

Web Decomp の紹介

——WWW 上で行う季節調整システム——

統計数理研究所 佐藤 整 尚

(1997 年 6 月 受付)

1. はじめに

本稿で紹介する“Web Decomp”は WWW 上に構築された統計解析システムである。ユーザーは季節調整モデルをはじめとした、様々な時系列モデルを用いてデータを解析することができる。中心となっている“Decomp”と呼ばれる手法は状態空間モデルを用いた季節調整法として有名であり、Kitagawa and Gersh (1984) によって提案されたものである。本システムでの計算はすべて、サーバーサイドで行われるため、ネットスケープ・ナビゲーターやインターネット・エクスプローラーなどのブラウザさえあれば、ソフトウェアのインストールなど特別な設定が一切不要な点が特長である。これまでも、統計数理研究所では TIMSAC シリーズなどの統計プログラム、ソフトウェアを開発してきたが、本システムは、それらのソフトウェアを気軽に利用してもらうことを主目的に作られている。そのため、本格的に大量のデータを処理するには不向きである。これらの用途には従来のソフトウェアや Windows 用に開発された“TIMSAC for Windows”を利用することをお勧めする。

WWW にアクセスできる読者は、まず以下の URL にアクセスし、本システムの雰囲気をつかむことをお勧めする。サンプルデータが用意されているので、何の準備もなしに試すことが可能である。

<http://www.ism.ac.jp/~sato/>

(または <http://ssnt.ism.ac.jp/inets/inets.html>)

以下の節では、本システムの概要を説明する。ただし、扱える手法の種類やその仕様は常に改善されており、以下で書かれる内容が古くなってしまう可能性がある。最新の情報に関してはオンラインヘルプのページを参照することをお勧めする。

2. Web Decomp のコンセプト

統計科学を研究するものにとって自己の研究成果をソフトウェアの形で公開することは重要な社会的責務であると考えられる。しかしながら、近年の急速なコンピュータの発達と普及は目覚ましいものがあり、それについていくことすら大変になりつつある。だからといって、世の中の流れを全く無視して、独りよがりのソフトウェアを作ってもほとんど使われない、といった状況に陥ってしまう。また、ソフトウェア作りにそれほど時間や労力をかけていられないのも事実である。では、どのような解決策が考えられるであろうか、考えてみる。

従来の方法：まず、従来の統計ソフトウェアの公開方法として以下のようなものがある。

- A) プログラムのソースを公開する方法
- B) 実行形式の形で公開する方法

問題点：これらの方法については以下のような問題点がある。

- A) ユーザーのコンピュータ環境は多種多様であり、1つ1つにソフトを移植または対応させるのは困難である。
- B) プログラムをソースの形で配布しても、インストールが大変である。
- C) ソフトの配布が大変である。
- D) 一度作って配布するとその後の修正は大変である。
- E) 実務家にとってはできれば GUI を使ったインターフェイスの方がわかりやすいがこれを実装するのはかなり面倒である。

これらの問題を改善する試みの例がここで紹介する Web Decomp である。

Web Decomp の特徴：Web Decomp は以下のような特徴を持っている。

- A) 計算はすべてサーバー側で行われるため、プログラムの修正等、メンテナンスが容易である。(故に最新の成果をすぐ公開することも可能)。
- B) すべての動作はインターネット経由で行われる。つまり、インターネットに接続されていればいつでもどこでも実行可能である。
- C) ユーザーは WWW ブラウザーを通じて操作するので、全く新しいインターフェイスにユーザーがとまどうことは少ない。
- D) 基本的に OS やマシンに依存しない。ブラウザーが動作する環境さえあれば十分である。

3. Web Decomp の仕組み

Web Decomp の仕組みについて、その概略を図1に示した。これを見ても分かるように、クライアント-サーバーシステムとして構成されている。ここでの、クライアントマシンは個々のユーザーのマシンであり、直接のインターフェイスは WWW ブラウザーになる。ユーザーはサーバーから送られたページの中で操作を行う。サーバーとのやり取りは、HTTP プロトコルを使いインターネットを介して行われる。

