

図3 6 . 3母数逆ガウス型分布へのあてはめの入力画面操作の図

結果として表示されるグラフは、

- (1) 発症日のヒストグラム（縦軸は全度数で割っている）
- (2) それに対応した推定した発症パターンの曲線
- (3) （原点を推定した後で）原点を移動することによって2母数の逆ガウス型分布に従う場合の確率紙を構成し全体のデータを確率紙に打点したグラフ
- (4) データを標本平均と比較して大きいグループと小さいグループに分けて確率紙上に打点しグラフ

である。

また、出力される推定値としては、各未知母数の推定値と、未知原点の推定値が表示される。図3 6、図3. 7、図3. 8は以上の出力例を示している。

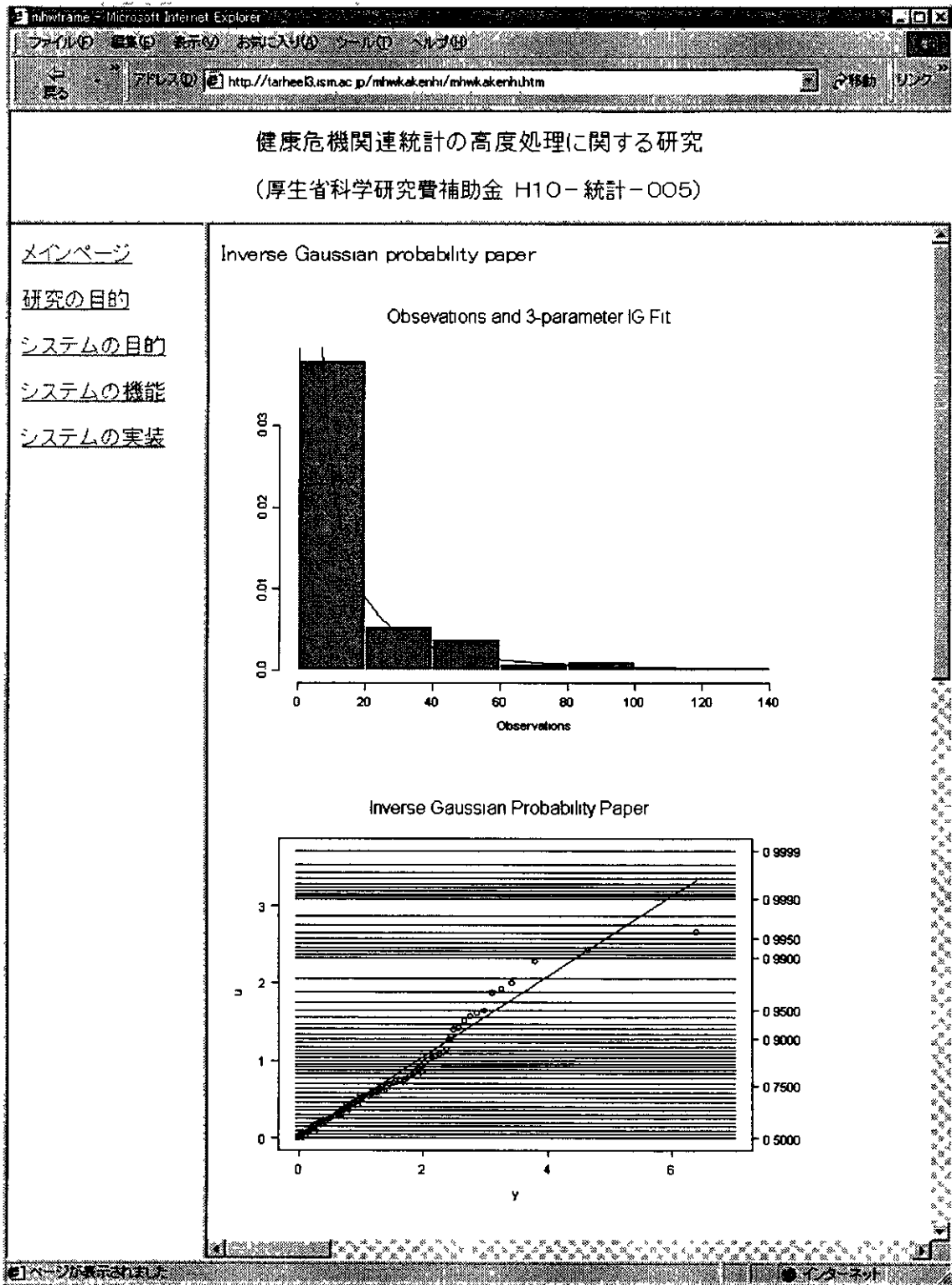
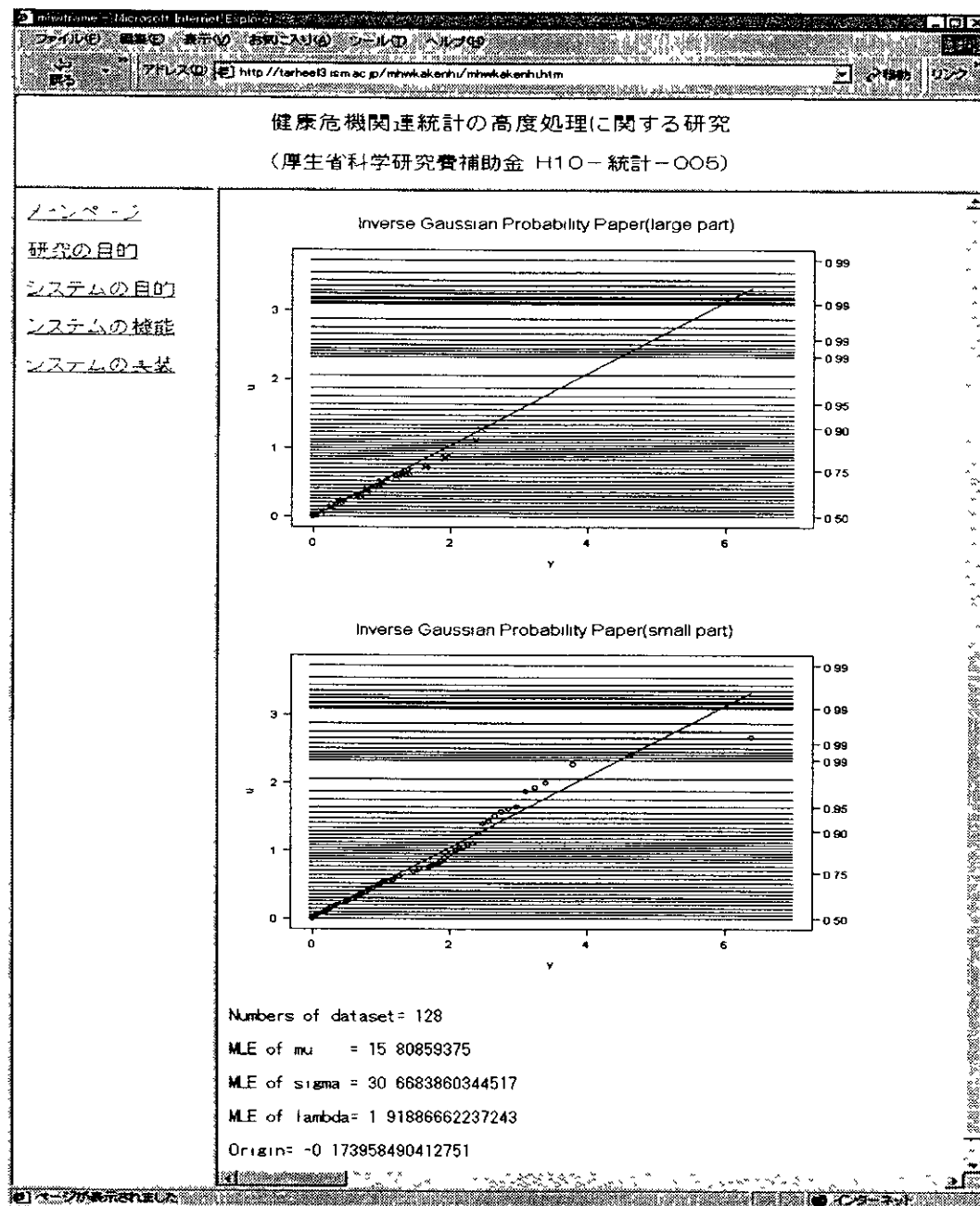


