

物の動きを読む数理—情報量規準 AIC 導入の歴史—

赤池 弘次

はじめに

情報量規準の考え方は、従来の正統的な統計学の展開に沿う形の研究で得られたものではなく、実際的な問題処理の必要から、道のない所を手探りで進むような研究を続ける過程で到達したものです。その意味では型破りな発想に見えます。このような研究の進め方をするようになった理由を、自身の物の見方の癖と育った環境や時代の影響などに関連させて見直しながら、情報量規準 AIC の導入とその効果について歴史的に眺めたいと思います。

生い立ち

富士山の南山麓の田舎に生まれましたが、叔父の一人が民間航空の草分け時代の操縦士で、この叔父から貰った機関車やモーターボートの模型をひっくり返して動きの仕組みを見るのが好きでした。この辺りから、出来上がった結果を眺めるよりは、何故そのような結果が生まれるのかを探ることに興味を感じていたように見えます。結果を予測しながらあれこれ試すわけです。何事につけてもこのスタイルで進む癖が、「三つ子の魂百まで」とそのまま現在まで継続しています。

小学生の頃は、書き取りなどの暗記物や珠算のような機械的な作業は不得意で、算術の応用問題を解くことは得意でした。飛行機の操縦士の叔父の家に寄宿して学校に通っていた兄の影響もあり、また、海軍の戦闘機乗りであったもう一人の叔父の影響もあって、中学に進む頃までは飛行機に関心がありました。この叔父が数学の話を聞かせてくれ、数学への関心が芽生えました。

中学では熱心な先生に英語を教えられましたが、単語の暗記は苦手で、実際の使用状況を単文の形で覚えていました。剣道と水泳の練習にも精を出して、結局当時の海軍兵学校に進学しました。

兵学校では、理工系の知識の初歩的な部分に触れることができ、確率や統計の初歩的な知識にも触れる機会が与えられました。すべての知識を有機的に利用することの有効性を体感したように思います。

終戦後は戦後の復興に貢献することを目指して故郷に帰りましたが、社会の価値観が音を立てて崩れるのを見て大いに悩み、あれこれ考えた末に、自他の生命の尊重が道徳の基本であると認識して心の平安を取り戻しました。旧制の第一高等学校理科に進学し将来の進路についても悩み、結局東京大学理学部数学科に進みましたが、これは私にとって、いわゆる狂瀾怒涛の時期でした。

1952年に統計学関連の数理の研究機関である当時の文部省統計数理研究所に入所し、ここで我が国の現実に根を置く研究を目指しました。10年間ほどは具体的問題の探索に時間を費やしましたが、戦後の復興を支える生産活動に直結する課題に触れる機会を得て、以来一貫して時間と共に変動する現象の解析と制御に関する課題を追求することになりました。これは、子供の頃からの動く物の仕組みに対する興味に導かれた結果のように見えます。この課題の選択があれこれの発想を要求してAICに到達することになったわけです。

これから、AICに関連する話題に進みます。

予測の数理

確率

AICを利用して観測データの有効利用を進めるには、当面の観測データがどのようにして得られるのかを説明するモデルを使います。データを生み出す仕組みは完全には記述できないのが普通で、この不確かさを表現するために確率を利用してモデルを組み立てます。

確率の考えはごく自然なもので、不確実な事柄の処理では誰でもほとんど無意識の内に利用しています。ある結果が現れることに対する見込みの確からしさを数字で表すものが確率で、明日雨が降る確率は0.7(70パーセント)などと表現します。この場合10中7の割合で明日は雨が降ると見込んでいるわけです。この考え方自体はごく単

純に見えますが、実際にある結果に対する確率を求めようとすると、いろいろ考えなくてはなりません。

これに対して、サイコロを振った時に6の目の出る確率のように、確率を決める仕組みが客観的に与えられる場合もあります。これは確率の考えの出発点となった、占いや賭で利用されるランダムな動きをする道具、すなわちランダマイザーの利用で与えられる確率です。

