

<研究発表>

プロセスデータの相関を利用した異常診断技術の下水処理プロセスへの適用

山中 理, 小原 卓巳, 山本 勝也

(株) 東芝(〒183-8511 東京都府中市東芝町 1 E-mail: osamu2.yamanaka@toshiba.co.jp)

概要

本論文では、水質・水量・機器の操作量など各種のプロセスデータ間の相関を定量的に解析する主成分分析(PCA)と呼ばれる手法を用いた多変量統計のプロセス管理(MSPC)に基づく異常診断システムの構築法を提案し、下水処理プロセスを対象として、その効果を評価する。特に、一定の異常診断性能を確保しながら、人間が介在するエンジニアリング要素を最小限にした異常診断システムを構築する方法について検討する。そのために、MSPCの異常検出用データの閾値を自動的に設定する方法を提案し、その有効性を下水処理プロセスの異常シナリオデータと実データに提案手法を適用して評価することによって示す。MSPCに提案手法を併用することにより、一定の異常診断性能を保ちながら、各種のプロセスデータを入力するだけで、ほぼ自動的に異常診断システムを構築することが可能になる。

キーワード: 異常診断, 閾値設定, 多変量統計のプロセス管理, 主成分分析, 下水処理プロセス

1. はじめに

下水処理場などのプラントでは、オンラインセンサで計測・収集されている水質・水量・機器の操作量などのプロセスデータを、運転員が監視画面で監視しながらプラント運用を行っている。このようなプロセス監視はプラント運用の基本であるが、高齢化社会に伴う熟練運転員の減少や多くの自治体の厳しい財政事情により、運転員が行っている判断の一部あるいは全部を自動的に行う自動化への潜在的な要求は大きい。

プロセスで生じる様々な異常あるいはその兆候を検出することはプラント運用の基本であるが、プラント運用の自動化のためには、異常検出や異常予知を自動的に行うことが重要になる。この方法の一つとして、各種のプロセスデータに統計的処理を施してプロセスの異常診断を行う多変量統計のプロセス管理(MSPC: Multivariate Statistical Process Control)がある^{1), 2)}。MSPCは、各種のプロセスデータ間に存在する相関を利用してこれらのプロセスデータのみから異常診断を行う、特定のプラントに依存しない汎用的な手法である。

MSPCを下水処理プロセスなどの具体的なプロセスに適用する際に考慮すべきことは、異常診断システムの構築時に人間が介在するエンジニアリング要素を極力排除しながら、異常検出率や異常誤検出率で評価される異常診断性能を一定レベル以上に保つことである。このためには、MSPCで用いられる異常検出用のデータにおいて、正常と異常を判断するための閾値を、異常診断性能が最良となると同時に試行錯誤を必要とせず自動的に設定することが一つのポイントとなる。

本稿では、異常検出用データの閾値設定法を中心として、MSPCによる異常診断システムの構築法を提案し、下水処理プロセスを対象としてその有効性を評価する。まず、MSPCの原理とアルゴリズムを、異常検出用データの閾値を自動的に設定する方法の提案を交えながら説明する。そして、下水処