一方、サーバーではクライアントからデータ等を受け取り、指定された解析を実行し、結果を送信する。サーバーでの処理はいくつかの工程に分かれている。まず、はじめに HTTP サーバーによってクライアントのブラウザーからのリクエストを受け付ける。次に CGI (Common Gateway Interface) が実行され、リクエストの内容、及びデータの読み込みがなされる。この後の処理は S 言語で作られたシステムで行われる。S では結果のグラフの作成が主で、計算本体は Fortran プログラムを呼び出すことで行われる。必要な計算が終了すると、計算結果とともに次の解析用のページが出力され、クライアントのブラウザーへ送信される。なお、S は商用ソフトであるので、このサーバーシステム自身を移植する際にはライセンスが必要である。(クライアントからの利用については、ライセンスは不要である。)なお、サーバーで動くプログラムは常駐型ではなく、リクエストを受けるたびに起動し、実行される。(よって、各回の実行はまったく独立である。)

クライアントはサーバーから結果を受け取ると、ブラウザーを使って表示する。したがって、非力なマシンでも分析が可能になる。また、それまでの計算結果はブラウザー内にキャッシュとして貯えられるので、前の結果に戻る事も簡単である。解析の途中で結果をページごとファ

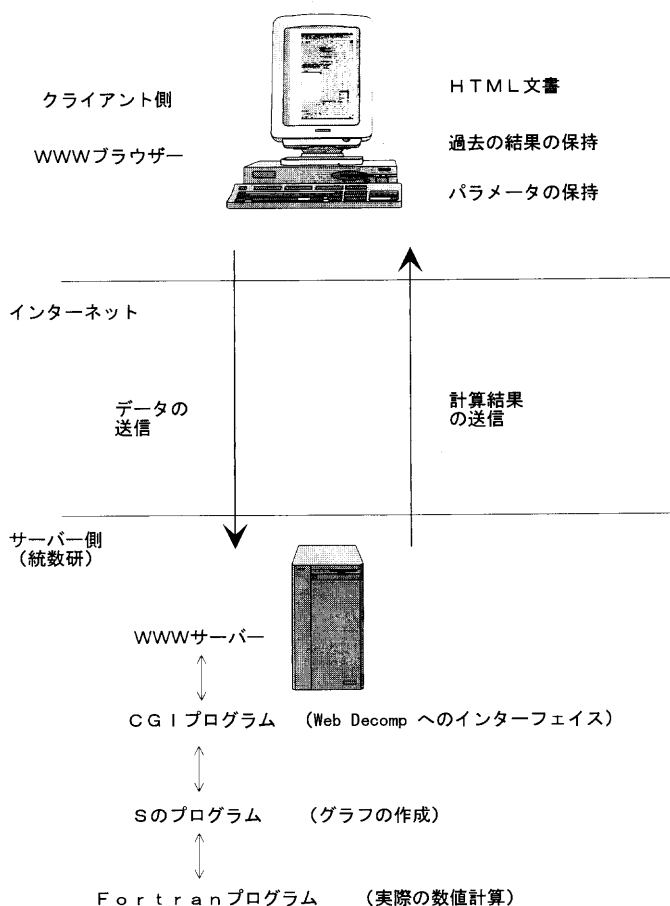


図1. Web Decomp の仕組み。

イルに保存しておけば、次回そこから解析の続きをすることも可能である。

このように Web Decomp のようなシステムが実現されたのはインターネットの発展があったからである。しかしながら、このことが、特長であると同時に欠点にもなる。インターネットを介するので、ネットワーク経路の混雑度によっては大幅に時間がかかることがある。また、クライアントから送られるデータやサーバーからの結果には暗号化などは行われないので、機密データを解析するのは危険である。ネットワークの負荷も考えて、扱えるデータ数にも制限がある。(現在は約 1000 位だが、今後変更する可能性がある)。

今後、インターネットの普及とともに、この Web Decomp のようなサイトも増えてくると予想される。その際、できれば、あるサイトで計算した結果を別のサイトでも利用できる仕組みが確立されれば、さらに使いやすくなると思われる。そのためにはフォーム形式やデータフォーマットの共通化が必要である。

