

解 説

UDC 681.3 : 669.1

新しいデータ処理法 (GMDH) の紹介とその製鉄プロセスへの応用*

榎 木 義 一**・小 西 正 躬***

Group Method of Data Handling (GMDH) and Its Applications
to the Metallurgical Processes of Iron and Steels

Yoshikazu SAWARAGI and Masami KONISHI

1. ま え が き

近年、日本鉄鋼業は目覚ましい発展をとげ、強い競争力を有するに至ったが、更にその地位を不動のものとするために技術者による地道な改善、改良が続けられている。この日本鉄鋼業の発展は最新鋭の設備による徹底した合理化にささえられているが、この合理化の過程で導入された電子計算機の円滑な稼動とそのためのソフトウェアの開発に多大なるマンパワーが動員され製鉄プロセスの計算機制御技術として結実しつつある。現在では新鋭製鉄所における円滑な運営において計算機が欠かせぬのは常識となつている。

製鉄所における計算機制御は原料ヤードにはじまり、焼結、高炉、転炉、連铸、圧延などのほとんどすべてのプロセスに適用されているが、いずれのプロセスにおいても各プロセスに固有の数式モデルが開発され計算機制御の用途に利用されている。これらの数式モデルには、物理現象を定量的に記述した物理モデルの場合と、物理現象を考慮しつつモデルの構造を決定したのち定量的に係数を決定するタイプのもの、すなわち、統計モデルの場合とがある。プロセスの特性が未知に近いものでは統計モデルが用いられ、特性が明らかなものについては物理モデルが用いられている。前者では数多くえられている操業データに基づく重回帰分析手法が普通用いられている。

さて、ソ連の A. G. IVAKNENKO により開発された GMDH (Group Method of Data Handling) もやはり前者の統計モデルの範疇に属するものである。これは、発見的自己組織化 (Heuristic Self Organization) の原理に基づき、本質的に複雑なシステムに対して一般的に

モデル作りを行なう方法である¹⁾。GMDH に関してはここ数年間に数多くの研究者により検討され種々の発表がなされているが、重回帰分析手法にくらべるとすみずみまで検討しつくされておらず、手法としてはまだ完全に体系化されているとは言い難い。また、応用範囲も社会経済問題や環境問題に限られていて製鉄プロセスへの適用は少ない。

ここでは、GMDHの基本的な思想とそのアルゴリズムを紹介する。さらには、製鉄プロセスへの応用に関して検討を加える。

2. 複雑な非線形系のGMDHによる
モデル同定

入出力データによる線形系のモデル同定に関しては重回帰分析に代表されるような多数の研究実績があり、学問的にも確立、完成の観がある。一方、非線形系のモデリングに関しては構造が与えられていて係数パラメータを推定する問題に帰着される場合を除いてはほとんど検討されていない。ただ非パラメトリックな一般的手法として Volterra 級数による同定手法が用いられている²⁾。しかし、この場合、大規模系を対象とすると決定する係数の数が非常に多くなり、また、したがって莫大な計算量となるなど実用上むづかしい問題を抱えている。

さらに、非線形大規模系のモデリングに際しては、①適切な変数の選択と②非適切な変数の棄却、③モデルとしてどのような構造式を採用するか、④パラメータを決定する方法、⑤予測時の精度を保証する方法などの点が重要な課題である。上記のような多次元の場合の次元の呪いや定式化のむづかしさに対し、IVAKNENKO の提案した自己組織化法によるGMDHアルゴリズムは1つの

* 昭和 52 年 8 月 29 日受付 (Received Aug. 29, 1977)

** 京都大学工学部 工博 (Faculty of Engineering, Kyoto University, Yoshidahonmachi Sakyo-ku Kyoto 606)

*** (株)神戸製鋼所浅田研究所 工博 (Asada Research Laboratory, Kobe Steel, Ltd.)

有力な解決方法を示している

GMDHの特色をこれらの観点から見ると、

- (1) 少ないデータによる大規模な非線形系のモデル同定が可能である。
- (2) 少ない計算量でモデル同定ができる。
- (3) 最適な複雑さを持つモデルを作ることができる。

