時間領域におけるシステム同定結果の評価指標について

室井 秀夫*・竹下 侑**・足立 修一*

Model Validation Criteria for System Identification in Time Domain

Hideo MUROI*, Yu TAKESHITA** and Shuichi ADACHI*

Fit ratio is one of the most popular criteria that evaluates a result of system identification in the time domain. This criterion is constructed by Root Mean Squared Error (RMSE) divided by the standard deviation of the measured signal. However the validity of fit ratio have not been discussed actively in the field of system identification. In this paper, we introduce some discussion about criteria like fit ratio from the field of physical geography. Then, we evaluated these criteria through two case studies. As a result, we found that fit ratio was not adequate for evaluation of the result of system identification.

Key Words: system identification, model validation, time domain, goodness of fit, fit ratio

1. はじめに

制御対象のモデルを利用した制御系設計¹⁾や状態推定²⁾は広く行なわれており、これらの実現には制御対象のモデリングが必須である。さまざまなモデリング法が存在するが、なかでも制御のためのモデリングとしては第一原理モデリングとシステム同定が広く利用されている³⁾。

モデリング後には、モデルが対象を適切に表現しているのかを評価する必要がある。この手順をモデルバリデーションと呼ぶ。モデルバリデーションは、時間領域、周波数領域、 s または z 領域など、さまざまな面から行なうことが好ましい。それぞれ具体的には

- モデル出力とプラント出力を比較する（時間領域）
- スペクトル解析法と予測誤差法でモデリングを行ない、それぞれの周波数応答を比較する（周波数領域）
- 極零相殺の有無を確認する（ s , z 領域）

などが挙げられる。これに加えて、残差の白色性や、残差と過去の入力との相互相関を解析し、モデリングできていない部分を検証する残差解析や、第一原理モデリングのように、パラメータに明確な物理的意味がある場合は、事前情報からその妥当性を検証することも有用である⁴⁾。

なかでも、時間領域での評価では、測定値とモデル出力の

二つの波形の適合度からモデルの妥当性を判断するものであり、直感的にわかりやすく、対象が非線形システムの場合にも利用できるため、応用の場面ではしばしば用いられる。ここで適合度を測る指標には MSE (Mean Square Error) や RMSE (Root Mean Square Error) がよく用いられるが、これらはモデリング対象となる物理単位での誤差に基づいた指標である。そのため、対象によって用いられる単位系が異なり、どの程度二つの波形が一致しているのかを一般的に論じる指標としては適切ではないと考えられる。

モデル性能を時間領域で評価する際には、対象の単位系によらず適合度を評価できる無次元な指標（以後、無次元指標と呼ぶ）を利用すると直感的にわかりやすい。そこで、本論文ではシステム同定結果を時間領域で評価するための無次元指標について検討する。

無次元指標は、主に気候学、水文学の分野（以降、この二分野をまとめて自然地理学分野と呼ぶ）において 1970 年代ごろから考えられてきた。最初に、“Coefficient of Efficiency E ” が、降雨流出モデル（降雨量から川の流量を予測するモデルで洪水予測などに利用される）の評価指標として、J.E. Nash らにより提案された^{5), 6)}。この指標は出力測定値（以下、測定値と略記）とモデルの出力予測値（以下、予測値と略記）の MSE を測定値の分散で規格化したものであり、MSE が 0 となるときに最大値である 1 をとる。しかし、この指標は MSE に応じて単調減少するため、最小値が存在しない。そこで、MSE を 0~100 となるように規格化した “Index of agreement d_j ” が C.J. Willmott らによって提案された^{7), 8)}。その後、この二種類の指標についての優劣や、これらの更新に関する議論が今日まで続いている^{9)~13)}。なお、自然地理学分野ではこのほかにも多数の指標が提案されているが、その多くは E または d_j のどちらかに類似しており、本質的な違いはない。そ

* 慶應義塾大学理工学部 横浜市港北区日吉 3-14-1

** ソニーグローバルソリューションズ (株)

東京都品川区東品川 4-10-27

* Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama

** Sony Global Solutions Inc., 4-10-27 Higashishinagawa, Shinagawa-ku, Tokyo

(Received December 25, 2014)

(Revised July 10, 2015)

のため、本論文ではこの二つについてのみ紹介し、ほかの指標については Moriasi らのガイドライン¹¹⁾を参照されたい。

システム同定分野においても、無次元指標として測定値と予測値の RMSE を測定値の標準偏差で規格化した“適合率”(FIT)がよく用いられる^{3), 14)}。適合率は前述の E の平方根をとったものであり、同様の特性をもつ。しかし、適合率がシステム同定結果の評価に対して適切な指標であるのか、ほかの指標が不適切であるか、といった議論は自然地理学分野と異なり活発には展開されていない。

