

# オンラインデータの異常値検出システム

## Unusual Data Online Detection System

平林 和也\*、藤田 逸朗\*、金谷 利憲\*  
Kazuya Hirabayashi\* Itsuro Fujita\* Toshinori Kanaya\*

津村 和志\*\*  
Kazushi Tsumura\*\*

\* (株)安川電機基礎研究所基盤技術研究室

Research Laboratory, YASKAWA Electric Corporation  
Yahata-nishi-ku Kitakyushu 806 Japan

\*\* 京都大学工学部環境地球工学教室

Division of Global Environmental Engineering, Kyoto University  
Yoshida, Sakyo-ku, Kyoto 606 Japan

### Abstract

The latest sensor technology development has introduced a number of automatic measuring instruments for wastewater treatment plant. However, it is rather difficult to measure data and automatically control the process over a long time. This is because the wastewater treatment is under adverse conditions, such as dirt or sediment in the measuring part, high humidity and corrosive gas. The authors have developed a detection system for unusual data in order to assure reliability of measured data in the wastewater process. The system confirms the validity of the measured data by six methods; upper/lower limit, excessive change rate, insufficient change rate, prediction error (relation to other instruments), comparison with similar instruments, and trends. The measured data are integrated to make an overall judgment. In this report, we introduce the outline for an unusual data detection system and application example of measuring actual water quality data.

**Key words :** wastewater treatment process, unusual data detection, principal component analysis, online data

## 1. はじめに

近年の計算機技術の進歩により、下水処理プロセスの計装システムは高度化し、多数のプロセス情報をCRTによる集中監視システムによって監視している。すなわち、大規模処理場では、処理場内の多数の計測器からの情報、さらには、ポンプ場などの遠方設備からも情報を得て運転員が監視・運転を行っている。一方、小規模処理場では、少人数の運転員で多くの情報を、常時、監視・運転しなければならない。また、センサー技術の進歩に

より下水処理場では多くの自動計測器が導入され、より多くの情報を得られるようになってきた。例えば、リン・窒素など今まで計測困難だった項目が、自動計測されるようになってきている。このように今後も、プロセスをより良く運転することを目指して、新しい計測器の導入、計測ポイントの増大が予想され、その結果オペレータが監視しなければならない項目は更に増えてくる。しかし、現実には、運転員が常時(24時間)、多数の情報を監視することは困難なことである。

そこで筆者らは、運転支援システムの開発が必要と考

え、先ず、下水処理場における計測値の信頼性を向上させることを目的として、オンラインデータの異常値検出の研究を行ってきた。すなわち、種々の角度(上下限值、変化量過大、変化量過小、他の計測器との関係、類似計測器との照合、トレンドなど)から計測値の妥当性をチェックし、その結果を総合的に判定するシステムの開発である。本報告では、この異常値検出システムの概要と実処理場で収集した4種の水質計測値に適用した例について述べる。

## 2. 異常値検出システム

異常値検出システムにおける重要な課題としては、次の2点が挙げられる。1つは処理場、計測器、季節変動などの特性に自動的に対応できる判定基準値の更新方法であり、他の1つは様々な場面に対応できるデータの評価方法(チェック方法)である。ここでは、先ず、判定基準値の自動更新方法について述べ、つづいて、ソフトウェア構成、そして、ソフトウェアの具体的内容である異常値検出ロジックおよび総合判定について説明する。

### 2-1. 判定基準値の自動更新<sup>(1)</sup>

理想的には、その処理場を熟知している運転員が判定基準値を逐次設定することが望ましい。しかし、計測器が多数あること、及び季節変動に対応しなければならないことから、運転員が定期的な判定基準値を更新することは困難である。そこで、計算機によって自動的に判定基準値を更新する方法を開発した。通常、計算機による判定基準値の設定は、過去の計測値から標準偏差を求め、この値を基準に判定基準値を設定することが多い。しかし、下水処理プロセスで取り扱う計測値の頻度分布は必ずしも正規分布を示していない。Fig. 1に水質データの頻度分布の例を示す。Fig. 1から明らかなように、下方側に尾を引いた分布、上方側へ尾を引いた分布、ピークが2つある分布などがある。従って、正規分布を前提(標準偏差基準)にして判定基準値を設定する方法は適切でないと考えられる。そこで、正規分布を示さないデータ群の取り扱い方法として提案されている探索的データ解析法を用いた。この方法は、四分位数を基準に判定基準値を設定する。四分位数とは、相対累積度数を0~25%点、25~50%点、50~75%点、75~100%点の4つに分けることであり、通常は、25%点~75%(Fig. 2のH)点