サイコロのそれぞれの目に賞金が出る賭では、賞金にその目の出る確率を掛けて加え合わせれば期待値が求まります。この場合の考え方の要点は

期待値 = 確率に基づいて予想される値
ということです。そこで、賭の参加者は儲けの可能性について慎重に考えます。

確率と統計の繋がり

実際に確率を使って効果的に将来に対する計画を立てようとすると、確率のもつ複雑な内容に気がつきます。明日雨が降る確率が0.7と言う時は、何かの根拠に基づいてこの数字を取り上げています。たとえば、過去の記録で今日の気象状態と似た状態の日の翌日が雨であった場合の割合、というような根拠に基づいて決められます。

そこで、これまでの同様な状況の日の記録を並べてみると、過去の100日の例の中で70日が翌日雨であったというような結果が得られます。この場合、日常の言葉の使い方に従えば、過去の統計に基づく、全数に対する雨の比率70/100(=0.7)、すなわち平均的な発生率で明日が雨になる確率が決まることとなります。

これを簡単に表現すれば

過去の統計の平均値 = 将来に対する期待値
となります。実用上の基本的な考え方はこのようなもので、確率が統計的なデータの利用と自然に結びつくことが分かります。

この考え方の基礎にあるのは、サイコロを何回も投げるように、データを無限に長く記録し続ければ確率が求まるという考え方です。ところが、実際にはこれは不可能で、

逆に今手許にあるデータをもとに、将来の観測値を生み出す確率的な仕組み、すなわちモデルを作り上げ、これを使って予測を実行することになります。

このように、利用可能なデータを有効利用して、将来の値を予測する仕組みをつくるのが統計的データ処理の本質なのです。「温故知新」を具体化する形の仕事です。以下、その実際的利用の例について話を続けます。

実際問題への適用

生糸繰糸工程の統計的管理

最初の例は生糸の繰糸工程の統計的管理で、当時の農林省蚕糸試験場の島崎昭典さんとの協力で得られたものです。戦前は生糸が日本の主要な輸出品で、歴史的に見ても長い伝統のある産業です。しかし、その生産方法は経験に大きく依存しておりましたが、戦後の統計的工程管理手法の導入に伴い、時々刻々の観測値に基づいて工程の異常の検出を行う管理図法の導入が検討されていました。

当時は一定個数の繭の繊維を撚り合わせて一本の生糸に紡いでいました。管理の対象は一定の時間内に繊維が切れて繭が落ちる落緒の回数で、これが異常に増加すれば工程の異常と判断されます。通常管理図法の教えでは、定常状態の運転記録から平均的な落緒数と、許容される変動幅を決定し、観測値がこの巾の外に出た時点で異常と判定することになります。

ところが、繭から生糸の繊維を引き出しやすくするための煮方を決める試験繰糸の結果として、一本の繊維の長さの統計的分布のデータが得られます。これを利用すると、次々に繊維を繋ぐ繰糸の流れの中での切れ目の現れ方が確率論的に決まります。これで理想的な運転状態の特性が確認され、管理に必要な数値も理論的に求まります (Fig. 1)。こうして実用上著しい成功が得られたのです。落緒の発生という将来の事象の確率的な構造を考えることにより、試験繰糸で得られた過去の知識の有効利用が実現したのです。

パワースペクトルの推定

これは、ランダム振動の統計的解析ということで、1960年代に多くの関心を惹いた問題ですが、不規則変動の特性を計測する方法の実用化です。当時のいすゞ自動車株式会社の兼重一郎さんと共に、この方法の実用化を進めました。

時間的に変動する時系列データを周期的な振動成分に分解すると、周波数(単位時間内の振動数)を示す横軸上に成分の強さを示すギザギザと激しく変動する系列が得られます。これはピリオドグラムと呼ばれ、その中にひととき目立つ成分が見出されることがあります。地球物理学的な観測値に含まれる周期的成分の検出に、古くから利用されてきたものです。

ところが、この解析法では、自動車の振動のような不規則変動のデータの場合は、全面的にギザギザの動きが現れるだけです。そこで当時の統計学では実用性が疑われていました。しかし、周波数軸上のデータを局所的に平均化すると、パワースペクトルと呼ばれる滑らかな変動のパターンが現れ、光のスペクトルのようにどの辺りの周期成分が優勢であるかを見ることができます(Fig. 2)。

周波数応答関数の推定

さらに一歩進めると、ランダムに変動するハンドル入力を利用して自動車の操縦性能を測るといような、システム特性の計測法が得られます。この場合のシステム特性を表現するものが周波数応答関数です(Fig. 3)。

