得ることを目的とするものである。

グラフによってはその縦横比のとりかたによって視覚的印象を大きく変えるものがある。そのためにScalingの機能も必要である。

Interpolationは表示された(あるいは表示される)viewとviewの間を動的に補間するものである。ただし，この際の2つのviewは同種のviewである必要がある。ある意味ではRotationやParameter controlはInterpolationの1つとしてとらえることができる。典型的な例としては多変量データの解析の際ある3変数の組で定義された空間と別の3変数の組で定義された空間を考え，その空間の間を補間する3次元空間を求めてデータを表現し，多次元構造を把握しようとする方法がある。このような動的表現はguided tour (Hurley and Buja (1990))と呼ばれている。

Alternationはクループ化された情報を強調表示するために，交互にクループごとの情報表示を行う方法である。

これらの動的な手法のほかに良く知られたものとして射影追跡やgrand tour (Asimov(1985), Hurley and Buja (1990), Cook and Buja(1997))などがあるが，詳細についてはこれらの文献のほかBecker, Cleveland and Wilks (1987)やCleveland and McGill (1988), Eto, Kanazawa and Ochi(1995)等を参照されたい。

4．3　2　動的手法適用の際の留意点

このような動的な手法を導入することによって，多様なグラフによるデータの調査が可能になり，

1）特徴のあるデータ点の把握，
2）データの基本的（低次元）構造の把握，
3）クループ化（クラスタ）情報の把握，
4）変数間の相互依存関係の把握，

などのグラフ認識の強化が期待できる。

しかし，一方で動的な表示のために配慮すべき問題点もある。その一つは人の認知的側面への配慮である。動的な表示は基本的に表示を切替えることによって動的なイメージをつくり出している。その際の切替え速度が遅いと，その表示はぎくしゃくした不自

然なものになる。また，ユーザとシステムとの対話時間のずれは作業や思考の妨げとなる。現在のテレビや映画では1秒間に24-30コマ程度の画像が流れている。データ解析では必ずしもこの水準が必要はないかもしれないが，表示が非常に低速になる場合には注意が必要である。

また，逆に一秒間に表示するコマ数が多いということは一枚の画面を見る時間が少なくなることを意味する。したがって，静的グラフ以上に画面における情報の量に注意しなければならない。過度に盛り込まれた情報は視覚的混乱を招いたり誤解の原因となる。また，人の注意は動くものに引き付けられる特性がある。このことは特に大きく動く対象物を調べる時には注意しなければならない。このように，視覚の認知特性について十分注意することが必要である。

また，特徴を動きとして把握しようとする場合，静的なグラフで特徴を把握するほどに容易ではない。動きのイメージは基本的に視覚残像の記憶によるものであるからである。したがって，その動き自身を同定するための支援機能が重要である。

あるデータ解析のセッション自身をオブジェクトとしてとらえ，それを記録，再生，あるいはデータ解析自身として編集できる能力がシステムに求められる。さらに，これまでに述べた動的手法は同時にいくつかのグラフを参照しながらデータの分析を行っていくことを前提としている。このことによって，単一グラフでは得られない多くの様相と知見を得ることができるが，そのような複数グラフの参照とリンクによる相乗効果を確認したり，同時に各グラフの派生履歴を追跡するための機能が必要となる。このことから，上に述べた基本的な単一グラフの動的様相の把握に加えて，データ解析の経過の認識と分析，さらに記録された過程の編集に関わる支援機能が望まれる(井美，越智，橋本(1998))。

5　O157データへの動的グラフィカルシステムの適用

上に述べてきたような観点から，今回のO157データの動的なグラフィカル表示の可能性について検討する。

5　1 動的なグラフィカルデータ解析システム

動的なグラフィカルデータ解析システムとしては，今回は当該研究者の一人（越智）の研究室で開発中のシステム dynamic+の適用を考えた。このシステムは，SunOS,

Solaris, Linux などの UnixOS のもとで，X ウィンドウ上で動作するプロトタイプシステムである。開発言語は C, C++であり，ライブラリとしては Xlib を使用している(柿内，越智，森重(1995))。

5　2　海岸線データ

O157 データの地理的情報をグラフィカル表示する上で，単に保健所の緯度，経度情報を表示するだけでは，その位置や特性を把握することは難しい。このため，背景となる地理情報を合わせ表示する機能が求められた。

このため，今回は慶応大学，柴田里程氏によって S 関数として作成されカーネギー・メロン大学の統計ソフト，統計データのデータベース Statlib (http //lib stat.cmu.edu) に登録された関数 jpn に含まれる日本の海岸線データを利用することとし，緯度・経度をもとにする散布図の背景としてこの海岸線データを描き，その地理的な位置関係を把握する上での参考とした。以下この view は散布地図と呼ぶことにする。ただし，このデータはあくまで日本全体の概形を描くために作成されたものであり，詳細な精度を要求するものでないため，このデータをそのまま適用すると，実際の保健所データを布置した場合に不整合が出る個所があったため数ヶ所の海岸線については補正を行った。（付録表４－２参照）

5　3　O157 データの動的表示の実際

以下当該システムでの表示状況の実際について紹介する。

dynamic+で解析を行うためのデータファイルは，通常のテキストファイルであり，

```
C C C N C C C C C N
経度 緯度 保健所探知日 都道府県名 有症者数 無症者数 入院者数 死亡者数 年齢 性別
千歳保健所 141 39258 42 49125 9 430137 北海道 0 1 0 0 -1 女
紋別保健所 143 21310 44 20347 9 438356 北海道 1 0 1 0 5 女
遠軽保健所 143 31503 44 03484 9 512329 北海道 1 0 0 0 2 女
旭川保健所 142 22111 43 46051 9 528768 北海道 1 0 1 0 22 女
釧路保健所 144 23148 42 59542 9 528768 北海道 1 0 1 0 73 男
帯広保健所 143 12433 42 55379 9 564384 北海道 1 0 1 0 4 女
網走保健所 144 15541 44 01264 9 586302 北海道 0 4 0 0 -1 ?
・・・
```

のような形態をしている。基本的にデータはスペースで区切られ，その属性と変数名がデータファイルの先頭で定義される。データファイルの第1行目はデータの属性を示す行である。ここで C は数値データを持つ変数，N は名義変数を示している。各レコードの先頭フィールドはデータ名となるように固定されているので，この属性，変数名は指定しない。時間情報のデータについては，現在のところこれを直接表示するようになっているため，10進表現で指定している。（このデータは平成9年のデータである。）

先に厚生省のホームページから入手したO157 データについて先に述べたような地理的情報の整理を行った上で，上記のようなデータファイルを用意すればUnix のX ウィンドウ環境下で

```
>dynamic+ “データファイル名”
```

とすれば，システムを起動し，表示すべき view を選択するメニューが現れる。その後は所望の view と動的手法を選択して動的なグラフィカルデータ解析を進めることができる。典型的には図4．6のような形態で本システムでのデータの視覚化表示がなされる。

