

研究メモ

# ソフトセンサー/バーチャルセンサーの設計

- 測定困難な製品特性を推定する技術 -

京都大学大学院工学研究科化学工学専攻

プロセスシステム工学研究室

加納 学

<http://www-pse.cheme.kyoto-u.ac.jp/~kano/>

2009年03月 初版作成

Copyright ©2009 Manabu Kano. All rights reserved.

本資料の全部または一部を無断で複写複製すること、および配布することを禁じます。

## 1 はじめに

ソフトセンサー (soft-sensor) とは、通常の計測機器のようにハードウェアとしての実態があるセンサーとの対比で用いられる用語であり、ソフトウェアによって実現されるセンサーを指す。化学、半導体、製薬など様々な産業界において、顧客に送り出す製品の品質がリアルタイムに計測されることは希である。多くの場合、抜き取り検査で、かつ分析時間が長いために、分析結果に基づくフィードバック制御が困難とされる。そのような状況において、仕様を満たす製品を高効率に生産するためには、どうしても製品品質を推定する必要がある。そこで、比較的容易にオンライン測定される変数の測定値から、製品品質を推定できるモデルを何らかの方法で構築し、その予測値を活用しようという着想を得る。このモデルこそがソフトセンサーである。ソフトセンサーは、推定モデルやバーチャルセンサーとも呼ばれ、分野や人によって呼び方は様々なようだ。本報では、以後、ソフトセンサーと呼ぶことにする。

厳密に考えると、ハードウェアセンサーとソフトセンサーの区別は実に曖昧であることに気付く。例えば、果物の糖度を非破壊検査するため、あるいは化学製品の濃度を計測するために利用される近赤外分光分析装置 (通称 NIR) は、どちらに分類されるだろうか。高価なハードウェアセンサーと捉えることもできるが、その内部では、照射した近赤外線各波長における吸光度が、統計的に求めた回帰式 (検量線と呼ぶ) によって、糖度や濃度に変換されている。高価なハードウェアが存在するが、ソフトセンサーとも言える。吸光度から濃度を求められる理論的根拠は、ランベルト・ベールの法則に求めることができるが、近年、製薬分野では Process Analytical Technology (PAT) という新たな概念の導入により、混合などの操作終了点をはじめ、製品とプロセスに関する実に様々な情報を NIR から得ようという取り組みが盛んに行われている。そうすると、もはや、ハードウェアセンサーとソフトセンサーを明確に区別することができなくなる。そこで、以下では、オンラインで得られる情報を用いて、製品品質など対象とする変数を推定するためのモデルをソフトセンサーと捉えることにする。

もちろん、ソフトセンサーは、物理モデルから構築することもできるし、統計的方法によって構築することもできる。対象プロセスの深い理解に基づいており、誰もが納得して利用できるという意味では、物理モデルに基づくソフトセンサーが望ましい。しかし、現実には、実用に耐えるソフトセンサーを物理モデルのみに基づいて構築することは困難である場合が多く、全部あるいは一部に統計的方法の力を借りることとなる。統計的方法のみを使用するモデルはブラックボックスモデル、物理モデルと統計的方法を組み合わせるモデルはグレイボックスモデルとも呼ばれる。本報では、ブラックボックスモデルに分類される、統計的なソフトセンサーを対象とする。

ソフトセンサーは様々な産業界で重要な役割を果たしているが、解決しなければならない課題も多い。大雑把に課題を分類すると、構築に関する課題と保守に関する課題にわけられる。産業応用上、現在深刻な問題になっているのは、ソフトセンサーの保守である。つまり、実用に耐えうる推定精度のソフトセンサーが構築できたとしても、プロセス特性や運転条件の変化によって、推定精度が劣化してしまうという問題が顕在化している。モデルを作り直せばよいというのは現場を無視

した意見であり，現実には，そのエンジニアリングコストは許容されそうにない．

そこで，本報では，ソフトセンサーの概要を紹介し，その構築に利用されている様々な統計的方法について解説した後，克服すべき課題を指摘し，問題解決に向けての取り組みをいくつか紹介したい．特に，各手法の特徴（長所と短所）および保守負荷低減に向けての取り組みに重点を置く．

## 2 統計的方法とその特徴

本節では，ソフトセンサー構築に広く利用されている統計的方法を概説する．取り上げるのは線形回帰手法に限定するが，これは，産業応用上，線形モデルで十分な場合が多いためである．