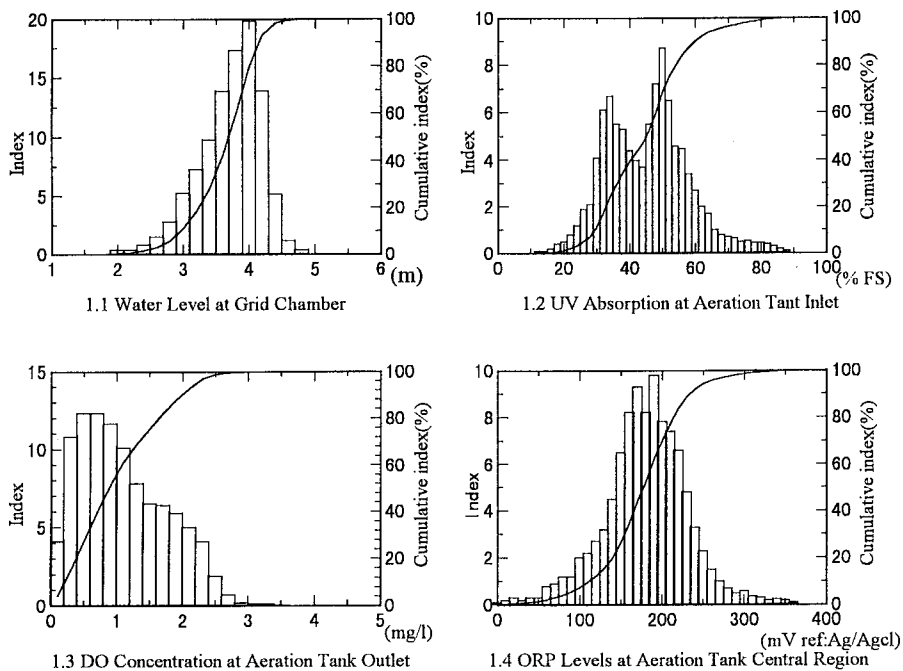


Fig. 1 Example of frequency distribution

を基準に判定基準値を設定する。しかし、Fig. 1に示すように、データの頻度分布は左に尾を引いたものもあれば、右に尾を引いたものもある。このようなデータの判定基準値を設定するために、四分位数に重みづけする必要があると考えた。重みづけによる判定基準値は、従来の基準(25%~75%)を25%点~50%点(Fig. 2のHl)、50%点~75%(Fig. 2のHu)点に分け、それを基準に設定した。標準偏差基準、四分位数基準、重みづけ四分位数基準による計測値の正常範囲をFig.2に示す。標準偏差基準の判定基準値に対し、四分位数基準の判定基準値は少し正常範囲を広くとっていることが分かる。また、重みづけ四分位数の正常範囲は頻度分布が上方側に重みがあるので、四分位数に比べ、少し上方側に正常値を広くとっている。ただし、四分位数と重みづけ四分位数の幅は等しく、重みのある方にシフトしたことになる。この方法により、処理場、計測器によって頻度分布の異なる様々なデータ(正規分布ではない)に対応できる判定基準値の設定が可能となると考えられる。また、計測値は季節によって変化するため、判定基準値は、直近30日のデータを使用して毎日更新することにした。Fig. 2の判定基準値を4種の計測器のデータに適用した結果をFig. 3に示す。この図は、計測値と判定基準値(上限値、下限値)および異常検出点(●点)を記している。このようにどの計測器にも自動的に対応した判定基準値を設定できていることから、この設定方法は妥当な方法と考えられる。