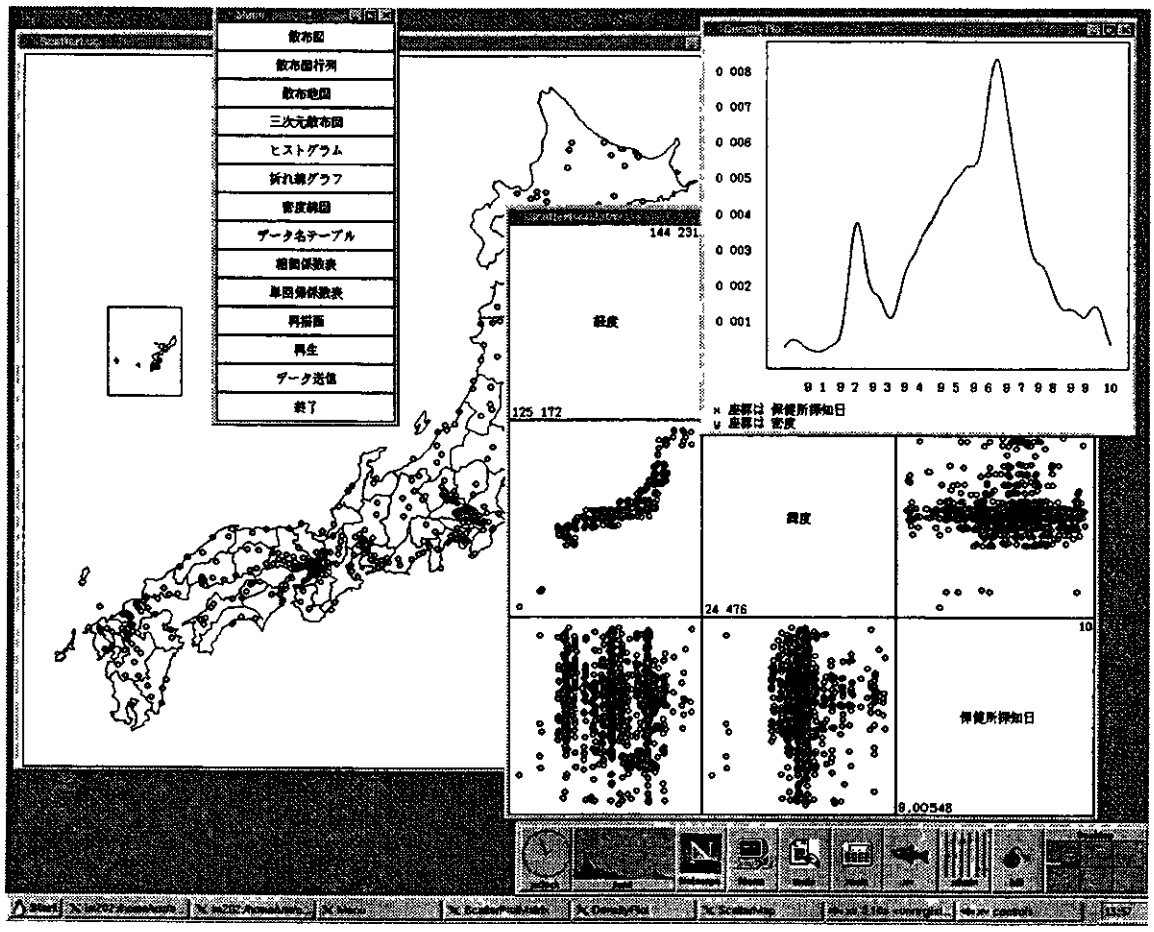


図4．6　動的グラフィカルデータ解析システムでの表示状況

ここでます，今回のデータをもとにして本システムで作成した平成９年から平成１１年までのO157の全体的な発生状況を示すことにする。それが図４　７－４　９である。ただし，ここでいう発生状況は保健所からの探知報告がなされたケースと言う意味である。いずれの年においても東京，大阪，北部九州などの大都市圏に多い様子がこの表示から見て取ることができる。

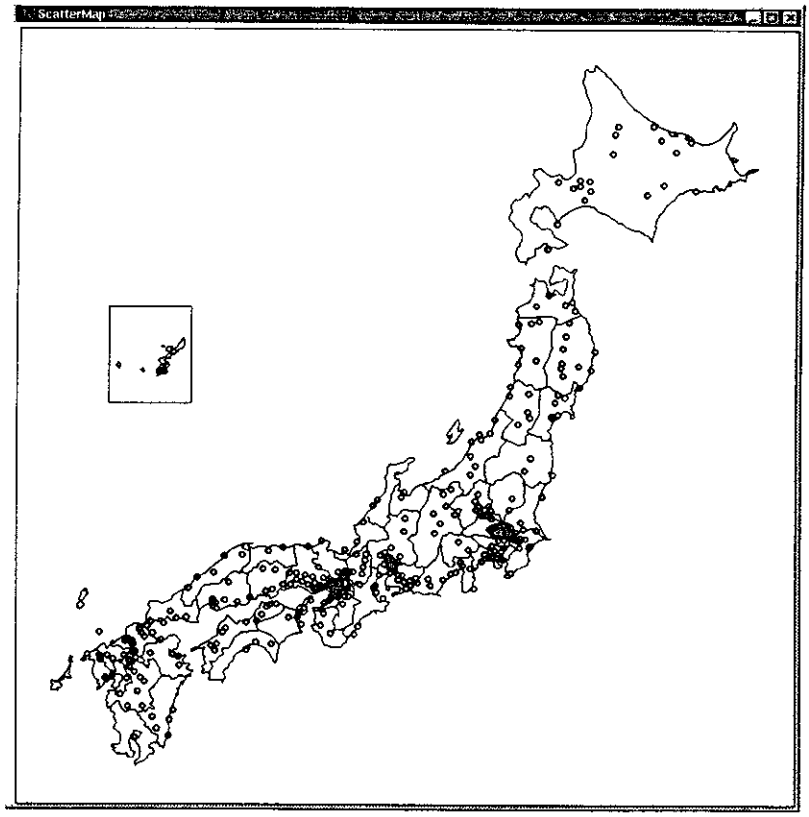


図４．７　平成９年のO157発生状況

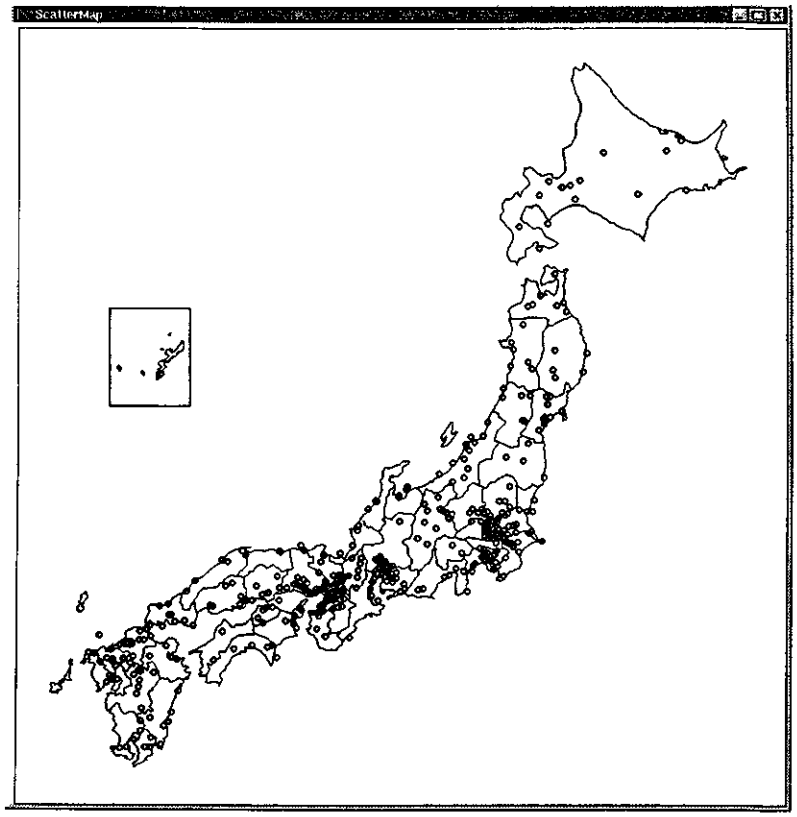


図４　８　平成１０年のO157発生状況

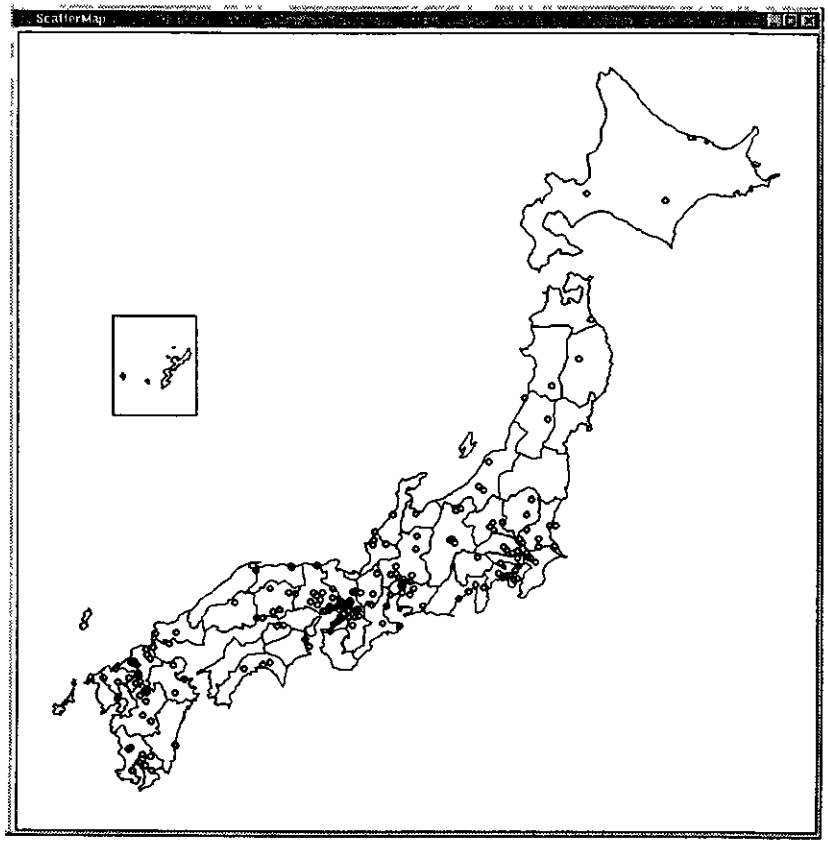


図４　９　平成１１年のO157発生状況

これらのデータでは，あくまで静的なグラフ表示であり，O157 発生の各年度の状況がわかるだけでその詳細がわかるわけではない。

以下動的なグラフィカル表示に関してその表示状況を紹介する。まず動的な表示に関する Identification としての機能であるが，これについてはたとえば図４　１０のように図中の特定の点指定し，これをデータ名テーブルと言う view で確認することでその名前を確認することができる。あるいは逆にデータ名テーブルから図上の点を確認することも可能である（図４　１１）。

