

〈特集〉

上水処理プロセスにおける将来の膜差圧予測

金子 弘 昌¹⁾¹⁾ 明治大学 理工学部 応用化学科

(〒 204-0004 川崎市多摩区東三田 1-1-1 E-mail: hkaneko@meiji.ac.jp)

概 要

上水処理プラントにおいてセンサー等で測定されている温度・流量・圧力等のプロセスデータを活用して、将来の膜差圧を予測する技術について紹介する。プロセスデータが与えられたら、genetic algorithm-based partial least squares (GAPLS) により将来の膜差圧を関係の深い重要なプロセス変数を選択し、選択されたプロセス変数と膜差圧との間で部分的最小二乗法 (partial least squares, PLS) により膜差圧予測モデルを構築する。実際の上水処理プラントで膜差圧予測モデルの検証をしたところ、7日後や30日後の膜差圧を精度良く予測可能であった。膜差圧予測モデルはプラントのプロセス管理のみならず膜開発の効率化にも貢献できる。

キーワード：上水処理, 膜差圧予測, プロセスインフォマティクス, プロセス管理, 回帰分析

原稿受付 2018.1.9

EICA: 23(4) 40-43

1. はじめに

産業プラントを運転するためにはプロセスを適切に管理しなければならない。プラントで製造する製品を高品質かつ安定にすることだけでなく、環境基準等を満たしたりコストを抑えたりすることもプロセス制御・プロセス管理の目的の一つである。例えば排煙脱硝装置・排煙脱硫装置では、それぞれ排出されるNO_x, SO_xの濃度を定められた基準以下に抑えなければならない。また、下水処理における膜分離活性汚泥法 (membrane bioreactor, MBR) では同じ処理量を達成するための膜差圧が高いままになると薬品で膜洗浄をする必要がある。

MBRにおいては膜洗浄のスケジューリングをするために将来の膜差圧を予測することが望まれる。例えば一週間後の膜差圧を予測できれば、膜差圧予測値が基準値を超えてから薬品洗浄の準備をすることで、一週間前からスケジューリングできるようになる。

MBRにおいて日々測定されているプロセスデータを活用することで将来の膜差圧を予測することが可能となる¹⁾。現在の膜差圧・処理流量・水温等のプロセス変数の値やその一日前の値を入力変数 X, 一週間後の膜差圧を出力変数 y として、X と y との間で測定データを用いて統計モデル $y=f(X)$ を学習させる。モデルに X の値を入力することで、y の値を予測できる。これにより、y の予測値をモニタリングすることで、膜洗浄時期のスケジューリングが可能となる。

今回は上水処理プロセスを対象にして、2つのプラントにおいてそれぞれ7日後・30日後の膜差圧を統

計的に予測した例を紹介する。

2. 手 法

$y=f(X)$ を統計的に構築する手法として部分的最小二乗法 (partial least squares, PLS)^{2,3)} を、y にとって重要は X の変数のみ選択する手法として genetic algorithm-based partial least squares (GAPLS)^{4,5)} を用いた。

2.1 部分的最小二乗法 (partial least squares, PLS)

線形回帰分析手法の中では、最小二乗法による重回帰分析 (ordinary least squares, OLS) が一般的である。OLSにより以下のような回帰式が得られる。

$$y=Xb+c+f \quad (1)$$

ここで、y が目的変数であり今回は将来の膜差圧に対応する。X は説明変数であり温度・圧力・流量等に対応する。b, c, f がそれぞれ回帰係数、定数項、誤差である。モデル構築用データを用いて b, c を計算することで、新しい時刻の X の値を用いて y の値を予測できる。

しかし OLS は変数の数がサンプルの数より大きい場合に回帰モデルを構築することができない。さらに、変数の間の相関関係が強い場合に、回帰係数が不安定になってしまう。膜差圧予測を行う際に扱うプロセスデータは、様々な箇所のセンサーで取得されたデータであり一般的に変数が多い。さらに、プロセスの動特性 (ダイナミクス) を考慮するために時間遅れ変

数^{6,7)}を追加するとさらに変数の数が大きくなってしまふ。そのような時間遅れ変数間は一般的に類似しており相関も高い。OLSの問題点とプロセスデータの特徴とを考慮すると、OLSはプロセスデータの扱いが不得手といえる。