この計測方法では、現実の動きに類似した状況で、システム特性を簡単に計測できます。この問題の取り扱い、当時の運輸省運輸技術研究所の山内保文さんとの協力で実用化に成功したものです。この周波数応答関数の推定法を様々な分野の研究者が実例に適用した結果が、1964年に統計数理研究所の英文報告に発表されましたが、これは当時世界的に見ても先端的な成果でした。この事例には、自動車、船舶、鉄道車両、航空機、配管系、水力発電所の水管系などへの応用結果が含まれています。

生産プロセスの最適制御

周波数応答関数の推定法を、当時秩父セメント株式会社の中川東一郎さんとの協力を通じて、セメント焼成炉の自動運転の実現に応用しようとしたのですが、難しい問題が浮かび上がり、その代わりに計算機制御に適した新しい最適制御理論の適用を考えることになりました (Fig. 4、5)。

制御系の設計の基礎となるルドルフ・カルマン教授 (第1回 (1985) 京都賞先端技術部門受賞者) のシステム理論では、伊藤清先生 (第14回 (1998) 京都賞基礎科学部門受賞者) の研究成果による確率論的な表現が利用され、理論が展開されていましたが、変動の大きな実プロセスでの実用化には種々の困難がありました。これを克服するために、まずプロセスの動きの予測式を作り、これを利用して制御を実行する方法を考えました。

自己回帰モデル

関係する変数全体の時系列について、現在の値を過去の何時点かの値に適切な係数を掛けて加えたもので予測する自己回帰モデルを採用し、その構造を時系列データから推定し、その結果から各変数間の繋がりを解析するという方法を採用することにしました。この場合の予測誤差は、白色雑音と呼ばれる、単純なランダムイザの出力として表現される構造を持ち、これを利用して実際のシステムの動きのシミュレーションが可能になります。これを利用すれば、時々刻々将来の動きを予測し、適切な制御入力を決定する制御が可能になります (Fig. 6)。

ところがここでさらに問題が発生しました。予測式に過去の何時点前までのデータを取り込むかという、いわゆるモデルの次数の決定の問題です。当時この問題についての実用的な方法は全く見当りませんでした。次数が低すぎると予測力がなくなり、高すぎるとデータの不足から推定の精度が落ちてしまいます。そこで、次数の異なるモデルの比較評価の方法を考案し、これを使って最適モデルを選ぶことで制御系の設計が可能になりました。

最適制御の実現

多変量自己回帰モデルの利用で、実際のシステムの動きの実態が簡単に把握でき

ようになり、不規則な変動を示すセメント焼成炉の制御での最適制御理論の実用化に成功しました。

この方法の実用に必要な計算プログラムを公表した結果、火力発電所ボイラー温度制御、船舶の自動操舵、生体の環境維持機能(ホメオスターシス)の解析、脳波解析、経済時系列解析、原子炉雑音解析など、従来は変量間の動きの関連が確認困難であった様々な研究対象について、実測データによる構造の確認とシミュレーションによる検討が可能になり、それぞれの分野で、驚きをもって迎えられるような結果を生み出しました。

火力発電所の蒸気温度制御の実用化は、当時九州電力の中村秀雄さんが電力中央研究所の研究者の協力を得て実現しました(Fig. 7、8)。定期点検後の運転再開時の模様を見学したことがあります。この時現場の運転責任者からこの制御を歓迎する発言がありました。後日、制御理論の大家、高橋安人先生(1912—1996)が、「光は東方から、実プロセスの最適制御は日本から」と、この制御を高く評価されたという話も中村さんから聞いております。

この制御はその後国内外のボイラーメーカーによって利用され、国内、中国、カナダで運用中とのこと。セメント焼成炉の自動運転の実現と共に、変動の大きな現実のプロセスで最適制御理論の適用を実現した、理論と現実を繋ぐ貴重な成功例と言えます。

実は、この場合、基礎になるモデルの次数を選ぶ際の評価値として何を取るのかは、簡単には決まらなかったのです。この問題を追究する過程で、これからお話をする尤度の概念の利用に引きつけられ、情報量規準 AIC に到達することになりました。