4. 解析手法の一覧

ここでとりあげた統計手法のほとんどは統計数理研究所において開発されたものである。そ

のため、ここでの詳しい説明は省略するが、適時参考文献を参照されたい。

- decomp ——— 状態空間モデルを使った季節調整法 [北川 (1986)]
- plot ——— データの時系列プロット
- autocor ——— 自己相関係数をプロット [赤池・中川 (1972)]
- spectrum ——— 自己共分散から求めたスペクトラム [赤池・中川 (1972)]
- ARfit ——— ユール・ウォーカー法による AR モデルのあてはめ (AIC 最小の次数を自動設定) [赤池・中川 (1972)]

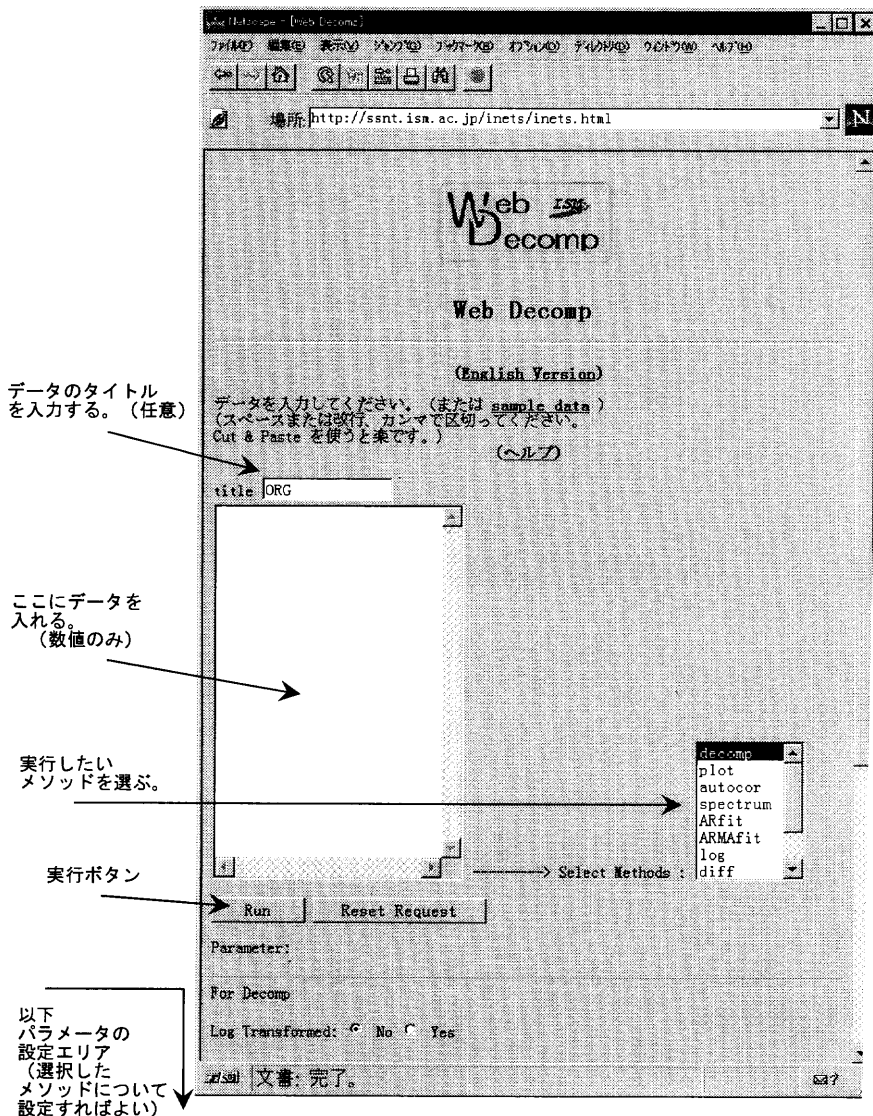


図2. 最初のページ。

- ARMAfit —— 状態空間モデルに基づく ARMA モデルのあてはめ (AR 及び MA の次数は手動設定) [北川 (1993)]
- log —— データの対数変換
- diff —— データの 1 階差分
- diff 4 —— データの季節差分 (四半期データ用)
- diff 12 —— データの季節差分 (月次データ用)
- VARest —— 分散変動型データの等分散化 [北川 (1993)]