そこで、本論文ではシステム同定結果の評価を模擬した簡単な例題を用いてこれらの指標がどのような場面で有用か、あるいは誤った判断を促す可能性があるかについてを考察する。

本論文の構成は以下のとおりである。まず、第 2 章において、本論文の議論の対象である適合率と、自然地理学分野で行われてきた無次元指標に関する議論について紹介する。第 3 章ではシステム同定結果の評価を模擬した例題により、第 2 章で紹介した指標の特徴を把握し、それらが有用な場面について考察する。最後に、第 4 章で結論を示す。

2. 無次元指標に関する議論

まず本論文で議論する適合率について紹介する。その後、自然地理学で提案された指標と提案されてきた背景に関する議論をまとめる。

2.1 適合率

システム同定分野で広く利用されている適合率 (以降, FIT と記載) は

$$\text{FIT} = \left\{ 1 - \frac{\sqrt{\sum_{k=1}^N (y(k) - \hat{y}(k))^2}}{\sqrt{\sum_{k=1}^N (y(k) - \bar{y})^2}} \right\} \times 100[\%] \quad (1)$$

で定義される。ただし、 $y(k)$ と $\hat{y}(k)$ はそれぞれ時刻 k での測定値、予測値であり、 $\bar{y}$ は測定値の平均値である。また、 N はサンプル数である。(1) 式の右辺第二項は、RMSE を測定値の標準偏差で規格化したものであり、NRMSE (Normalized-RMSE) と呼ばれる。そして、FIT 自体も NRMSE fitness value と呼ばれることがある¹⁴⁾。この指標は RMSE が 0 となると、すなわち完全に y と $\hat{y}$ が一致したときに最大値 1 をとる。しかし、RMSE の増加に対して値が単調減少するため、最小値が存在しない。特に、FIT は測定値の標準偏差を上回った場合に負値を取る。

システム同定分野において適合率は利用される一方、この指標の妥当性について述べた文献は筆者の知る限りない。そこで、自然地理学分野で用いる指標に関する議論を追いながら、指標の妥当性について考える。

2.2 自然地理学での無次元指標に関する議論

自然地理学分野では古くからモデルの評価方法が課題となっている。特に、降雨流出モデルでは、降雨量から対象となる

川の評価地点での流量の増加にかかる時間、増加した流量の両方についての予測性能を評価する必要がある。評価において無次元指標を導入することで、さまざまな川において同様の議論を展開できることが期待される。本章では自然地理学で行われてきた無次元指標についての議論をまとめる。

2.2.1 相関係数

多くの文献では、無次元指標として有名な相関係数がまず最初に議論される^{9), 11)}。相関係数 (以降, r と記載) は、

$$r = \frac{\sum_{k=1}^N (y(k) - \bar{y})(\hat{y}(k) - \bar{\hat{y}})}{\sqrt{\sum_{k=1}^N (y(k) - \bar{y})^2 \sum_{k=1}^N (\hat{y}(k) - \bar{\hat{y}})^2}} \quad (2)$$

で定義される。ただし、 $\bar{\hat{y}}$ は予測値の平均値である。(2) 式をベクトルを用いて書き直すと、

$$r = \frac{\langle \mathbf{y} - \bar{\mathbf{y}}, \mathbf{\hat{y}} - \bar{\mathbf{\hat{y}}} \rangle}{\|\mathbf{y} - \bar{\mathbf{y}}\| \|\mathbf{\hat{y}} - \bar{\mathbf{\hat{y}}}\|} \quad (3)$$

のように内積の形で記述できる。ただし、

$$\mathbf{y} = [y(1) \ y(2) \ \dots \ y(N)], \quad \mathbf{\hat{y}} = [\hat{y}(1) \ \hat{y}(2) \ \dots \ \hat{y}(N)] \\ \bar{\mathbf{y}} = [\bar{y}, \bar{y}, \dots, \bar{y}] \in \mathbb{R}^N, \quad \bar{\mathbf{\hat{y}}} = [\bar{\hat{y}}, \bar{\hat{y}}, \dots, \bar{\hat{y}}] \in \mathbb{R}^N$$