尤度(ゆうど:尤もらしさ)の解明

尤度とは？

ここで、日頃聞き慣れない言葉と思われませんが、確率的なモデルの優劣の評価を与える、尤度(ゆうど)について簡単な説明を試みます。これは非常に面白い概念で、情報量規準はこれに基づいて定義されます。

手許のサイコロを振って6の目が出たとします。ここからが問題です (Fig. 9)。A、B 二つのサイコロがあり、その中の一つが振られことは確かですが、実際に振られたサイコロがどちらかは分からない、というのです。そこで、A か B かの判断が求められます。探偵が6という観測データを手懸かりに、犯人を A、B の何れかに特定しようとする時の状況です。

AとBのサイコロで6の目の出る確率を、それぞれ $P(6/A)$ 、 $P(6/B)$ とします。今6の目が出たという条件のもとで、 $P(6/A)$ が $P(6/B)$ よりも遙かに大きければ、Aのサイコロが使われたと考えるのが自然に見えます。この考えを一般化して、観測データ x に対する $P(x/A)$ 、 $P(x/B)$ の値を、それぞれサイコロA、Bのデータ x に関する“尤もらしさ”、すなわち尤度(ゆうど)と呼びます。

この場合、 x というデータは既に観測され、確定していることに注意する必要があります。 $P(6/A)$ 、 $P(6/B)$ は、それぞれ6に関するサイコロA、Bの尤度であり、6の目の出る確率ではありません。

確率は過去の知識や経験から将来のデータの見方を与えるものですが、尤度は、現在のデータを用いて、過去の時点にこれを生み出した仕組みを評価しようとしています。一見「温故知新」を超える知識の取り扱いの議論に入っているかのように見えます。この問題の本質は、次の情報量の議論で明らかになります。

情報量規準

情報量

尤度をモデルの評価値として利用することの根拠として、情報量と呼ばれる量の利用が考えられます。情報量 $I(Q:P)$ は、モデルPから見て真の構造Qの離れ具合を測る量です。ここでP、Qは観測値の現れ方を決める確率的な仕組みを表し、それぞれの仕組みからデータ x が得られる確率が、 $P(x)$ と $Q(x)$ で与えられます。

観測データ x が得られた場合、これを利用して $I(P:Q)$ の値を測るには、Pの尤度 $P(x)$ の自然対数 $\log P(x)$ を利用します。これを対数尤度と呼びます。この変換により、二つのモデルの比較では、真の構造Qを知らなくとも、「対数尤度の大きい方がQに近

い」と評価するのが合理的である、ということが確率の立場から説明できるのです。感覚的に表現すれば、海拔の高い山が「天」に近いように、対数尤度の高いモデルが「真理」に近いと見なされることとなります (Fig. 10)。これが情報量規準の意味で、これにより尤度の利用の根拠が明らかになったのです。

パラメータを含むモデル

実際に尤度を利用して推論を展開する状況では、いくつかの調節可能な変数、すなわちパラメータを含んだモデルが利用されます。パラメータを調節するとモデルの構造が変わり、現在の観測データと同じ結果を生み出す確率、すなわち尤度を調節することができます。

簡単で具体的なイメージとしては、サイコロの重心部分に重い鉄球があると考え、その位置を前後、左右、上下に移動させる、三つの調節用のねじ、すなわちパラメータを考えます。これを動かせば、それぞれ鉄球の近づく面が下向きになる確率が高まるでしょう。このような方法で、確率分布を調節できるモデルが得られます (Fig. 11)。

AIC

いくつかのパラメータを持つモデルを用い、与えられた観測データに関する尤度を最大化するようにパラメータを調節する最尤法を適用すると、このモデルの情報量規準による評価は

$$\text{AIC} = (-2) (\text{最大対数尤度}) + 2 (\text{パラメータ数})$$

で与えられます。尤度そのものではなく対数尤度を使うところが、情報量規準の微妙な点です。

最大対数尤度に掛かる(-2)という係数のために、AIC が大きいほど悪いモデルと見なされます。第2項の2 (パラメータ数) は、パラメータの調節による尤度の上昇分の補正と、あてはめられたモデルによる予測の誤差分の評価を加え合わせたものです。この項は無用なパラメータの数の増加を抑えます。