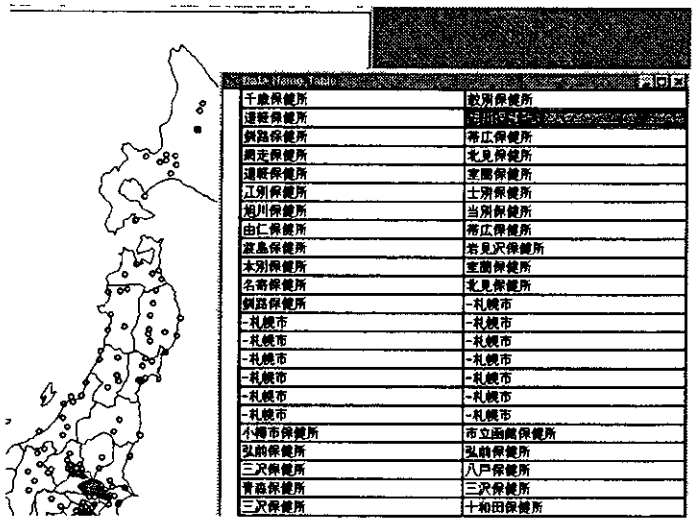


図４．１０　図中の保健所を確認

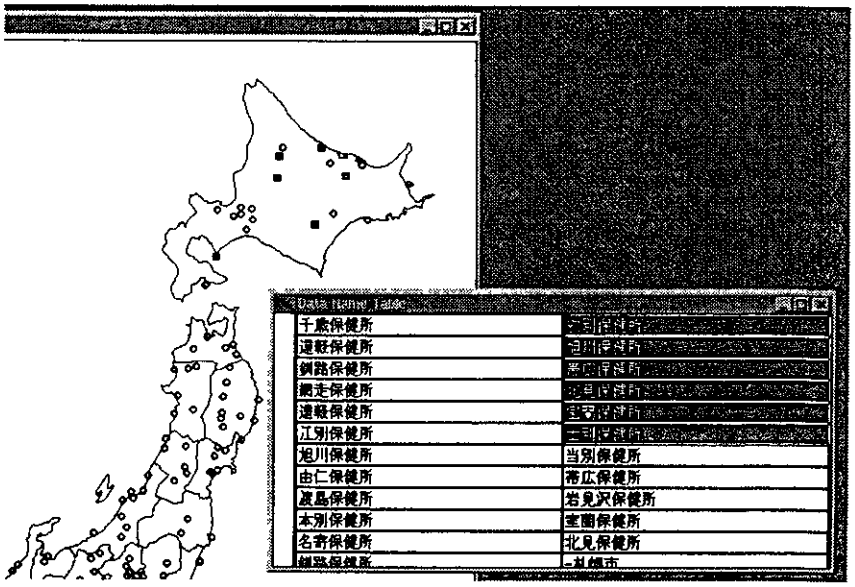


図４．１１　データ名テーブルから図中の位置確認

View 間のリンク機能により，散布地図だけでなくその他の view との間でも点の確認が行えることになる。

また，複数の点の識別操作としての Brushing を利用することにより，データの様相を把握することも可能である。たとえば次の保健所探知日と年齢の散布図ならびに散布図行列を見るとO157 の発生状況について年齢に３つのクラスタがあることが読み取れる。つまり１０歳未満の若い年齢層と２０歳前後の年齢層，そして４０歳から６０歳の間の高年齢層での発生数が多い。特に，そのうち２０歳前後の年齢層について夏場に集中して多く発生し，この年齢層に比べて高年齢層は年間通じて発生状況が散らばっている様子がわかる。さらに，緯度と年齢に関する散布図（散布図行列の左下）から，特に西日本で年齢分布の２分化が起きている様子が示唆されている。

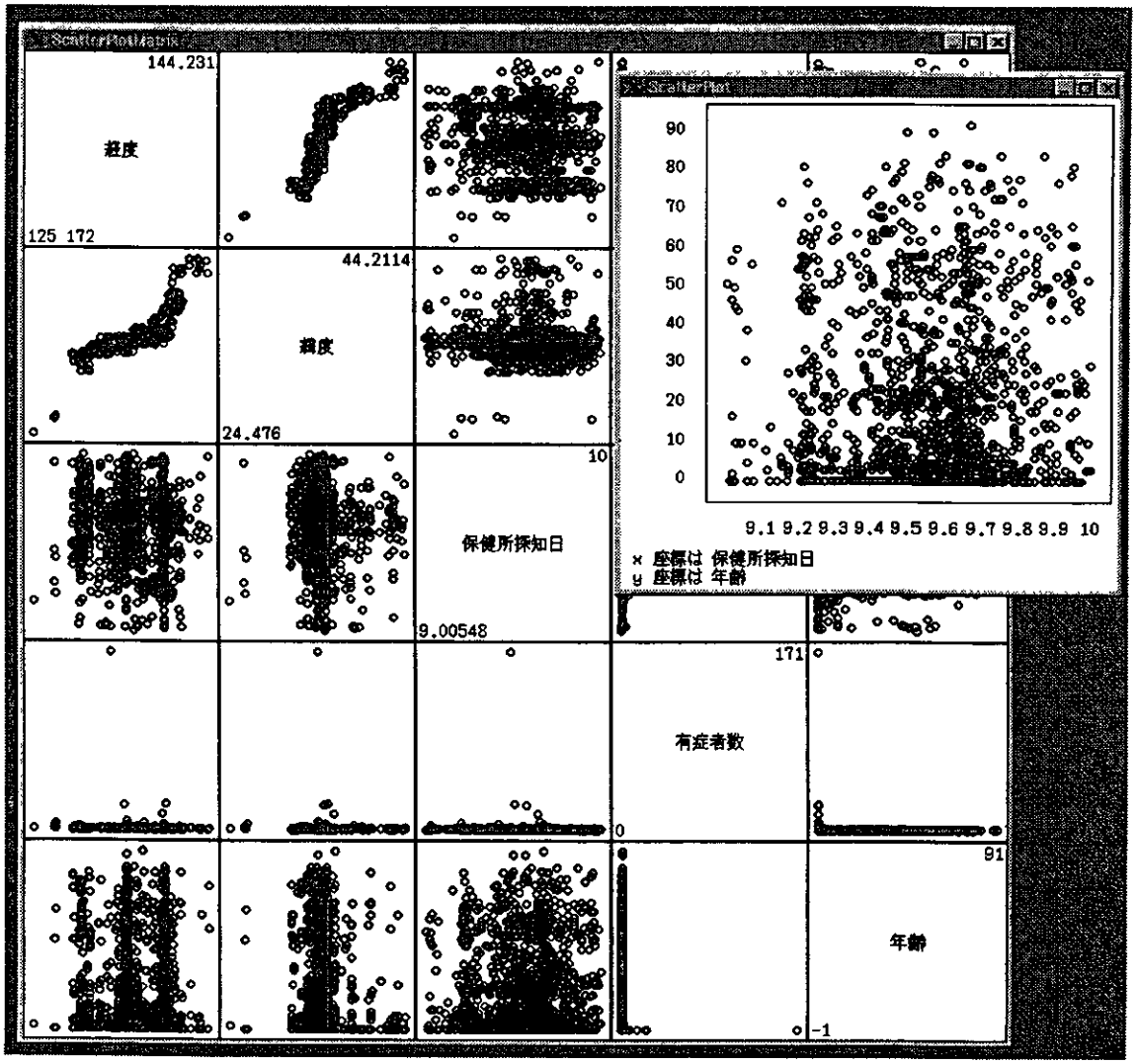


図４．１２　緯度，経度，保健所探知日，有症者数，年齢に関する散布図行列。保健所探知日と年齢については別に散布図を用意した。（平成９年）

このことから，たとえば夏季の２０歳台の発生状況を示すためにその年齢群をBrushingにより選択表示したものが図４　１３てある。選択された年齢層のデータはハイライトされており，特に首都圏，関西に集中して多いことが確認できる。ただ，今回のデータのようにデータ数が多く，図中に多くの点が布置されている場合には，この図のようにすへての点を表示しながら選択された点をハイライト表示するよりは，シャドウハイライトとして，選択された点のみを表示する方が効果的な表示となる場合が多い。図４．１４は同時期の４０歳から６０歳の年齢層について同様のBrushingをシャドウハイライトを用いて表示したものである。もちろん首都圏，関西地区での発生は多いものの九州地区での発生が先の年齢層の場合と比較して目立つことが確認できる。