図 3. 7 : 3 母数逆ガウス型分布へのあてはめと  
2 母数逆ガウス型分布確率紙へのデータのプロット



図・3 8 : 3母数逆ガウス型分布へのあてはめ結果

図3 9は、散在的集団発生が2ヶ所で起こっていると仮定して、2つの集団での発症日のデータファイル(一つのファイルに結合したもの)を入力したところである。図3 10はテストデータとして素性がわかっている(2母数逆ガウス型分布に従う)2組のデータをまとめて一つのファイルに結合したファイルを作成し、2組における母平均値、母変動係数の推定値と、3つの未知母数の仮説に対するp値を表示している。

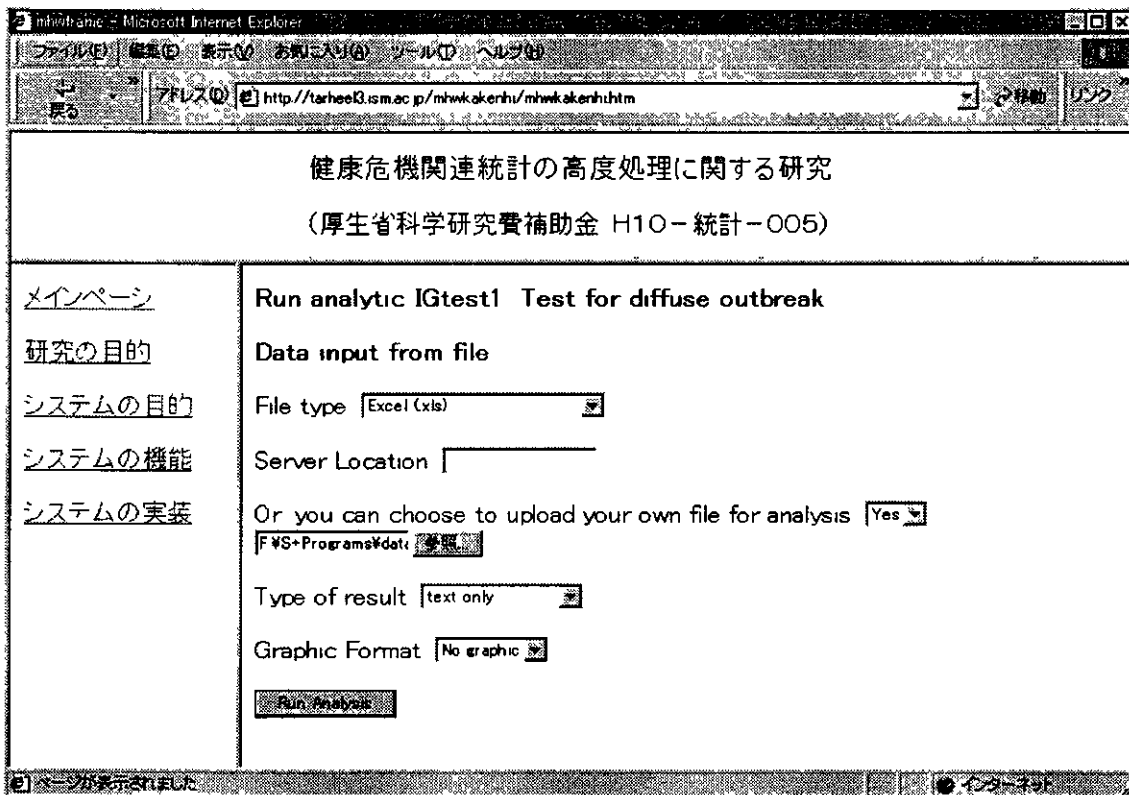


図 3. 9 : 入力例

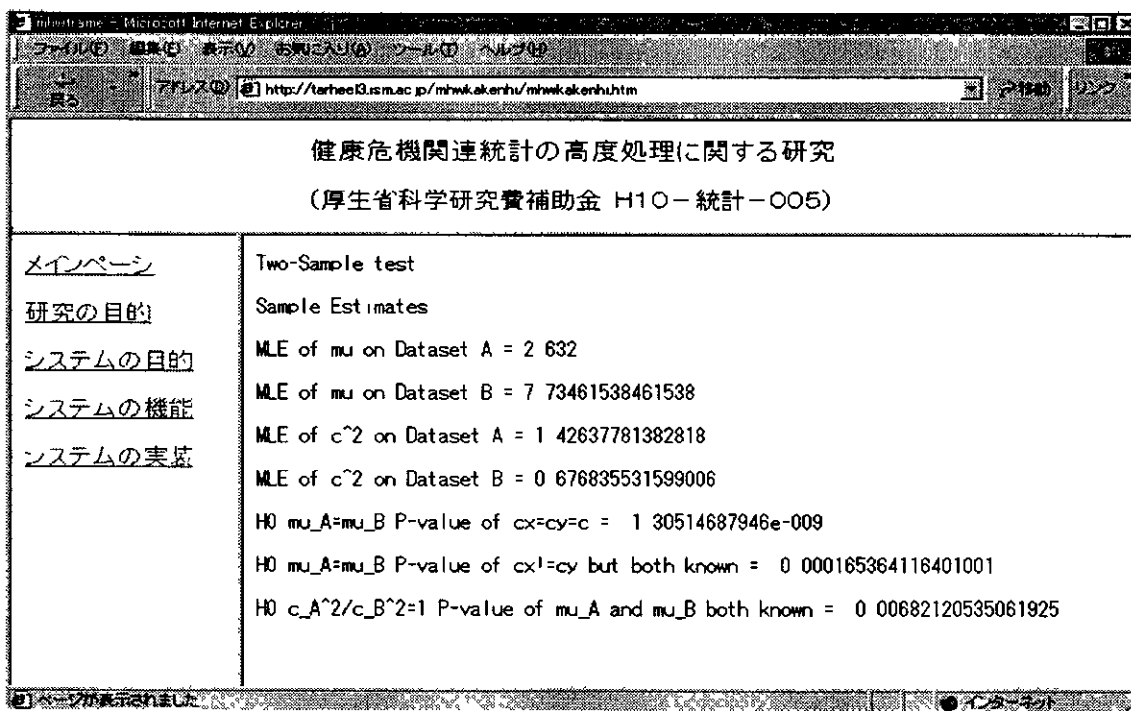


図 3. 10 : 2母数逆ガウス型分布の母数に関する検定の結果の表示例

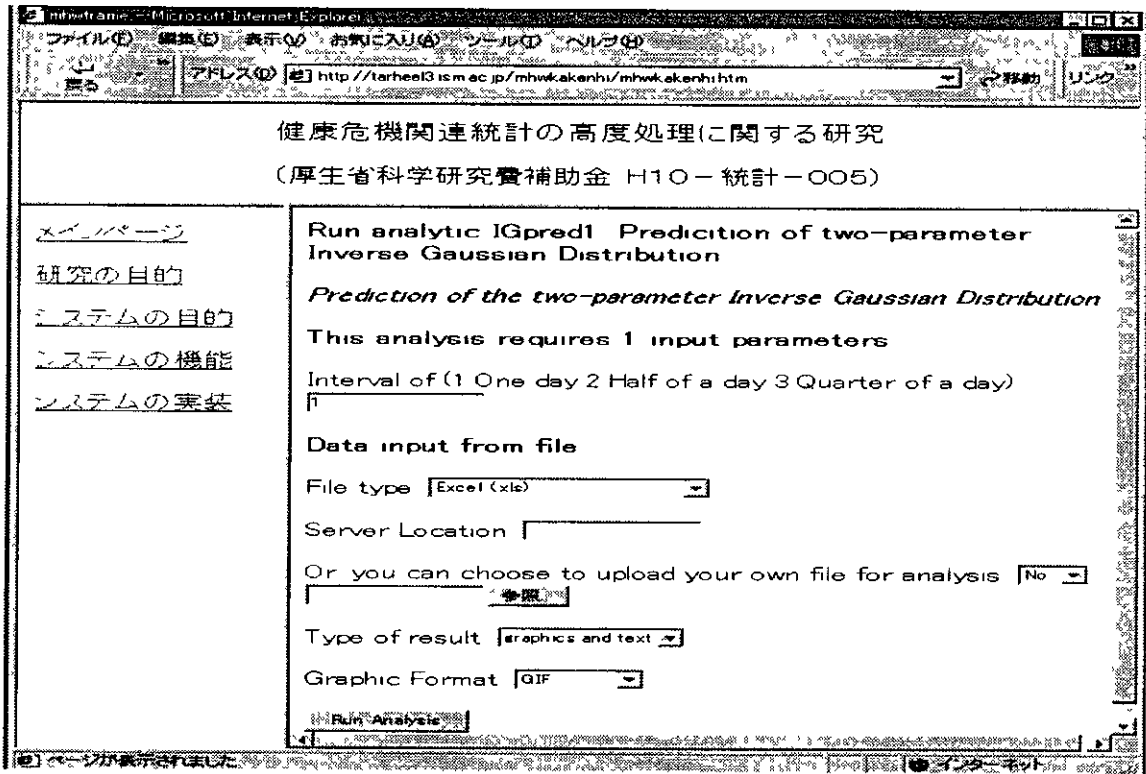


図 3. 1 1 2 母数逆ガウス型分布を仮定した予測プログラムの画面

