

長野県植物誌データベース始末記

金井弘夫

Hiroo Kanai: Database retrieval of the flora of Nagano Prefecture

長野県植物誌の刊行は追い込みに入った。私も分布図作りの他に、「長野県植物誌とコンピュータ」と題する一章を担当しており、その中でデータベースの維持や訂正の問題点を述べておいた。しかし限られたスペースには入り切れない雑多な問題も多く、また送稿以後にわかった問題も少なくない。とくにレコードのすべてを、人の目に触れるような形に整えることについては、分布図の見かけを良くするようなキレイごとでは済まされない問題がある。こういうことは長野県に限らず、データベースを構築維持しようとするれば必ずぶつかる問題で、一般性が大きいと考えるので、植物誌の中身と一部重複する点もあるが、トピックごとに記しておく。

1. メッシュシステムと境界線の問題

植物誌分布図の表示メッシュは5倍メッシュである。このサイズは1/5万図の1/16で、私が提唱した Locality Index に相当する。一方、地図に描かれた県界、海岸線、等高線は、位置を分の小数まで計算した値をそのまま描画している。

植物産地の位置表示にメッシュシステムを採用する得失については、ここでは議論しない。理解して欲しいのは、真の位置がメッシュのどこにあっても、分布点はメッシュの中央に表示されるということである。長野県での5倍メッシュの大体の大きさは、東西5km、南北4.6kmである。分布マークの周囲に、これだけのボカシを入れて考える必要がある。メッシュを小さくすればこのボケは小さくなるが、だからといって「正確さ」が増すわけではない。位置決めの誤差が大きくなるからである。

分布図の修正要求の中に「分布点が県界の

外に出ているから、中へ移動させよ」というものが結構多かったのだが、これには応じられなかった。前記の事情を考えれば、これはメッシュシステムを採用した必然の結果なのである。研究者の中には、「自分の縄張りを守る」「他人の縄張りを侵さない」という自己規制にこだわる人が少なくない。県界の外の分布点は「不自然」と見るより、「モラルに反する」と意識されたのではあるまいか？手作業で分布図を作るならば、こういう分布点を県内に移動させることは難なくできる。実際、メッシュシステムを採用したと称する分布図の中には、境界外に分布点が全く現れないものが見られる。私に言わせれば、これはきわめて不自然である。

生物の分布は人為的な境界線には無関係である。もし隣の県の分布図も上記のような「配慮」に基づいてつくられていたら、二つの県の分布図をつなぎ合わせたとき、県界に沿って分布の空白地帯ができるに違いないのである。県分布図だから県界しか描かれていないが、市町村界が描かれるような分布図だとどうなるだろう？

メッシュ分布図は、分布パターンを示すためにだけあるのではない。パターンを示すのなら、手作業による「精密な」分布図の方がずっともっともらしいものが出来る。しかしメッシュ分布図は、それを使って分布を数量的に扱うことができるということの方が、はるかに大きな利点なのである。新潟大学の石沢進氏のお仕事をはじめ、新潟県植物分布図集に挿入された多くの報文を見れば、このことは理解できるだろう。メッシュ分布図がなければ、こういう研究は不可能なのである。長野県植物誌の分布図は、刊行が終わればオシマイな

のではない。それを使ってこれからどんな研究をするのか、会員の方々の工夫が期待されるのである。

2. 端点情報

前項の「県外の点」は、メッシュシステムに由来する宿命である。しかし分布図を描くと、「県外の点」はやたらにたくさん現れる。県界より1メッシュ以上離れていれば、位置座標のつけ誤りにちがいないので、訂正せねばならない。これについては植物誌の中で触れておいたし、あらためて別に書くこともあろう。では点が県内に納まればそれでよいかというと、そうはゆかない。

ためにしに地区コードごとにまとめて分布図を描いてみた。地区コードは市町村を意味するので、分布点は小さい範囲に集中するはずなのに、どの地区コードについてもおそろしく拡散してしまった。もちろんエラーであり、地区コードか位置座標のどちらか（ことによると両方）が誤っているはずである。わざわざこういうことをしてみるからエラーがわかるのであって、実際には分布図上でもデータリスト上でも、見つけようがない。植物誌の段階ではこれを放っておいても頼かむりできる。環境庁(1985)がこの類である。しかし全件リストを公表しようとする、あるいは市町村の植物リストを作るとなると、ボロが出てしまう。実はチェックリストで既にボロが出ているのだが、誰も気づかないだけのはなしである。