AIC の発表

AIC についての最初の発表は、1971年の日本統計学会での報告です。これに次いで、同じ年にアルメニヤ共和国で開催された第2回情報理論国際シンポジウムで詳しい内容を発表しました。これが1973年に公刊され、その内容の解説を1974年にアメリカ電気電子学会制御システム部会紀要に発表しました。

当初、応用分野からは積極的な反応がありましたが、正統的な統計学の分野からの反応は消極的でした。しかし、1992年になると、1973年の論文が現代の統計学への顕著な理論的貢献のいくつかを纏めた書物に収録されました。AIC がゆっくりと受け入れられて行った様子が窺われます。

現在では、哲学者も AIC に対して関心を示すようになりました。論理学では「オッカムの剃刀」と呼ばれる原理、すなわち不要な要素は使わないという原理が知られています。AIC の2(パラメータ数)の項は、無用なパラメータ数の増加を抑えます。その意味で、AIC をこの原理の具体化と捉えて議論を展開している人もいます。

しかし、AIC の利用の前提はモデルの提案です。モデルの構成には、客観的知識、経験的知識、観測データなどを有機的に利用し、仮説の提案と検証を限りなく続ける努力が要求されます。このような仮説の提案の過程を通じて真理を目指す推論の展開を議論した C. S. パースは、これをアブダクション、あるいはリトロダクションと呼んで重視しています。AIC はこの過程の実行を助けるものです。

情報量規準導入の効果

最尤法は、真の構造の中のパラメータの値を、観測値を用いて推定する方法として、これまで理論が展開されてきました。しかし、現実の適用では、これは人工的な実験の場合以外にはあり得ません。実用場面で最尤法が適用されるのは、データからの有効な情報の抽出を目指した人工的な構造物、すなわちモデルなのです。このように考えると、伝統的な統計学が強調した主観性の排除が、実は柔軟な発想によるモデルの展開とその利用を制約し、有効なデータ処理法の実現を妨げてきたことが明らかになります。

情報量規準の導入がもたらす自由な発想により、モデル構成の自由度が高まり、心

理的なイメージに基づくモデルをも含め、利用できるモデルの種類が急激に拡大されました。その効果は、ベイズ・モデルと呼ばれるモデルにより、経済時系列から季節変動の成分の影響を分離する、季節調整法の実現の例で確認できます (Fig. 12)。ベイズ・モデルの発想の様子を図に示します (Fig. 13)。

このモデルは、地球潮汐の解析などの地球物理学的観測値の有効な処理法の開発にも利用されています。統計数理研究所の石黒真木夫さんと国立天文台水沢VERA 観測所の佐藤忠広さん他の皆さんが纏め上げた計算プログラムによる出力例をご覧に入れます (Fig. 14)。

終わりに

話の終わりに、知識の有効利用について触れたいと思います。飛行機の実用化に際しては、科学者が科学的知識を利用して飛行機は実用にならないことを示し、ライト兄弟は逆に空を飛ぶことを目指して知識を活用し、飛行機の実用化に成功したことが知られています。AIC の場合も、既存の理論に基づくデータ処理法に限界を感じ、問題の要求に従って既存の理論を検討することで結果に到達しました。

私の AIC の話は、子供の頃の飛行機との関わりに始まり、飛行機の研究との類比で終わる感じがしますが、ここで最後に強調したいことは、何かを実現しようとする目的意識の重要性です。セメント焼成炉あるいは火力発電所ボイラーの制御の実現では、関係する研究者の粘りと迫力には目覚ましいものがありました。この点についての理解と関心の深まりを願って、私の話を終わりたいと思います。