もちろん，現存する生産プロセスのほぼすべてが非線形システムである．単純な液タンクにしても，ベルヌーイの定理に従って，タンクからの流出流量がタンク内の液高さの平方根に比例するならば，非線形システムである．しかし，わざわざタンクを非線形モデルで表現する必要はない場合が多いだろう．実は，より複雑なプロセスであっても，非線形モデルに手を出す前に，線形モデルで十分に近似できるのではないかと考えてみるべきである．仕様を満たす製品を製造するために，ある限定された領域で運転されているプロセスを対象とするのであるなら，線形近似が有効な場合は多い．線形モデルが有効でない場合には，非線形モデルを試すことになる．ニューラルネットワークの他，近年では，サポートベクトル回帰も注目されている．

### 2.1 重回帰分析

ソフトセンサーを構築するに際して，まず最初に思い浮かぶ手法は，重回帰分析である．モデルの入力変数を  $x_m (m = 1, 2, \dots, M)$ ，出力変数（簡単のため 1 変数とする）を  $y$  とすると，重回帰モデルは

$$y = \sum_{m=1}^M a_m x_m + e \quad (1)$$

で与えられる．ただし，事前に各変数の平均値  $\bar{x}_m, \bar{y}$  を 0 に合わせておくと約束する． $x_m$  によって  $y$  の変動のすべてを表現できないために，誤差  $e$  が生じる．モデル中の定数  $a_m$  は偏回帰係数と呼ばれ，入力から出力への影響の程度を表す． $N (\geq M)$  をサンプル数として，入出力の測定データ行列をそれぞれ  $X \in \mathbb{R}^{N \times M}$ ,  $y \in \mathbb{R}^N$  とおくと，偏回帰係数  $a = [a_1, a_2, \dots, a_M]^T$  は，誤差の二乗和を最小化するための必要条件から導かれる正規方程式の解として，次式で求められる．

$$a = (X^T X)^{-1} X^T y \quad (2)$$

この結果は単純明解であり，データさえ用意すれば，様々なソフトウェアを用いて容易に計算することができる．しかし，それが故に，誤用する危険性も高い．

重回帰分析を理解し，より高度な線形回帰手法へと進むためには， $N$  次元線形空間の元である

$y, \hat{y}, x_m$  の関係を知る必要がある．出力の推定値は

$$\hat{y} = \sum_{m=1}^M a_m x_m = \mathbf{X} \mathbf{a} \quad (3)$$

で与えられる．つまり，推定値  $\hat{y}$  は  $M$  個の入力  $x_m$  によって張られる  $M$  次元線形部分空間  $S$  の元である．残念ながら，一般に測定値  $y$  はこの部分空間  $S$  内に存在しないため，推定誤差  $e = y - \hat{y}$  が生じる．誤差の大きさ  $\|e\|$  を最小とする  $\hat{y}$  は， $y$  から部分空間  $S$  へ垂線を降ろせば求められる．このとき，推定値  $\hat{y}$  と推定誤差  $e$  は直交するため，

$$\langle \hat{y}, e \rangle = \mathbf{a}^T (\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \mathbf{a}) = 0 \quad (4)$$

が必ず成り立ち，再び正規方程式が導かれる．ところで， $\|e\|$  が最小となるとき， $y$  と  $\hat{y}$  のなす角  $\theta$  も最小となる．したがって，関係式

$$\cos \theta = \frac{\mathbf{y}^T \hat{\mathbf{y}}}{\|\mathbf{y}\| \|\hat{\mathbf{y}}\|} = \frac{\sigma_{y\hat{y}}^2}{\sigma_y \sigma_{\hat{y}}} = \rho_{y\hat{y}} \quad (5)$$

より，重相関係数  $\rho_{y\hat{y}}$  が最大となる．つまり，重回帰分析とは，測定値  $y$  と推定値  $\hat{y}$  の相関係数を最大化する手法とも言える．なお， $\sigma_{y\hat{y}}^2$  は共分散， $\sigma_y, \sigma_{\hat{y}}$  は標準偏差である．

## 2.2 多重共線性

偏回帰係数  $a$  が一意に定まるためには， $x_m$  が線形独立でなければならない．そうでなければ，共分散行列  $\mathbf{X}^T \mathbf{X} / (N-1)$  が逆行列を持たず，正規方程式が解けない．実際には，測定ノイズの影響もあり， $x_m$  が厳密に線形従属になることはないが，線形従属に近い状態にはなりえる．注意すべきは，まさにこの状況である．逆行列が存在するため， $a$  は計算できるが，その信頼性は極めて低い．測定ノイズがほんの少し異なるだけで，全く異なる偏回帰係数  $a$  が得られる．ときに符号すら変化してしまう．そのような状態では，推定にはとても使えない．これがいわゆる多重共線性の問題であり，重回帰分析を利用する際には常に注意しておかなければならない．