植物誌の私の文章の中に「地区の端点」の表がのせてある。これは国土地理院(1995)によるもので、市町村の東西南北の端の位置と役場の位置、それに市町村の「重心」の位置がリストされており、この中から長野県関係を抽出したものである。たとえば堀金村は

東端(137° 53' 54" E, 36° 17' 17" N)

西端(137° 43' 27" E, 36° 17' 36" N)

南端(137° 48' 07" E, 36° 15' 42" N)

北端(137° 45' 33" E, 36° 19' 28" N)

役場(137° 52' 45" E, 36° 17' 30" N)

重心(137° 48' 18" E, 36° 17' 37" N)

(これは表にはのせていない)

である。だから堀金村(MAD)と自称するレコードの位置座標が、上記の範囲に納まらなければ訂正する必要がある。うるさくいえばこれだけでは済まないのだが、さし当たり大きなエラーのチェックはできる。この端点情報をデータファイルに作っておいて、プログラムによってすべてのレコードと比較し、範囲外の産地をリストしたら約1500件あった。これは地名数であって、関係するレコードは12000件程度あると思われる。

分布図作りのために送られてくるデータの中には、位置座標がついていなかったり、明らかに誤っているものが少なくない。しかしそれらを一々資料で調べたり、執筆者に問い合わせている余裕はないので、あらかじめデータベースの全ての地名、位置座標、地区コードを抽出ソートしたリストを作っておいたが、これはなかなか便利な代物だった。しかし上のような操作をやってみると、この地名リストの中に正誤とり混ぜて入っていることになる。たいていはエラー産地と共に「正しい産地」情報も含まれているので、それらを比較すればかなりの訂正はできると考えた。やってみたらいろいろな種類のエラーがあった。

産地を代表するデータとしては、地区コード、地名(よみ)、位置座標の項目がある。原資料(たとえば標本ラベル)を読み取って調査票を作り、データとして入力が終わるまでに、エラーが起こり得るチャンスは次のようにたくさんある。

1. 原資料の誤り
2. 原資料の読み誤り
3. コード表の誤り
4. コード表の読み誤り
5. 調査票に記入するときの誤り
6. 調査票の読み誤り
7. 入力時のミスタッチ

エラーデータがみつかったとき、その回復

に三項目のうちどれを最も頼りにしたらよいかということである。私は地名（よみ）＞位置座標＞地区コードの順に頼りにすることにした。産地のようにコードを用いない場合には、3と4は省かれる。地名の場合には、片仮名といえどもその文字列の素子にはある程度の前後関係があるので、エラーの発見も訂正もかなりできるのに対して、コードの場合には、字列の素子の前後関係がきわめて希薄だから、発見も訂正も困難なことが多い。これが地名項目に一番の信頼度を置いた理由である。

コード化は一見合理的に思えるけれど、人間が扱う限りエラーが起り易く直しにくく、おまけに一々コード表を参照しないといけないうので手間がかかるという点で、そんなに便利なものではない。コードはコンピュータがデータを扱うときに、（頭が悪いから）やむを得ず人間が手間をかけて与えてやるのであって、人間がデータを扱うときにはむしろ邪魔である。ただ、人間は大量のデータを短時間に処理する能力がないので、機械にやらせるのである。

地名よみは、原資料をそのまま読んで調査票に記入すればよいし、読み取りの際読み誤る可能性が最も低いと考えられるので、一番頼りにしてよいだろう。片仮名なのでキーボードのミスタッチは起り易いが、たとえ間違っても、前後の文字のつながり具合から誤りの箇所を見つけ、復元することは困難ではない。

位置座標は数字なので、誤るチャンスはたくさんある。一つは地図なり索引なりから位置座標を読み取るときに、数字を読み誤ることで（ケース2）、これはベテランでも起り得る。本当はその前に、地図や資料に記入された位置座標やコード自体が誤っていることもあり得る（ケース1, 3）。ケース6は極めてポピュラーな問題で、記入者が何気なく記した文字が、入力者に読めないならよいのだが、読み誤ってしかもそれに気づかないとい