### 2-2. ソフトウェア構成

このシステムのソフトウェア構成は、大きく二つのブロックに分けられる。その一つは、計測値を様々な角度からチェックする異常値検出ブロック、他の一つは、検出ロジックの結果を総合的に判定する総合判定ブロックである。このようなシステムでは、検出ロジックの追加、変更及びプログラムの変更が頻繁に起こることが予想される。そこで、Fig. 4に示すように、プログラムの追加、

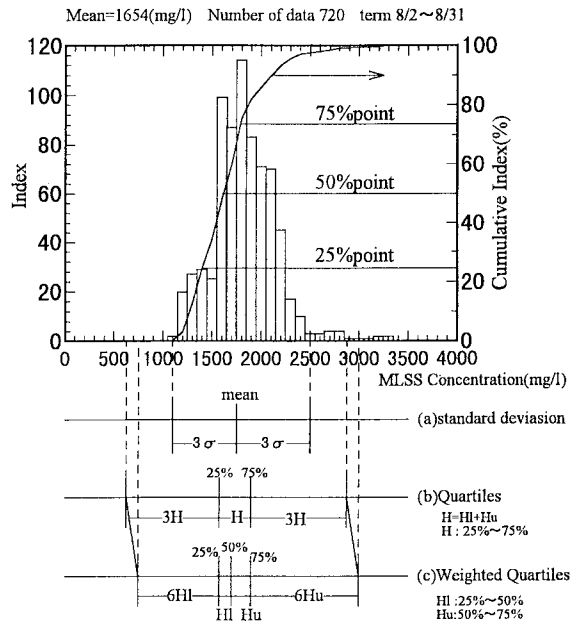


Fig. 2 Range of criteria by the used methods

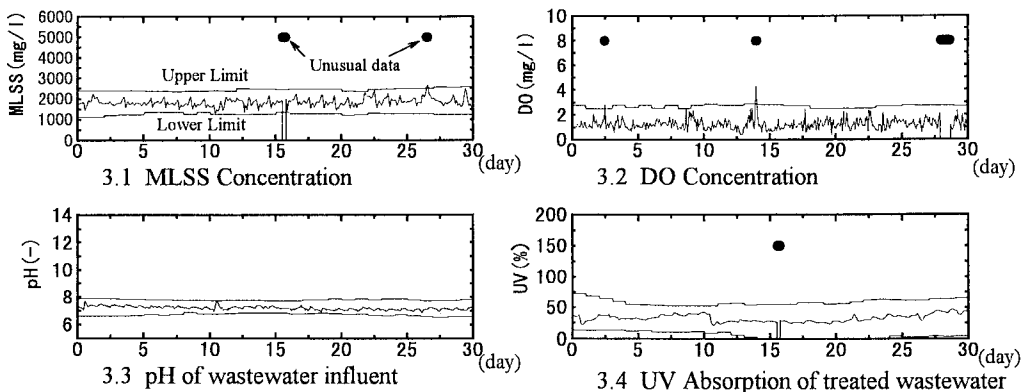


Fig. 3 Results of upper/lower limit

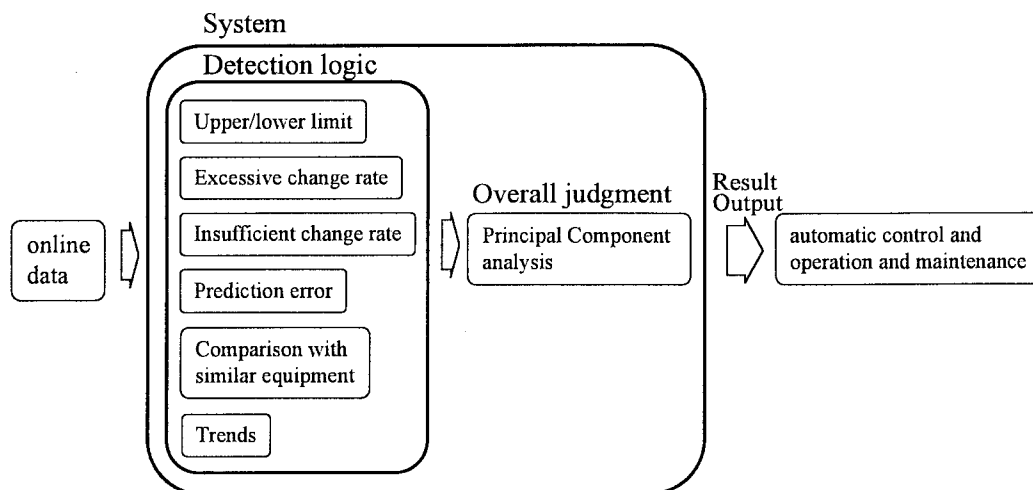


Fig. 4 Configuration of Unusual Data Detection System

変更が容易になるように検出ロジックを独立に構成し、また、判定基準値の設定、各検出ロジックの検出作業、総合判定、CRT表示の流れを独立に構成した。

### 2-3. 異常値検出ブロック

各検出ロジックのチェック方法を記すと次のようになる。

- (1) 上下限值による異常値検出<sup>(1)</sup>  
計測値の判定基準値を設定し、計測値がこの値を越えた時に異常とする。
- (2) 変化量過大による異常値検出<sup>(1)</sup>  
変化量過大の判定基準値を設定し、変化量がこの値を越えた時に異常とする。
- (3) 変化量過小による異常値検出  
変化量過小の判定基準値を設定し、変化量とその判定基準値内に連続して数点入った時に異常とする。
- (4) 他の計測器の影響を考慮にいたれた異常値検出(予測誤差)<sup>(1)</sup>  
計測値は操作量によって大きく変化する。例えば、曝気風量を増大させれば、DO計の値は上昇し、上下限值異常、変化量過大異常を示すことがある。しかし、この値の上昇は曝気風量を増大させたために起こった現象であり計測器は正常に動いていると考えられる。そこで、自己回帰モデルを使用して、検出対象の計測器に影響を及ぼす要因を考慮に入れた予測を行い、その予測誤差が判定基準値を越えた時異常とする検出口

ジックを構築した。自己回帰モデルによる予測式を下記に示す。

$$Y_p(n) = \sum_{m=1}^M A(m) Y(n-m) \quad \dots \dots (1)$$