理プロセスシミュレータを用いた異常シナリオデータと、実際の下水処理場で収集したプロセスデータに対して、提案手法を適用し、評価することによって、その有効性を示す。

2. プロセスデータの相関を利用した異常診断

2.1 プロセスデータの相関と異常診断の原理

Fig.1は、ある下水処理場における複数の計測変数のプロセスデータをプロットしたものである。

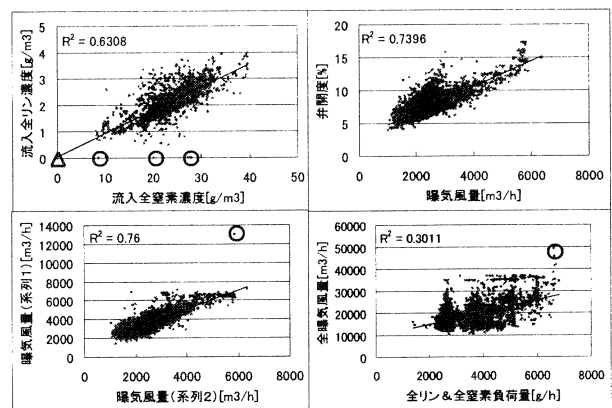


Fig.1 Correlation between Various Process Data

Fig.1に示すように、各種のプロセスデータ間には何らかの相関関係があることが多く、そのため、○や△で囲んだデータは異常である可能性が高い。○で囲んだデータは、相関関係が著しく崩れており、△で囲んだデータは、相関関係は崩れていないが、他のデータから大きく乖離している。MSPCは、このような2つのタイプの異常を、2変数以上の通常数十点～数千点におよぶ複数の変数が互いに複雑な相関を持つ場合にも、多変量解析手法を用いて検出し、その異常要因となっている計測変数を推定する方法である。

2.2 MSPCによる異常診断モデルの構築

MSPCでは、上述した原理を利用して、Fig.2の様な構成でプロセスの異常診断を行う。

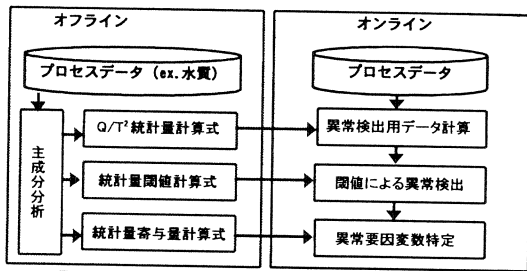


Fig.2 Outline of Fault Diagnosis based on MSPC

MSPC では、オンラインで異常診断を行う前に、予め過去のプロセスデータを利用して、異常診断に用いる各種の計算式を導出しておく。これらの計算式を以下では異常診断モデルと呼ぶが、異常診断モデル構築の際、MSPC では数十～数千変数の間の相関を計算するために、主成分分析 (PCA:Principal Component Analysis)と呼ばれる手法を用いている。以下では、PCA を用いた異常診断モデル構築の概要を示し、その中でエンジニアリング要素を低減しつつ、異常検出性能向上を目指した新しい閾値設定法を提案する。

(1) 主成分分析によるプロセスデータの相関解析

主成分分析 (PCA) では、まず、M 個の計測変数を行方向にとり、列方向に各計測変数の N 個の時系列データを代入した行列 **X** を用意する。但し、各変数の値はその物理量に応じて範囲が大きく異なるため、予めでスケールしておく。そして、この行列 **X** を次式の様に分解する。

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \dots \dots \dots (1)$$

ここで、**P** はローディング行列と呼ばれ、Fig.3 の2変数の例では、変数 A と変数 B の相関方向の軸を表現する行列である。**T** はスコア行列と呼ばれ、相関軸上でのデータの位置を表す行列である。**E** は誤差行列であり、相関軸からの乖離を表す。また、右肩添え字の **T** は、行列の転置を表す。(1) 式の様な分解を行うアルゴリズムが PCA である。

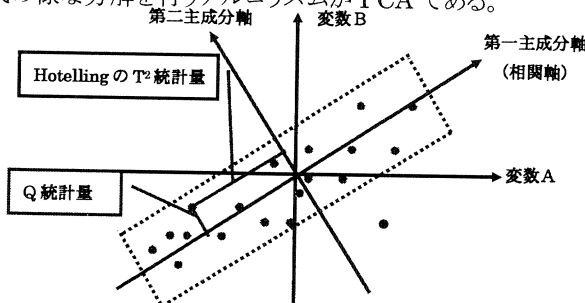


Fig. 3 Principle of Fault Detection and Isolation (FDI)

(2) 異常検出のための統計量計算モデル

PCA を利用した異常診断では、異常検出を行うために、(1) 式の変数を用いて、Q 統計量および Hotelling の T² 統計量と呼ばれる以下の異常検出用データを定義する。

Q 統計量:

$$Q(\mathbf{k}) = \mathbf{x}^T(\mathbf{k})(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\mathbf{x}(\mathbf{k}) \dots \dots \dots (2)$$

Hotelling の T² 統計量:

$$T^2(\mathbf{k}) = \mathbf{x}^T(\mathbf{k})\mathbf{P}^T\mathbf{W}^{-1}\mathbf{P}\mathbf{x}(\mathbf{k}) = \mathbf{t}^T(\mathbf{k})\mathbf{W}^{-1}\mathbf{t}(\mathbf{k}) \dots \dots \dots (3)$$