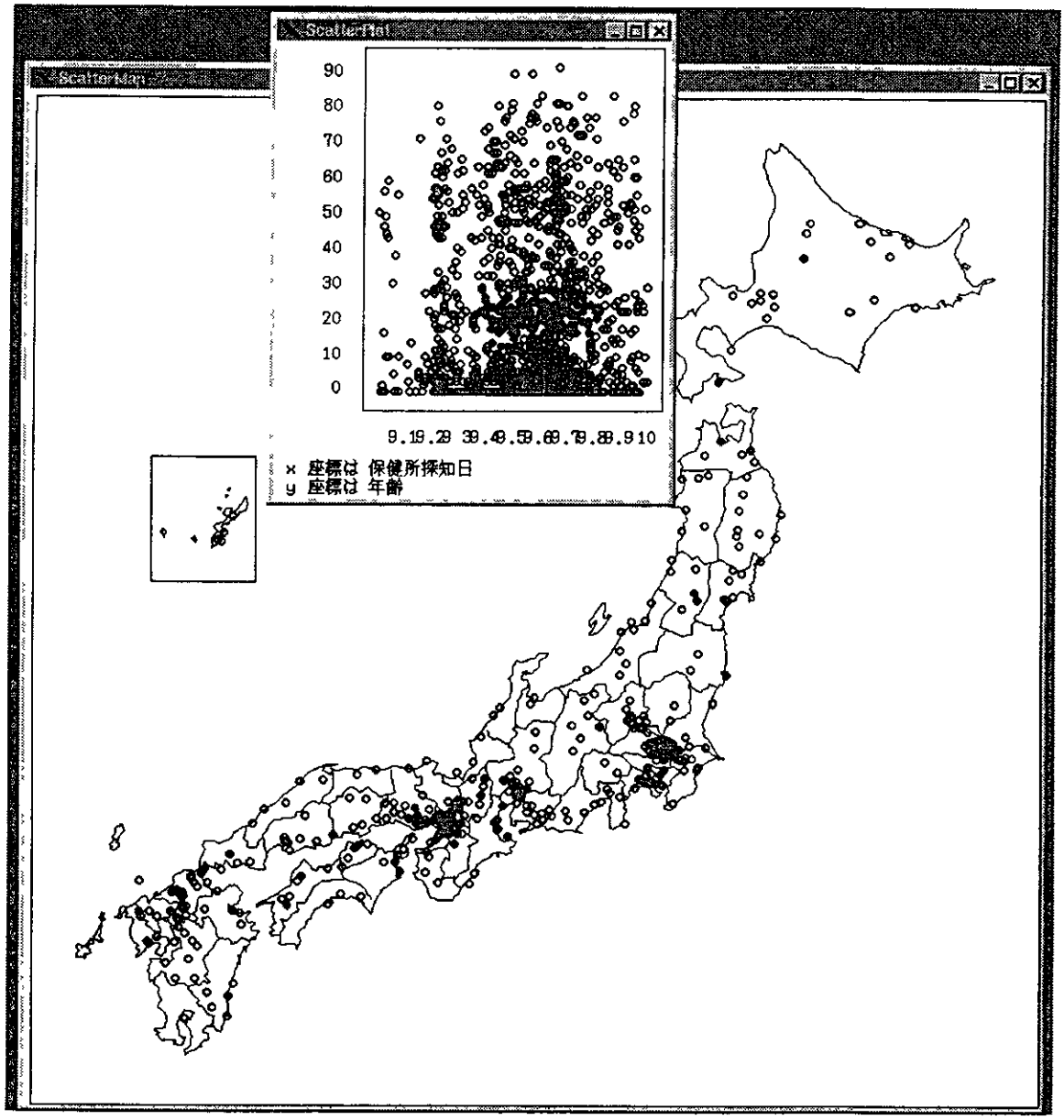


図４　１３　年齢２０歳前後の夏季のO157の発生状況をBrushingで確認。
年齢と保健所探知日に関する散布図でBrushしてそれを散布地図で表示。
（平成９年）

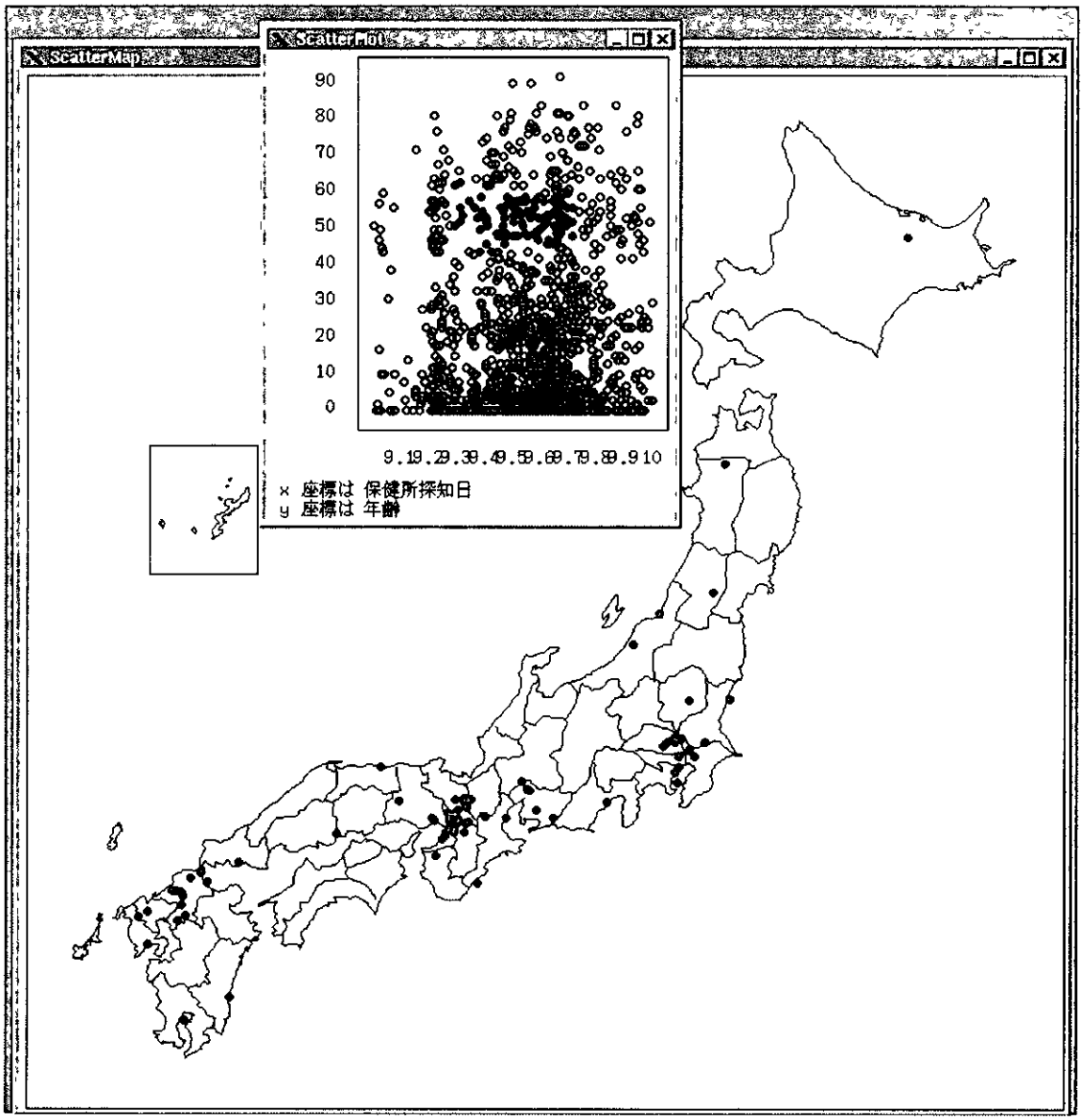


図４．１４　年齢４０歳－６０歳前後の夏季のO157の発生状況を確認。
年齢と保健所探知日に関する散布図でBrushしてそれを散布地図上でシャドウハイライトとして表示。(平成９年)

このような年齢の３カテゴリ化の傾向は平成１０年, 平成１１年のデータにも見られている。

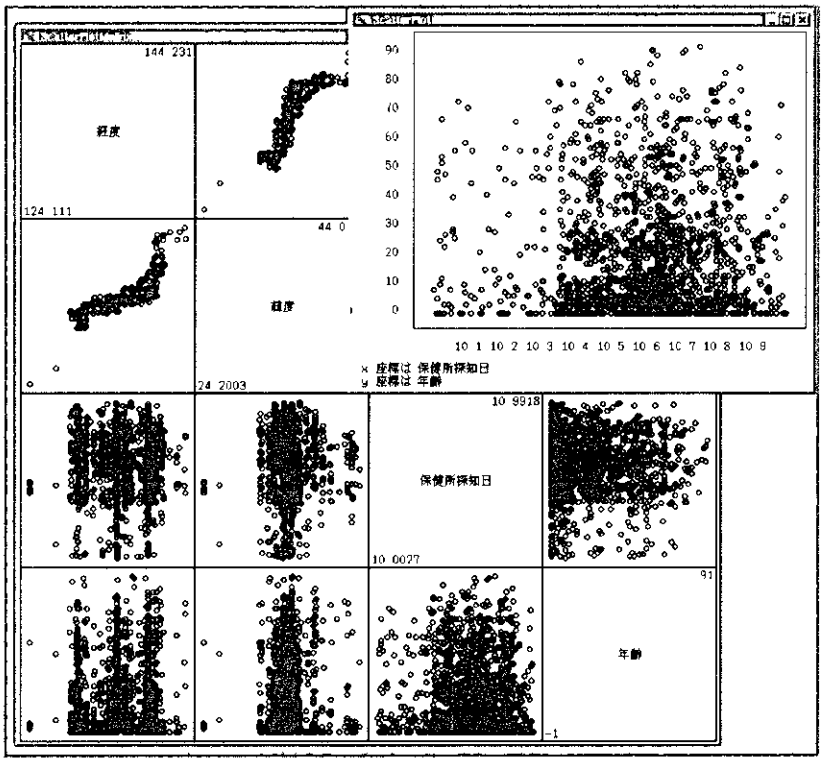


図４　１５　緯度，経度，保健所探知日，年齢に関する散布図行列。(平成１０年)