多重共線性を克服するための方法はいくつかある．1つ目は，偏回帰係数が不必要に大きな値を取ることが問題であるとの立場から，誤差と同時に偏回帰係数も最小化する，つまり  $\|e\|^2 + \lambda \|\mathbf{a}\|^2$  を最小化する方法である．この手法はリッジ回帰と呼ばれ，

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

が解となる．ただし， $\lambda$  は定数である．2つ目は，入力変数を取捨選択することによって，互いに無相関な入力変数を用いて重回帰分析を実施する方法である．取捨選択には，変数についての知識を利用してよいし，変数増加法・減少法，ステップワイズ法などの統計的方法を利用してよい．3つ目は，互いに無相関となるような新たな変数を，入力変数の線形結合として作り出す方法である．このとき，新たな変数は元の変数の情報をできるだけ多く持つべきであると考えらるなら，主成分分析 (PCA) を採用するのが自然である．PCA は最もよく知られた多変量解析の1つで，入力

$x_m$  の線形結合である主成分  $z_r (r = 1, 2, \dots, R \leq M)$  を、互いに無相関になるように、かつ分散が最大になるように、求める方法である。こうして得られた主成分を入力変数として重回帰分析を実施する方法は主成分回帰 (PCR) と呼ばれる。

### 2.3 係数 影響度

上記いずれの方法を利用しても良いが、知っておくべきことがある。そもそも重回帰分析は、入力変数が無相関であることを前提としている。偏回帰係数が入力から出力への影響度と見なせるのも、入力変数が無相関であれば、影響を分離して把握できるためである。入力変数に相関があれば、程度の差はあるにしても、多重共線性の問題が生じる。リッジ回帰や PCR は巧妙にこの問題を回避しているが、そのために、係数にバイアスがのるという犠牲を払っている。つまり、もはや、係数 = 影響度ではない。したがって、ソフトセンサーを用いて出力を推定することには何の問題もないが、構築したモデルの係数から、各入力出力に及ぼす影響を把握しようとするれば、それは間違いである。この観点からは、ステップワイズ法などによる入力変数の絞り込みが望ましい。しかし、入力変数の絞り込みも万能ではない。機械的に選択される入力変数が運転員の知見と合致しないという指摘がなされることは少なくない。また、入力変数の数を少なくすれば、より安定した偏回帰係数が算出されるが、測定データに含まれる有用な情報をも捨てることになるため、推定精度が低下する恐れがある。実際そのような事例は存在し、それが次に紹介する Partial Least Squares (PLS) が好んで利用される理由にもなっている。

なお、多重共線性の問題を回避し、かつ影響度を正しく求めるためには、変数間の相関関係を正確に捉えた上で、各入力変数の変化が他の入力変数にどのような変化を促すか、そしてその全体の結果として出力変数がどのように変化するかを算出する必要がある [1]。

### 2.4 Partial Least Squares (PLS)

前述の PCR は多重共線性に対処できる手法として有効であるが、出力の推定という観点からは最善といえない。入力変数のみに基づいて導出される主成分が必ずしも出力と高い相関を持たないためである。この問題に対処するために考案された方法が PLS である。

重回帰分析は、重相関係数  $\rho_{y\hat{y}} = \cos \theta$  を最大化する方法であった。一方、PCR/PCA は、主成分の分散  $\|z_r\|^2$  を最大化する方法であった。PLS は、入力変数の線形結合である潜在変数  $z_r$  と出力  $y$  の内積

$$\langle y, z_r \rangle = \|y\| \|z_r\| \cos \theta_r \quad (7)$$

を最大化する方法である。これは、 $\|z_r\|$  と  $\cos \theta_r = \rho_{yz_r}$  の積を最大化しているということであり、PLS は重回帰分析と PCR を組み合わせた手法であると解釈できる。より具体的には、入力変数間の相関関係と入出力間の相関関係を同時に考慮して潜在変数を決定し、その潜在変数を入力変数とする線形回帰モデルを構築するのが PLS である。一般に、PLS は PCR と比較して、より少数の潜在変数を用いて、より高い推定精度を実現できるとされる。なお、PLS においても、係数が