というのがその特色である。これらの特色はGMDHのプロセス制御への応用という観点から見てもまことに都合の良いものと言わざるをえない。

非線形系に限らず一般のプロセスのモデル同定とその制御への応用を業務とする者ならばすべて経験することであるが、モデリングに際しては、えられたモデル式あるいは予測制御式が実用に耐え得るように種々の試みがなされる。ある場合には物理的に推定できるパラメータを固定したり他の場合には実験的に特定のパラメータを決定したり種々の試行により実用的なモデルが確立されるのである。

GMDHにおいても妥当な解をえるための工夫が必要であり、これをヒューリスティックスと呼んでいる。GMDHで用いられる種々のヒューリスティックのうちの主たるものの1つは、与えられたデータを構造決定のための Training データと、より有効な変数の組み合わせ選択のための Checking データとに分けることである。これにより同定時に入出力データでのオーバーフィッティングを避けて新しい入出力データに対してもモデルとして適度な滑らかさを持たせることができる。

上記のデータ分割以外のヒューリスティックスとしては、(1)与えられている入力データの変換、(2)有用な変数の自己選択のための評価判断規範、(3)変数の自己選択のための閾値設定が主たるものである。これらについては次節で詳述するが、この操作はGMDH使用者の経験と個性とにより千差万別で、経験を積み使い込むにつれてますます有効にその機能が発揮できるというのがGMDHの特色である。

3. GMDHのアルゴリズム

GMDHのアルゴリズムは図1のフローチャートで示すことができる。以下順を追って紹介する。

3.1 基礎関数

GMDHでは入力変数のランダムな結合を発生し発見的規範を用いて多層構造で最良のものを閾値的に自己選択していく。したがって、GMDHでは中間変数発生のための基礎関数の選び方で結果が左右されるため、前述のヒューリスティックスに加えて基礎関数の選択という新たなヒューリスティックスが入ってくるようになる。最もよく用いられる基礎関数は KOLMOGOROV-GABOR の完全多項式表現

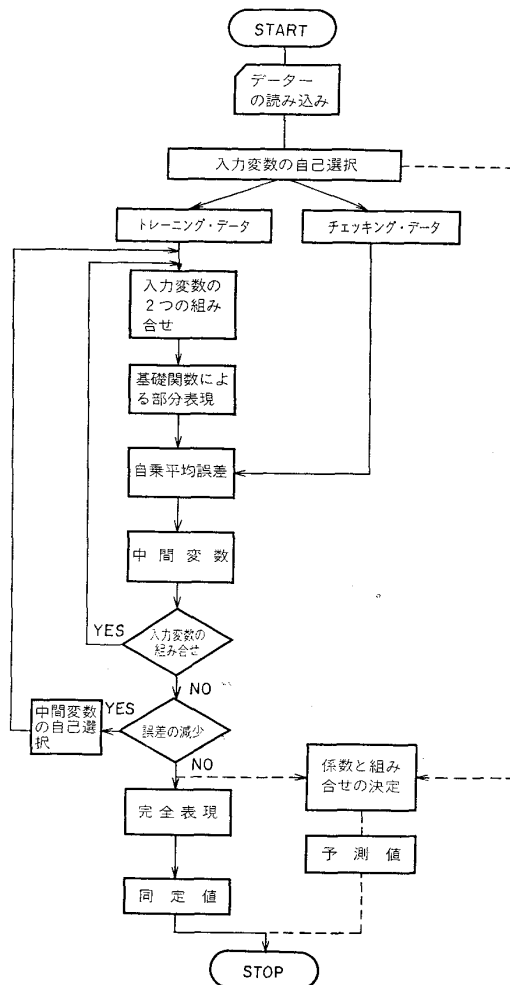


図1 GMDHアルゴリズムのフロー

$$y = a_i + \sum_i a_i x_i + \sum_i \sum_j a_{ij} x_i x_j + \sum_i \sum_j \sum_k a_{ijk} x_i x_j x_k \dots\dots\dots (1)$$

を用いて、

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \dots\dots (2)$$

あるいは、単に

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j \dots\dots\dots (3)$$

の表現式である。

(2)式の2次の多項式を用いた場合、各層でのアルゴリズムは2次式の回帰を行なっていることになり、 N 層を通過すると 2^N 次の多項式となる。しかし、各層で(2)式の係数を決定するので、決定のためのデータ数はたかだか6個あれば良いことになり、少ないデータの時にも高次の多項式近似ができることになる。