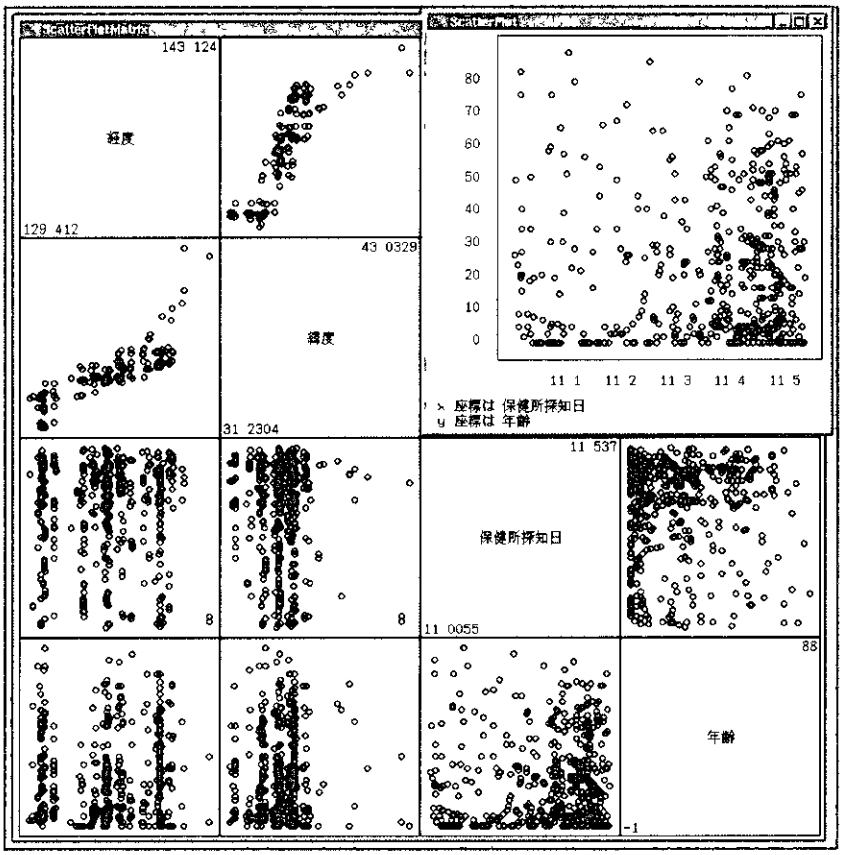


図４　１６　緯度，経度，保健所探知日，年齢に関する散布図行列。(平成１１年)

また，これらの散布図から平成１０年では発生数の立ち上がりが夏季に集中していたことも分かる。この様子を密度関数として捉えると，下の図のような密度関数推定を得ることができる。ここでは密度推定の際の計算領域（近傍に含めるデータの範囲：個数）をパラメータコントロールを用いて制御，３年とも約 20 76%となるように揃えている。

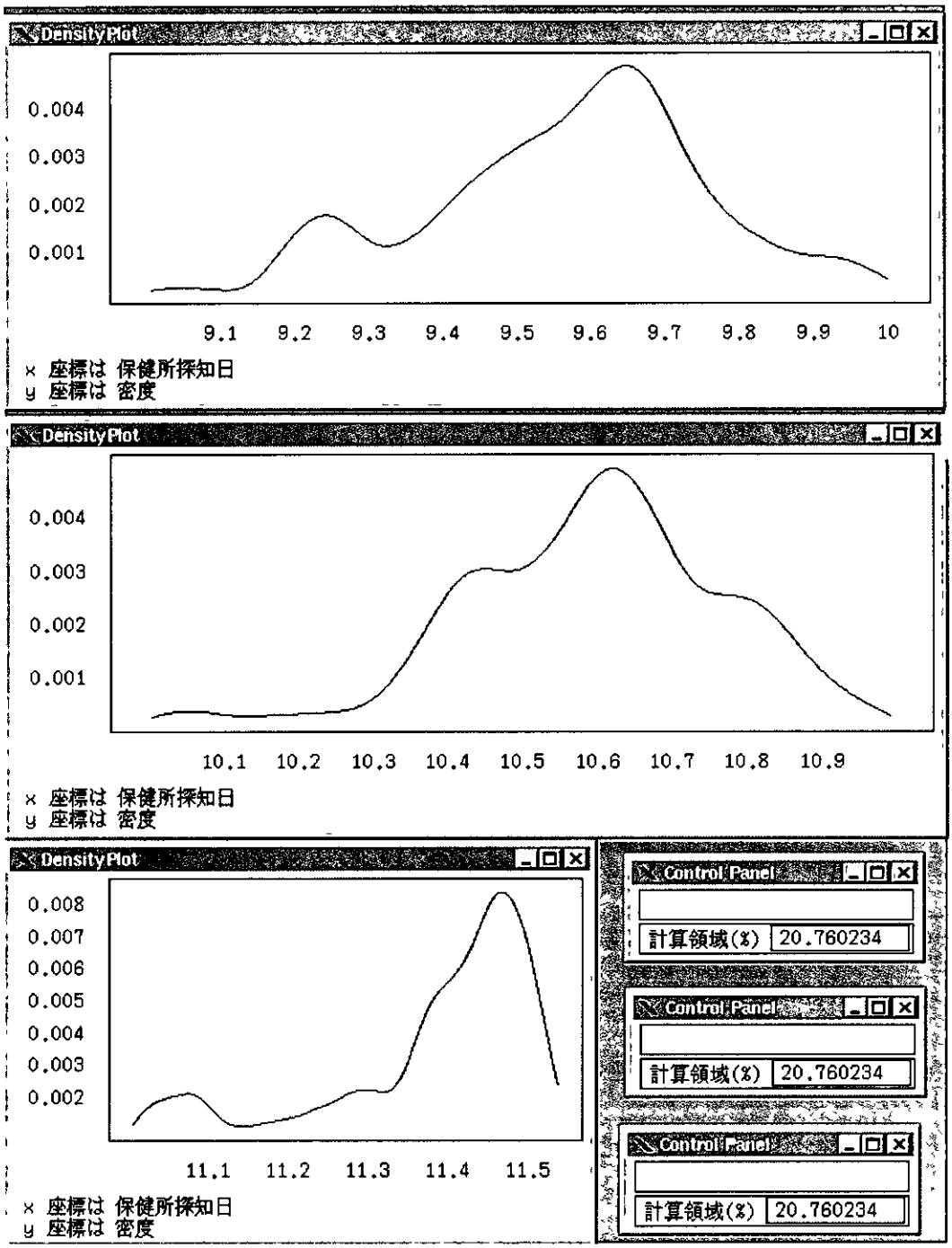


図４．１７　上から順に平成９年，１０年，１１年の保健所探知日に関する密度推定を表示。その近傍の計算領域は 20 76%に設定。平成１１年の密度の右端の減少は計算手続きによる効果。観測期間が異なるので平成１１年の縦軸の高さはそのままでは他の２年とは比較できない。

この図から平成９年には春に１つのピークがあったことが読み取れる。一方平成１０年，平成１１年にはその傾向は見受けられない。ただ平成１０年の年末から平成１１年の年始にかけて，あまり強くは現れてはいないものの，O157 発生の増加傾向があったことをうかがわせる様子も見て取れる。また平成１１年は平成１０年とほぼ同様な夏季にかけての発生の立ち上がりを示していることが分かる。ただし，平成１１年の密度推定に関しては，ピークから右にかけての密度関数の下がり方は推定方式の特性によるもので実際の発生数が減少していることを示唆しているのではない。

この平成９年の春のピークについては，やはり Brushing によって春のピークの立ち上がり時期のデータを選択することによって，それが主に関東地方での発生からきていることをグラフから読み取ることができる。

