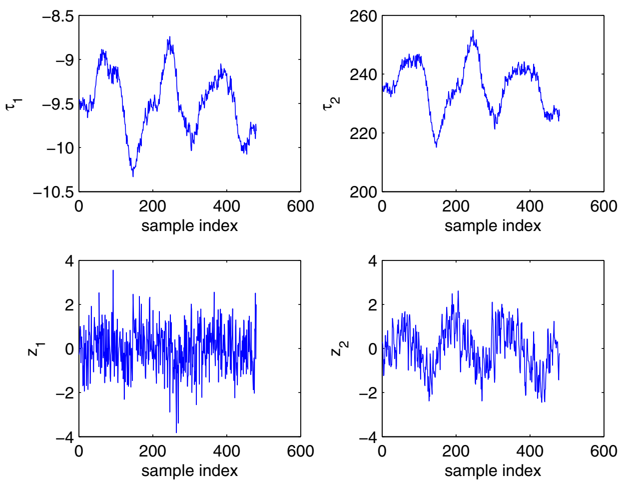

Fig. 2. Common trends τ_1, τ_2 and error process z_1, z_2

1. The results show that ADF and PP test outperform KPSS test in unit root test for industrial process. Let $x = [x_{18}, x_{19}, x_{20}, x_{31}]^T$. We then adopted the Johansen test on $\mathbf{x}_k$, and had $r = 2$ as the test result. The cointegration matrix and the intercept are estimated as

$$B = \begin{bmatrix} -5.3796 & -10.3143 \\ -0.5358 & 0.2777 \\ -0.1131 & -0.5579 \\ 2.8858 & 0.6203 \end{bmatrix}, \mu = \begin{bmatrix} 378.6013 \\ 774.9191 \end{bmatrix}$$

Following that, common trends τ_k and the error process z_k are calculated by (18) and (19), respectively and plotted in the Fig. 2. The nonstationary tests for z_k, τ_k are listed in Table 3, which is consistent with the cointegration test.

There are 15 known faults in TE process (Yin et al., 2012). Take the IDV(1) as an example. When IDV (1) is introduced to testing data at the 160^{th} sample, a step change is induced in the A/C feed ratio in Stream 4. The proposed statistics can detect the fault effectively, as plotted in Fig. 3. From Fig. 3, it is observed that T_z^2 outperforms T_τ^2, because T_τ^2 utilizes the differencing series of original data, which loses much information.

For other stationary variables, traditional PCA or DPCA can be applied to model and monitor the process effectively. However, for these four nonstationary variables, cointegration analysis is more efficient to process the nonstationary series with common trends.

Table 2. Nonstationary tests for 22 measurements and 11 manipulated variables

Variable	Description	ADF test	PP test	KPSS test
Xmeas(1)	A feed (stream 1)	1	1	1
Xmeas(2)	D feed (stream 2)	1	1	1
Xmeas(3)	E feed (stream 3)	1	1	1
Xmeas(4)	Total feed (stream 4)	1	1	1
Xmeas(5)	Recycle flow (stream 8)	1	1	1
Xmeas(6)	Reactor feed rate(stream 6)	1	1	1
Xmeas(7)	Reactor pressure	1	1	1
Xmeas(8)	Reactor level	1	1	1
Xmeas(9)	Reactor temperature	1	1	1
Xmeas(10)	Purge rate (stream 9)	1	1	1
Xmeas(11)	separator temperature	1	1	1
Xmeas(12)	separator level	1	1	1
Xmeas(13)	separator pressure	1	1	1
Xmeas(14)	separator underflow	1	1	1
Xmeas(15)	stripper level	1	1	1
Xmeas(16)	stripper pressure	1	1	1
Xmeas(17)	stripper underflow	1	1	0
Xmeas(18)	stripper temperature	0	0	0
Xmeas(19)	stripper steam flow	0	0	0
Xmeas(20)	compressor work	0	1	0
Xmeas(21)	reactor cooling water outlet temp	1	1	1
Xmeas(22)	condenser cooling water outlet temp	1	1	1
Xmv(1)	D feed flow (stream 2)	1	1	1
Xmv(2)	E feed flow (stream 3)	1	1	1
Xmv(3)	A feed flow (stream 1)	1	1	1
Xmv(4)	total feed flow (stream 4)	1	1	1
Xmv(5)	compressor recycle valve	1	1	0
Xmv(6)	purge valve (stream 9)	1	1	1
Xmv(7)	separator pot liquid flow	1	1	1
Xmv(8)	stripper liquid product flow	1	1	1
Xmv(9)	tripper steam valve	0	0	0
Xmv(10)	reactor cooling water flow	1	1	1
Xmv(11)	condenser cooling water flow	1	1	0