ここで、

- $Y_p(n)$  : n 時点の予測値ベクトル
- $A(m)$  : 自己回帰係数行列
- $Y(n-m)$  : (n-m) 時点の計測値ベクトル
- M : モデル次数

#### (5) 類似計測器比較による異常値検出<sup>(2)</sup>

下水プラントにおいて運転員は、ある系列の MLSS 濃度が急変した場合、他の処理系の MLSS 濃度がどのような動きを示しているかを確認する。また、処理水の紫外線可視吸光度差計(以下 UV 計と略記する)の値が変化した場合は、数時間前の流入水 UV 計の動きを確認する。このように運転員が通常行っていると思われる類似計測器の照合方法を、この検出ロジックで構築した。その手法は、類似計測器間の判定基準値を設定し、類似計測器の計測値の差がこの値を越えた時に異常とするものである。

#### (6) トレンドによる異常値検出<sup>(2)</sup>

計測器の洗浄装置が故障するとセンサー部に汚れが蓄積し、データが一方向的に上昇または下降する。このような場合、データがゆっくりと変化するために運転員も発見することが困難である。このような異常を検出するためにトレンドによる異常値検出を考えた。

各検出ロジックの適用例を MLSS 濃度計、DO 計、pH 計、UV 計を例にあげて Fig. 5 に示す。これらの計測器を検証に用いたのは、水質計測器の中で、応答の速い計測器、遅い計測器など様々な計測器に、このシステムが対応できるかどうかをチェックするためである。サンプリング周期は、すべて60分で行っている。この図から、従来の上下限值による異常値検出だけでは検出できなかった点も異常値として検出している。また、Fig. 5.1の(a)、Fig. 5.2の(b)のように異常の確率が高いと思われる点は、各検出ロジックで異常を検出している。しかし、それぞれの検出ロジックで異常値として検出された点は、異なった結果を出力している。例えば、Fig. 5.1のa'点では、変化量過大は異常、上下限值、予測誤差、類似計測器の照合などは正常と出力されている。またFig. 5.2のb'点では、変化量過大、類似計測器の照合は異常、上下限值、予測誤差などは正常と、異なった結果を出力している。このように6種の検出ロジックはそれぞれ違う角度からデータをチェックしているため、異なった結果が出力される。そのため、検出ロジック毎に異常警報を出力すると、実際の運転監視の場面では煩雑となる。したがって、これらの結果を計算機が総合的に判定することが要求される。