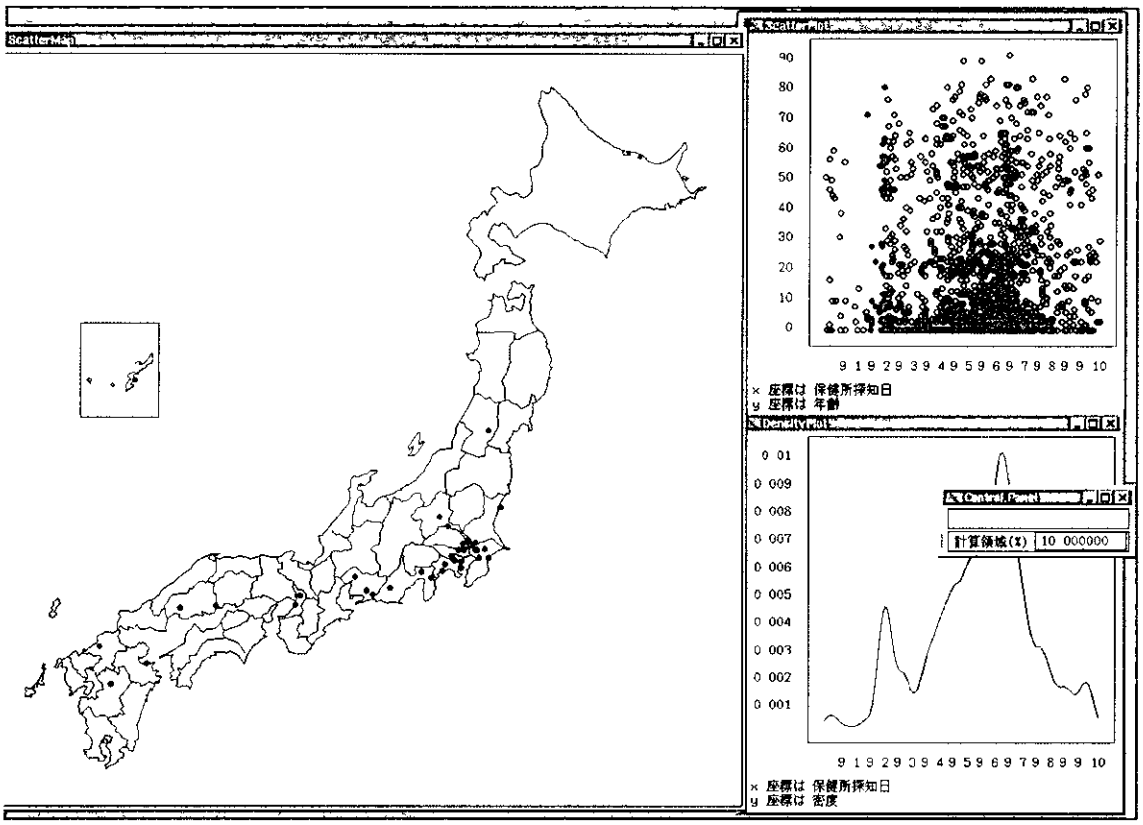
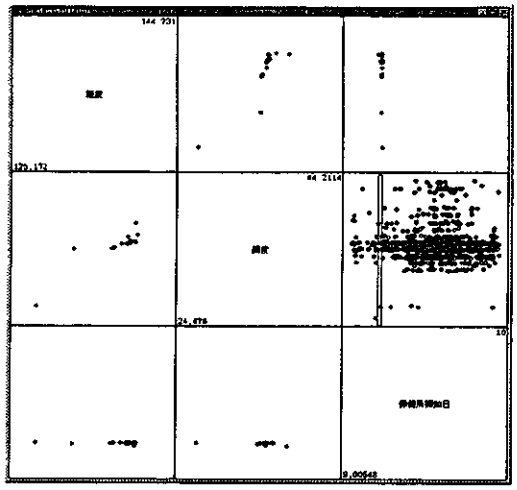


図４．１８　平成９年の春時期のO157 の発生について，その立ち上がり時点での発生場所を特定。(Brushing により，シャトウハイライト表示)

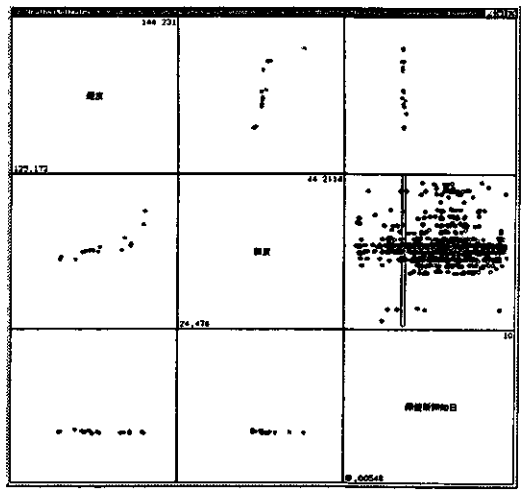
また，上のような手続きを動的に Brush を変化させることによって，時間経過に伴ってO157 がとのように発生しているかをいわゆる動画的な表示を用いて確認することもできる。

紙面上ではその効果を十分に伝えることは難しいが，次の図４　１９は散布図行列上て

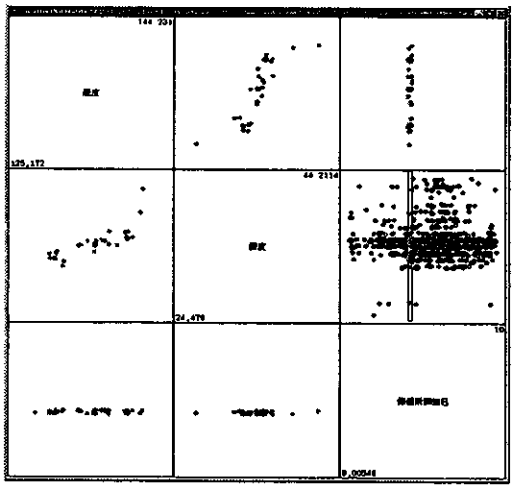
Brushing を用いて保健所探知日に関する条件付けを動的に変化させ，シャドウハイライトによって，発生個所を逐時的に確認した模様を示したものである。



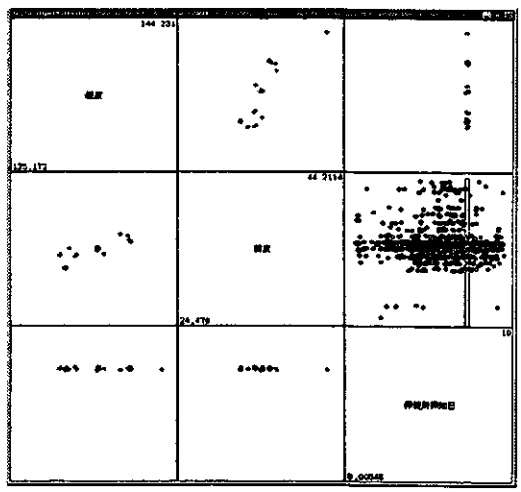
（1）



（2）



（3）



（4）

図４．１９　保健所探知日で条件付けし動的に Brush を変化させることで O157 の発生個所を動的表示。（1）（2）（3）（4）の順に散布図行列の２行１列目のパネルに注意。(平成９年のデータ)

これまでBrushingを中心に動的なデータの表現を見てきたが，今回のデータではその他の方法は必ずしも効果的な表現とは考えにくいものかあることが確認された。たとえばRotationについては，その位置情報と時間情報による3次元表現によって，何らかのデータの特性が明らかになることが期待されたが，それはたとえば散布図行列から得られる情報よりも有効とは考えにくいものであった。