影響度に一致しないことを指摘しておく。

## 2.5 動特性を捉えたモデリング

PLS に代表される多変量解析あるいはケモメトリクス (Chemistry と Metrics からの造語で計量化学とも呼ばれる) 手法をソフトセンサー設計に利用するためには、変数の静的な相関関係のみならず、プロセスの動特性をモデル化する必要がある。そのための方法が Dynamic PLS (DPLS) である。DPLS とは、入力変数の現時刻の測定値のみならず、その過去の測定値も利用する方法であり、プロセスの動特性を捉えたソフトセンサーの構築に適している。もちろん、PLS 以外の手法を用いることもできる。しかし、複数の時刻における測定値を列挙して入力変数とみなす方法では、入力変数の数が膨大になってしまう。そこで、動特性を持つ多変数プロセスのモデル化に有効である部分空間同定 (SSID) のソフトセンサー設計への応用が考えられる [2]。ところが、状態空間モデルとカルマンフィルタに基づく推定法は、システムが白色雑音によってドライブされ、かつ測定されていない外乱 (非観測外乱) の特性が時間的に変化しないという前提条件の下で最適であるものの、現実のプロセスではこれらの前提条件が満たされないことも多く、結果としてソフトセンサーの推定精度が低下するという問題がある。この問題を克服するために、入出力変数の測定データから非観測外乱を推定すると共に、非観測外乱が製品品質に及ぼす影響をモデル化することによって、精度の高いソフトセンサーを構築する方法が提案されている。この方法は 2 段階部分空間同定 (TS-SSID) と呼ばれ、1) SSID による観測入出力変数に対する状態空間モデルの構築、2) 出力残差変数からの非観測外乱の推定、3) SSID による製品品質推定用状態空間モデルの構築、という 3 ステップから構成される [3]。

## 3 ソフトセンサーの実際

本節では、設計・運用・保守の観点から、産業応用事例を通して、ソフトセンサーの現状を紹介する。

### 3.1 ソフトセンサーの設計

昭和電工でのソフトセンサー構築事例 [4] を紹介する。対象はエチレンプラントで主要なタワーの一つであるエチレン精留塔 (T431/2) である。その概略図を図 1 に示す。本精留塔は下段の T431 および上段の T432 の 2 塔から成り、原料は下段の T431 に供給されている。原料は主としてエチレン、エタンの 2 成分である。分離により得られる製品エチレンの主なスペックは製品中エタン濃度であり、省エネルギーを図るためには不純物であるエタン濃度をできるだけ製品スペック上限ぎりぎりまで運転する必要がある。このため、本プロセスには、高度制御の一種である多変数モデル予測制御が導入されている。ここでは、製品エチレン中のエタン濃度を正確に推定するソフトセンサーを構築し、エタン濃度をオンライン制御することが目的となる。