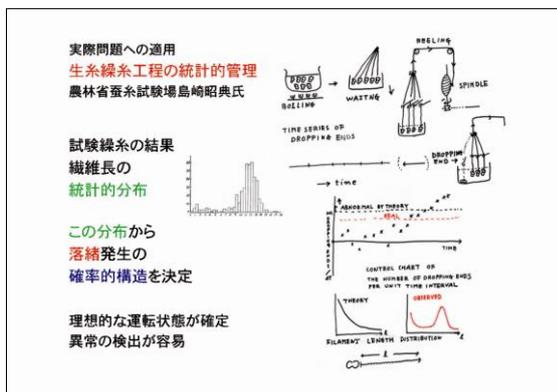


Fig. 1

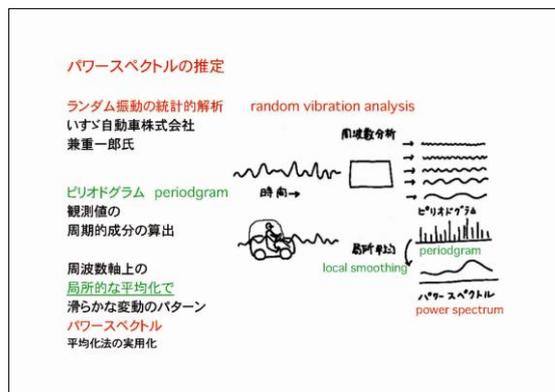


Fig. 2

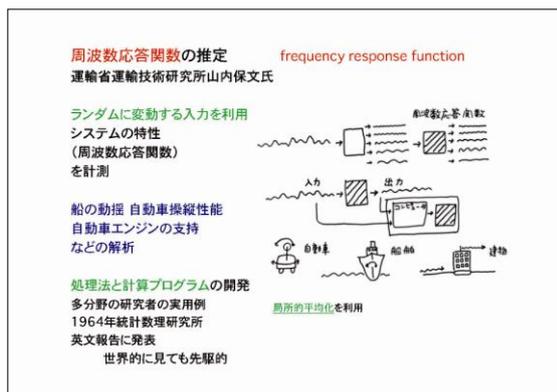


Fig. 3

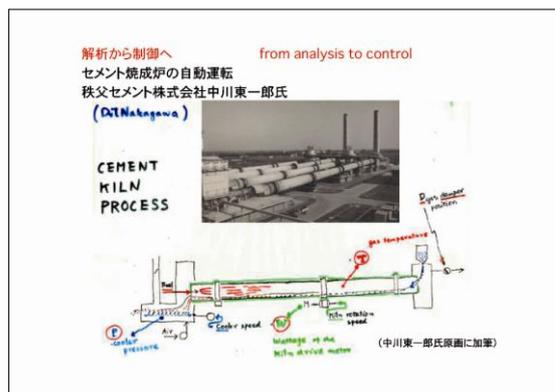


Fig. 4

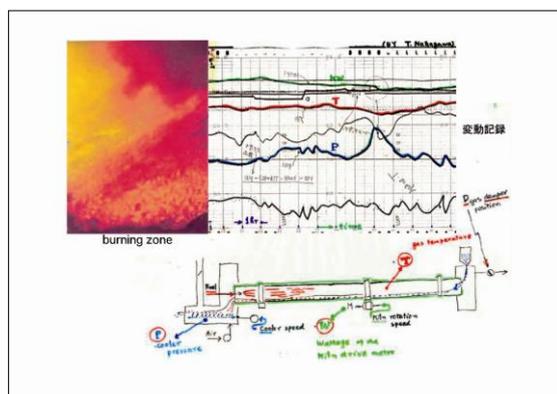


Fig. 5

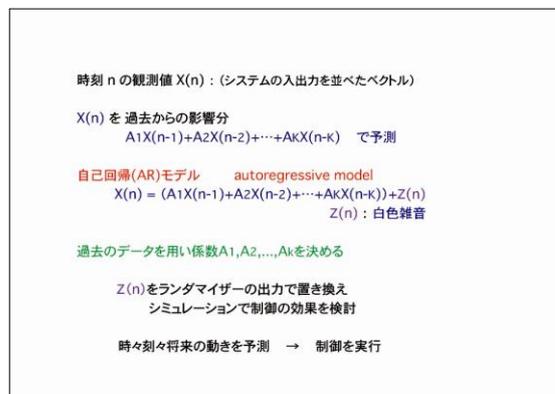


Fig. 6