ここで、**x(k)** は診断対象の M 個の計測変数から成るデータベクトル、**I** は単位行列、**W** は **X^TX** の固有値を対角要素に持つ対角行列である²⁾。また、**k** はある時刻 (サンプル) を表す変数である。(2) 式の Q 統計量は、Fig.3 に示す様に相関軸

からの乖離の大きさを表すスカラ量であり、(3) 式の Hotelling の T² 統計量は、相関軸方向の分布を表すスカラ量である。これらの値が過大になる点が異常と見なされる。

(3) 閾値設定のための計算モデル

異常検出を行うためには、(2) 式や (3) 式に対して、適切な閾値を設定する必要がある。閾値の理論値は知られているが²⁾、(i) データが多量にある場合に理論値を適用すると、異常誤検出が増える場合がある²⁾、(ii) 文献 3) の様に、異常検出用データ自身に処理を加えた場合、理論値は原理的に適用できない、という問題がある。そのため、文献 2) では、異常診断モデル構築に用いる過去のプロセスデータから異常検出用データを計算し、その上位数%の値を閾値とする代替案を示しているが、何%に設定すべきかの具体的指針が無く、試行錯誤が必要になる。そこで、ここでは、任意の異常検出用データに適用でき、試行錯誤を必要としない閾値設定の計算方法として、2つの集合 (正常データ集合と異常データ集合) を分離する2クラスのクラスタリング法に付加的な処理を加えた方法を提案する。このフローチャートを Fig.4 に示す。

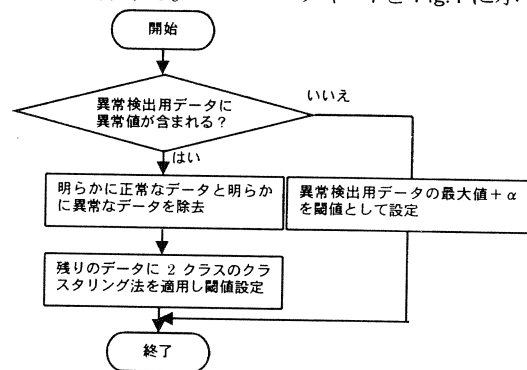


Fig.4 Proposed Thresholding Procedure

Fig.4 は単純なフローチャートであるが、二つの重要な点を含んでいる。一つは、異常検出用データに異常値を含むか否かの判断を要求している点である。クラスタリング法では2つの集合が存在していることを前提としているため、異常検出用データに異常値が含まれない場合にクラスタリング法を適用すると不適切な閾値を設定することになる。もう一つは、異常検出用データから、明らかに正常なデータと明らかに異常なデータを予め除去する処理を加えている点である。この具体的手段は様々なものが考えられるが紙面の都合上省略する。この処理を行う理由は、クラスタリング法を単純に適用すると、Fig.5 に示す様に不適切な点 (「不可」の点) に閾値を設定することが生じるためである。また、2クラスのクラスタリング法としては数多くのものが知られているので⁴⁾、適切なものを用いれば良いが、ここでは、2つのクラスのクラス間分散を最大にする大津の閾値設定法⁴⁾と呼ばれる方法を用いた。

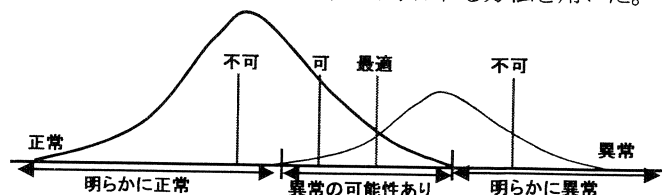


Fig.5 Proper Choice of Threshold Level in Distribution of Statistics

(4) 異常要因変数特定のための寄与量計算モデル

異常検出を行った後に異常要因である変量を特定・推定するために、Q 統計量および Hotelling の T² 統計量に対する各計測変数の寄与量と呼ばれる量が次式で定義される。