Kano ら [5] は、下記 3 種類の蒸留塔製品組成推定モデルについて検討し、定常状態モデルや静

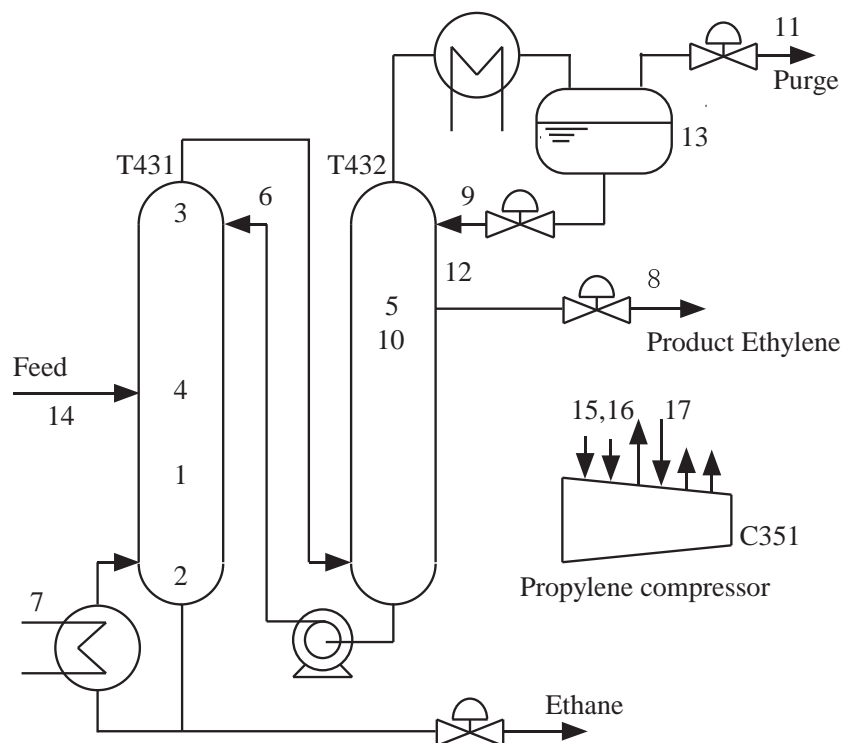


図 1 エチレン精留塔 (T431/2) 概略図

的モデルと比較して、動的モデルの推定精度が非常に良いことを示している。

- 定常状態モデル：定常状態データを使用する
- 静的モデル：時系列データの過去のデータを使用しない
- 動的モデル：時系列データの過去のデータを使用する

そこで、プラントでオンライン測定されている流量、圧力、温度などのプロセス変数を利用して、DPLS モデルを構築した。モデルの構築には約 3 ヶ月間の運転データを使用した。入力変数として、タワー周りのプロセス変数 (32 変数) から推定精度に影響の大きい 17 変数を選択した。選択した変数の番号が図 1 に記載されている。番号 1, 2, 3, 4, 5, 16 の 6 変数は温度、6, 7, 8, 9, 10, 11 の 6 変数は流量、12 は還流比、13, 15, 17 の 3 変数は圧力、そして 14 が原料中エタン濃度である。なお、モデル構築には 5 分平均の運転データを使用した。

エタン濃度推定結果を図 2 (最上段：実測値との比較、2 段目：推定誤差) に示す。推定は定期的に異なる 3 期間で行っている。なお、推定結果およびその誤差はスケールリングしてある。

図 2 (上段) よりエタン濃度の推定が良好に行われていることが確認できる。この精留塔では、エチレン製品中のエタン濃度は PPM オーダーで管理されている。したがって、その相対誤差が 10 % 以下で推定できていることは、非常に精度の高いモデルが構築できたことを意味しており、モデル予測制御への適用も十分可能である。ここで、1100~1200 ステップ付近で大きな推定誤差が

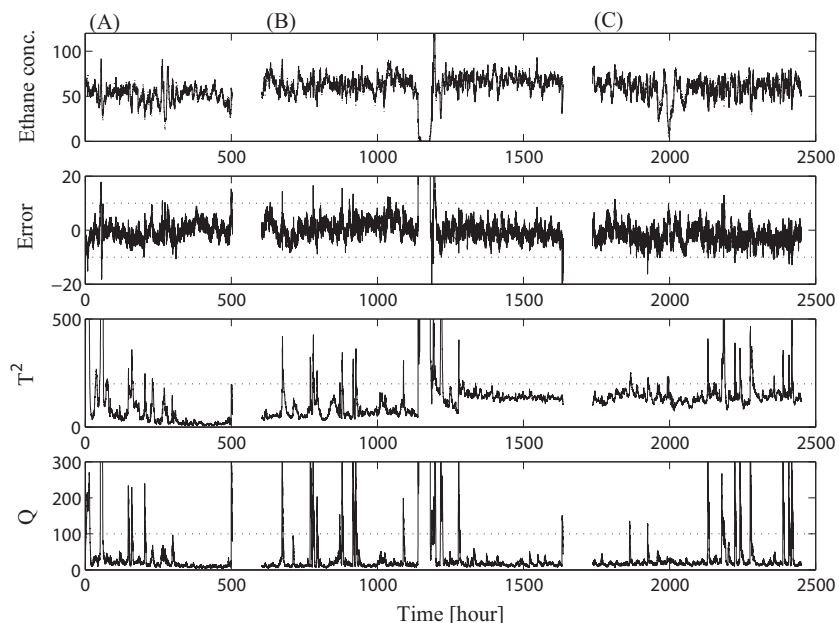


図 2 Dynamic PLS による組成推定および監視結果

表 1 DPLS と TS-SSID の推定精度比較

	$r$		RMSE	
	期間 D	期間 E	期間 D	期間 E
DPLS	0.75	0.74	5.34	4.18
TS-SSID	0.90	0.88	4.28	2.97

生じているのは、この期間で製品スペックを変更した非正常な運転が行われたためである。

さらに推定精度を向上させるために、TS-SSID を適用することもできる。先とは異なる 2 期間を対象に、重相関係数  $r$  と根平均二乗予測誤差 (RMSE) を指標として、DPLS と TS-SSID の推定精度を比較した結果を表 1 に示す [3]。この結果は、TS-SSID の有効性を示している。