基礎関数の形は上記の(2)、(3)式に限ったものではなく多くの形が提案されている。たとえば、入力 x_i, x_j に対し

$$y = b_0 + b_1 z_i + b_2 z_i^2 \dots\dots\dots (4)$$

$$z_i = w x_i + (1-w) x_j \dots\dots\dots (5)$$

w : 重み係数

の形によるフィルター効果を持たせたものも考案されて

いる³⁾。

3.2 GMDHの基本的アルゴリズム

IVAKNENKO は基礎関数として (2) 式を用い基本アルゴリズムを開発している。この手順を以下で紹介する。

入力変数を $x_i (i=1, 2, \dots)$, 出力変数を y とし入力と出力の間には

$$y=f(x_1, x_2, \dots) \quad (6)$$

なる関数関係があるとすると、関数 f の同定はつぎのように行なわれる。

[ステップ 1]

y に関係する入力変数を選択し、必要があればデータの前処理をする。出力 y と各入力変数の相関を取り、相関関係の大きいものを有用な入力変数として残し、それ以外のものをすてる。

[ステップ 2]

残った入力変数と出力変数に関する時系列データを training データと checking データの 2 組に分割する。分割の方法は解析の目的に応じ、training データと checking データを交互に取る場合もあれば、分散の大きいほうを training データに選ぶ場合もある。

[ステップ 3]

入力変数 $x_k (k=1, 2, \dots, M)$ の任意の 2 個の組合せ x_i, x_j に対して中間変数 $Z_k (k=1, 2, \dots, M(M-1)/2)$ を

$$Z_k = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad \dots (7)$$

により作る。ただし (7) 式の右辺の係数 $a_0 \sim a_5$ は training データによる 2 乗平均誤差

$$\epsilon_k^2 = \frac{1}{N} \sum_{i=1}^N (y(i) - Z_k(i))^2 \quad (i \text{ はデータ番号}) \quad \dots (8)$$

を最小にするように決定される。

[ステップ 4]

ステップ 3 で決定された係数 $a_0 \sim a_5$ と checking データを用いて (8) 式の ϵ_k^2 を計算し、 ϵ_k^2 の小さいものから順に M_j 個 (j は層番号) だけ中間変数に採用し、残りはすてる。

[ステップ 5]

$x_i = Z_i, x_j = Z_j$ としてステップ 3 へいき、つぎの中間変数を求める。

以下、ステップ 3～5 をくり返す。

[ステップ 6]

ステップ 4 において checking データに対して計算された 2 乗平均誤差の最小のものが前回の最小な 2 乗平均誤差を越えた場合計算を停止する。

[ステップ 7]

前回までに計算された (7) 式の 2 次式をつぎつぎと代入することにより関数 f が求まる。

以上のアルゴリズムを図示したのが図 2 である。

なお、上記のアルゴリズムは基本的なものであるが、

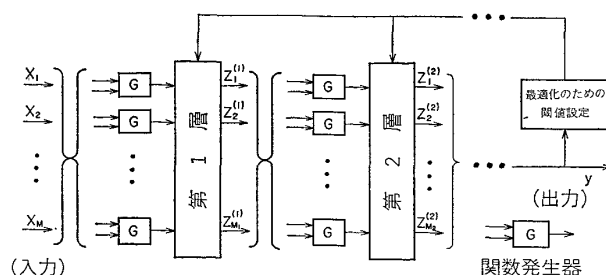


図 2 GMDHの入出力関係

最近、GMDHの予測精度を改善するための工夫が種々なされている。たとえば赤池の開発した情報量規範 (AIC)⁴⁾ を 2 乗平均誤差のかわりに用いる試み⁵⁾ や、多変量分析で良く用いられる残差平方和 (PSS)⁶⁾ を 2 乗平均誤差のかわりに用いる試みがなされている⁷⁾。

4. GMDHによる製鉄プロセス解析・同定

これまで述べてきたように統計手法であるため、GMDHでは重回帰分析と同様にデータの母集団の統計的性質によりえられる結果が異なる。しかし、重回帰分析ではモデルの構造をあらかじめ決める必要があるのに対して、GMDHでは入力変数を決めるだけで、あとは自動的にGMDHが入力変数の取捨選択およびモデル構造を決定するので反応の解析、要因分析が有効に行なえる。