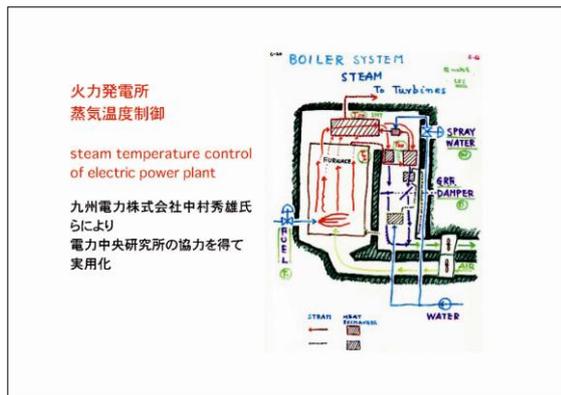


Fig. 7

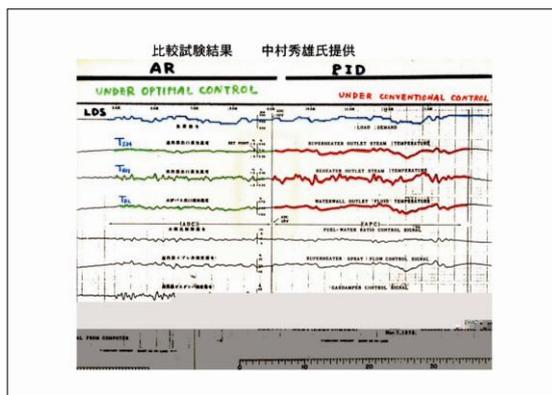


Fig. 8



Fig. 9

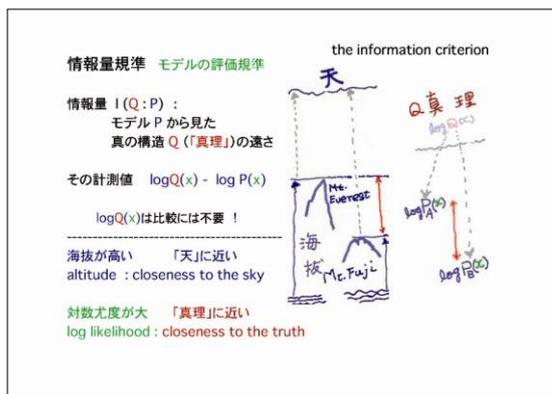


Fig. 10



Fig. 11

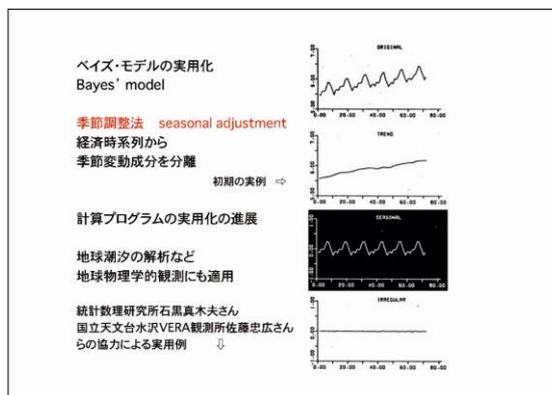


Fig. 12

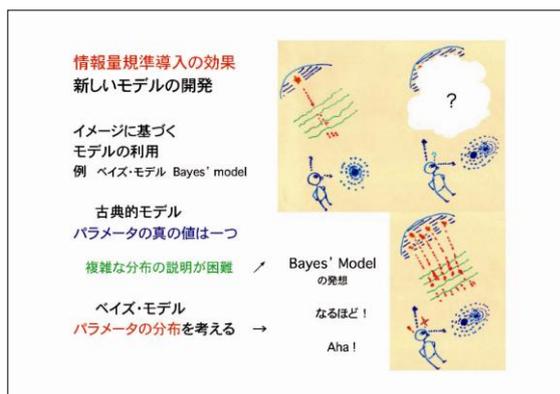


Fig. 13

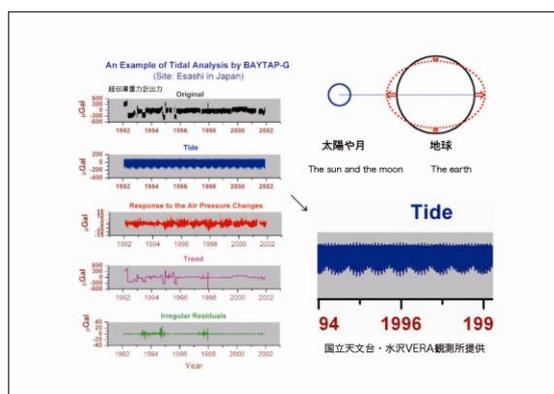


Fig. 14