### 3.2 ソフトセンサーの運用

ソフトセンサーは、モデル構築期間のデータを使用して、その期間のオンラインガスクロ分析値に合うように（実測値と推定値の誤差が最小になるように）構築した。つまり、運転状態がモデル構築期間とほとんど同じであるならば、その推定値の信頼度は高く、逆に、運転状態がモデル構築期間と大きく異なるなら、推定値の信頼度は低くなる。したがって、ソフトセンサーの妥当性をオンラインで評価するために、運転状態の監視が必要となる。そこで、モデル構築に使用した DPLS と多変量統計的プロセス管理 (MSPC) [6] を組み合わせて、運転状態の監視を行うこととした。

MSPC は、変数間の相関関係を考慮して高度なプロセス管理を実現するための方法であり、相関



関係をモデル化するために、PCA/PLS などの多変量解析が利用される。古典的な SPC が監視変数に管理限界を設定し、測定値が管理限界を超えると異常が発生したとみなすのと同様に、MSPC では監視変数から計算される 2 種類の統計量  $T^2$  と  $Q$  に対して管理限界を設定する。PCA を用いる MSPC においては、Hotelling's  $T^2$  統計量は主成分が張る部分空間内での原点からの規格化された距離（マハラノビス距離）を、 $Q$  統計量はその直交補空間における原点からの距離を表す。このため、 $T^2$  か  $Q$  のいずれかの統計量が管理限界を超えると、異常発生と判断できる。MSPC を利用することによって、現在の運転状態がモデル構築用データの内挿か外挿かを判断できる。

本検討における  $T^2$  統計量および  $Q$  統計量の管理限界は、概ね 99 % 信頼限界となるように、それぞれ 200, 100 と決定した。運転監視結果を図 2 (下段) に示す。1100~1200 時間付近の運転が非定常な期間では、明らかに監視指標が管理限界を超えている。さらに、その他の期間においても、 $T^2$  または  $Q$  統計量が管理限界を超える時刻において、推定誤差が大きくなっている。すなわち、構築した運転監視システムによって、ソフトセンサーの妥当性をオンラインで監視できることが確かめられた。

構築した運転状態監視システムは、オンライン分析計の異常監視にも利用できる。つまり、運転状態が参照期間の運転状態に近く、監視指標が管理限界内に存在する状態においては、推定値とガスクロ分析値に大きな誤差が生じていないはずであり、誤差が生じている場合は推定値よりもむしろオンライン分析計の異常である可能性が高いと判断できる。

### 3.3 ソフトセンサーの保守

ソフトセンサーは様々な産業プロセスで活用されているが、プロセス特性の変化により、推定性能が劣化するケースが少なくない。例えば化学プロセスでは、触媒の劣化や装置の汚れによって反応活性や伝熱抵抗が変化する。半導体プロセスでは装置メンテナンスの前後で装置特性が大きく異なる。このため、プロセス特性の変化に応じて、ソフトセンサーを更新する仕組みが必要である。ところが、モデル更新は非常に困難な作業であり、ソフトセンサーの保守が実用上大きな課題となっているのが現状である。

これまでに、モデルを逐次的に更新する Recursive PLS (RPLS) [7] が提案されている。RPLS では、新たなサンプル  $x_{new}$ ,  $y_{new}$  が得られると、次式に従って入出力データ行列を更新し、モデルを再構築する。

$$X_{new} = \begin{bmatrix} X \\ x_{new}^T \end{bmatrix}, Y_{new} = \begin{bmatrix} Y \\ y_{new}^T \end{bmatrix} \quad (8)$$

このままではサンプルが蓄積されるにつれて行列のサイズが増加してしまうが、このデータ行列から計算される PLS モデルは、

$$X_{new} = \begin{bmatrix} P^T \\ x_{new}^T \end{bmatrix}, Y_{new} = \begin{bmatrix} Q^T \\ y_{new}^T \end{bmatrix} \quad (9)$$

から得られる PLS モデルと等しいことが知られている [7]。ここで、 $P \in \mathbb{R}^{M \times R}$ ,  $Q \in \mathbb{R}^{L \times R}$  はそれぞれ  $X$ ,  $Y$  の負荷量行列であり、 $R$  は採用する潜在変数の数である。さらに、より迅速にプロ

セス特性の変化に追従するために、過去のデータ（あるいは (9) 式中の  $P$  および  $Q$ ）に忘却係数  $\beta(0 < \beta \leq 1)$  を乗じて、現在に近いサンプルを重視してモデルを構築することもできる。

RPLS は、プロセス特性の変化が穏やかであれば、その変化にモデルが十分に追従し、高い推定精度を維持できる。しかし、プロセスが長期間にわたりある狭い領域で運転される場合、モデルがその領域のみに適応してしまい、運転条件が大きく変化したときに推定精度が低下するなどの問題がある。