Q 統計量の寄与量:

$$Q_{cont}(n,k) = x^T(k,n)F(:,n)TF^T(:,n)x(k,n)1, F = I - PP^T \quad (4)$$

Hotelling の T² 統計量の寄与量:

$$T^2_{cont}(n,k) = x^T(k)P^TW^{-1}P^T(:,n)x(k,n) \dots \dots \dots (5)$$

ここで、n は n 番目の計測変数という意味であり、(:,n)は、行列の n 列目を取り出すことを意味し、(k,n)は、サンプル k の n 番目の要素を取り出すことを意味する。異常検出時には、寄与量の大きい計測変数、例えば、(i)最大の寄与量を与える計測変数、(ii)寄与量とその平均値よりある一定値以上大きい計測変数、などが異常要因変数として推定される。

以上の一連の手順によって、異常診断モデルを構築した後、オンラインでプロセスデータを(2)式~(5)式などのモデルに入力することにより、異常診断を実施することができる。

3. 下水処理プロセスデータを用いた評価

3.1 下水処理場の異常シナリオデータを用いた評価

活性汚泥モデルを用いた下水処理プロセスシミュレータ³⁾に 10 個の異常シナリオを組み込み、異常シナリオデータを生成し、これに提案手法を適用して評価を行った。シナリオを利用することにより、異常検出率や異常誤検出率を正確に計算でき、定量評価や異なる手法間の比較が可能になる。

(1) 対象プロセスと異常シナリオの定義

対象プロセスを、A2O(嫌気-無酸素-好気)法を適用した下水処理プロセスとし、計測変数を Table 1 の項目とした。

Table 1 Measurement Variables for FDI (Simulated Process)

| No | 計測変数 | 単位 | No | 計測変数 | 単位 |
|----|---------------------|---------------------|----|-----------|---------------------|
| 1 | 流入 TP | [g/m ³] | 10 | 放流 COD | [g/m ³] |
| 2 | 流入 TN | [g/m ³] | 11 | 放流 SS | [g/m ³] |
| 3 | 流入 SS | [g/m ³] | 12 | 余剰汚泥濃度 | [g/m ³] |
| 4 | 好気槽 DO | [g/m ³] | 13 | 流入量 | [m ³ /d] |
| 5 | 好気槽 MLSS | [g/m ³] | 14 | 循環量 | [m ³ /d] |
| 6 | 好気槽 NH ₄ | [g/m ³] | 15 | 返送量 | [m ³ /d] |
| 7 | 好気槽 PO ₄ | [g/m ³] | 16 | 余剰流量 | [m ³ /d] |
| 8 | 放流 TP | [g/m ³] | 17 | 放流量 | [m ³ /d] |
| 9 | 放流 TN | [g/m ³] | 18 | 曝気量 (KLa) | [1/d] |

また、異常シナリオとして、(A)センサ異常、(B)操作量異常、(C)制御異常、(D)プロセス異常、(E)突発外乱、の5タイプの異常の中から Table 2 に示すシナリオを定義した。

Table 2 Definition of Fault Scenario

| No. | シナリオ | シナリオ定義 |
|-----|--------------|--------------------------------------|
| 1 | MLSS センサドリフト | ドリフト率 200/14 [g/m ³ /day] |
| 2 | DO センサドリフト | ドリフト率 1/14 [g/m ³ /day] |
| 3 | DO 制御ハンチング | DO センサ計測無駄時間 2分 |
| 4 | バルキング | フロック沈降速度通常時の 1/8 |
| 5 | 硝化阻害 | 硝化菌増殖速度 1→0.125(1/day) |
| 6 | 無酸素槽酸素混入 | 無酸素槽 DO=0.5[g/m ³] |
| 7 | 流量過負荷 | 通常の流量パターンの 2倍 |
| 8 | 散気管目詰まり | 風量 (空気倍率) 低下 10/14[%/day] |
| 9 | 毒物混入 | 微生物の増殖に関わる係数 1→1/4 倍 |
| 10 | 雨天時水質負荷低下 | 雨天時負荷低下パターン入力 |