とおいた。(2) 式はコーシー＝シュワルツの不等式 (以下、シュワルツの不等式と略記) により規格化されており、 $-1 \sim 1$ の値をとる。一般的に 1 の場合を正の相関、 -1 の場合を負の相関と呼び、両者を区別する。一方、 r は y と $\hat{y}$ が定数倍の関係にある場合は常に 1 となり、区別がつかない。また、 $\mathbf{\hat{y}} = \mathbf{0}$ に対して、(2) 式はゼロ割をおこし値が算出できないことに注意する。

r はさまざまな分野で用いられているが、自然地理学分野において、相関係数は推奨されていない^{9), 11), 15)}。この理由は、

- アウトライアに敏感である
 - オフセットが存在しても値が変化しない
 - 定数倍の関係に対して値が変化しない
- というものである。

そのため、自然地理学分野においては、つぎに紹介する Coefficient of Efficiency が一般的に使われる。

2.2.2 Coefficient of Efficiency

Coefficient of Efficiency (以降, E と記載) は自然地理学分野、特に水文学分野で一般的に利用される指標で、

$$E = 1 - \frac{\sum_{k=1}^N (y(k) - \hat{y}(k))^2}{\sum_{k=1}^N (y(k) - \bar{y})^2} \quad (4)$$

で定義される。ただし、 $y(k)$ と $\hat{y}(k)$ はそれぞれ時刻 k での測定値、予測値であり、 $\bar{y}$ は測定値の平均値である。また、 N はサンプル数である。なお、文献によっては E を “Nash-Sutcliffe model efficiency coefficient” (NSE) と表記するも

のもある¹¹⁾。

この指標は水文学分野で扱われる降雨流出モデルの評価指標として米国土木学会 (ASCE) から推奨されている¹⁶⁾。その理由としては、 $E \leq 0$ の場合は、モデルの予測結果が測定値の平均をモデルとして採用した場合よりも良くないと解釈でき、モデルが妥当でないと判断できることを挙げている。

この指標は MSE が 0 となると、すなわち完全に y と $\hat{y}$ が一致したときに最大値 1 をとる。しかし、MSE の増加に対して値が単調減少するため、最小値が存在しない。特に、MSE が測定値の分散を上回ると (4) 式は負値をとることについて評価指標として議論となる部分であった。これを受けて提案された最小値が存在する指標がつぎに示す Index of agreement である。

2.2.3 Index of agreement

Index of agreement (以降、 d_j と記載) は

$$d_j = 1 - \frac{\sum_{k=1}^N |y(k) - \hat{y}(k)|^j}{\sum_{k=1}^N (|y(k) - \bar{y}| + |\hat{y}(k) - \bar{y}|)^j} \quad (5)$$

で定義される⁸⁾。ただし、 j は任意の正整数であり、特に $j = 1$ としたものは “Modified index of agreement” (d_1) と呼ばれる。

d_j が規格化されていることは簡単に示すことができる。(5) 式はつぎのように書き直せる。

$$\begin{aligned} 1 - d_j &= \frac{\sum_{k=1}^N |y(k) - \hat{y}(k)|^j}{\sum_{k=1}^N (|y(k) - \bar{y}| + |\hat{y}(k) - \bar{y}|)^j} \\ &= \frac{\sum_{k=1}^N |y'(k) - \hat{y}'(k)|^j}{\sum_{k=1}^N (|y'(k)| + |\hat{y}'(k)|)^j} \end{aligned} \quad (6)$$

ただし、 $y'(k) = y(k) - \bar{y}$ 、 $\hat{y}'(k) = \hat{y}(k) - \bar{y}$ とおいた。ここで、三角不等式より

$$\begin{aligned} |y'(k) - \hat{y}'(k)| &\leq |y'(k)| + |\hat{y}'(k)| \\ &= |y'(k)| + |\hat{y}'(k)| \end{aligned} \quad (7)$$

が成り立つ。なお、(7) 式の等号成立条件は $y'(k)$ か $\hat{y}'(k)$ のいずれかが 0 か、任意の正数 λ に対して、 $y'(k) = -\lambda \hat{y}'(k)$ となるときである。また、 $y'(k) = \hat{y}'(k)$ の場合、(7) 式左辺が 0 となる。(7) 式より、

$$0 \leq \frac{\sum_{k=1}^N |y'(k) - \hat{y}'(k)|}{\sum_{k=1}^N (|y'(k)| + |\hat{y}'(k)|)} \leq 1 \quad (8)$$