今後も他の手法が追加される予定である。

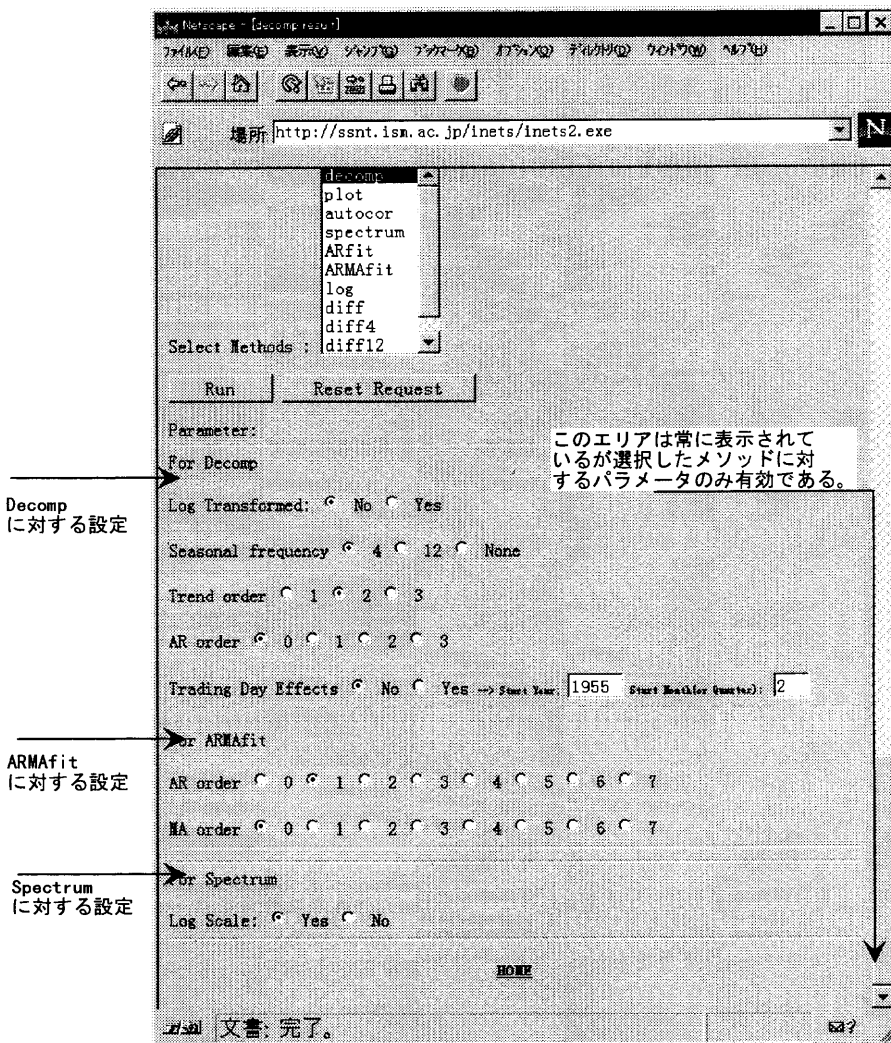


図 3. パラメータ設定エリア (各ページ共通)。

5. 解析の手順

A) データの入力

- ・今のところ1変量データのみ扱うことができる。手で入力する場合はデータ入力フィールドに数値のみを入力する。(全角のスペースを含めてはならない。) データとデータの区切りは空白またはカンマ, 改行に限られている。データ数の多い場合はマウスによる Copy & Paste の利用をお勧めする。あらかじめ, エクセル等でデータを表示させておき, マウスで「範囲指定」-「コピー」-ブラウザのデータフィールドで「ペースト」, をすれば完了である。(図2)