(2) 評価方法

提案手法の効果の有無を検証するために、以下の 3 つの閾値設定法を比較評価した。

- (A) 各統計量の信頼区間に基づく理論値を用いた方法
- (B) 2クラスのクラスタリング手法(大津の閾値設定法)
- (C) (B)に修正を加えた提案手法

なお、提案手法では、理論値が誤検出を増加させる傾向にあることを考慮し、この値が下限値となる処理を行った。

また、各手法の比較のため、以下の評価指標を定義した。

$$\text{評価指標} : (0.5A - 0.5B + 0.5) \log_{1/2} / \log_{3/4} \dots \dots (6)$$

ここで、A は異常検出率 = 異常検出点数 / シナリオ異常点数、B は異常誤検出率 = 異常誤検出点数 / シナリオ正常点数である。この指標は、異常検出率と異常誤検出率を総合評価するものであり、値域が 0~1、異常検出率 1 で異常誤検出率 0.5 の場合に指標値が 0.5 となるように設計した。また、参考として、異常発生から検出までの時間も評価した。

(3) 評価結果と考察

各シナリオに対する評価結果を Table 3~Table 5 に示す。

Table 3 Evaluation of FDI Results (Confidence Limit)

| 異常シナリオ | 指標 | 検出率 | 誤検出率 | 検出時間 |
|-------------|-------|-------|-------|--------|
| MLSSセンサドリフト | 0.605 | 0.629 | 0.005 | 1365分 |
| DOセンサドリフト | 0.226 | 0.157 | 0.078 | — |
| DO制御ハンチング | 0.797 | 0.827 | 0.007 | 15分 |
| バルキング | 0.459 | 0.464 | 0.016 | 30分 |
| 硝化阻害 | 0.727 | 0.830 | 0.077 | 10分 |
| 無酸素槽酸素混入 | 0.787 | 0.875 | 0.065 | 70分 |
| 流量過負荷 | 0.562 | 1.000 | 0.425 | 5分 |
| 散気管目詰まり | 0.287 | 0.195 | 0.003 | 12190分 |
| 毒物混入 | 0.876 | 0.990 | 0.097 | 15分 |
| 雨天時水質負荷低下 | 0.408 | 0.839 | 0.460 | 5分 |
| 平均 | 0.573 | 0.681 | 0.123 | 1523分 |

Table 4 Evaluation of FDI Results (Clustering)

| | 指標 | 検出率 | 誤検出率 | 検出時間 |
|-------------|-------|-------|-------|--------|
| MLSSセンサドリフト | 0.621 | 0.650 | 0.009 | 1365分 |
| DOセンサドリフト | 0.234 | 0.186 | 0.092 | — |
| DO制御ハンチング | 0.804 | 0.839 | 0.012 | 15分 |
| バルキング | 0.472 | 0.488 | 0.023 | 25分 |
| 硝化阻害 | 0.725 | 0.837 | 0.087 | 10分 |
| 無酸素槽酸素混入 | 0.798 | 0.893 | 0.072 | 70分 |
| 流量過負荷 | 0.549 | 1.000 | 0.441 | 5分 |
| 散気管目詰まり | 0.206 | 0.037 | 0.000 | 32295分 |
| 毒物混入 | 0.909 | 0.985 | 0.063 | 20分 |
| 雨天時水質負荷低下 | 0.500 | 0.722 | 0.221 | 5分 |
| 平均 | 0.582 | 0.664 | 0.102 | 3757分 |

Table 5 Evaluation of FDI Results (Proposed)

| | 指標 | 検出率 | 誤検出率 | 検出時間 |
|-------------|-------|-------|-------|--------|
| MLSSセンサドリフト | 0.605 | 0.629 | 0.005 | 1365分 |
| DOセンサドリフト | 0.226 | 0.157 | 0.078 | — |
| DO制御ハンチング | 0.797 | 0.827 | 0.007 | 15分 |
| バルキング | 0.459 | 0.464 | 0.016 | 30分 |
| 硝化阻害 | 0.726 | 0.828 | 0.077 | 10分 |
| 無酸素槽酸素混入 | 0.787 | 0.875 | 0.065 | 70分 |
| 流量過負荷 | 0.560 | 0.994 | 0.422 | 5分 |
| 散気管目詰まり | 0.285 | 0.190 | 0.003 | 12190分 |
| 毒物混入 | 0.879 | 0.990 | 0.095 | 15分 |
| 雨天時水質負荷低下 | 0.413 | 0.810 | 0.424 | 5分 |
| 平均 | 0.574 | 0.676 | 0.119 | 1523分 |