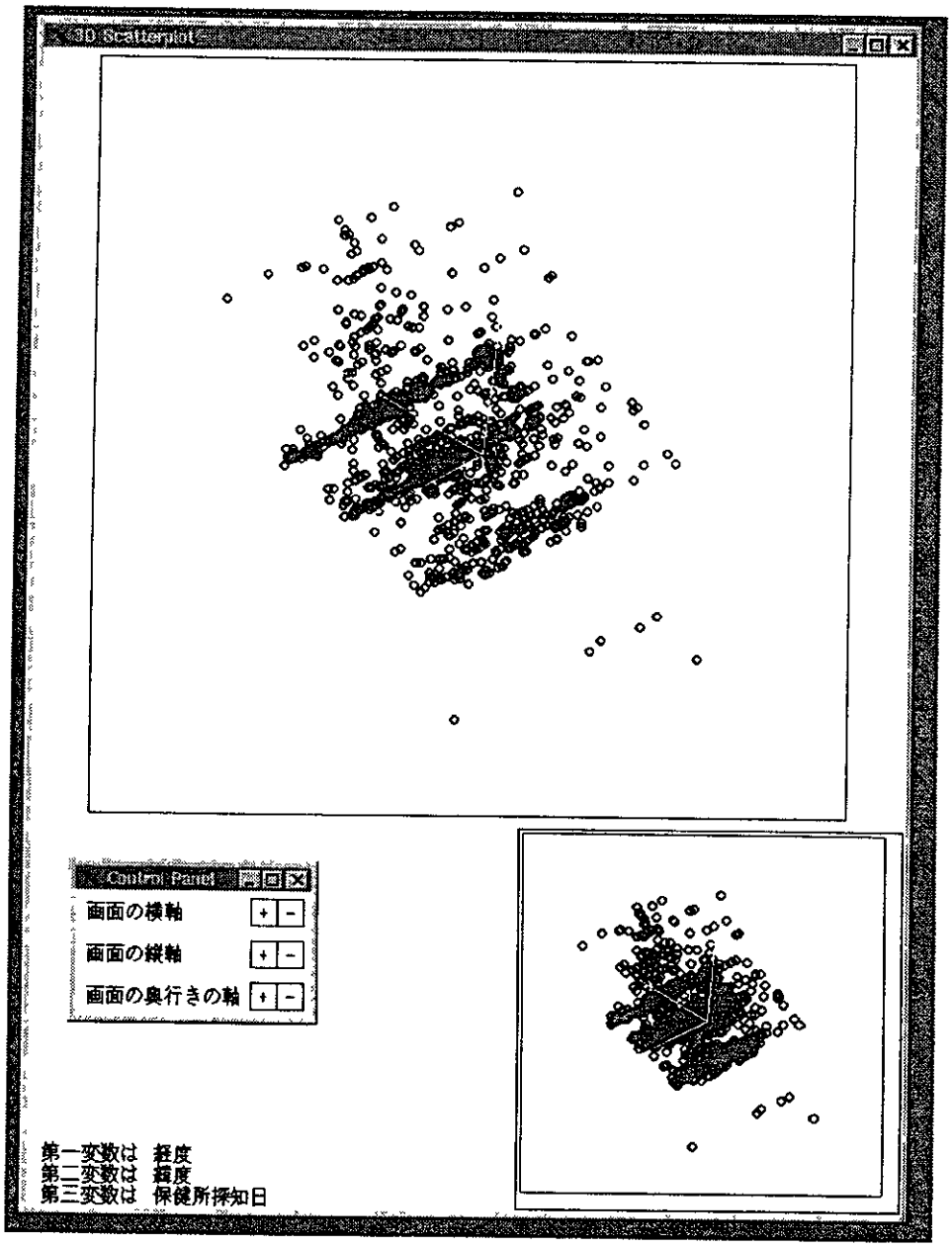


図4　20　緯度，経度，保健所探知日に関する3次元表現。
実際にはこの図は画面中央を中心に軸を決定し回転運動しており，データ点雲(Point Cloud)の形状・構造的配置はその表現より把握しやすい。

たとえば，図４．２０は平成９年のデータの上記３次元表現であるが，この図は実際には回転運動をしており，その効果から紙面から受ける印象以上に立体としてデータ点がどのように配置されているかといった構造は把握しやすい。ただし，今回のデータの場合，保健所は位置的には固定されており，時間的変化に伴いO157の発生状況が現れることから，その点雲（Point Cloud）としての立体構造は，板状の点の塊が日本の形状に折れ曲がったものとなっている。その様子は３次元表現から把握しやすい。また図４　２０からも分かるようにデータは３本の線上に集中している様子も読み取れる。ただし，この３本の線は散布図行列でも明確であった関東，関西，北部九州を示しているにすぎない。また，その板状の構造から，これまでに調べてきた以上に位置情報と時間情報に関してデータ解析に有益な情報や解釈は得られてはいない。

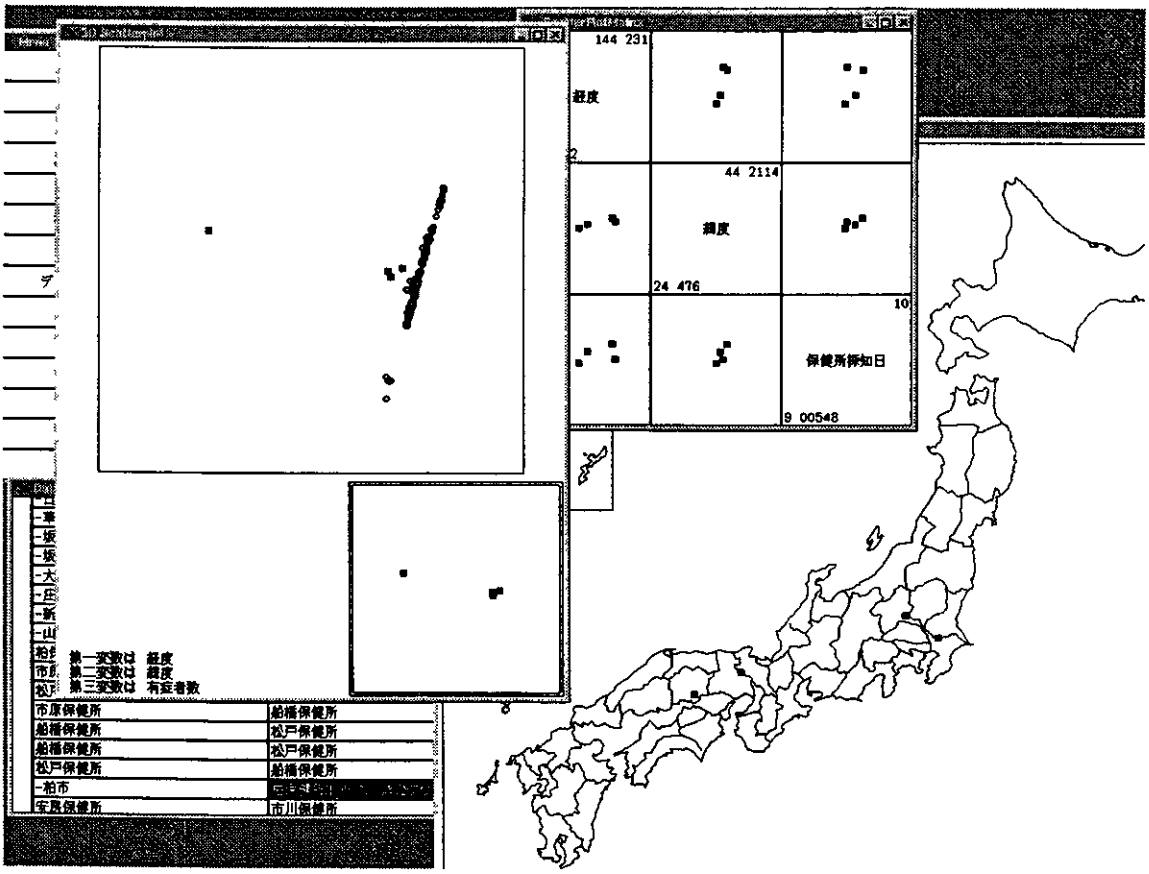


図４．２１　緯度，経度と有症者数によるローテーションで特異データを認識。（平成９年データ）

また，これまでのグラフィカル表示は基本的に保健所のO157情報の報告を基礎とするのものであった。大半のデータは個人単位に報告されているが，場合によってはまとめられた形で報告されて，グループとして報告されている場合もある。図４．２１では緯度，経度と有症者数による３次元散布図によりそのような観点からの特異データを把握し他のviewで確認している。このことにより，その報告のあった時期ならびに保健

所とその位置が把握できた。しかし，ここで呼ぶ特異データとは，３次元構造表現の中での特異点を指してはいるが，実のところ単に有症者数が多いケースにすぎない。したがって，このことが問題であれば，別に３次元表現を用いなくとも散布図などの２次元表現あるいはもっと低次元の表現での認識とリンクの手法により関連データの把握は可能である。

このことは先に Rotation のところで述べたように，軸に明確な解釈を与えてその解釈の上で現象を理解する上では Rotation は不向きであることを反映しているものと考えられる。Rotation では回転と言う操作により各軸の持つ意味を同等化しており，そのようなデータ構造を持つデータではその構造を立体的に把握し解析に反映させる上では有効と言えるが，そうでない場合には，構造的な把握から解釈への移行は必ずしも容易ではなく，また有益でない場合もあることに注意が必要である。したがって，今回のような位置情報をもつデータの直接的な３次元表現としての Rotation の効果については，今後より有効な利用法について検討を加えていく必要かある。

また今回のデータでは詳細な関連データを用いた分析については考えていないので，空間補間に関する検討は行わなかった。より詳細な個人データと疾患の状態や病態に関する分析あるいは可能な原因食材等のデータからの因果関係の調査などの際には，なんらかの効果が期待できるもしれないが，この点については未検討である。

さらに Alternation については基本的に Brushing と同様な効果を生むものであり，実際当該システムではこの機能は Brushing と同一視して実装しているので，この点については先の Identification, Brushing に関する手法と同様の効果が得られるものと考えられる。

５　４　他のシステムとの連携について

本システムは基本的にグラフィカル表示に特化したシステムであるが，前述のようにその基礎データは極力シンプルな構造にしている。また，動的な手法で選択されたデータについてはその表示フラグ情報と共にそのテータレコードの ID がファイルに保存される。またその手法選択が終了すると，このファイルはシステムから切り離されて他のシステムからアクセス可能になる。このことを利用することによって選択されたデータについて他のシステムで数値的なデータ解析を実行することが可能である。