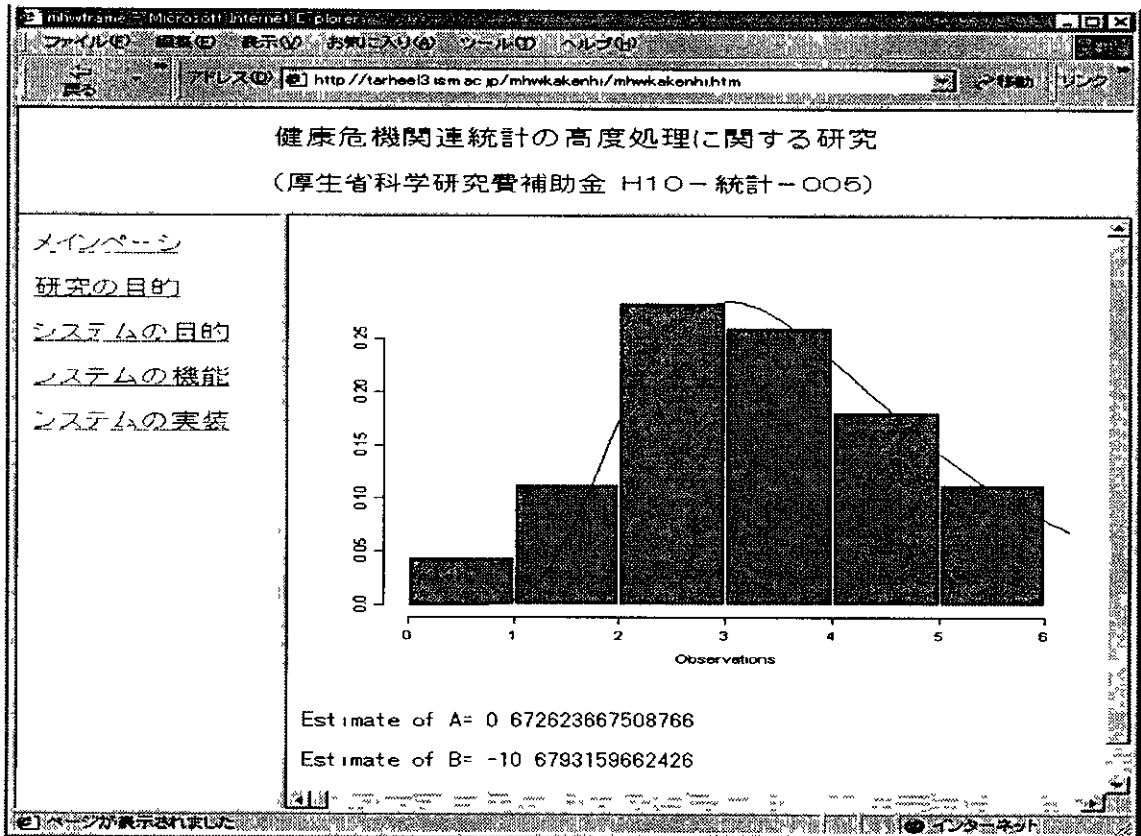


図 3. 1 2 サンプルデータでの予測結果

図3 11は、発生規模の予測を行うプログラムの実行例である。ここでは、食中毒等が発生した場合、発症日に関する度数分布表が作られることを前提としている。ここでの目的は、観測が行われた日までの度数分布表からそれ以降の発生規模を推定することである。図3 12は、サンプルデータとして初期発生日から6日間のデータの観測に基づいて度数分布が作られた場合、7日目以降の発生状況を予測（推定）した例である。ここでは、既に観測されているヒストグラムと予測（推定）した発生強度曲線とその元になる曲線の母数の推定値が表示される。

## 5 システムの評価

本研究の目的の一つに、統計的解析手法の構築のみならず、得られた統計解析手法の実装（イントラネット上での公開を念頭に置いた）に関する基礎的研究があげられる。プログラム実装を単体のコンピュータのみならずネットワーク上での利用を考えた場合、当初の予想として全体のシステム規模が大きくなると考えた。しかし、現状の計算機環境では、それほど規模を必要としないことがわかった。