Table 3~Table 5 の評価指標を見ると、平均的な異常診断性能は(B), (C), (A)の順で優れている。しかし、(B)のクラスタリング法では、8 番目の「散気管目詰まり」のシナリオの様に、他手法よりも格段に劣化する閾値を与える場合があり、必ずしも安定的に閾値を設定できない。一方、(C)の提案手法は、平均的には若干(A)の理論値よりも優れているが、No.5, No.7, No.8のシナリオで理論値よりも悪く、No.9とNo.10

のシナリオで理論値より優れているので、理論値とはほぼ同等の性能であると考えられる。異常診断性能は同等であるが、提案手法は、理論値よりも異常検出感度を下げる(異常検出率と異常誤検出率を共に下げる)様に閾値を設定する傾向を持っている。この特徴の違いが実データの異常診断において、どのような診断結果を与えるかを次節で検討する。

3.2 下水処理場の実プロセスデータを用いた評価

ここでは、実際の下水処理場のプロセスデータに対して提案手法を適用し、実際のプロセスデータへの提案手法の適用可能性を評価する。但し、実データでは「異常状態」を明確に定義できないため、主観的な判断に基づく評価となる。

(1) 対象プロセス

対象プロセスを、A2O法で処理されている下水処理場とし、曝気風量制御関連の Table 6 の計測変数を用いた。

Table 6 Measurement Variables for FDI (Real Process)

| No | 計測変数 | 単位 | No | 計測変数 | 単位 | No | 計測変数 | 単位 |
|----|-----------|---------------------|----|-----------|---------------------|----|----------|---------------------|
| 1 | No.1DO 濃度 | [g/m ³] | 12 | No.3 流入量 | [m ³ /h] | 23 | No.6 弁開度 | [%] |
| 2 | No.1 空気量 | [m ³ /h] | 13 | No.4DO 濃度 | [g/m ³] | 24 | No.6 流入量 | [m ³ /h] |
| 3 | No.1 弁開度 | [%] | 14 | No.4 空気量 | [m ³ /h] | 25 | 流入 TN 濃度 | [g/m ³] |
| 4 | No.1 流入量 | [m ³ /h] | 15 | No.4 弁開度 | [%] | 26 | 流入 TP 濃度 | [g/m ³] |
| 5 | No.2DO 濃度 | [g/m ³] | 16 | No.4 流入量 | [m ³ /h] | 27 | ブロウ1 弁開度 | [%] |
| 6 | No.2 空気量 | [m ³ /h] | 17 | No.5DO 濃度 | [g/m ³] | 28 | ブロウ2 弁開度 | [%] |
| 7 | No.2 弁開度 | [%] | 18 | No.5 空気量 | [m ³ /h] | 29 | ブロウ3 弁開度 | [%] |
| 8 | No.2 流入量 | [m ³ /h] | 19 | No.5 弁開度 | [%] | 30 | ブロウ1 曝気量 | [m ³ /h] |
| 9 | No.3DO 濃度 | [g/m ³] | 20 | No.5 流入量 | [m ³ /h] | 31 | ブロウ2 曝気量 | [m ³ /h] |
| 10 | No.3 空気量 | [m ³ /h] | 21 | No.6DO 濃度 | [g/m ³] | 32 | ブロウ3 曝気量 | [m ³ /h] |
| 11 | No.3 弁開度 | [%] | 22 | No.6 空気量 | [m ³ /h] | 33 | 散気管圧力 | [MPa] |

(2) 評価方法

シナリオデータを用いた場合と同様に、3つの手法を比較評価する。但し、性能を定量評価できないため、結果の評価は、評価者の主観的な判断に依存する。

(3) 評価結果と考察

Q統計量および Hotelling の T²統計量に対する(A)~(C)の方法による閾値の設定値を Fig.6 に示す。なお、Fig.6 では、各統計量の値を対数で表示している。また、異常診断結果例として、13番目の No. 4DO 濃度と 29番目のブロウ3弁開度の診断データを Fig.7 に示す。Fig.7 の●印の箇所は、異常検出後にその変数が要因であると推定された箇所である。