## 2-4 総合判定

### (1) 総合判定の必要性

各検出ロジックから得られた結果を、総合的に評価するために主成分分析<sup>(5)</sup>を適用することを試みた。ただし、6種の検出ロジックの中で、変化量過小による異常値検出は計測器の故障などに原因があると考えられ、また、トレンドによる異常値検出は他の検出ロジックとチェック周期が異なる。そこで、この種の検出結果については、総合判定を行わず計測値異常として出力することにした。そして、その他の4つの判定結果を対象とした主成分分析を行い総合判定を試みた。<sup>(3)</sup>

### (2) 主成分分析

主成分分析とは、多くの変量の相関関係を解析し、そ

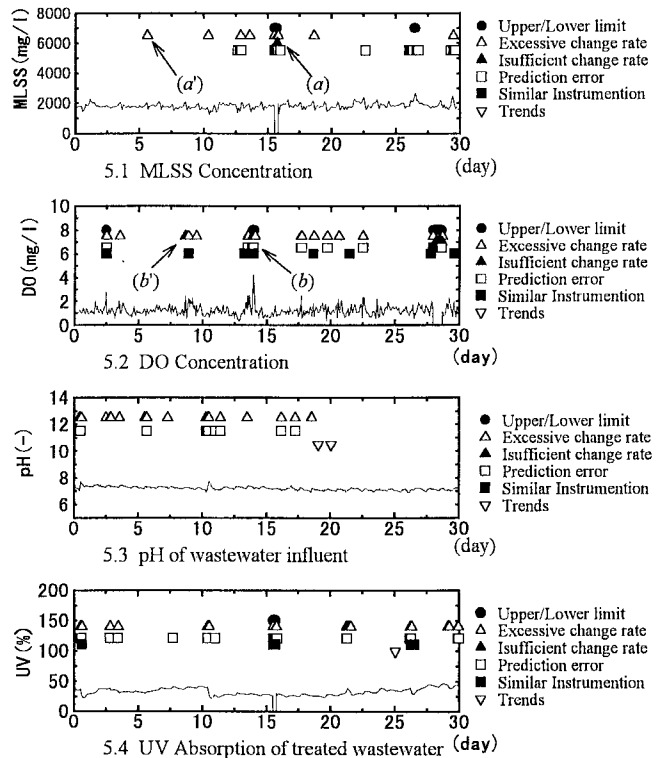


Fig. 5 Results of Each Unusual Value Detection Logic

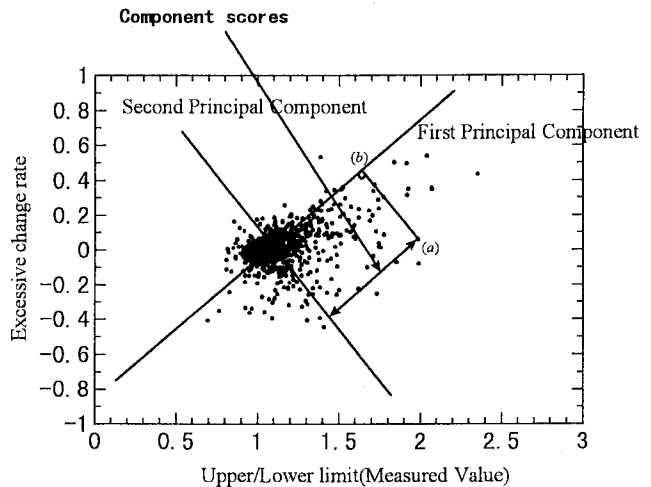


Fig. 6 Results of Component Scores

れぞれの変量の持つ変動を1個または少数個の総合指標で説明しようとするもので、情報の圧縮を意図した解析方法である。つまり、変量  $X_1, X_2, \dots, X_p$  の値を