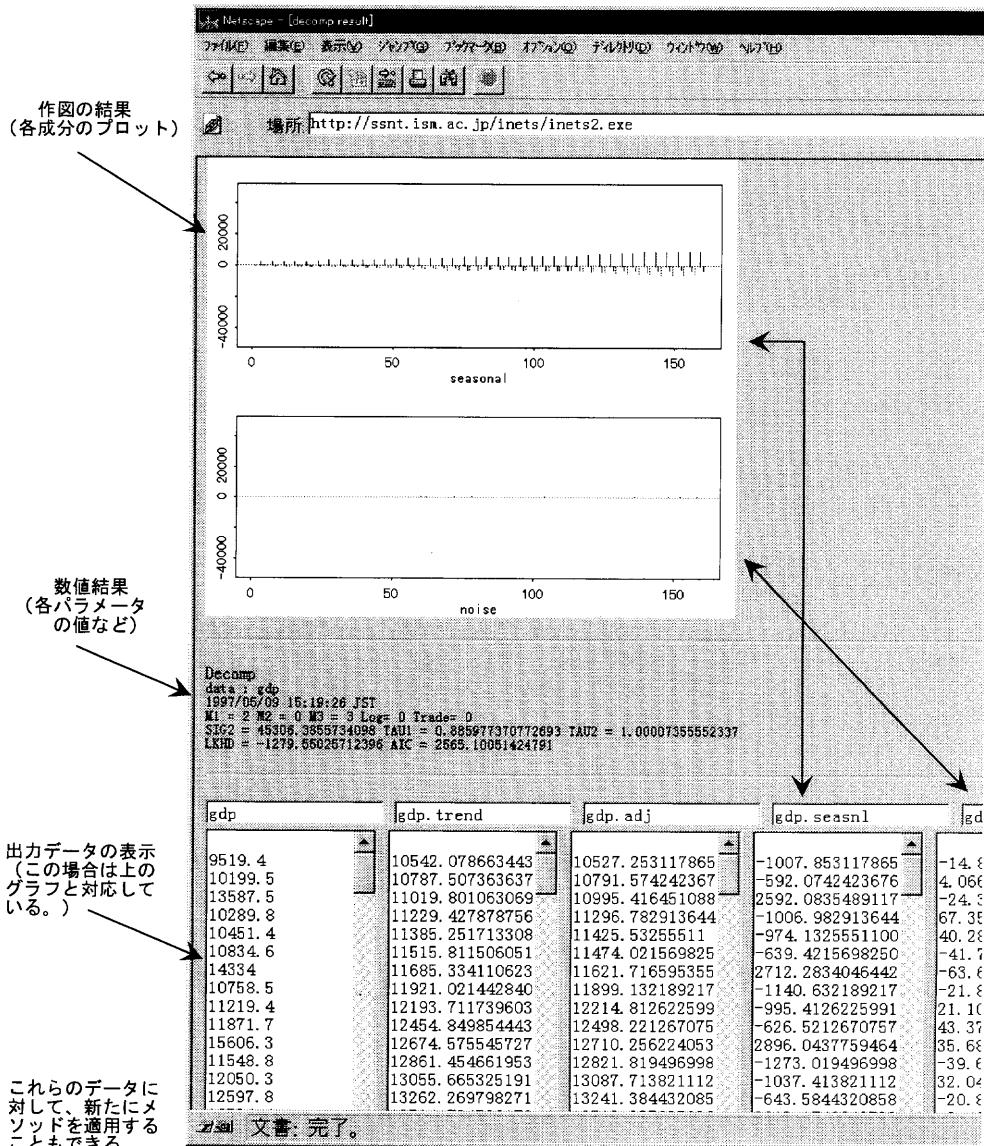


図4. 結果のページ (Decomp の例)。

- 任意であるが、テキストフィールドにはデータ名を入力することができる。ただし、特殊文字を含むことはできない。(デフォルトでは“ORG”になっている。) (図2)
- Sample Data のページまたは解析結果のページでは単に次の解析の対象となるデータを選ぶだけである。(データ名の編集はできない。) (図5)