また、ここで実験したシステムは、イントラネットに限らずインターネット上での運用も可能である。例えば、国立感染症研究所に1台の解析サーバを設置し、地方の保健所等が手元にある発症者の情報を用いて解析することも現状のネットワーク環境でも可能であると思われる。

実験したシステムの実行速度に関しては、クライアントからファイルを入力し、それを解析サーバで処理した場合においても解析サーバ上のファイルを指定した解析を行った場合でもそれほど差は無い。実験では、利用した StatServer での実行時間が全体の計算時間を支配している状況であった。

ファイル形式に関して、一つのメーカ（Microsoft 社の Excel）のファイルにした理由は、クライアントが Windows や Macintosh であっても現時点ではファイルの互換性がある点とテキストファイルの場合、クライアントの違いによって入力ファイルに差が生じる。現時点では、クライアント側はパーソナルコンピュータを考えているが、それは本質的な制約ではない。

## 6 まとめ

本研究で、ネットワーク（イントラネット）上での統計解析サーバの有効性が認識された。しかし、その利用方法に関しては注意が必要である。本研究で提案されたシステムにおいては、利用者を制限するべきであると考えている。当然ではあるが、システム

はどのようなデータに対しても解を求める。本研究では、データの素性についてよく理解した研究者が解析結果についてもその妥当性を判断できた状況で初めて求めた解の利用を考えるべきである。そういった意味でも解析結果についての妥当性が検討できる機関での利用が必要である。

計算機を取り巻く環境は、日進月歩である。当然、システムの実行時間は問題とならない状況である。このような状況ではより利用しやすい（誤用をなくす・解析の仮定等のチェック・表示結果の解釈に誤りをなくす）システムが必要である。本研究は、そのような点についても考慮する基礎的システムの構築を一部試みている。

研究課題全体のシステムとしては、第4章で述べる可視化システムと第3章で述べた統計解析サーバシステムを合わせたものである。最後に、このようなシステムが有効に機能するためには、やはり迅速なデータの収集が必要である。

## 4章 データの視覚化に関する検討

### 1 はじめに

本研究は健康危機関連統計情報について、その情報を分析し適切に状況を把握することを支援するためのシステムの構築に関する検討とその提案ならびにその問題点を把握することを目的とする。厚生省では健康情報の一つとしてO157などの食中毒に関する情報をホームページ上で公開していた。ここではこのようなデータが入手可能である場合にどのような形態でデータを把握し、それをどのような形で分析し情報の把握を行うことが可能であるかについて検討を加える。特に本章ではこれらのデータの視覚的表示の可能性について考える。

O157他食中毒等のデータでは、それら疾患の同定と発生経緯・原因の確認、さらにその予防などが問題となる。したがって分析の観点をどこに置くかによって、そのデータとしての扱い方が変わってくるが、その時間情報、地理情報は非常に重要な情報といえる。前章の統計的なデータ解析に関する検討は、その時間情報に関する分析を主眼に置いたものであったが、それら解析対象となる対象者の選定やあるいは発生状況の確認のために、まず探索的な分析が必要とされる。探索的な分析では、データを視覚化表現し、解析者の視覚的な情報処理能力を利用することが有効であることが知られている(Tukey(1977), Chambers, et al. (1983))。また本研究の対象となるような地理的な情報を含むデータの分析では、地理的特性やその地理的関連性を把握する上でその情報を視覚的に表現することが不可欠な要素となる。

ここではこのような地理的情報を含んだデータをグラフィカル表示することにより、経時変化・地域性などの特性を考慮に入れてこうした疾患の発生状況を分析し、その情報を提供することを目的とするシステムの構想について議論する。

本研究ではその視覚化について2つの観点から検討を加えた。その1つは動的なグラフィカル表示を用いてデータの視覚化であり、もう1つは地理情報システムを用いたデータ表現である。前者についてはグラフィカルなデータの解析について動的な観点をその表現形式に加えることにより効果的な分析結果を得ることを目的とするものであり



(Becker, Cleveland and Wilks(1987), Cleveland and MacGill(1988), etc), 後者については特に地理的な情報の処理に特化したシステム(シオマチックス研究会(1999))に関する検討である。今回はこれらのシステムを用いた際のデータ表現の可能性について検討し、さらにこれらの情報あるいはシステムの利用形態に関する問題点についても考察を加えた。

## 2 基礎データとしてのO157 データ

健康危機関連情報では従来型の単なる要約統計の公開ばかりではなく、即時的で多角的な情報の提供が求められる。たとえば厚生省ではO157・食中毒関連情報について“腸管出血性大腸菌O157による食中毒等の発生状況”を緊急情報・速報として平成8年度からのデータを平成11年7月までの間逐次インターネット上のホームページで公開していた。図4-1は平成11年7月に公開されていた厚生省のホームページのトップページである。