一方、プロセスの特性変化や非線形性に対応できる手法として、Lazy Learning または Just-In-Time (JIT) モデリングと呼ばれる方法が提案されている。これは、出力の推定値が要求されたときに限り、データベースに蓄えられたデータからクエリ点近傍のサンプルを選択し、局所モデルの構築と出力の推定を行う手法である。JIT モデリングは大域的なモデルの適用が困難な場合に有効であり、国内でも鉄鋼プロセスを中心に様々な産業応用事例が報告されている [8, 9, 10, 11]。JIT モデルの一種である局所回帰モデルを厚鋼板の材質制御および幅寸法制御へ適用した事例においては、既の実績値が存在する前工程の情報から、これから操作することが可能な後工程の最適な操作を決定することによって、製品品質制御精度の向上、品質不良発生リスクの低減、製造コストの低減が達成された他、従来スタッフが行う必要のあった膨大なモデルパラメータテーブルの保守作業負荷が大幅に低減され、従業員満足度の向上にも大きく貢献していると報告されている [12]。

従来の JIT モデリングでは、局所モデル構築用サンプルはクエリ点からの「距離」に基づいて選択されている。ところで、我々が線形回帰モデルを構築する場合、モデル構築用サンプルに期待する性質は何であろうか。恐らく、サンプルが互いに近くにあることではなく、サンプルがある 1 つの「相関関係」に支配されていることであろう。どれほど離れていようとも、同じ相関関係に支配されている限り、信頼するに足る良いモデルが構築できる。このような発想に基づいて提案されているのが、相関型 JIT モデリング (C-JIT) [13] である。触媒の劣化と再生を繰り返す連続槽型反応器を対象として、C-JIT の適用により、従来の RPLS や JIT モデリングでは対応が困難であったプロセス特性の急激な変化にも十分追従し、推定性能を大幅に改善することができると報告されている。さらに、C-JIT の分解ガソリン精留塔への適用についての検討結果も報告されている [14]。RPLS と C-JIT を用いて、分解ガソリン中のアロマ濃度を推定した結果を図 3 に示す。この結果は、C-JIT の実用的かつ高い推定能力を示している。

## 4 おわりに

本報では、製品品質管理に不可欠な技術となっているソフトセンサーについて、いくつかの統計的なモデル構築方法を概説すると共に、ソフトセンサーの運用や保守における現実的な課題とその解決策を紹介した。ここでは取り上げなかったが、統計的モデル構築の前段階において必要となるデータの前処理の重要性を指摘しておきたい。測定ノイズ、外れ値、欠損値への対応を含めて、モデル構築に利用すべきデータであるかどうかの判断を適切に行わなければならない。実際、ソフトセンサーの構築に際しては、この前処理にほとんどの時間と労力を費やしていると言っても過言ではない。この事実が、ソフトセンサーの保守が困難とされる原因でもある。つまり、自動的にモデ