B) 解析の実行

- 選んだデータに適用したい手法を“Select Method”の中から選ぶ。(図2)
- 必要であればパラメータの設定をする。(パラメータの設定の必要がない手法もある。) (図3)
- “Run”を押して実行する。(手法によっては数分時間がかかるものもある。) (図2)

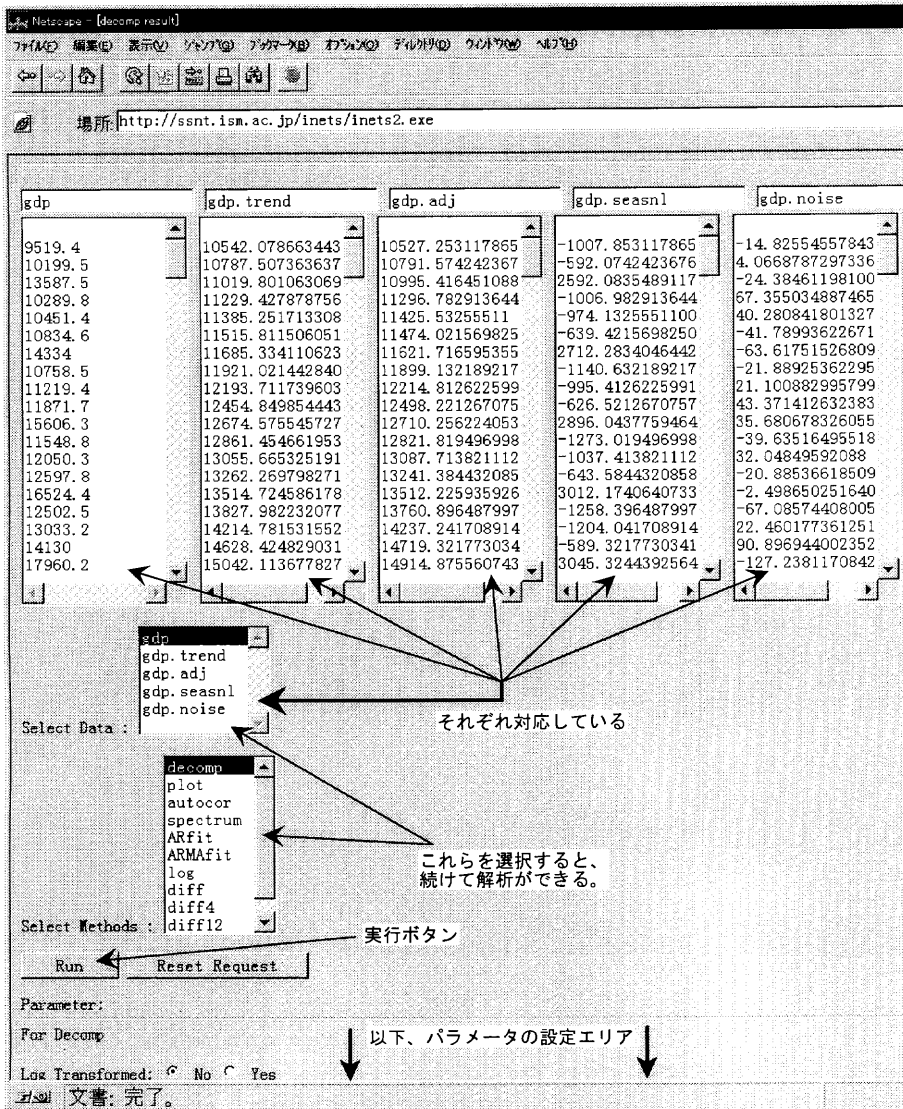


図5. 結果のページ (続き).