できるだけ情報の損失なしに、次式の総合指標  $Z_m$  で代表する手法である。

$$Z_m = \sum_{p=1}^N A_{mp} X_p \quad \dots \dots \dots (2)$$

ここで、

$Z_m$  は総合指標である第  $m$  成分

$X_p$  は各変量の値

$A_{mp}$  は変量  $X_p$  の第  $m$  主成分係数

主成分得点の考え方を Fig. 6 に示す。説明を分かりやすくするために、2次元(上下限值、変化量)で説明する。この2変数の散布図から分散の1番大きい方に引いた軸を第1主成分という。また、その次に分散が大きく、第1主成分と直交する軸を第2主成分という。第1主成分得点は、計測された(a)点から第1主成分の軸に垂直におろした(b)点と重心(第1主成分と第2主成分の交

る点)の距離として求めることができる。

### (3) 判定方法

主成分分析を利用した総合判定には、次のデータを用いることにした。上下限は計測値、変化量過大は変化量、予測誤差は予測誤差、類似計測器の差は類似計測値の差を使用した。このデータ群の直近30日のデータ(サンプリング周期60分の場合:データ数720)から主成分分析を入力するデータを算出する。この4種のデータから主成分分析を行い、固有値、寄与率、主成分係数、主成分得点を計算する。この結果を基に、総合指標算出の検討を行った。Table. 1 に MLSS 濃度に適用した場合の主成分分析の結果を示す。第1主成分の寄与率は43%となっており、検出ロジックの基本部分を1つの指標にまとめていると考え、第1主成分得点を総合指標に適用することにした。

Table. 1 Eigenvalue, Proportion, Cumulate proportion by correlation vector				
	1st principal component	2nd principal component	3rd principal component	4th principal component
Eigenvalue	1.721305	1.171357	0.592310	0.515028
Proportion	0.430326	0.292839	0.148077	0.128757
Cumulate proportion	0.430326	0.723166	0.871243	1.000000

## 3. 異常値検出システムの検証

総合判定の検証に適用したデータは、Fig. 5 と同じ MLSS 濃度計、DO計、pH計、UV計である。Fig. 7 には、各計測器の主成分得点と判定基準値及びオリジナルデータと異常検出点(●点)を記している。また、判定基準値の設定は、直近30日のデータを使用して毎日更新している。各計測器の異常値検出結果を見てみると、応答の早い計測器と遅い計測器いずれについても、少し過敏に計測値異常を検出しているようである。しかし、pH計のa点(Fig. 7. 3. 2の1日、10日頃)、UV計のb点(Fig. 7. 4. 2の1日、10日、22日、27日頃)などは、異常な動きといえる。運転員にとって、日常の運転の中でこのような変動を発見することは困難と思われるが、このシステムではこれらのデータを異常値として検出している。この4種の計測器においては、このように異常と思われる点をほとんど検出している。この結果から、主成分分析による総合判定は、各検出ロジックで計算された結果を主成分得点として1つにまとめ、また、様々

な要因がからみあった計測値の動きをうまく抽出しているといえる。このように、異常値として検出された値は、警報によって運転員に知らせる。また異常の現象を知りたい場合は、Fig. 5 に相当するグラフも見ることができる。

本システムの検出結果は、全体的に過敏な検出をしているので、主成分得点の値と判定基準値の関係について検出してみた。例えば、Fig. 7. 1 の MLSS 濃度計では26日頃、Fig. 7. 2 の DO 計では9日頃に検出された異常は、判定基準値をわずかに越えたものである。これに対して MLSS 濃度計の16日頃、DO 計の3日、14日、18日頃に検出された異常は、判定基準値を大きく逸脱している。従って実用的には、異常度を大、中、小とレベル分けすることなどが考えられる。異常度のレベル分けは、プロセスによって異なるため、各下水処理プロセスに対応した設定をすることになろう。