が成り立つ。したがって、任意の $j > 0$ に対して、(6) 式は

0~1 の間に規格化されている。

d_j のもう一つの大きな特徴は、(5) 式右辺第二項の分母において、測定値 y 、予測値 $\hat{y}$ それぞれから、測定値の平均値 $\bar{y}$ を差し引いていることが挙げられる。これにより、測定値と予測値の平均値の差まで含めて評価できる。具体例を挙げてこの効果を説明する。測定値と予測値が以下の関係をもつとする。

$$\hat{y}(k) = y(k) + f, \quad \bar{y} = 0, \quad f > 0 \quad (9)$$

ただし、 f は予測値のオフセットである。このとき、 d_j は

$$d_j = 1 - \frac{N |f|^j}{\sum_{k=1}^N (|y(k)| + |y(k) + f|)^j} \quad (10)$$

となる。 $y(k) > 0$ の場合、 $|y(k)| + |y(k) + f| = 2y(k) + f$ 、 $y(k) \leq 0$ の場合、 $|y(k)| + |y(k) + f| = f$ であるので、

$$d_j = 1 - \frac{N f^j}{\sum_{k=1}^{N_1} (2y(k) + f)^j + (N - N_1 + 1) f^j} \quad (11)$$

となる。ただし、簡単のため $1 \leq k \leq N_1$ のとき $y(k) > 0$ 、 $N_1 + 1 \leq k \leq N$ のとき、 $y(k) \leq 0$ とした。すると、(11) 式は

$$d_j = 1 - \frac{N f^j}{g(f) + N f^j} = 1 - \frac{f^j}{g(f)/N + f^j} \quad (12)$$

と書き表わせる。ただし、 $g(f) = (2y(k) + f)^j - f^j = g_1 f^{j-1} + g_2 f^{j-2} + \dots$ とした。(12) 式を f で偏微分して、

$$\frac{\partial d_j}{\partial f} = -\frac{f^{j-1} g(f)}{N(g(f)/N + f^j)^2} \quad (13)$$

を得る。 $g(f)$ の各要素 $g_1, g_2, \dots$ と f は正なので、(13) 式は常に負値となる。よって、オフセット f の増加に応じて値が単調減少することがわかる。

d_j は E の欠点であった最小値が存在しない課題を克服しており、この指標をもとにした新たな指標を作る動きが最近も続いている¹²⁾。一方で、この指標は Coefficient of Efficiency のような解釈ができない欠点があるとされ、Coefficient of Efficiency は標準的に用いられていない⁹⁾。

2.3 適合率の妥当性について

適合率は前節で述べた E と平方根の有無を除いて同様の形状である。そのため、 $FIT < 0$ の場合はモデルの予測結果が測定値の平均をモデルとして採用した場合よりも RMSE の意味で良くないという解釈はそのまま適用でき、ある程度の妥当性はあると思われる。

3. 数値例による無次元指標の特徴把握

線形システム同定においては、推定されたモデルの振幅特性、位相特性それぞれを適切に評価できることが望ましい。FIT による RMSE の意味での評価結果がこれらの特性の評価とどのように関係するかは自然地理学分野の議論ではなさ

れていない。そこで、線形システム同定の評価を意識した二つのケーススタディをもとに、適合率や、自然地理学分野で提案されている無次元指標の妥当性について考察する。本章では、まずシステム同定結果評価を単純化した正弦波の波形を比較することによって、前章までに紹介した指標それぞれの特徴を把握する。つぎに、より現実的な例での各指標の振る舞いを観察し、それぞれの指標が有用である場合について考察する。

3.1 正弦波例題

3.1.1 問題設定

簡単な場合として、二つの正弦波

$$y_1(\theta) = \sin \theta, \quad y_2(\theta) = a \sin(\theta + \phi) \quad (14)$$

を仮定する。ただし、 $a \geq 0$ とし、 y_1 を測定値、 y_2 を予測値とする。また、 $0 \leq \theta < 2\pi$ とし、 a と ϕ を変化させたときの各指標値 (FIT, r , E , d_1) を観察する。なお、このケーススタディでは簡単のため、連続的な変量を扱うこととした。そのため、各指標の和分を積分に書き換えて評価を行なった。