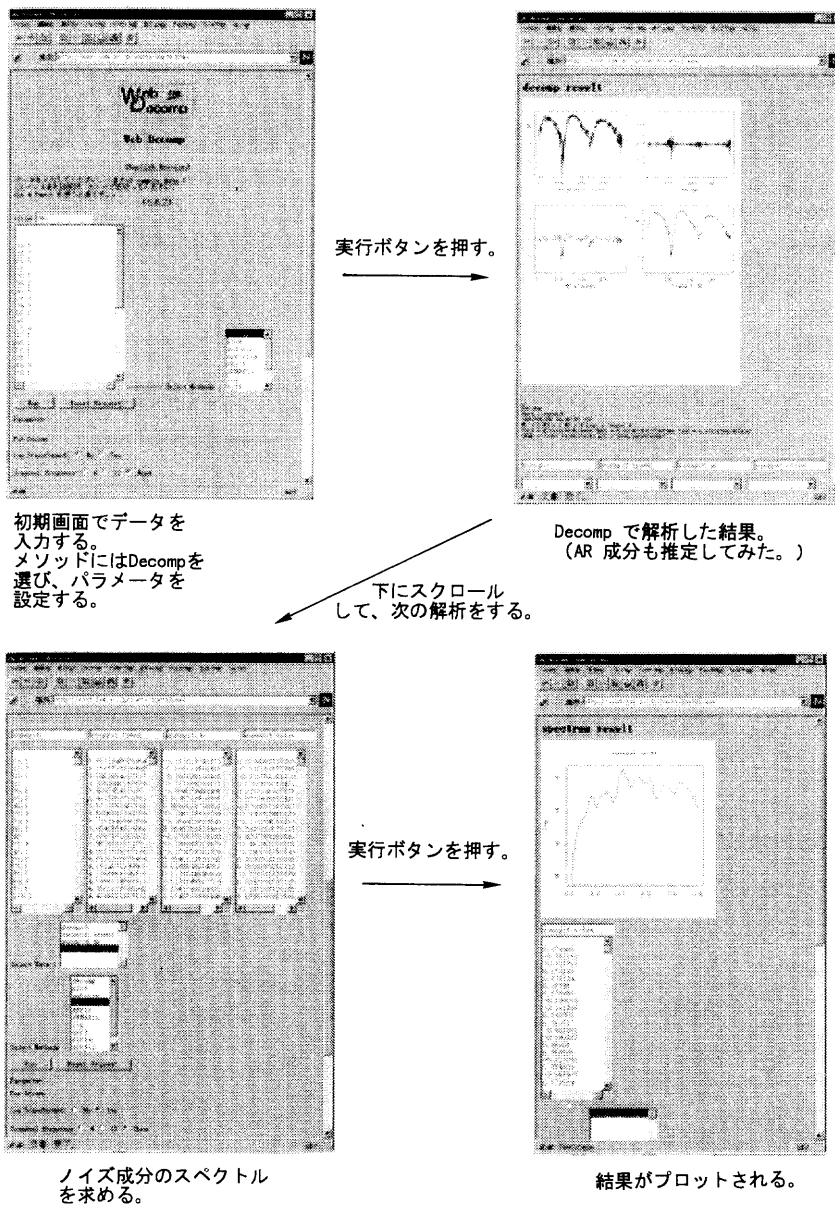


図6. 操作の一例。

C) 結果の表示

- ・正常に実行されると自動的に結果が表示される。
- ・結果のページにはグラフと数値結果が表示される。またデータを加工した場合は、その結果が下記の解析に利用できるようにデータフィールドに挿入される。(図4)。ここに表示されたデータはやはり、マウスを使った Copy & Paste により他のアプリケーションに持っていくことも可能である。
- ・結果に続いて次の解析用のフォームがついてくるので、すぐ次の分析が始められる。(パラ

メータ等は前のものが保存されている。) (図5)

図6に一連の操作を例示しておいたのので、参考にしてほしい。解析結果を使って、すぐ次の分析・診断が始められるため、データに適した解析をしているかどうかを簡単にチェックできる。