Table 3. Nonstationary tests for common trends and error process

series	ADF test	PP test	KPSS test
τ_1	0	0	0
τ_2	0	0	0
z_1	1	1	1
z_2	1	1	0

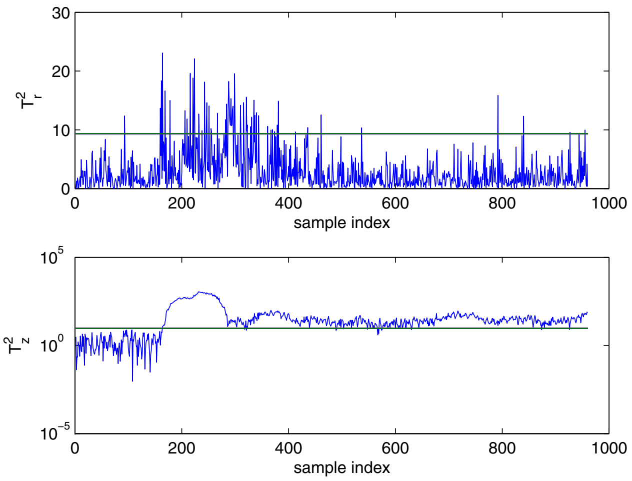

Fig. 3. Fault detection result with T_z^2 and T_τ^2

6. CONCLUSIONS

In this paper, the process monitoring problem for nonstationary multivariate series is addressed. As the statistical properties of nonstationary variables vary along time, traditional multivariate statistical process monitoring techniques are inadequate to deal with the modeling and monitoring of nonstationary processes. This paper utilizes nonstationary tests to identify the nonstationary variables in a dynamic process, and then adopts cointegration analysis to describe the common trends among cointegrated variables. Based on the survey of these methods and models, a new fault detection policy is established for detecting the fault during multivariate nonstationary processes. The proposed method is demonstrated by the case study on TE process. The results show the effectiveness of the proposed method. In the future, the results will be extended to the monitoring of nonstationary processes with a deterministic trend.

REFERENCES

Banerjee, A. and Hendry, D.F. (1992). Testing integration and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, 54(3), 225–255.

Chiang, L.H., Russell, E.L., and Braatz, R.D. (2000). Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50(2), 243–252.

Ding, S., Zhang, P., Naik, A., Ding, E., and Huang, B. (2009). Subspace method aided data-driven design of fault detection and isolation systems. *Journal of Process Control*, 19(9), 1496–1510.

Engle, R.F. and Granger, C.W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 55(2), 251–276.

Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. *New York*.

Ku, W., Storer, R., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1), 179–196.

Kwiatkowski, D., Phillips, P.C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1), 159–178.

Li, G., Liu, B., Qin, S.J., and Zhou, D. (2011). Dynamic latent variable modeling for statistical process monitoring. In *Proc. IFAC World Congress, Milan, Italy*, 12886–12891.

Li, W. and Qin, S. (2001). Consistent dynamic PCA based on errors-in-variables subspace identification. *Journal of Process Control*, 11(6), 661–678.

Negiz, A. and Çinar, A. (1997). Statistical monitoring of multivariable dynamic processes with state-space models. *AIChE Journal*, 43(8), 2002–2020.

Phillips, P.C. and Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys*, 12(5), 423–470.

Qin, S.J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36, 220–234.

Reimers, H.E. (1992). Comparisons of tests for multivariate cointegration. *Statistical papers*, 33(1), 335–359.

Said, S.E. and Dickey, D.A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599–607.

Shumway, R.H. and Stoffer, D.S. (2011). *Time series analysis and its applications: with R examples*. Springer.

Stock, J.H. and Watson, M.W. (1988). Testing for common trends. *Journal of the American statistical Association*, 83(404), 1097–1107.

Turner, P. (2009). Testing for cointegration using the johansen approach: are we using the correct critical values? *Journal of Applied Econometrics*, 24(5), 825–831.

Yin, S., Ding, S.X., Haghani, A., Hao, H., and Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*.