この例題は、正弦波を印加した際の線形システム同定結果の比較問題を単純化したものである。まず、システム同定結果の比較問題として、以下を仮定する。

$$y(t) = G(p)u(t) + v(t)$$

$$\hat{y}(t) = \hat{G}(p)u(t)$$

ただし、 $G(p)$, $\hat{G}(p)$ はそれぞれ真のシステム、モデルであり、 u , y , $\hat{y}$, v はそれぞれ、入力、出力測定値、出力予測値、雑音とする。また、 p は微分オペレータである。 u や v はさまざまな周波数成分を含んでいるが、ある周波数 ω に注目するとそれぞれ正弦波で表わすことができる。すなわち、

$$y(t) = G(p)a \sin(\omega t) + b \sin(\omega t + \phi) \quad (15)$$

$$\hat{y}(t) = \hat{G}(p)a \sin(\omega t) \quad (16)$$

で表わされる。ただし、 a , b はそれぞれ周波数 ω における入力 u と雑音 v の振幅、 ϕ は位相差を示す適当な値である。 G の ω におけるゲイン、位相遅れをそれぞれ g_a , g_ϕ , $\hat{G}$ の ω におけるゲイン、位相遅れをそれぞれ $\hat{g}_a$, $\hat{g}_\phi$ とすれば、 y , $\hat{y}$ は

$$y(t) = c \sin(\omega t - g_\phi + \psi)$$

$$\hat{y}(t) = a \hat{g}_a \sin(\omega t - \hat{g}_\phi)$$

の形になる。ただし、

$$c = \sqrt{(ag_a)^2 + b^2 + 2ab \cos(\phi + g_\phi)}$$

$$\psi = \arctan \frac{b \sin(\phi + g_\phi)}{a + b \cos(\phi + g_\phi)}$$

である。このように、正弦波入力という限定された条件についてではあるが、システム同定結果の時間領域での比較例となる。

Table 1 Representing values of case study 1

	a	ϕ	Condition
Case 1	1	0	$y_2 = y_1$
Case 2	0	0	$y_2 = 0$
Case 3	2	0	$y_2 = 2 y_1$
Case 4	1/2	0	$y_2 = 1/2 y_1$
Case 5	1	$\pi/2$	$\pi/2$ phase lag
Case 6	1	π	Opposite phase

Table 2 Results of representing values of case study 1

	FIT	r	E	d_1
Case 1	100	1	1	1
Case 2	0	NaN	0	0
Case 3	0	1	0	0.67
Case 4	50	1	0.75	0.67
Case 5	-41.5	0	-1	0.29
Case 6	-100	-1	-3	0

3.1.2 計算結果

まず、 a と ϕ を **Table 1** に示す代表的な 6 条件で変化させ、それぞれの指標を計算した。その計算結果を **Table 2** にまとめた。ただし、NaN はゼロ割りのため値が算出できないことを示す。

そして、この例に対しておのおのの指標を解析的に計算した結果をつぎに示す。

適合率

$$\begin{aligned} \text{FIT} &= \left(1 - \frac{\sqrt{\int_0^{2\pi} (\sin \theta - a \sin(\theta + \phi))^2 d\theta}}{\sqrt{\int_0^{2\pi} \sin^2 \theta d\theta}} \right) \times 100 [\%] \\ &= \left(1 - \sqrt{a^2 - 2a \cos \phi + 1} \right) \times 100 [\%] \end{aligned} \quad (17)$$

相関係数

$$\begin{aligned} r &= \frac{\int_0^{2\pi} a \sin \theta \sin(\theta + \phi) d\theta}{\sqrt{\int_0^{2\pi} \sin^2 \theta d\theta} \sqrt{\int_0^{2\pi} a^2 \sin^2(\theta + \phi) d\theta}} \\ &= \cos \phi \end{aligned} \quad (18)$$

Coefficient of Efficiency

$$\begin{aligned} E &= 1 - \frac{\int_0^{2\pi} (\sin \theta - a \sin(\theta + \phi))^2 d\theta}{\int_0^{2\pi} \sin^2 \theta d\theta} \\ &= 2a \cos \phi - a^2 \end{aligned} \quad (19)$$

Modified index of Agreement

$$\begin{aligned} d_1 &= 1 - \frac{\int_0^{2\pi} |\sin \theta - a \sin(\theta + \phi)| d\theta}{\int_0^{2\pi} |\sin \theta| d\theta + \int_0^{2\pi} |a \sin(\theta + \phi)| d\theta} \\ &= 1 - \frac{\sqrt{a^2 - 2a \cos \phi + 1}}{a + 1} \end{aligned} \quad (20)$$

Fig. 1 Visualization of FIT on the example (black shows values less than -100)

Fig. 2 Visualization of E on the example (black shows values less than -1)

3.1.3 計算結果の考察

以上で示した各指標の計算結果 (17)~(20) を、横軸 ϕ 、縦軸 a として、Fig. 1~Fig. 4 に示した。なお、FIT は下限が存在しないため、-100 以下は同色で示している。