6. Decomp について

ここでの中心的手法である Decomp についてやや詳しく説明しておく。Decomp は北川源四郎氏によって開発されたもので、以下の状態空間モデルにもとづいて、時系列を Trend 成分、季節成分、AR 成分、Noise 成分に分解する手法である (北川 (1986))。

$$\begin{aligned} \Delta^d T(t) &= e_1(t) \\ S(t) &= -S(t-1) - \dots - S(t-p) + e_2(t) \\ A(t) &= a_1 A(t-1) + \dots + a_q A(t-q) + e_3(t) \\ y(t) &= T(t) + S(t) + A(t) + TD(t) + e_4(t) \end{aligned}$$

T : トレンド成分, S : 季節変動成分, A : AR 成分 (短期循環変動),

TD : 曜日効果, y : 観測値, e_1, e_2, e_3, e_4 : 互いに独立な正規ノイズ

パラメータ:

- Log transform: データに対してあらかじめ log 変換を行うかどうか。(0 または負のデータがあるとき, Yes をチェックするとエラーになる。)
- Seasonal frequency: 季節性の周期。月次データの時は 12, 4 半期データの時は 4, 季節性を仮定しないときは None を選ぶ。
- Trend Order: 上式において d に当たる値。大きいほどなめらかなトレンドが推定される。
- AR Order: AR 成分の次数。0 ならば AR 成分を推定しない。
- Trading Day Effects: 曜日効果の調整を行うかどうか。Yes なら横にデータの開始時期を入力する。

出力結果:

- グラフ出力: 各成分の時系列プロットが表示される。スケールはすべて同じである。
- データ出力: 元のデータと各成分の値及び季節調整値がセットされる。これらは次の解析に利用できる。(log 変換した場合でも季節調整値のみは逆変換した値になる。)
- パラメータ及び統計量: ノイズの分散や AIC の値が表示される。

SIG2: e_4 の分散

TAU1: e_1 の分散 (SIG2 にたいする割合で表示)

TAU2: e_2 または (AR の次数が 1 以上のとき) e_3 の分散 (SIG2 にたいする割合)

TAU3: e_2 の分散 (SIG2 にたいする割合)

7. 注 意 点

最後に幾つかの注意点を述べたい。

- Web Decomp を使うことにより、直接・間接を問わず損失や損害が生じても著者及び統計数理研究所はいかなる責任も負わない。
- バグのないように最善を尽くしているが、万が一、おかしい所があった場合は、著者までご連絡願いたい。

- ・このサイトへのリンクを張る場合は著者までご一報を願う。
- ・サービスの形態についてはこちらの事情により予告なく変更される可能性がある。
- ・このシステムの利用は、学術研究の目的のみに限定されている。

謝 辞

統計数理研究所・北川源四郎教授には多方面にわたるご教示をいただきました。また、多数の方がテストに協力して下さり、貴重なご意見を述べて下さいました。あわせて、感謝いたします。このシステムの開発にあたっては科学研究費補助金(基盤研究(A)(2))「時系列解析ソフトウェアの組織化の研究」研究代表者 北川源四郎 課題番号 08558021)の援助を受けています。レフェリーの方には貴重なコメントをいただきました。この場を借りて、感謝の意を表します。

参 考 文 献

- 赤池弘次, 中川東一郎 (1972). 『ダイナミックシステムの統計的解析と制御』, サイエンス社, 東京.
- 北川源四郎 (1986). 時系列の分解——プログラム DECOMP の紹介——, 統計数理, **34**, 255-271.
- 北川源四郎 (1993). 『時系列解析プログラミング』, 岩波書店, 東京.
- Kitagawa, G. and Gersh, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *J. Amer. Statist. Assoc.*, **79** (386), 378-389.

Introduction to “Web Decomp” —Seasonal Adjustment System on WWW—

Seisho Sato

(The Institute of Statistical Mathematics)

In this note, we describe how to use “Web decomp” System. Main aim of this system is to do a seasonal adjustment for user own’s data. An important feature of “Web Decomp” is that this system is a WWW site of the Internet and all computation can be done by server machine, not by users’ computers. The URL of “Web Decomp” is

<http://www.ism.ac.jp/~sato/>

or

http://ssnt.ism.ac.jp/inets/inets_eng.html

“Decomp” which is the main program of this system is a popular seasonal adjustment method by using a state space model (Kitagawa and Gersh (1984)).