

大気汚染データ解析におけるサンプリングの問題点

統計数理研究所

鈴木 達三

大隅 昇

1 序

数多くの大気汚染データ解析の方法論、シミュレーションなどが行なわれているわりには、確としたものがみられないことの原因の一つは、解析に使用するデータの観察されたいきさつ（測定方法、測定環境、気象条件など）をほとんど考えずに、不用意にデータ処理を行なうところにあると思われる。比較的その性質が捕えやすいといわれているSDxを例にとってみても、ほとんどの場合、導電率法による一時間累積値、あるいは酸化鉛法などによる暴露テストにもとづく積分量を用い、また風向、風速なども、ある一定時間内の（たとえば、一時間中の最終10分間の平均値）を利用している。これらの測定法は、特に深い根拠があるとは思えない（事実15分～数時間の単位で記録している例もある）。つまり、こうした測定法は便宜的に決められたもので、実際の測定は、

一種の系統サンプリングを行なう。このことになり、しかも手にしたデータはすでに若干の加工がほどこされていることにもなる。そこで、われわれは1時間値を利用するという立場を一時保留して、汚染質(とくに $SO_x$ )の大气中における状態はどうあるか、実際に測定器はそれをどの様な状態で捕えているかを考えこみた。1時間値の履歴を連続的に結んで累積曲線として観察すると、その変化は1時間という間隔にかかわらず様々に変動することがわかる。いいかえると、連続的に汚染の瞬間値が捕えられたと考えると、それは様々な変動を持つ時系列量として現われるはずである。そこで一方では、瞬間値の測定記録をおこない、汚染質の分布の連続的な変化の様相を把握し、他方、実際に使われている導電率法テークとの対比を行なった。これらをもとにして、できるだけ単純なモデルを考え、実際の汚染質が測定されている状態を餞刺し、1時間値としてサンプリングされている現実のデータが、どのような性質を持っているかということを検討するのための若干の試みを報告するのがこの目的である。

## 2. 調査対象と大まかな前提

まず調査対象として、千葉市を選び、1973年度今の全観測地点のデータを使い、2様々の立場から集計分析した(別刷)。この結果観察された性質の主なものとして、

- i) 風向は非常に支配的である。
  - ii) 風向、風速などは 短時間内で眺めるとかなりランダムに変動するが、相対的に長い時間間隔にわたって平均的にみると安定しており、しかも各地点とも非常に類似していることがわかる。
  - iii) したがって、汚染のパターンも若干の違いはあるが、各地点とも、おおむね似たような様相を示す。
  - iv) 1時間値の系列相関は、若干の傾向がみられることもある(12時間単位の周期など)が大きなものではなく、むしろ、気象の様々の要因(雨、強風、季節変動など)に支配され、これが各地点とも実に類似して現われる。
- などのことがわかる。つまり、各地点の様子は、短時間内で眺めると大気はかなりランダムに乱れた動きを有する、しかし、長期にわたって眺めると、かなり類似した状況にあると考えるよさそうである(大気汚染は、地域性が強いので千葉市のように、比較的安定した地理条件の下での話)。このことから、観測値によって測定される汚染質の状態は、長期的にはほぼ類似した条件におかれているとみてよいであろう。精密かつパラメータの多い拡散モデルなどによって汚染源との因果関係を必ずしも都市部の複雑な構造を反映しているとはいいがたく、それより前に、現実のデー

夕の質とその限界に依じた、かなり簡単なモデルを想定して、  
“汚染質が、実際に観測されている状態”を記述してみるこ  
 とも必要と考えられる。このために、われわれは、野外実験  
 (1975.年3月中旬自~4月中旬自まで実施)により $\beta\text{O}_2$ の瞬間値  
 を連続的に記録し、同時に、同一地点で導電率法による測定  
 も行ない、この両者のデータを比較した。