前節ではGMDHのアルゴリズムを紹介したので、本節では数値例をあげてプロセス解析への応用を検討する。

4.1 高炉反応の解析

高炉操業における操業要因は装入原料、炉頂ガス成分、炉頂圧、羽口からの重油の吹込量、送風温度、送風湿度等で高炉を安定に操業しつつ出鉄量、出鉄成分、出鉄温度等を目標値に制御している。

ここでは、一定時間ごとに操業データがサンプリングされている場合の高炉要因分析について検討する。検討の目的は重油流量、送風温度、送風湿度が溶鉄成分 (Si%) に及ぼす効果を知ることである。

GMDHの入力変数として以下の 30 個をえらぶ。

- X_1 : 現時点の重油流量
- X_2 : 1 時点前の重油流量
- X_3 : 2 " "
- X_4 : 3 " "
- X_5 : 4 " "
- X_6 : 5 " "
- X_7 : 6 " "
- X_8 : 7 " "
- X_9 : 8 " "
- X_{10} : 9 " "
- X_{11} : 現時点の送風温度
- X_{12} : 1 時点前の送風温度

X_{13}	: 2	〃	〃
X_{14}	: 3	〃	〃
X_{15}	: 4	〃	〃
X_{16}	: 5	〃	〃
X_{17}	: 6	〃	〃
X_{18}	: 7	〃	〃
X_{19}	: 8	〃	〃
X_{20}	: 9	〃	〃
X_{21}	現時点の送風湿分		
X_{22}	1 時点前の送風湿分		
X_{23}	: 2	〃	〃
X_{24}	: 3	〃	〃
X_{25}	: 4	〃	〃
X_{26}	: 5	〃	〃
X_{27}	: 6	〃	〃
X_{28}	: 7	〃	〃
X_{29}	: 8	〃	〃
X_{30}	: 9	〃	〃

また、出力変数は現時点の溶銑中の Si% とする。

これらの入出力変数の 500 時点の連続データを GMDH プログラムへの入力とし、交互に training データと checking データとしたとする、計算結果として

$$y = f(X_6, X_{14}, X_{22}) \dots \dots \dots (9)$$

がえられた。すなわち、5 時点前の重油、3 時点前の送風温度、1 時点前の送風湿分が現時点での溶銑中の Si% に寄与しているということで、溶銑温度 (Si%) の要因抽出とこれらの要因と結果の間の時間的反應おくれを知ることができる。(9)式で求めた y 値と実測の y 値の比較を図 3 に示した。

4.2 圧延荷重式の同定

ホットストリップミルの圧延荷重式にはシムスの圧延荷重式が広く用いられているが、さらに高精度の圧延を行なうためには補正式が必要である。GMDH により補正式を作成する試みは以下のように考えることができる。

すなわち、

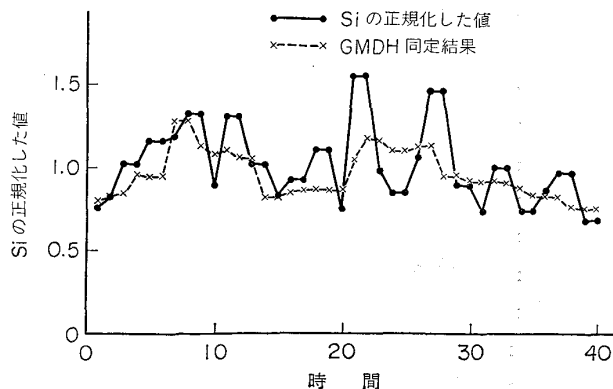


図 3 GMDHによる溶銑中の Si の同定

(圧延荷重) = (シムス式による計算圧延荷重)

$$\times (1.0 + \text{補正項}) \dots \dots \dots (10)$$

とする。(10)式の右辺の第2項中の補正項をGMDHで同定するのであるが、この場合の入力変数は以下のよう

X_1	: 正規化された入側板厚
X_2	: 〃 出側板厚
X_3	: 〃 圧延温度
X_4	: 〃 圧延速度
X_5	: 〃 板幅
X_6	: 〃 成分元素含有量
X_7	: 〃 鋼種コード
X_8	: 〃 ロールクラウン
X_9	: 〃 ロール径