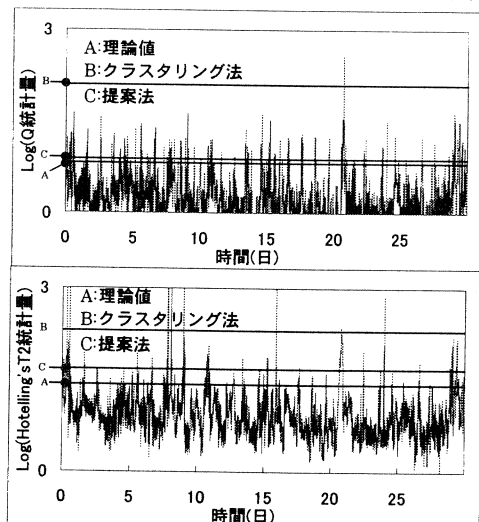


Fig.6 Comparison of Obtained Threshold Levels

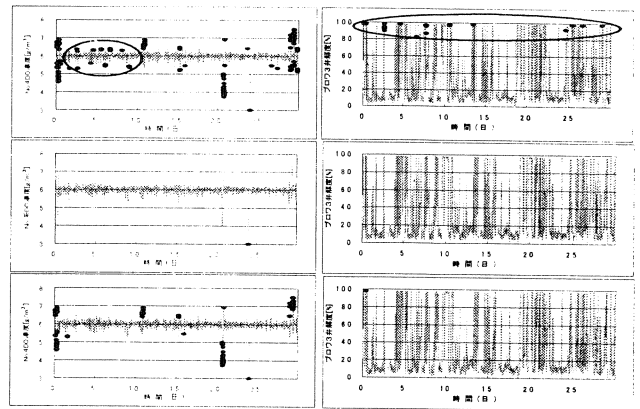


Fig.7 Examples of Fault Diagnosis

Fig.6 より、両統計量の閾値は、異常検出感度の高い順に、(A)、(C)、(B)となっている。一方、Fig.7 を見ると、(B)のクラスタリング法は異常をほとんど捉えておらず、(A)の理論値は(C)の提案手法よりも不必要に異常を検出する傾向が読み取れる(囲み部分)。この結果より、提案手法は異常検出の適切な閾値を自動的に設定するための有効な方法であると推測される。但し、評価の確立のためには、多くの評価者の判断を収集する必要がある。

4. まとめ

本論文では、各種のプロセスデータ間の相関情報を利用した MSPC と呼ばれる手法を用いた異常診断システムの構築法を提案し、下水処理プロセスを対象として、その有効性を評価した。特に、一定の異常診断性能を確保しながら人間が介在するエンジニアリングを最小にする異常診断システムの構築のために、MSPC で用いられる異常検出用データの閾値を自動的に設定する方法を提案した。

まず、シナリオデータによる評価の結果、提案手法は、閾値の理論値よりも異常検出感度を若干下げながら、異常診断性能をほぼ同等に保つように閾値を設定することがわかった。理論値を用いると、異常誤検出が増加する場合があることが指摘されているため、この結果から提案手法の有用性が示唆された。そこで、次に実データを用いて評価したところ、提案手法は理論値よりも適切な閾値を与えると推定された。さらに、提案手法は、文献3)の様に理論値が存在しない任意の異常検出用データに適用できるため、一定の精度を持つ異常診断システムの自動構築に役立つと考えられる。

[参考文献]

- 1) 加納:「多変量統計のプロセス管理」,
<http://tech.chase-dream.com/spc/book.html>
- 2) C.Rosen “Monitoring Wastewater Treatment Systems”, Lic.Thesis, IEA, Lund University, Lund, Sweden 1998
- 3) 山中理:「主成分分析と離散ウェーブレット変換を用いた下水処理プロセスの異常診断」, H17年電気学会 D 部門大会予稿集, II-73-76,2005
- 4) M.Sezgin *et al* “Survey over Image Thresholding Techniques and Quantitative Performance Evaluation”, J. of Electric Imaging 13(1), 146-165, 2004