### 3. 野外実験データと既存データの対比

3.1.  $\beta\text{O}_2$ の瞬間値の分布. まず、 $\beta\text{O}_2$ の瞬間値を連続的に  
 捕えて、汚染質の出現パターンがどんなものであるかを観測  
 した。データは瞬間値と、これと同時に記録される1時間累  
 積値(Y)、および、同一地点に設置した導電率法による1時  
 間累積値(Z)、(つまり一般に使用されている自動記録デー  
 タにあたるもの)である。瞬間値データはさらに1時間内の  
 面積をアラシノータで5回読みとり、これの平均値(X)を算  
 出した。こうして求めたX, Y, Zの間には、きれいな一次関  
 係がみられ、相関係数は、それぞれ $r(X, Y) = 0.984$ ,  $r(X, Z) = 0.943$   
 $r(Y, Z) = 0.944$ であった。また導電率法による測定器を上記  
 データを観測した地点より約1.5kmほど離れた地点にも設置  
 して、この2地点間の1時間値の間の相関を求めたところ、  
 かなり高い相関がみられた。これから、瞬間値データは、ほ  
 ぼ導電率法と比例関係にあると判断される。瞬間値のパター

ンを検討したところ、次のような特徴がみられた。

- i) 何も感知されない状態から初めて汚染を感知して汚染の記録を始め、再びもとの状態に復元するまでを一つの汚染の集団(「パフ」と呼ぶことにする)と見ることになると、パフの出現している状態と、そうでない状態との弁別は、かなり明確に読みとれる(いいかえると、パフが一旦出現すると激しい変動がおこるが、そうでない平坦な時間もありということ)。
- ii) パフ内での汚染レベルの変動はかなり激しく、チャート上で大きく振動するが、これは、同時測定した汚染物質吸込口の近傍の風向、風速などの変化と、かなりよく対応している傾向がみられる(つまり、風向、風速は瞬間的には、かなりランダムに変動している様子がみられるわけで、これによる大気とその中に含まれる汚染物の乱れをかなりうまく反映して記録していることになる)。いいかえると、大気中に適当にかく乱、希釈された汚染物の測定器近傍の濃度分布を、かなりの確率に捕えているとみてよい。
- iii) こうした瞬間値の変動(パフの有無、汚染レベルの変動中など)を1時間値の単位で眺めると、累積値の勾配の変化として、確かに認めることが出来る。
- iv) 結局、パフがどのように出現するか、その様子を捕える

わけであるが、パフは一旦出現すると、必ずある時間持続して通過し、汚染質の感知されない状態に復元することが特徴である。

3.2 導電率法データによる解析。上の野升実験から予想される汚染質出現のパターンとして、いくつかの場合が考えられようである。たとえば、濃度の高いパフが断続的に出現する場合、あるいは長時間持続する場合、さらに薄い濃度の状態が比較的長く持続する場合、等々である。こうした各種の状況を把握するには、1時間値だけでは十分とはいえない。導電率法によりチャートに記録された累積曲線を接続して連続データと見立て、パフの出現の分布を大まかに調べた。

いま、 $A_n$  = パフが出現してから次のパフが出現するまでの時間間隔

$T_n$  = パフが出現してから退去する時点までの時間間隔(持続時間)

$D_n$  = パフが退去してから次のパフが消えるまでの時間間隔

$W_n$  =  $T_n - A_n$  (パフが認められない、つまり汚染のないとみてよい状態)

$S_n$  = パフ内の汚染質の総和(面積)

とおく。

パフ内では複雑な変動が起きているであろうが、ここでは粗く、あるパフが出現したかどうか、また、出現した場合の持続時間を調べる。数地点について、1973年度7月上旬のデータをを用い、上の諸量の各種統計量を算出したものが(表1)。