圧延プロセスでは圧延理論が充分適用可能であり、わずかにこれを補正する意味でGMDHにより補正項を同定するのであるが、プロセス制御の観点から見ると、補正式の次数が高いことは計算の複雑さにつながるため望ましくない。IVAKNENKO が提唱した基本的な GMDH アルゴリズムでは多数の入力変数を含み、しかも次数が比較的低いシステムとしてのモデリングには不向きである。そこで、たとえばつぎのような修正アルゴリズムを用いると便利である。

すなわち、中間変数の部分表現のための基本式は(2)式あるいは(3)式の形とし、checking データに対する2乗平均誤差の小さいものを次の層への中間変数として残すという基本的アルゴリズムは保存しつつ図4に示ように出力誤差に対し再度GMDHを適用する。このGMDHの逐次適用は数回でも良く、checking データに対する2乗平均誤差が前回のGMDHによるそれより大となる時停止しても良い。

このアルゴリズムによると、

$$(\text{補正項}) = f^*(X_1, X_2, \dots) + f^*(X_1, X_2, \dots) + \dots$$

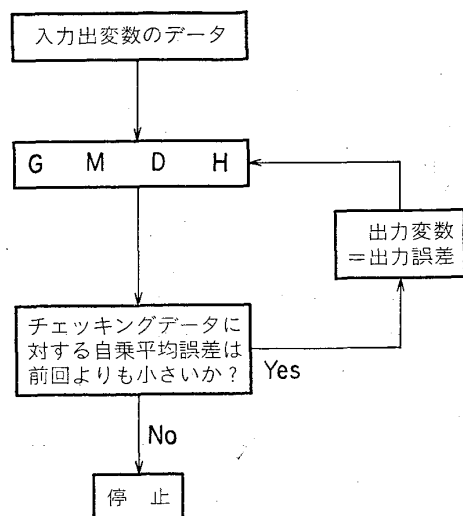


図 4 修正型の GMDH アルゴリズム

.....(11)

なる表現で低次元でかつ多くの要因を取り込むことができる。ただし、 f_i^* は第 i 回目の GMDH の適用時のモデル誤差の近似表現である。このモデルによる同定予測をした時には従来のシムス式単独の場合に比して予測誤差を 3 割方減少させることができる⁸⁾。

以上のべてきた 2 つの例からわかるように、GMDH の適用方法は未確立であるというよりむしろ自由にとれるということで、別に決った方法がある訳ではなく、今のところひとえに解析者の工夫にまかされているが、ここに、前述のヒューリスティックスの活用余地が生まれるのである。この意味からも重回帰分析による要因分析やモデル同定によつてすでに得られた結果を GMDH により再解析することにより、新しい技術情報が得られる可能性が残されている。

5. GMDH によるプロセスの予測、 制御の検討

これまでに述べて来たように、充分な入出力データの蓄積があれば GMDH によりプロセスモデルを同定することができる。本節では、このようにしてえられたモデルをプロセスでの予測、制御に使用する場合の問題と対策を検討する。それらは以下のように要約できる。

第 1 に制御のためのプロセスモデルは基本的には物理モデルを主とし、これまでの各プロセスでの研究成果を活用すべきであろう。ただし、圧延プロセスでの例についてのべたように、物理モデルで記述されていない内容を補充する意味で GMDH の利用は有効である。しかしこの場合、補正項の次数を上げぬように工夫がされなければならない。

第 2 に高炉、転炉、焼結プロセス等のブラックボックスに近い構造のシステムでは GMDH によるモデル同定が有効である。ただし、同定時に用いるデータの統計的性質を良く吟味しつつモデル化にかかる必要がある。これは、予測制御時に同定時の母データと異なる統計的性質のデータ群が与えられるとモデルで外挿をすることとなり予測精度が著しく劣化し、モデルとしてのタフネスが失われるおそれがあるからである。

第 3 に予測精度を追求する余りに複雑な関数形にならぬよう注意が必要である。たとえば圧延プロセスの例でのべたようにくり返し型のアルゴリズムを利用したり基本関数として簡単な記述式を採用し次数を下げる努力がなされなければならない。

第 4 に GMDH によるモデル同定・予測精度が重回帰分析手法による同定予測精度と同等かこれを上まわるものとするためには図 2 のアルゴリズムの閾値設定値を重回帰分析時の最小 2 乗平均誤差の値に設定する必要がある。転炉制御に GMDH を用いた文献⁹⁾ によれば、重回帰分析時の同定精度にくらべ GMDH では 1.8 倍の精度