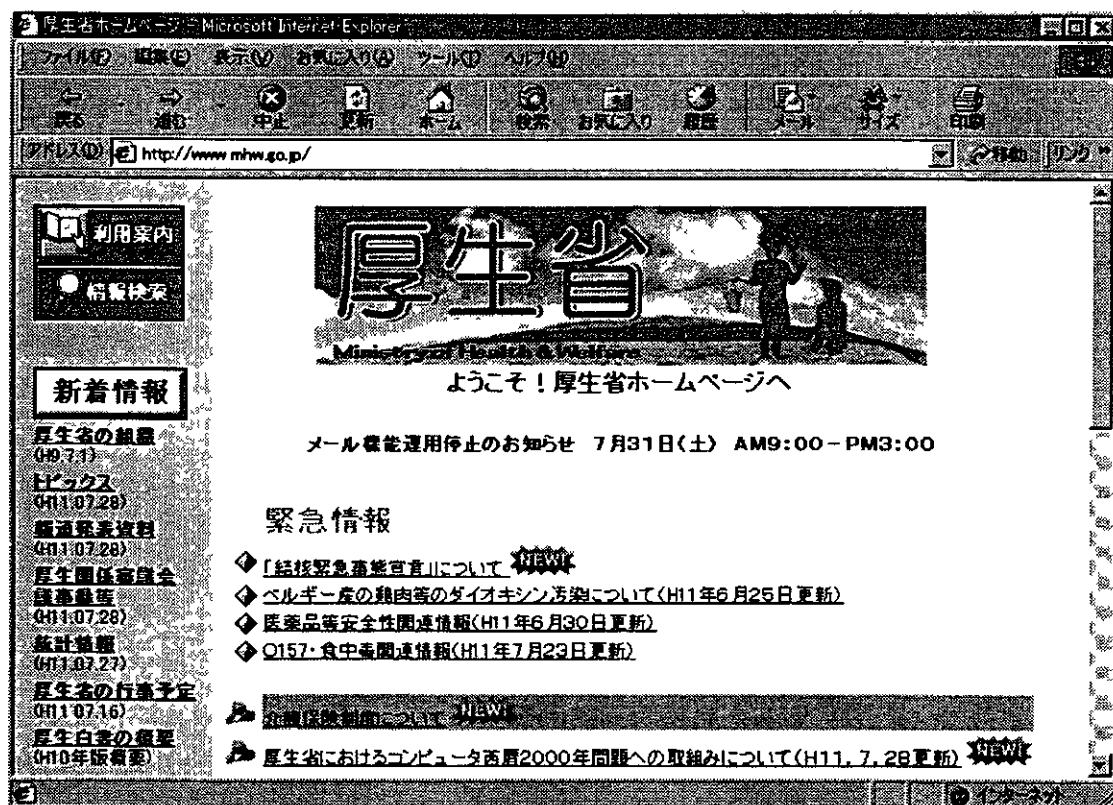


図4-1 厚生省のホームページ1999年7月29日

その冒頭の個所で、「O157・食中毒関連情報」を緊急情報として提供していたことが分かる。年によって公開されていたデータの情報について多少の違いはあるものの、

このときに公開されていたO157 関連情報は基本的には図4.2のような形式であり、都道府県におかれた保健所からの情報をもとに、

都道府県名	政令市名	郡名	市区町村名
届け出保健所	保健所探知日	患者発生日	有症者数
無症者数	入院者数	死亡者数年齢	性別
発症日	病原菌	検出日	原因食品
HUS	備考		

などの情報が公開され、そのデータがホームページからダウンロード可能であった。

The screenshot shows a web browser window with the URL [http://www.nih.go.jp/o-157/h1to157/kan\\_13.html](http://www.nih.go.jp/o-157/h1to157/kan_13.html). The page title is "都道府県発生状況詳細" (Detailed Occurrence Status by Prefecture). The table contains the following data:

都道府県名	政令市名	郡名	市区町村名	届け出保健所	保健所探知日	患者発生日	有症者数	無症者数	入院者数	死亡者数	年齢	性別	発症日	病原菌	検出日	原因食品等	HUS	備考
01北海道				帯広保健所	1月11日	1月9日	1	0	1	0	1	男	1月4日	O157 (VT)	1月9日			
01北海道				帯広保健所	1月14日	1月14日	0	1	0	0	34	女		O157 (VT)	1月14日			1月11日探知の1歳男児の母
06山形県			酒田市	酒田保健所	2月6日	2月6日	1	0	1	0	57	男	2月9日	O157 (VT)	2月5日			
08茨城県				竜ヶ崎保健所	3月1日	3月1日	1	0	0	0	0		2月23日	O157 (VT1,2)	3月1日			小学生
09栃木県			宇都宮市	宇都宮市保健所	1月14日	1月14日	1	0	1	1	2	男	1月5日	O157 (VT1,2)	1月14日			5日発症、7日入院、11日死亡、14日ベロ毒素確認
10群馬県				吾妻保健所	2月22日	2月22日	1	1	1	0	0		2月14日	O157 (VT1)	2月22日			有症者 21歳女性、無症者 16歳女性 (妹、26日書検出)
11埼玉県			蕨市		1月21日	1月21日	0	1	0	0	20	男		O157 (VT1,2)	1月21日			
11埼玉県			新座市		2月27日	2月27日	1	0	0	0	34	女	2月21日	O157 (VT2)	2月27日			
12千葉県			鎌ヶ谷市	給養保健所	1月8日	1月8日	1	0	0	0	27	男	1月4日	O157 (VT)	1月8日			
12千葉県			浦安市	市川保健所	2月4日	2月4日	1	0	1	0	4	女	1月30日	O157 (VT)	2月4日			
12千葉県			松戸市	松戸保健所	2月5日	2月5日	0	1	0	0	39	女		O157 (VT)	2月5日			

図4.2 上記ホームページで公開されていたO157データ

この研究では、このホームページから入手可能であった、平成9年、平成10年、平成11年のデータを用いてデータのグラフィカルな表示に関する検討を行うことにした。平成8年のデータについてもホームページで公開されていたが、これら公開されたO157のデータについては平成8年の岡山県新見市や邑久町、大阪府堺市の大発生事件以降のデータが多く、静岡、香川で4月、5月のデータの記載があったものの、ほとんどの県で6月以降のデータしか記載が無く、また保健所探知日のフィールドも無かったため、今回は対象データとしなかった。また平成11年のデータについては7月までのデータとなっている。

## 2 1 位置情報について

上記の〇157に関するデータを視覚化するにあたり、まず対象イベントの生起場所に関する情報を決定する必要がある。ここではこれらのデータが基本的に保健所単位の情報として構成されていることから、基本情報としては保健所の緯度、経度情報をその位置情報として用いることにした。

そのために、まず厚生省から平成11年2月現在での保健所の登録住所情報を入手し、その緯度、経度を(株)センリンの電子地図ソフトZを用いて確認した。その際このソフトのデータベースから電話番号などで検索できる保健所についてはそのデータベース検索により、またデータベースなどで検索できない保健所については住所から検索し該当する住所の近隣の緯度、経度情報を用いる事にした。付録表4-1はこの作業の結果得られた各保健所の緯度、経度情報である。

さらにこれらの保健所の位置情報を得られた〇157データとマッチングを行い保健所データでマッチできるものについてはマッチさせた。その後保健所に記載があるものの、用意した保健所位置情報ファイルでマッチできなかった保健所については

- 1) 〇157データの方で記載のある保健所名について簡単に保健所名が類推できる場合あるいは支所の記載のある場合には該当保健所に適合させ、
- 2) 保健所位置情報ファイルとの対応がとれないものについてはその都道府県、市、区、郡、町村データの位置情報とマッチさせるものとした。