部分的最小二乗法 (partial least squares, PLS)^{2,3)} も線形回帰分析手法の一つであるが、Xの線形結合で表される少数の主成分Tとyとの間で回帰モデルを構築することで、上記のOLSの問題点を克服できる。PLSにより効果的にプロセスデータを解析可能といえる。

PLSは以下の2つの基本式からなる。

$$X = TP^T + E \quad (2)$$

$$y = Tq + f \quad (3)$$

Tはyとの共分散が最大になるように計算され、Xの中からyを予測するために主要な成分が抽出される。情報の重複がないようにTが抽出されるため、Tの間は無相関である。これにより最小二乗法により安定的にqを計算できる。式(2)(3)を組み合わせることで式(1)に対応する回帰係数も計算可能である。

最終的に膜差圧予測に用いるPLSモデルを構築するためには、用いる成分数を決めなければならない。一般的にはPLSの成分数の決定にクロスバリデーションが用いられる。例えば5-foldクロスバリデーションを用いる際、データセットを5つのグループに等分割する。そして、4つのグループのみでPLSモデルを構築して残りの1つのグループのyの値を予測することを、すべてのグループが予測用のグループになるまで5回繰り返す。これによりyのクロスバリデーション推定値が得られる。成分数ごとにyの実測値とyのクロスバリデーション推定値との間で決定係数 r^2 を計算し、 r^2 が最大となる成分数とする。

2.2 Genetic Algorithm-based Partial Least Squares (GAPLS)

PLSはXの変数の数がサンプルの数より大きい時にも回帰モデルを構築可能であり、Xの変数間の相関係数が大きい状況でも対処可能ではあるが、Xの変数の数が大きくなると、ノイズ等の影響によりモデルの推定性能が低下しやすい。そこでyと関係のあるXの変数のみを用いてPLSモデルを構築することが望ましい。しかし、例えばXの変数の数が100のとき、すべてのXの変数の組み合わせを検討しようとすると、 $2^{100}-1$ 通りでモデリングしなければならない現実的ではない。

そこで、最適化アルゴリズムの一つであり生物の進化の過程を模倣したアルゴリズムである遺伝的アルゴリズム (genetic algorithm, GA) を用いて、効率的に

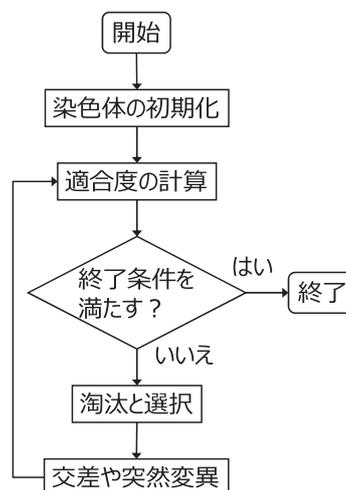


Fig. 1 GAPLSの計算の流れ

最適なXの変数の組み合わせを探索する。この変数選択手法を genetic algorithm-based partial least squares (GAPLS)^{4,5)} と呼ぶ。GAPLSの計算の流れをFig. 1に示す。選択されたXの変数を1、選択されなかったXの変数を0とした染色体をランダムに多数準備し、それぞれの染色体の適合度を計算する。適合度はPLSでクロスバリデーションを行ったときの r^2 である。適合度の高い染色体ほど生き残るように染色体が淘汰され、残った染色体で交差や突然変異といった遺伝的操作が行われ、次の世代の染色体が生成される。このように、適合度の計算・淘汰と選択・遺伝的操作によって、適合度の高い染色体、つまりクロスバリデーション後の r^2 が高くなるXの変数の組み合わせを得ることが可能となる。

3. 上水処理プラントにおける膜差圧予測の結果

2つの上水処理プラントにおいて、GAPLSにより有効なプロセス変数を選択し、PLSによって将来の膜差圧を予測した結果を示す。

1つ目のプラントでは7日後の膜差圧の予測を試みた。Xの変数は温度・圧力・流量などの46のプロセス変数およびそれらの24時間前の46変数の合計92変数である。この中からGAPLSにより最終的に使用する変数を選択した。

yの変数は7日後の膜差圧である。データの前処理としてSavitzky-Golay (SG) 法^{8,9)}により平滑化を行い、平滑化後の値を解析に用いた。SG法は元々スペクトル解析における平滑化手法であり、各種の移動平均法より良好に平滑化できるといわれている。今回はyの時系列データに対してSG法により平滑化を行った。