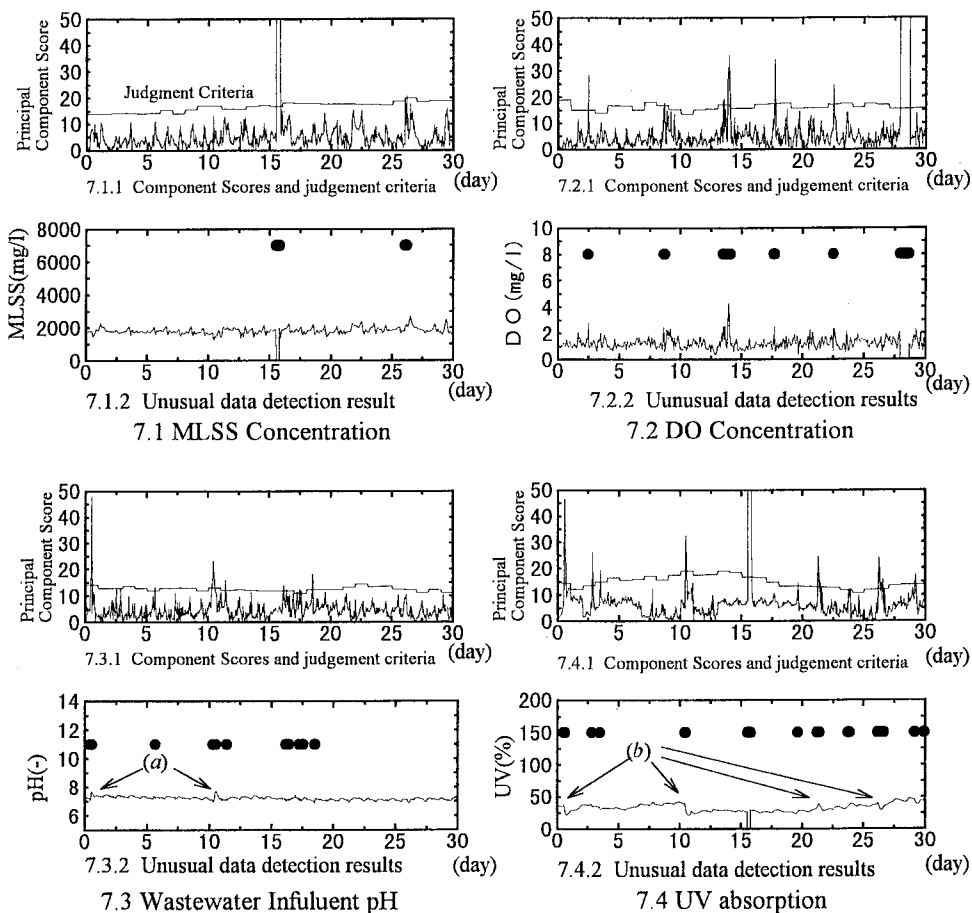


Fig. 7 Unusual Data Detection Results for Measurement Instrument

#### 4. 成果と今後の課題

計測値の妥当性をチェックするために、様々な角度からデータをチェックした。そして、各検出口ジックから得られた結果を総合的に判定する手法を構築することができた。今後は、実処理場においてテストを行うことによって、このシステムの適応性、異常度のレベル分けについて検討する。そして、運転員に使いやすいシステムを構築していく予定である。

#### 【参考文献】

- (1) 平林、金谷、津村. 水質計測器の異常検出技術の研究. 第31回下水道研究発表会講演集、pp. 725-727, 1994.
- (2) 平林、藤田、金谷、津村. 下水処理場におけるオンラインデータの異常検出. 第32回下水道研究発表会講演集、pp. 820-822, 1995.
- (3) 平林、藤田、金谷、津村. 主成分分析を用いたオンラインデータの異常検出システム. 第33回下水道研究発表会講演集、pp. 1000-1002, 1996.
- (4) T. Kanaya, K. Hirabayashi, I. Fujita, K. Tsumura. Detection of Unusual Data in on-line Monitoring of Wastewater Processing. Water Science and Technology, vol.33 No.1, pp. 71-79, 1996
- (5) 田中、垂水、脇本：パソコン統計解析ハンドブックⅡ 多変量解析編、1984、共立出版

(受付 1996.12.20)

(受理 1997. 2. 4)