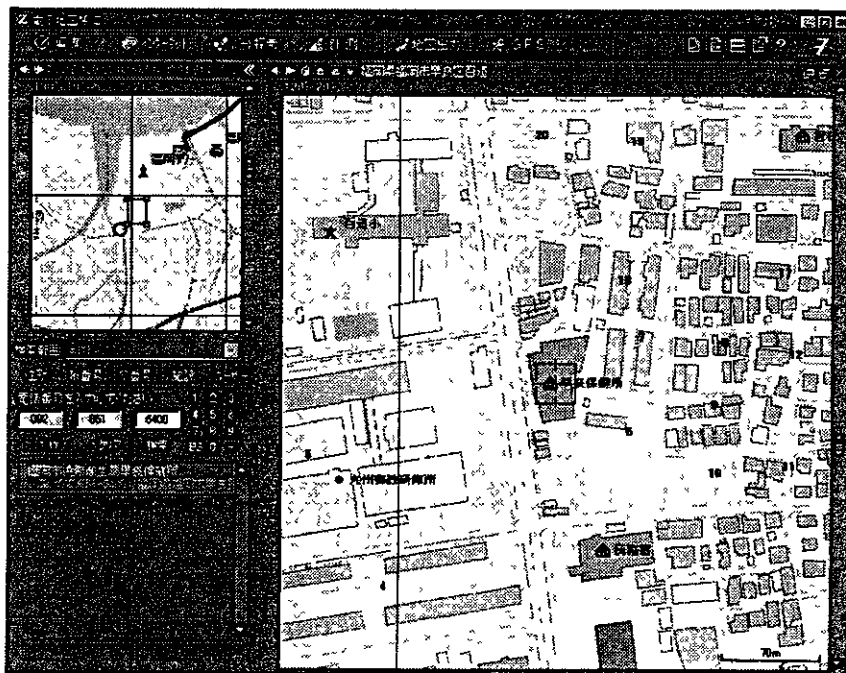


図4 3 福岡市早良保健所を電話番号(092-851-6400)から検索

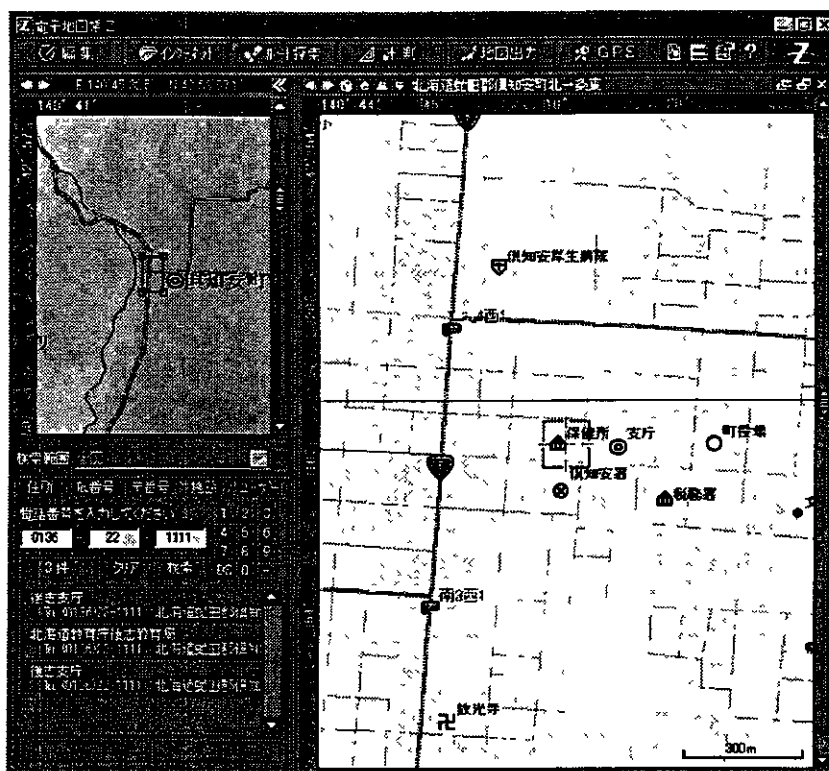


図4. 4 俱知安保健所の電話番号(0136-22-1111)で検索。電話番号データベースには明記されていないが地図上にマークがある場合はその緯度、経度を採用。

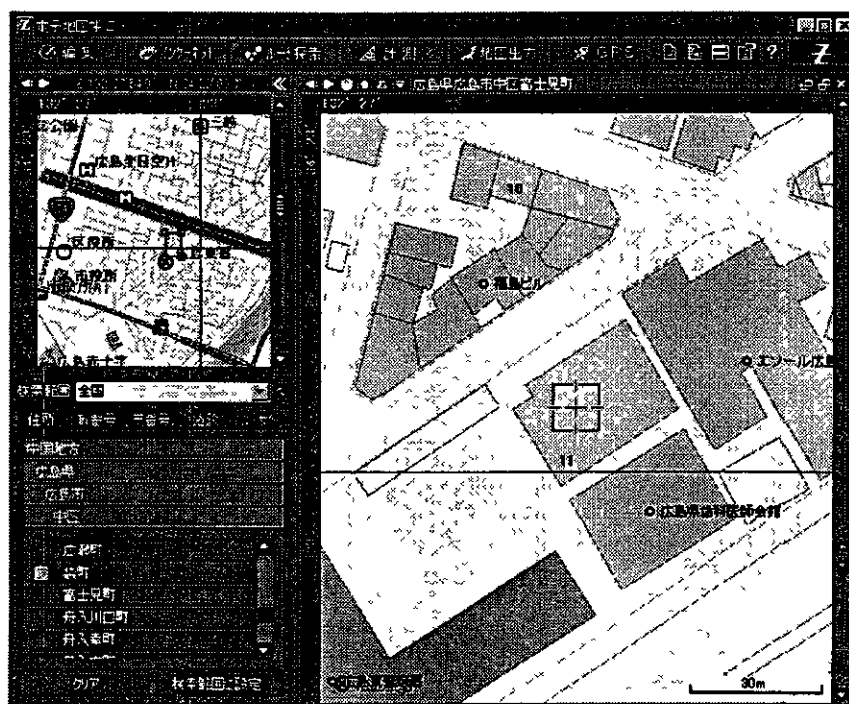


図4 5 広島市保健所を住所(広島市中区富士見町 11-27)から検索

今回のデータセットでは速報という意味合いもあり、保健所に記載のないデータも数多く見うけられた。平成9年のデータで約30%、平成10年のデータで約36.8%、平成11年のデータで約34.4%のデータについて保健所に関する記載がなかった。これらのデータについては、市町村の情報が取れないものもあり、都道府県、市、区、郡、町村の記載のうち詳細な情報を選びその地区の代表点を選択するものとした。

表4 1 厚生省のホームページから入手したデータにおける位置情報等

	平成9年	平成10年	平成11年
全データ件数	1513	1727	488
保健所名記載あり	1056	1091	320
検索可能	799	1022	235
検索不可能 *1	257	69	85
保健所名未記入	457	636	168
(住所) ArcView 検索可能	132	232	63
(住所) ArcView 検索不可能	325	404	105
*1のうち 保健所マスターに無し	34	5	5
入力不整合その他 *2	223	64	80

平成9年から11年のデータで、上記\*2（保健所名は入力されているが、入力データとの不整合等で検索不可能であったもの）を調査したところ、以下のような問題が見られた。

#### <データ不整合>

- 1 大文字と小文字の違い
- 2 ○○市、市立○○などの「市」や「市立」等統一されていない。  
(入力されていたりされていなかったりしている)
- 3 保健所名に市名が入力されている
- 4 漢字の入力ミス  
  - 例) 四条畷(マスク) → 四条綴(データ)
  - 鷹栖(マスク) → 鷹巢(データ)
  - 刈谷(マスク) → 荻谷(データ)
  - 観音寺(マスク) → 観音時(データ)
- 5 「名古屋市」「京都市」「大阪市」「堺市」「福岡市」「東大阪市」等の比較的大きな市のデータが簡略して入力されている  
  - 例) 京都市上京(マスク) → 上京(データ)
  - 福岡市博多(マスク) → 博多(データ)
  - 堺市堺(マスク) → 堺(データ)

### ＜保健所マスタでの問題点＞

- 1 「保健所」以外に「健康福祉センター」「保健福祉センター」「保健総合センター」「衛生環境研究センター」「保健福祉事務所」等のデータが登録されていない
- 2 「支所」のデータが登録されていない

これらの都道府県，市，区，郡，町村データのマッチングに関しては，まず後述の ArcView の市町村位置情報マッチング機能(アドレスマッチング/ジオコーディング機能)を用いることにより緯度，経度情報の入手を試みた。この機能を用いることにより，東京都23区，横浜市，川崎市（番地レベルの代表点），および全国（市区町村レベルの代表点）までマッチングが可能である。ただし，今回のデータについては残念ながら東京都については詳細な地理情報は得られていない。