(表2)である。次のような特徴がみられる。

- i) いずれの分布もかなり変動係数が大である。
- ii) パフの長さは短い(15~20分位)ものから相当長いもの(数時間~数日)まで含まれる。1時間値で評価した場合、短かいものは見落とされがちである。
- iii)  $T_n$ と $\beta_n$ の間の相関は高い。パフの持続時間が長ければ $\beta$ も大きくなるというもっともな結果ともいえるが、(表2)のように回帰分析を行なうと、回帰係数と地点の特徴が現われているように見える。
- iv) 地点Ⅰ, Ⅱ, Ⅲは地理的にみてほぼ同一線上にあり、地点間の距離と関連させて考えると、各数値に意味がありそうである。
- v) 地点Ⅰ, Ⅱは比較的濃度の高いパフが断続的に表われ、濃度の高低の激しいタイフ、地点Ⅲ, Ⅳなどは濃度の低いパフが長く滞留している場合といえる。またⅣは他と様相が違ふ地点といえそうである。
- vi) 地点Ⅵは野外実験によるもので、他の地点と時期が異なるが、地理的にはⅠ, Ⅱに近いところに設置された。各値がⅠ, Ⅱにかなり似ている点が興味あると思われる。
- vii) パフが存在しないと思われる汚染レベルに対応する時間区間の1時間値を眺めて、その時の濃度レベルをみるとおよ

そ 0.005 ppm/h. 以下程度であることがわかる。これは丁度測定器の直示可能な精度内の数値であり、パフの単位で汚染状況を捕えることは、それなりに意味があるといえる。しかしかえると、1時間値を利用している場合には、汚染濃度の高い地点にと、ては持続時間は長めに、また低濃度が長く続くような地点にと、ては滞留が認められないと判断されるという現象がおこりうる。1時間内の短かい低濃度のパフは見落とされ、さらに、実際の1時間値のデータにみられる急激な変化が随所に現れる現象もパフの概念で説明がつきそうである。

以上はおくまで1つの試みにすぎないが、1時間値が観測時間間隔として、各当な地点とそうではない地点とがあり、パフの生起する過程を地点ごとには捕えて、1時間値との対応を関連づけて、始めて1時間値が明確な意味をもつと考えられる。

## 参考資料:

### 4 公害問題におけるデータ解析の方法論に関する研究

(大気汚染データのとり方とその解析)

文部省統計数理研究所

第2研究部第1研究室 鈴木達三

第4研究部第1研究室 大隅昇



(表1)

地点	A (分)	D (分)	T (分)	W (分)	S	$\bar{T}/\bar{A}$	$\bar{S}/\bar{T}$	$r(T,S)$	$r(W,T)$
I	417.5	456.8	367.9	84.7	87.4	0.7705	0.2376	0.7696	-0.091
	420.7	467.8	415.4	99.1	226.4				
II	399.4	444.7	305.0	139.7	40.9	0.7636	0.1341	0.812	-0.240
	389.5	398.2	414.0	177.1	92.7				
III	1003.9	968.4	829.7	136.7	141.4	0.8265	0.1704	0.870	-0.189
	902.8	854.4	887.5	168.8	179.6				
IV	291.7	288.4	167.6	121.3	15.9	0.5746	0.0948	0.951	0.548
	314.2	232.5	198.8	150.3	27.7				
V	1690.0	1564.5	1515.5	138.6	302.9	0.8967	0.1998	0.984	-0.290
	2786.9	2579.6	2763.5	185.6	688.2				
VI	679.5	1021.1	502.4	163.6	(109.7)	0.7394	—	0.803	-0.074
	653.9	1532.3	511.1	154.5	(147.8)				

注1. 上段は平均, 下段は標準偏差

注2. 地点VIは野外実験によるもの

(表2) SとTの回帰

I	$\hat{S} = 0.419T - 66.86$	$\hat{T} = 1.41S + 245.82$
II	$\hat{S} = 0.182T - 14.57$	$\hat{T} = 3.63S + 156.55$
III	$\hat{S} = 0.176T - 4.74$	$\hat{T} = 4.30S + 222.79$
IV	$\hat{S} = 0.129T - 5.79$	$\hat{T} = 6.66S + 61.81$
V	$\hat{S} = 0.245T - 68.52$	$\hat{T} = 3.95S + 318.56$
VI	$\hat{S} = 0.23T - 15.55$	—