用いたデータセットには、膜洗浄から次の膜洗浄ま

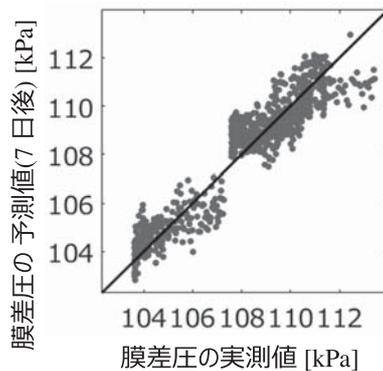


Fig. 2 1つ目の上水処理プラントにおける膜差圧の実測値 vs. 予測値プロット

でを1区間とすると4区間存在したため、3区間で膜差圧予測モデルを構築し、残りの1区間における7日後の膜差圧予測を行い、実際の膜差圧の値と比較した。これを4つのすべての区間が予測用の区間になるまで4回繰り返し実施した。

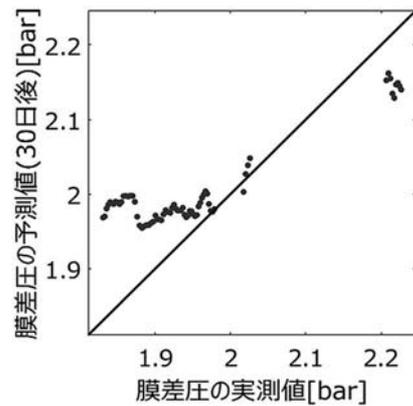
4区間すべての膜差圧予測の結果を Fig. 2 に示す。横軸が膜差圧の実測値であり、縦軸が膜差圧の7日後の予測値である。 r^2 は0.885であり、7日後であっても膜差圧のおよそ90%の情報をPLSモデルにより予測できた。また誤差の絶対値の平均は0.699 kPaであった。Fig. 2からもPLSモデルにより精度良く7日後の膜差圧の値を予測できることを確認した。膜差圧の予測値を上水処理プラントのプロセス管理に活用することで効率的なプラント運転が可能となる。

2つ目のプラントでは30日（一ヶ月）後の膜差圧の予測を試みた。Xの変数は温度・圧力・流量などの55のプロセス変数およびそれらの24時間前の55変数の合計110変数である。この中からGAPLSにより最終的に使用する変数を選択した。

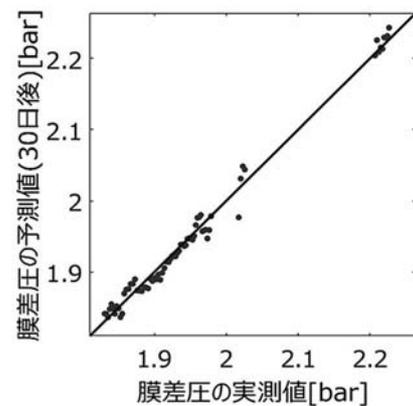
yの変数は30日後の膜差圧である。1つ目のプラントと同様にデータの事前処理としてSG法により平滑化を行い、平滑化後の値を解析に用いた。

用いたデータセットには1区間のみ存在していたため、最初の35日間のデータで膜差圧予測モデルを構築し、その後のデータにおける30日後の膜差圧予測を行い、実際の膜差圧の値と比較することで検証した。

PLSとGAPLSとを用いた際の、膜差圧の実測値 vs. 予測値プロット予測の結果を Fig. 3 に示す。横軸が膜差圧の実測値であり、縦軸が膜差圧の30日後の予測値である。 r^2 はPLSでは0.464であったが、GAPLSにより0.989と向上した。30日後であっても膜差圧のおよそ99%の情報をGAPLSモデルにより予測できた。またGAPLSを用いた際の誤差の絶対値の平均は0.009 barであった。Fig. 3からもGAPLSモデルによって精度良く30日後の膜差圧の値を予測できることを確認した。