まず、FIT は Table 2 の Case 3, 4 を比較すると、Case 4 のように a が小さい場合のほうが高い値をとっている。さらに、Fig. 1 より、FIT は a が小さいほど、 ϕ の変化に対して高い値をとりやすいことがわかる。特に、 $a \approx 0$ では位相変化によって負値をとりにくい、 $a = 1$ では $|\phi| > \pi/3$ において負値をとってしまう。この傾向から、FIT は実際のシステムよりもゲインの小さいモデルのほうが、ゲインの大きいモデルよりも良い値を与えることが示唆される。また、ゲインが小さいモデルほど、位相のずれに対して値は減少しにくいことも示唆される。これは、(17) 式を a で偏微分すると

$$\frac{\partial \text{FIT}}{\partial a} = -\frac{a - \cos \phi}{\sqrt{a^2 - 2a \cos \phi + 1}} \quad (21)$$

となり、 $a = \cos \phi \leq 1$ で (21) 式が極値をとることからも確かめられる。

また、E は Table 2 や Fig. 2 でもわかるとおり、FIT と同

Fig. 3 Visualization of r on the example

Fig. 4 Visualization of d_1 on the example

傾向を示す。

r は Fig. 3 より、 a の値には無関係で ϕ が直接指標値に現れる。これより、 r は位相のずれに対しての評価に有用でありそうである。

Fig. 4 は Fig. 1 を a 軸に対して引き伸ばしたような形状をしている。特に、 $\phi = 0$ において、 $a = 1$ から $a = 0$ に向かうときに急激に指標値が 0 に近づく点は FIT と大きく異なる。実際、Table 2 の Case 3, 4 は同じ値を示しており、ゲインの小さいモデルに対して良い値を与えるものではないことが示唆される。これについても (20) 式を a で偏微分すると

$$\frac{\partial d_1}{\partial a} = -\frac{(1-a)(1-\cos \phi)}{(a+1)^2 \sqrt{a^2 - 2a \cos \phi + 1}} \quad (22)$$

となり、(22) 式は $a = 1$ または $\phi = 0$ の場合に極値をとることから確かめられる。

3.2 システム同定結果の比較例

前節で正弦波を用いてそれぞれの指標の特徴の一部分を説明した。本節ではより現実的な例を示す。

3.2.1 問題設定

離散時間線形時不変システム

$$y(k) = G(q)u(k) \quad (23)$$

Fig. 5 Waveform of the example: gain shift case (solid: $y(k)$, dotted: $y_+(k)$, and dashed: $y_-(k)$)

を考える。ただし、 $G(q)$ は離散時間伝達関数、 $u(k)$ 、 $y(k)$ はそれぞれ時刻 k でのシステムの入力、出力であり、 q はシフトオペレータである。

同定入力 u は 1 と -1 の二値で構成される 1 周期 1023 サンプルの M 系列を 1 周期印加した。ここで、真のシステムとして

$$G(q) = \frac{q^{-1} + 0.5q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}}$$

を仮定した。いま、 $G(q)$ に対してゲインのみ異なる 2 種類のモデル

$$\hat{G}_+(q) = 2.01G(q), \quad \hat{G}_-(q) = 0.49G(q)$$

そして、位相のみ異なるモデル、位相とゲインの異なるモデル

$$\hat{G}_p(q) = G(q)q^{-3}, \quad \hat{G}_{p-}(q) = 0.49G(q)q^{-3}$$

が推定されたとする。出力 y に対し、前述の 4 種類のモデルそれぞれに u を印加した際に得られたモデル出力

$$\hat{y}_+(k) = \hat{G}_+(q)u(k), \quad \hat{y}_-(k) = \hat{G}_-(q)u(k)$$

$$\hat{y}_p(k) = \hat{G}_p(q)u(k), \quad \hat{y}_{p-}(k) = \hat{G}_{p-}(q)u(k)$$

そして、全時刻で 0 をとる出力

$$\hat{y}_0(k) = 0$$

の指標値をそれぞれ算出した。

3.2.2 計算結果

真の出力 y とゲインのみ異なる二種類のモデル出力 $\hat{y}_+$ 、 $\hat{y}_-$ の 0~200 サンプルを **Fig. 5** に示した。また、真の出力 y とゲイン、位相の異なる二種類のモデル出力 $\hat{y}_p$ 、 $\hat{y}_{p-}$ の 0~200 サンプルを **Fig. 6** に、算出した指標値を **Table 3** にまとめた。