さらにこのマッチングでもマッチできなかった情報のうち，まったく郡市区町村レベルの情報が無いもの，あるいは郡情報はあるもののその詳細情報が無いためにマッチができなかったものなどについては，ゼンリン電子地図ソフトZを用いて，前者については都道府県代表点，後者については郡としてデータベースの最初に得られる地区の代表点を採用することとし最終的に公開されていたO157データの全レコードについて緯度，経度情報を付加したデータファイルを構築した。ちなみに上に注意したように，トータルとしての発生数の最も多い東京都についてはその詳細情報はほとんど記載されておらず，今回のデータに関しては東京は1点として表現されてしまうことに注意が必要である。

したがって今回の基礎データについてはデータの性格上，正確な位置情報を詳細に表示する目的で利用するには不適と言える。このため，本研究での地理的な情報表示に関しては，あくまでその表示形態の可能性と問題点に関する研究の一つとしてとらえ，その表示内容の詳細については深く言及しないものとする。ただ，後述のグラフィカル表示の結果から判断すると，日本全国レベルでのO157の発生状況の地域的な概要を把握する目的のためには，ほぼ十分な精度が得られているものと考えられる。また，このデータの地理的な情報の正確性については，データの整理の問題あるいは基礎データの入手方法の改善などにより，実務システムとしては十分に改善可能な問題と考えられる。

## 2 2 時間情報について

O157データの発生状況をとらえる上でもう一つの重要な変数は時間情報である。公開されているデータでは，平成9年から11年までともに用いることができるデータとしては，保健所探知日，患者発生日，発症日，検出日などがある。データの分析の上で

は研究目的によってその情報の使われ方が異なるか、ほとんどのレコードについて記載のある保健所探知日を時間情報として視覚化に用いることにした。患者発生日も記載されている場合が多いが、検出日には欠測も少なくない。また、疾患の性質からO157感染はあるものの症状の無いケースも多く、この場合、当然発症日は存在しない。

### 3 視覚化について

2節で述べた基礎データをもとに、ここではデータの視覚化について2通りの方法での視覚化について検討した。

その1つ目の方法はこれらのデータを動的なグラフィカルデータ解析の一環としてとらえた場合のデータ表現形式の可能性について検討し、動的にデータ表現する際の利点ならびに問題点、さらにその実装化に関する技術的な問題点について検討を行った。2つ目の方法としては、今回のデータの特長としてその基礎データのなかで地理的な情報が大きなウェイトを占めることから、この表現形式に特化した地理情報システムによるデータ表現とそのデータハンドリングの可能性について検討した。

いずれの場合についても、この視覚化表現について単に1人の解析者の下で利用するのではなく、データ解析の環境を複数の解析者が共有できるものとしてとらえ、さらに解析結果の公開につなげる場合の形態についても検討した。ここでの共有化の際のキーワードはインターネット、イントラネットと言ったネットワーク技術の利用である。インターネットをはじめとするネットワークを利用した情報の交換技術は今日急速な広がりを見せており、先に述べた厚生省のホームページでの健康危機管理情報の公開もこのような技術的な進展抜きでは考えられない。ただし、現状のネットワークに関する社会的なインフラストラクチャは高度な情報交換に耐えうるものといえるかどうかについては疑問が残る。データの視覚化表現の観点からこの点についての検討も行った。

また、データの視覚化はそれ自身重要な解析手順ではあるが、これはより詳細な数値的なデータの解析と並行的に進められるべきである。その際、視覚化システムと数値解析システムとを単一の不可分なシステムとして構築する立場も考えられる。しかし、今日の解析技術の急速な進展に即応した形でシステムをとらえる場合、それぞれを比較的疎な結合方式でとらえ、両者のシステムを何らかのインタフェースで連結して考える方がより柔軟なシステムが構築できる可能性がある。ここではこのようなシステムの連携の可能性についても検討する。

以下順に動的なグラフィカルデータ解析手法、地理情報システムを利用した表現、などについてそれぞれの特性について議論し、解析環境の共有化、解析システムとの連携の問題についても検討する。

#### 4 動的なグラフィカルデータ解析手法について

本節ではまず動的なグラフィカルデータ解析手法について概説する。

多量のデータを高速かつ正確に処理するためにデータ解析では計算機は不可欠な道具であり、その数値的なデータの処理能力のため統計的なデータの分析の範囲は飛躍的に広がった。一方で、CPU の高速化、メモリやハードディスクなどの記憶媒体の大容量化、CRT の高解像度化と高機能化、グラフィカルアクセラレータの進歩、通信能力の向上など計算機は目覚ましい速さで進歩している。

CPU クロック(800MHz)、メインメモリ 256MB、ハードディスク容量 30GB、1677 万色、解像度 1600 x 1200 ドット、リフレッシュレート 85Hz で動作可能な 22 インチ CRT とグラフィカルボード、最大 100Mbps の計算機間通信能力。これは現在、平均的とは言えないかもしれないが、少なくとも最高水準ではない個人ユーザをターゲットにしたパーソナルコンピュータのスペックである。今日、現実に個人が購入可能な価格でこの程度の計算機システムが市場に出回っている。

このような計算機の変化はデータ解析自身のあり方にも影響を及ぼしてきている(Buja and Tukey(1991))。人の視覚システムや認知機能を支援用し積極的にデータの分析を進めるためのグラフィカルデータ解析もその一つである。

計算機技術の進歩により、グラフや絵は紙の上に描くのではなく、CRT の上で高品位な画像として表示することが可能となった。しかも通常のグラフや絵であれば CRT の上で瞬時に描くことができる。このことにより、人と計算機とのインタフェースとしてグラフィカル・ユーザ・インタフェース(GUI)が現実的なものとなり、今日では、Unix マシン上での X Window System や PC 上での MS Windows などの Window システム、あるいは Macintosh では OS 自身が GUI 環境として提供され、標準的な作業環境として定着している。当然データ解析環境もこのような GUI ベースの作業環境のもとで提供されるようになっている。

さらに、コンピュータ・グラフィクス(CG)の世界では、すでに平面的な CRT 上の像



から仮想的な3次元空間への立体像を構成する試みが行われている。現状では十分な立体像を構成するためには専用の機器を要する場合が多いが、研究レベルではこれらの機器に触れる機会も増えてきている。

#### 4 1 データ解析とグラフィクス

データ解析で使われるクラフとしては、散布図やヒストグラム、あるいは各種の集計値の表示に用いられる棒クラフや円クラフなど種々のクラフが従来から用いられてきた。

先に述べたような環境では、図として表示することに関する時間やコストあるいは労力などの多くの問題が軽減されて、クラフがより身近なものになってきた。さらに、探索的データ解析やデータ解析における診断の必要性が議論されるようになり、データ解析におけるグラフィクスの重要性が増してきている (Tukey(1977), Chambers, et al (1988))。

グラフィカルデータ解析の環境に望まれるいくつかの条件を整理してみると、

- 1) 所望のクラフが容易に描けること、
- 2) クラフの修正変更が容易であること、
- 3) 種々のクラフが同時に参照可能であること、
- 4) クラフ間に関連性を持たせることが可能であること、
- 5) 数値的なデータ解析と連動すること、