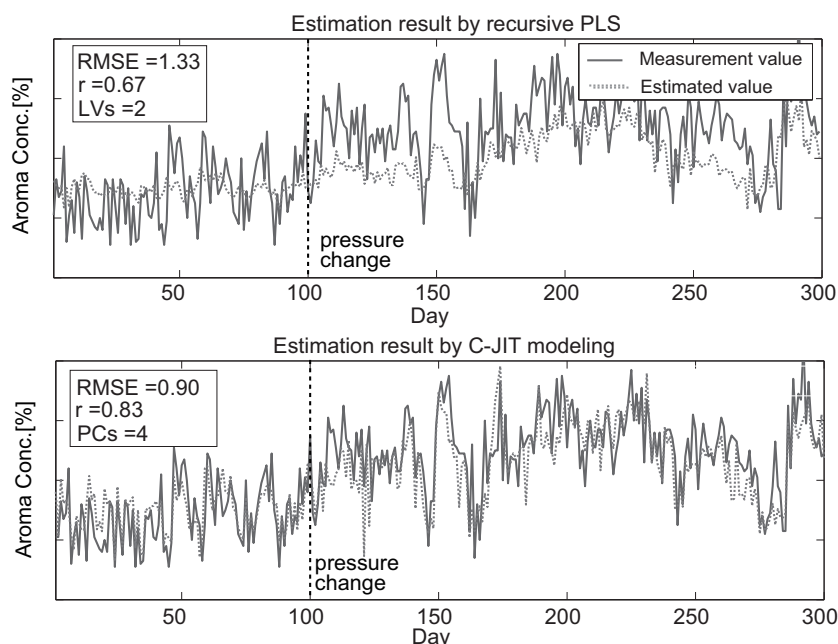


図3 分解ガソリン精留塔における製品中アロマ濃度推定

ル更新に利用すべきデータを取捨選択あるいは生成する方法が確立されていない。

また、本報では比較的古典的な方法を紹介するにとどめた。これはソフトセンサーの設計・運用・保守に関連する問題の核心を平易に開示するためである。種々の非線形手法の適用のみならず、独立成分分析、ベイズ推定、ウェーブレット解析など様々な手法がその特徴を活かしてソフトセンサーの構築に役立てられている。ソフトセンサーの使用目的と対象の特徴に合わせて、適切な手法を活用したい。

最後に、ソフトセンサーを構築・運用するに際しての、いくつかの注意事項を示す。

- 出力推定に不必要な入力変数を用いてはならない。PLSなどの次元圧縮を伴う回帰手法を用いる場合でも、入力変数の取捨選択は必要である。
- ステップワイズ法などの変数選択法は有効であるが、万能ではない。パラメータの設定によって結果は大きく異なり、推定精度が低下する恐れもある。
- 線形回帰モデルの係数は、入力変数から出力変数への影響度に一致しない。一致するのは、入力変数が互いに無相関な場合のみである。モデルの解釈には細心の注意を払うべきである。
- 測定データから導く統計的なモデルである以上、内挿はできて、外挿はできない。モデルの信頼性確認を怠ってはならない。
- 外れ値や欠損値への対処など、測定データの前処理が結果を大きく左右する。

ソフトセンサーでの利用に限らず、統計で人を騙すのは簡単である。たとえ意図しなくても、使い方を誤れば、嘘をついてしまうこともある。統計的方法の前提条件を把握し、正しく利用し、成果

を挙げていくために、本報が多少なりとも読者の参考になれば幸甚である。

## 参考文献

- [1] 向井, 加納, 長谷部 : 入力変数間の相関関係を考慮した入力変数から出力変数への影響度算出法, 計測自動制御学会制御部門大会 (2008)
- [2] R.Amirthalingam and J.H.Lee : Subspace Identification Based Inferential Control Applied to a Continuous Pulp Digester, J. Proc. Cont., 9, 397/406 (1999)
- [3] M.Kano, S.Lee, and S.Hasebe : Two-Stage Subspace Identification for Softsensor Design and Disturbance Estimation, J. Proc. Cont., 19, 179/186 (2009)
- [4] H.Kamohara, A.Takinami, M.Takeda, M.Kano, S.Hasebe, and I.Hashimoto : Product Quality Estimation and Operating Condition Monitoring for Industrial Ethylene Fractionator, J. Chem. Eng. Japan, 37, 422/428 (2004)
- [5] M.Kano, K.Miyazaki, S.Hasebe, and I.Hashimoto : Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression, J. Proc. Cont., 10, 157/166 (2000)
- [6] 加納学 : プロセスケモメトリクスによる統計的プロセス管理, システム/制御/情報, 48-5, 165/170 (2004)
- [7] S.J.Qin : Recursive PLS Algorithms for Adaptive Data Modeling, Comput. Chem. Engng, 22, 503/514 (1998)
- [8] 筒井, 黒崎, 佐藤 : 履歴データを事例として使用する非線形モデリング技術, 計測自動制御学会論文集, 33-9, 947/954 (1997)
- [9] 鄭, 木村 : Just-In-Time モデリングの新しい手法とその圧延セットアップモデルへの応用, 計測自動制御学会論文集, 37-7, 640/646 (2001)
- [10] 伊藤, 松崎, 大貝, 大館, 内田, 斎藤, 佐々木 : 高炉操業における大規模データベースオンラインモデリング, 鉄と鋼, 90-11 (2004)
- [11] 岸, 木村, 山本 : Just-In-Time モデリング技術による鉄鋼製品の収縮率予測, 計測と制御, 44-2, 116/119 (2005)
- [12] 茂森, 長尾, 平田, 南部, 池田, 水島, 加納, 長谷部 : 局所回帰モデルを用いた鋼材の品質制御, 計測自動制御学会論文集, 44-4, 325/332 (2008)
- [13] 藤原, 加納, 長谷部 : 相関型 Just-In-Time モデリングによるソフトセンサの設計, 計測自動制御学会論文集, 44-4, 317/324 (2008)
- [14] K.Fujiwara, M.Kano, S.Hasebe, and A.Takinami : Soft-Sensor Development using Correlation-Based Just-In-Time Modeling, AIChE J., accepted (2009)