５　５　動的グラフィクス環境の提供について

これまで述べてきたように，今回のO157データのような位置を含むデータについても動的なグラフィカルデータ解析環境を適用することによって，データの持つ情報をより効果的に抽出し，データ解析に反映できることが確かめられた。ここでは，このような動的なグラフィカルデータ解析の結果を共有する問題について考える。

先に述べたように動的なグラフィカルデータ解析の本質はユーザが意図した解析結果を瞬時にその表示に反映できることにある。静的なグラフィカルな表示であればそれを静止画像として取り込み何らかの形で，提供することも可能である。また，動的なグラフィクスについてもデータ解析の結果をちょうどビデオで録画・再生するかのようにそのイメージを提供することも考えられる。しかし，ここでは先に述べたようにユーザの意図した解析結果を参照できるような形での表示の提供・共有の可能性について検討した。

そのためには実際の運用上での使い方は，セキュリティやデータ管理の観点などから，かなり制限されたものにせざるを得ない可能性があるものの，基本的にはこのシステム自身にユーザがアクセスできる環境を整備することが必要と考えた。

ただし，当該システムはUnix，Xウィンドウ上で動作しているシステムである。現状では多くのユーザがこのような環境を有しているとは限らない。ユーザの中にはUnixを常用している者もあればマイクロソフト社のWindows環境を使っているもの，マキントッシュを日常業務で使用しているものなどその環境は多様である。このような多様な環境に対して提供可能な一つの方法として，ここではWeb環境を利用してユーザにインターネット上でのWebブラウザを介してUnixのアプリケーションを使用可能にするSCO社のTarantella(Version1.1日本語版)というシステムによる運用の可能性について検討した。

今回の試験環境では，Tarantellaの運用プラットフォームとしては大分大学工学部知能情報システム工学科に置かれたSunマイクロシステムズ社製ワークステーションSun Ultra 1, SunOS 5 5 1（Solaris 2.5.1）を使用することとした。このマシンの上にまずWebサーバー環境としてThe Apache HTTP Server Project (http://www.apache org/) が公開しているHTTP (Web) サーバープログラム apache (ver. 1 3.3)をインストールし，その環境下で使用することとした。

ユーザはインターネットあるいはイントラネットなどでネットワークに接続されたリモートマシンでWebブラウザを起動し，上記のWebサーバー内に用意されたTarantellaの用意されたurlへアクセスする。そうすると図４．２２のような起動画面

か現れユーザ認証を行うステップに入る。

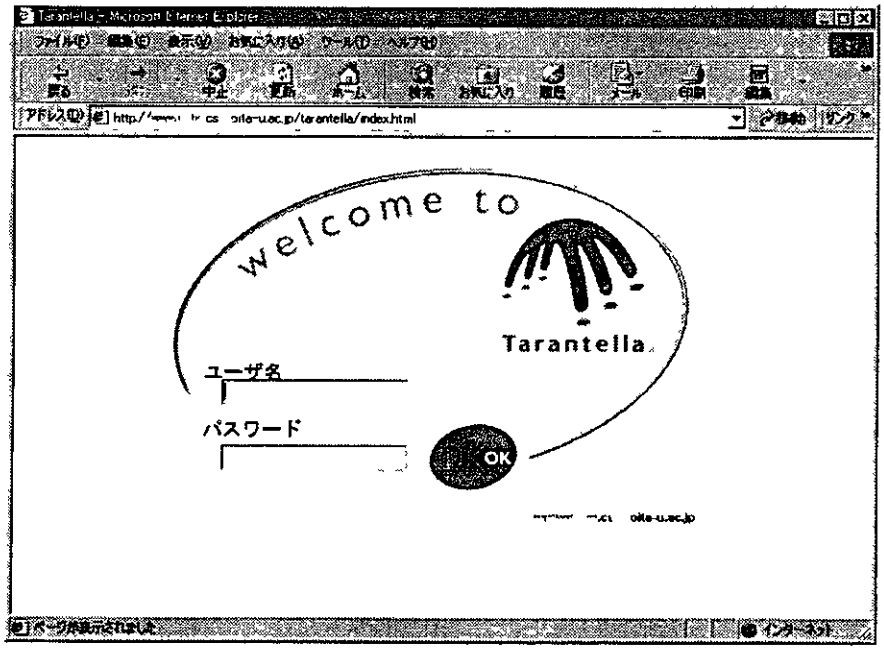


図４　２２　Tarantella の起動画面

この際ユーザは，基本的にこの Web サーバーとなるワークステーションで使用権を与えられたユーザてなければならない。ユーザ認証後はそのユーザが使用している Web ブラウザをちょうど X 端末の様に考えて Unix の環境が利用可能になる。

実際の使用状況は図４　２３のとおりである。

このことにより，ユーザ側の使用環境によらずに動的なグラフィカルデータ解析環境の提供は可能である。

実際の使用環境としては，同学科内におかれた計算機ネットワーク内の計算機からは，多少リモート側の計算機の処理能力に左右されはするが，　Pentium（166MHz），10BaseT 接続された Windows95 マシンでも，動作は多少緩慢となるが許容範囲内で利用可能であることが確認できた。この動作の緩慢さはリモートマシンを Unix 環境で運用し X ウィンドウ環境で先の Web サーバー上で dynamic+を使用して表示をリモート側に表示させた場合には感じられなかった。このことは Tarantella の通信形態（Tarantella では X ウィンドウの通信プロトコルを独自仕様の通信形態に変換してリモートの表示プログラムと交信する形態で実装が行われている）によるものであり，サーバー機・リモート機双方の処理能力が向上すれば解消できる問題と考えられる。

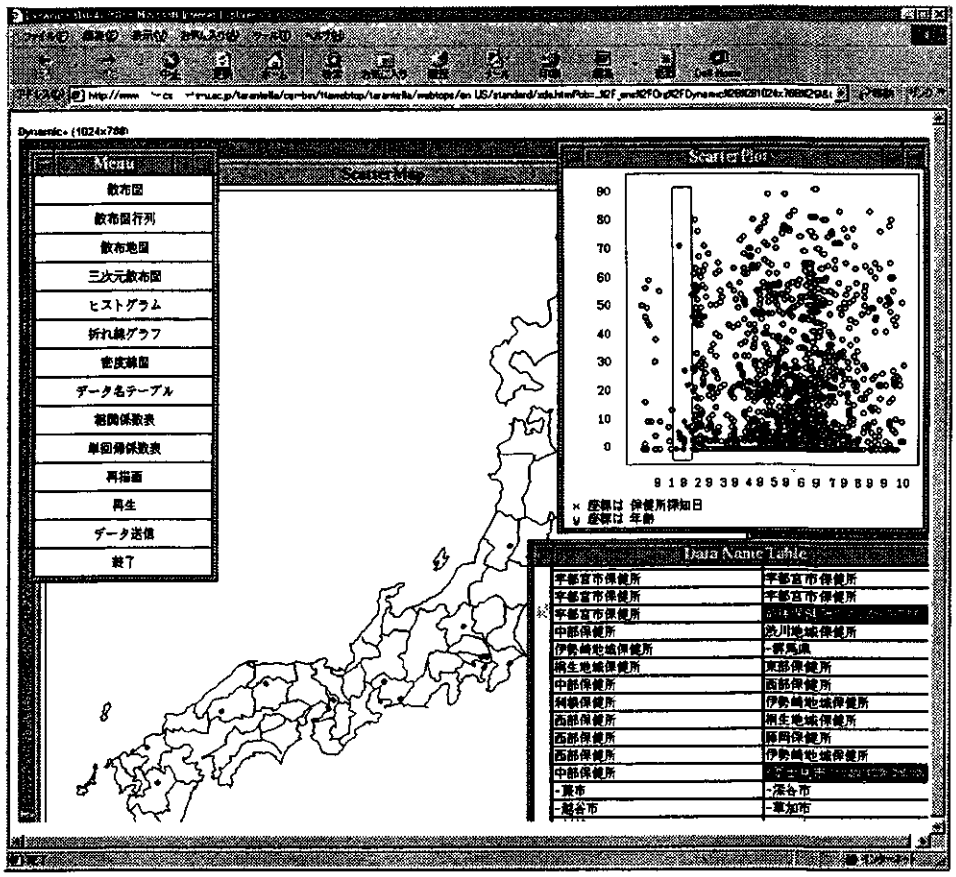


図４．２３　リモート（学科内）におかれたパーソナルコンピュータから Tarantella を経由して動的グラフィカルデータ解析システムを利用する。

しかし，大学外部からのアクセスについては，その動作は相当に遅くなり，それなりに動作する場合であっても動的な動作を行う際には，非常にぎこちないものとなった。この点は現在のインターネットの接続環境の問題にかかわる問題であると考えらるので，その通信状況について次に検討を加えた。