FIT は、出力 y のほぼ 2 倍である y_+ に対しては負値をとったが、ほぼ半分である y_- に対しては正値をとっている。また、 y と同振幅で位相のみ異なる y_p に対しては負値をとったが、 y_p と同じだけ位相遅れをもち、かつ y よりも振幅が小さい y_{p-} は正値をとっている。最も極端なケースである全時刻で 0 をとる y_0 に対して、FIT は、 y_{g+} 、 y_p よりも大きい

Fig. 6 Waveform of the example: phase shift case (solid: $y(k)$, dotted: $y_p(k)$, and dashed: $y_{p-}(k)$)

Table 3 Values of criteria for case study 2

	FIT	E	r	d_1
$\hat{G}_+$	-1.0%	-0.02	1	0.66
$\hat{G}_-$	49%	0.74	1	0.66
$\hat{G}_p$	-13%	-0.29	0.36	0.44
$\hat{G}_{p-}$	5.6%	0.11	0.36	0.38
$\hat{G} = 0$	0%	0	NaN	0

値である 0 をとっている。 y に対して位相のみが異なる波形や、単純に y を 2 倍した波形よりも、0 をとり続ける出力のほうが良い評価値を示す。

FIT はゲインが小さいモデルのほうが大きいモデルよりも良い値を示しやすい傾向があり、モデルの妥当性を評価する際に誤った判断を促す可能性がある。 E も同様である。

r は y_+ と y_- に対しては最大値である 1 を示す一方、 y_p と y_{p-} についてはこれらよりも小さな値を示している。これより、 r は振幅の一致度に対する評価には向かないと考えられる。一方、前節でも示唆したとおり位相の一致度を評価する指標として利用できると考えられる。

最後に、 d_1 は、 $\hat{y}_+$ と $\hat{y}_-$ ではほぼ同じ値をとり、 $\hat{y}_p$ と $\hat{y}_{p-}$ を比べると $\hat{y}_p$ のほうが大きな値を示している。また、 y_0 に対しては最小値である 0 を示している。このケーススタディにおいては d_1 は FIT であったような不自然な評価結果にはなっていない。

3.3 無次元指標によるシステム同定結果の評価について

二つの例題を通して考察したとおり、モデル出力 $\hat{y}$ の振幅が出力 y より小さい場合、FIT による評価結果のみをもってモデルの良し悪しを判断するべきではないと考えられる。一方、 r は位相についての評価に有用であると考えられ、 d_1 は FIT に代わる評価指標として利用できる可能性がある。

4. おわりに

システム同定において広く用いられる適合率 (FIT) について考察した。その際、自然地理学分野で行なわれた相関係数 (r)、Coefficient of Efficiency (E)、Index of Agreement (d_j) についての議論を紹介し、それらの特徴を簡単な数値例を用いて整理し、自然地理学分野においては適合率と同様な定義をもつ Coefficient of Efficiency が推奨されていること

を述べた．一方で，適合率がこれらと同じ議論ができないことを二つの数値例を通じて指摘した．今後はシステム同定結果の評価に有用な無次元指標について検討していく．

参 考 文 献

- 1) J.M. Maciejowski: モデル予測制御 制約のもとでの最適制御, 7/8, 東京電機大学出版局 (2005)
- 2) 足立, 丸田: カルマンフィルタの基礎, 1/14, 東京電機大学出版局 (2012)
- 3) 足立修一: システム同定の基礎, 27/28, 東京電機大学出版局 (2009)
- 4) L. Ljung: *System Identification — Theory for the User*, 2nd Edition, 509/516, PTR Prentice Hall (1999)
- 5) J.E. Nash and J.V. Sutcliffe: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of hydrology*, **10**-3, 282/290 (1970)
- 6) M. Garrick, C. Cunnane and J.E. Nash: A criterion of efficiency for rainfall-runoff models, *Journal of Hydrology*, **36**-3, 375/381 (1978)
- 7) C.J. Willmott: On the validation of models, *Physical Geography*, **2**-2, 184/194 (1981)
- 8) C.J. Willmott, S.G. Ackleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates and C.M. Rowe: Statistics for the evaluation and comparison of models, *Journal of Geophysical Research*, **90**-C5, 8995/9005 (1985)
- 9) D.R. Legates and G.J. McCabe: Evaluating the use of “goodness of fit” measures in hydrologic and hydroclimatic model validation, *Water Resources Research*, **35**-1, 233/241 (1999)
- 10) B. Schaeffli and H.V. Gupta: Do Nash values have value?, *Hydrological Processes*, **21**-15, 2075/2080 (2007)
- 11) D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel and T.L. Veith: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transaction of the ASABE*, **50**-3, 885/900 (2007)
- 12) C.J. Willmott, S.M. Robeson and K. Matsuura: A refined index of model performance, *International Journal of Climatology*, **32**-13, 2088/2094 (2012)
- 13) D.R. Legates and G.J. McCabe: A refined index of model performance: A rejoinder, *International Journal of Climatology*, **33**-4, 1053/1056 (2013)
- 14) Mathworks: Compare model output and measured output — MATLAB compare (online), available from <http://www.mathworks.co.jp/jp/help/ident/ref/compare.html> (accessed 2014-10-07)
- 15) C.J. Willmott: Some comments on the evaluation of model performance, *Bulletin of the American Meteorological Society*, **63**-11, 1309/1313 (1982)
- 16) ASCE Task Committee: Criteria for evaluation of watershed models, *Journal of Irrigation and Drainage Engineering*, **119**-3, 429/442 (1993)