がえられたとしている。

第 5 に制御モデルとしての要求はプロセスでの制御量を要因の関数関係で表現することであるが、被制御量に関係する同定モデルの要因に制御量を含ませ、モデルから制御量を逆算することは好ましくない。したがって、制御量に関する数式モデルを GMDH で同定する必要がある。たとえば、転炉のスタティックコントロール時に制御モデルを作成するには、

(吹止めまでの総酸素流量)

$$= f(\text{目標吹止 } C\%, \text{ 目標吹止温度}, \dots) \dots (12)$$

のようなモデルを GMDH で同定すればよい。

第 6 に GMDH で制御モデルを同定する際に母データの統計的性質によつては制御に必要なパラメータ（たとえば、(12) 式での目標吹止 $C\%$ ）がすてられることがある。これは制御モデルとしては不都合でありこのままでは使用できない。この問題に関しては必要最小限のパラメータは各層で優先的に残すという新しいヒューリスティックスを導入してカバーする必要がある。

第 7 に操業データが豊富にえられている場合にはできるだけ多くのデータの利用が望しい。これは、GMDH では数少ないデータで非線形モデルの同定が可能であるが、この場合ともするとデータに対するモデルのオーバーフィッティングが生じ予測に必要な安定性がそこなわれやすいからである。

第 8 に操業中に人間や計算機によるフィードバック制御が行なわれている場合のデータでのモデル同定には充分注意が必要である。この場合にはプロセス単独のモデル同定が困難となるからで、おくれが大きい反応プロセスの場合には作業標準を確立し、フィードバック効果によるデータへの外乱を除去した後に GMDH を用いないと再現性のあるモデル同定は困難と思われる。

以上のような種々の要件は GMDH のような統計モデルをプロセス制御に活用する際に基本的に必要と考えられるので、GMDH 特有のヒューリスティックスを駆使してそれぞれのケースに対し工夫し解決すべきものであろう。

6. あとがき

GMDH は重回帰分析ほどまだ充分には数学的に論理づけがなされず未完成であるが、多変数でブラックボックス的なプロセスのシステム同定、予測について非常に有効と考えられる。また、要因分析にも有効であるが、これは複雑なシステムの物理的あるいは現象論的なモデル作成に当つての仮説を提供することができる。

GMDH の適用に際しては、解析者の工夫によるヒューリスティックスの活用が重要な役目をはたすと考えられるが、この意味からも、本稿で紹介した例での取り扱いも筆者らの独断の感をまぬがれない。GMDH は今後大いに発展する可能性を秘めており、本稿が GMDH の

理解に少しでも役立てば幸いである。

読者諸賢におかれても、身近な製鉄プロセスへのGMDHの適用をどしどしはかられることを希望して止まない。

文 献

- 1) A. G. IVAKNENKO: Heuristic Self-Organization in Problems of Engineering Cybernetics; Automatica, 6 (1970), p. 207~219
- 2) G. E. P. Box and G. M. JENKINS: Time Series Analysis-Forecasting and Control. San Francisco, Holden-Day (1970)
- 3) 池田三郎, 藤重 悟, 榎木義一: 河川流量予測へのGMDHの応用, システム制御におけるモデリングとシミュレーションシンポジウム, (1974), p. 59~62
- 4) 赤池弘次: 情報量規準 AIC とは何か—その意味と将来への展望, 数理科学, サイエンス社, (1976), p. 5~11
- 5) NAKAJIMA and SAWARAGI: A Modified GMDH based on Akaike Information Criterion, (1977)
- 6) 奥野忠一, 久米 均, 芳賀敏郎, 吉澤 正: 多変量解析法, 日科技連, (1971)
- 7) 田村坦之, 近藤 正: IFAC 国際環境システムシンポジウム前刷集, (1977), p. 373~380
- 8) 能勢和夫: 修正 GMDH による圧延荷重モデルの同定, 鉄鋼協会第 63 回計測部会資料 (1977)
- 9) L. G. RYBALKO: Use of GMDH for the development of a model of the Oxygen-converter Process, Izvest. Vuz. Chern. Met, (1973) 6, p. 146~148