などが挙げられる。

このことは、“ユーザの意図を反映した表示が容易にできること”を意味している。ここでさらに重要なポイントは、データ分析の際にできるだけデータ解析者の思考を妨げることが少ないように表示が可能であることである。これは別の見方をすれば、“ユーザ・クラフ(システム)間の相互情報交換が瞬時である”ことを意味しており、この意味においてデータ解析では“動的な”グラフィクスが重要であるといえる。

この観点から考えると、それが静的なクラフであっても、ユーザの意図を反映した形で多様な表示や修正が瞬時に可能であれば十分動的なものと考えることができる。また逆に、アニメーションのような動画であっても単に再生が可能であるにすぎないような場合には、ユーザの意図を反映できないという意味で、動的とは言えないことになる。

## 4 2 グラフィカルデータ解析システムの開発環境

上で述べたような GUI 環境はユーザの側からすると、直接的で扱いやすい作業環境であるが、システムやアプリケーションの開発の側にとっては作成の際に逆に大きな負担を強いられることになる。

ソフトウェアの開発に関しては、GUI をベースにしたソフトウェアの開発ばかりでなく、他の大規模システムの開発と保守が問題となってきた。このような状況のもとで、オブジェクト指向にもとづいたシステムの開発技術が基本的なアプローチとしてとらえられるようになってきた。特にグラフィカルデータ解析の開発の際には、オブジェクト指向にもとづいたシステムの設計と実装は有効である (Hurley and Oldford(1991))。

オブジェクト指向では処理の対象となる“もの”をオブジェクトと呼び、その中にデータと手続きを埋め込み、オブジェクトの階層的抽象化と階層間継承ならびに情報隠蔽などの局所化を行う。このことによってオブジェクトの部品化と再利用が容易になり、同時にシステム全体の見通しもよくなる。

グラフィカルデータ解析システムでは、その主要なオブジェクトとしてデータの何らかの統計的情報を示した“もの”を対応させることができる。こうすると、これらをユーザの意図に合わせて組合せ、意味あるかたちにまとめたオブジェクトが最終的に画面に表示されるグラフに対応することになる。このような設計では、多くの場合グラフの変更や修正は単にオブジェクトを取り替えるだけで済むことになる。

また、上記の考え方を発展させて考えるとデータのある様相を表示するものがオブジェクトである。この意味で、グラフィカルデータ解析システムではオブジェクトは単にグラフやその構成要素だけではなく一般にデータの統計的な様相をあるフォーマットで表現したものとしてとらえることができ、数値的な分析結果も一つのオブジェクトとなる。ここではこのようなオブジェクトを“view”と呼ぶことにする。当該研究者の一人(越智)の研究室でも、この方向で動的なグラフィカルデータ解析支援システムのプロトタイプの実装を進めている(柿内, 越智, 森重(1995))。

#### 4 3 動的なグラフィカルデータ解析システムの手法と機能

##### 4 3 1 動的グラフィクスの機能

4節冒頭で述べたように、静的な view であってもその構成要素を取り替えること、つまり複数の view を同時に表示したり変更したりすることによって、ユーザの必要な情報がある意味で動的に表示可能である。また一方で、静的 view にいわゆる動的手法を付加することによってその表示を強化することが考えられる。表4. 2はその代表的な動的手法を示したものである。

表4 2 グラフィクスにおける動的手法

手法	説明
Identification	個々の標本に対する認識
Brushing	複数の標本の認識
Deletion	削除
Linking	個々の標本に対しての連結
Rotation	データ構造の視覚的なイメージの回転
Parameter control	グラフィカル表示に対するパラメータ制御
Scaling	縦横比の変更
Alternation	グループ化情報の動的視覚化
Interpolation	View 間の補間

このうち Identification, Brushing, Deletion, Linking は散布図あるいは散布図行列などで用いられる手法であるが、概念的には必ずしもそれだけに限定されるわけではない。

Identification はおもに単独の標本に対してその認識を行う手法であり、図中のある特定のデータ点がどういう属性を持っているかを確認するといった方法である。実際には図中の点や対応するデータの属性を表示するテキストなどがハイライトや点滅などによって示されることが多い。場合によっては、Identification は Labeling と Locating の2つの手法に分けて考えられる事もある。

この場合 Labeling では図中のデータを何らかの方法で指定することにより、対応するデータのデータ名などの関連情報や属性を参照する方法であり、Locating では逆にデータ名などの属性から選択した特定のデータが図の中のどこに存在するかについて検査することになる。この Labeling 操作では右上の図中に属性を直接書き込む操作を指すこともある。

Brushing は同時にいくつかのデータ点を把握するために Becker ら (Beker,

Cleveland and Weil(1988)) によって考案された手法である。散布図や散布図行列の中に Brush と呼ばれる矩形を用意し、これをマウスなどのポインタ装置を用いて画面上で動かし、その矩形内の点を認識しようとするものである。通常一旦矩形に含まれたデータはそのまま矩形から外れても選択され続けたままになる lasting モード、矩形内に含まれた場合にのみ選択されたことになり外れると自動的に選択外となる transient モード、逆に選択指定を Brush を用いてはずす操作をする undo モードなどのいくつかのモードと、すべてのデータを表示しつつ選択されたデータをハイライトする highlighting, 選択されたデータのみを表示しその他のデータは表示しない shadow-highlighting, 表示を消す操作に対応する deleting, などの操作を持ち、種々の形態のデータの認識と表示に対応している。Deletion は Brushing の一部と考えることも可能である

Brushing や Identification のようなデータ認識機能は単一のグラフ内だけで閉じて使われるものではなく、いくつかのグラフ表示間で関連を持たせながら利用されるべきである。つまりある図で選択表示されハイライトされたデータは、他の図の中でも何らかの形で選択表示されるように動作すべきである。このような関連性は link と呼ばれている。Link 自身は、認識されたあるいは認識されていない各標本をどのように各 view で表現するか定義することによって実現され、数値解析との linking も自然に定義可能である。

Rotation は CG などによく用いられる 3 次元での空間あるいは対象物の回転表現である。この場合 3 次元で表現されたデータの様相をいろいろな角度から眺めることによるメリットのほかに、空間の奥行き感を得る効果もある。ただし、図を静止させるとその奥行き感が失われるため立体視の応用も考えられている(Becker, Cleveland and Wilks(1987))。この方法により対象物体の形状など構造的特性の理解を助ける効果が期待できる。ただ一方で、空間の回転や移動はその座標の意味の把握を困難にする場合がある。特に特定の座標軸に意味があり、空間内の関連性をその軸をもとに考えなければならない場合には必ずしも適さないことも指摘されている(Becker, Cleveland and Wilks(1987))。

Parameter control はグラフの構成にかかわるパラメータを動的に変化させてその特性を把握しようとするものである。グラフには単にデータ値をそのまま図中に示す散布図のようなものばかりでなく、モデル適合や何らかの処理の後グラフ表示するものも少なくない。たとえばヒストグラムの表示の際にはまず度数分布を計算し、それを棒グラフとして表示することになる。その際にはまず各階級幅を決定しなければならないが、これを変化させることにより、ヒストグラムの形状が大きく変化することになる。このように図の作成の際のパラメータを動的にコントロールすることによって動的表示を