《 付 録 》

A. 不等式の適用対象による分類

紹介した無次元指標のうち最小値が規格化されている r と d_j を, 規格化に利用した不等式についてまとめると, **Table A.1** のようになる. なお, 表中の All samples は二つの波形それぞれの全体を一つのベクトルとみなし, これらのノルムや内積を評価するものであり, Each sample は二つの波形に対して, サンプル 1 点ずつのノルムを評価するものである. 具体的にはつぎに示すとおりである.

Table A.1 Classification of criteria by inequality

	Each sample	All samples
Schwarz inequality	-	r
Triangle inequality	d_j	d_1

A.1 シュワルツの不等式, Each sample

$$A_1 = \frac{1}{N} \sum_{k=j}^N \frac{\langle \mathbf{y}(k), \hat{\mathbf{y}}(k) \rangle}{\|\mathbf{y}(k)\| \|\hat{\mathbf{y}}(k)\|} \quad (\text{A.1})$$

$$A_2 = \frac{\sum_{k=j}^N \langle \mathbf{y}(k), \hat{\mathbf{y}}(k) \rangle}{\sum_{k=j}^N \|\mathbf{y}(k)\| \|\hat{\mathbf{y}}(k)\|} \quad (\text{A.2})$$

ただし, $\mathbf{y}(k) \in \mathbb{R}^j$, $\hat{\mathbf{y}}(k) \in \mathbb{R}^j$ で, j はベクトルの次元である.

A.2 シュワルツの不等式, All sample

(3) 式を参照.

A.3 三角不等式, Each sample

(5) 式を参照.

A.4 三角不等式, All sample

$$B = 1 - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|}{\|\mathbf{y}\| + \|\hat{\mathbf{y}}\|} \quad (\text{A.3})$$

である. ここで, (5) 式のうち d_1 は

$$\begin{aligned} d_1 &= 1 - \frac{\sum_{k=1}^N |y(k) - \hat{y}(k)|}{\sum_{k=1}^N (|y(k) - \bar{y}| + |\hat{y}(k) - \bar{y}|)} \\ &= 1 - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_1}{\|\mathbf{y} - \bar{\mathbf{y}}\|_1 + \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|_1} \end{aligned}$$

であるため, この分類では (A.3) 式にも含まれる. Modified index of Agreement と呼ばれる d_1 は d_j の中でも特殊な位置づけであることがわかる.

[著 者 紹 介]

室 井 秀 夫 (正会員)

2010 年慶應義塾大学大学院理工学研究科基礎理工学専攻博士前期課程修了。同年 (株) リコーに入社。2013 年慶應義塾大学大学院理工学研究科基礎理工学専攻博士後期課程入学。現在に至る。システムの同定に関する研究に興味をもつ。

竹 下 侑

2014 年慶應義塾大学大学院理工学研究科基礎理工学専攻博士前期課程修了。同年ソニーグローバルソリューションズ (株) に入社。現在に至る。在学中、システムの同定に関する研究に興味をもつ。

足 立 修 一 (正会員・フェロー)

1986 年慶應義塾大学大学院工学研究科博士課程電気工学専攻修了 (工学博士)。同年 (株) 東芝入社。総合研究所勤務。90 年宇都宮大学工学部電気電子工学科助教授。93~96 年科学技術庁航空宇宙技術研究所客員研究官兼務。2002 年宇都宮大学工学部電気電子工学科教授。2003~2004 年ケンブリッジ大学客員研究員。2006 年慶應義塾大学理工学部物理情報工学科教授。現在に至る。システム同定理論、制御理論とそれらの応用研究に従事。電気学会、日本機械学会、IEEE などの会員。