(a) PLS



(b) GAPLS

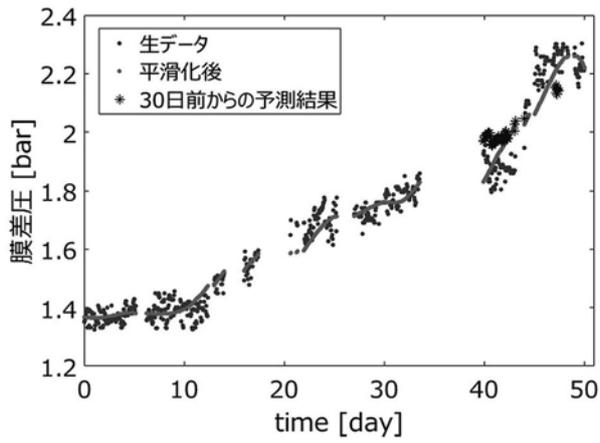
Fig. 3 2つ目の上水処理プラントにおける膜差圧の実測値 vs. 予測値プロット

膜差圧の予測結果を Fig. 4 に示す。膜差圧の時間プロットであり、黒のアスタリスクで示されている膜差圧の予測結果は、30日前から予測された結果である。Fig. 4を確認しても、特にGAPLSモデルにより膜差圧の上昇を30日前から適切に予測できていることがわかる。30日前から膜差圧を予測し、その結果を上水処理プラントのプロセス管理に活用することで効率的なプラント運転が可能となる。

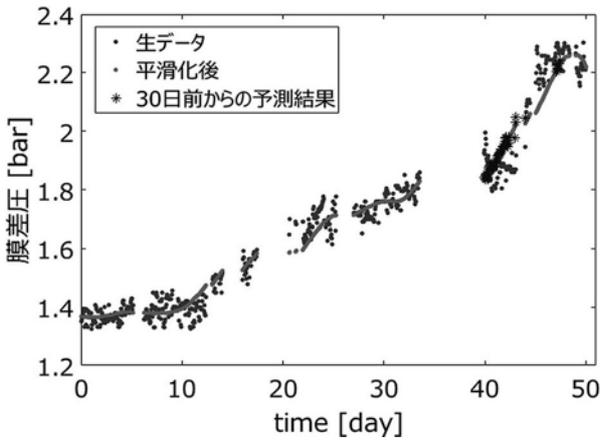
4. おわりに

今回は上水処理プラントにおいて測定されている運転データを活用することで、7日後の膜差圧や30日後の膜差圧を精度良く予測できる技術について紹介した。膜差圧予測モデルによる膜差圧の予測結果を活用して上水処理プラントを運転することで、効率的なプロセス管理が達成される。

今回はPLSやGAPLSといった線形の回帰分析手法を用いたが、サポートベクター回帰 (support vector regression, SVR)^{10,11)} 等の非線形回帰分析手法を活用することで膜差圧予測のさらなる精度向上も期待できる。さらに、膜開発において今回の膜差圧予測モデルを活用することで、膜の長期運転実験を行わなくても短期



(a) PLS



(b) GAPLS

Fig. 4 2つ目の上水処理プラントにおける膜差圧予測の結果

間における実験結果から膜性能を検討することが可能となり、膜開発の効率化を達成できると考えられる。

参考文献

- 1) H. Kaneko and K. Funatsu : Visualization of Models Predicting Transmembrane Pressure Jump for Membrane Bioreactor, Ind. Eng. Chem. Res., Vol. 51, No. 28, pp. 9679-9686 (2012)
- 2) S. Wold, M. Sjöström and L. Eriksson, PLS-regression : a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. Vol. 58, pp. 109-130 (2001)
- 3) <https://datachemeng.com/partialleastsquares/>
- 4) J. Ghasemi, A. Niazi and R. Leardi : Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS : application on copper and zinc mixture, Talanta, Vol. 59, pp. 311-317 (2003)
- 5) <https://datachemeng.com/gaplsgasvr/>
- 6) H. Kaneko and K. Funatsu : A New Process Variable and Dynamics Selection Method Based on a Genetic Algorithm-based Wavelength Selection Method, AIChE J., Vol. 58, No. 6, pp. 1829-1840 (2012)
- 7) <https://datachemeng.com/pointsoftimeseriesdataanalysis/>
- 8) A. Savitzky and M. J. E. Golay : Smoothing and Differentiation of Data by Simplified Least-Squares Procedures, Anal. Chem., Vol. 36, pp. 1627-1639 (1964)
- 9) <https://datachemeng.com/preprocessspectratimeseriesdata/>
- 10) C. M. Bishop : Pattern recognition and machine learning, Springer (2006)
- 11) <https://datachemeng.com/supportvectorregression/>