うことが頻繁に起こるのである。これは記入者の側の責任なのだが、よほど訓練されていない限りそういう配慮はなされない。とくに丸文字族のバイトさんには泣かされるのである。たとえば1と2と7、7と9、6と0、4と9と7、4と0、4と1などは、誤読されかねない字体になり勝ちである。入力時のミスにはキーの打ち違いのほか、数字の順序を誤る例もかなりあった。数字というものは前述のとおり、文字と違って前後の数字とのつながり具合で訂正するということは原則として不可能である。しかし長野県の位置座標、それも Locality Index となると、その変動の範囲は決まっているので、それをはみだした値は見つけることができる。たとえばコードの第1、4字目には3以外は出てこないし、7、8字目には1～4以外はあり得ない。それに、前記のような地名索引ができていれば、同じ読みに対して与えられる位置座標の組み合わせは知ることが出来るので、それを使って妥当な数値に直すことはかなりの場合できる。

地区コードの誤りには最も問題が多かった。このコードは産地から与えられるものなので、ラベルに市町村名が入っていればコードは一義的に決まるはずである。データベースはスペース節約のため、市町村はコードのみで記録されているから、それだけを見ても正誤はわからない。原票をたどれば訂正できるはずだが、そんなことは理屈であって実行できまい。またラベルの記録者が、産地の市町村名を正しく把握していなかったと思われる、まとまったエラーも少なくない。それよりは、元資料に市町村名が入っていないため、調査票記入者の段階でそれを入れる場合の方が多かったのではあるまいか。そうすると調査票記入者の知識に左右されることになる。たとえば浅間山とあれば、直観的に軽井沢と連想してコードをつけやすいだろうが、県内に浅間山が一つしかないとは限らない。それに地区コードを軽井沢(KSB) とつけてよいかどうか問題なのだが、とにかくコードをつけ

なければ、作業が先へ進めないのである。私のように現地の地理に暗い人間には、こういうことの判断はつかない。けれどもたとえばハクバダケが SIE (大鹿村) となっていれば、首をひねらざるを得ない。こういう例は非常に多かった。これについても、他の「正しい」地名・位置座標と比較して、地区コードを書き換えた例は少なくない。たとえば、「ノリクラダケ」に対して以下のような地区コードと位置座標のセットがあった。

KAC37436513(小谷村),

KAD37436513(白馬村),

MAB37336113(安曇村)

MAB37336123(安曇村),

MAC37436513(穂高村)

最後の穂高村が誤りで、どう訂正するかは位置座標を他の同じ地名のレコードと比較した結果地区コードをKAC(小谷村)に書き替えた。エラーの原因の一つは、調査票のローマ字がまぎらわしかった可能性が高い(ケース6)。それともう一つの心配は、記録時と現在では町村合併の結果地区名が変わっている、あるいは地区名が同じでもその範囲が変わっているケースがあるかもしれないのである。そういう記録に現況をあてはめてエラーを見つけ、訂正してよいものか迷ってしまう。地区コードのとくに先頭二字は、位置を大きく変えてしまうので、後日地区別の一覧表を作ろうとすると、問題が生ずるだろう。KOCがKACになれば、王滝村が小谷村になってしまうのである。数字と同様、読み誤りが多かったものに、CとE、OとD、DとB、CとO、MとNとH、TとIとJなどがあつた。これも位置座標と同じく、記入者が配慮すべきことなのである。

ついでに書いておくが、私のところで今回問題にはなっていないが、片仮名ではもっと問題が大きい。書く人間が書き方とくに書き順を教えられていないようである。だから学生が植物名を書くとき、シとツとミ、ンとソとり、アとマとスとヌ、チとケ、ワとフ、ク

とケ、コとユなどの区別がなくなりつつある。教育にワープロが導入されるようになれば、救いがたい状態になるだろう。先日東京は青山の大通りで、「〇〇スポーシ」というのれんが堂々とさがっているのを見た。

3. 一つの分布点にはたくさんのレコードが隠されている

メッシュシステムによって表示される分布点は、その単位メッシュ内にどんなにたくさん産地があっても一つしか表示されない。もちろん、分布量を示したり、レコードの種別を示したいときにはそれなりの工夫がされるが、「複数の産地(重複を含む)を一点で代表させる」というのがメッシュ分布図の要諦である。そうすることによって、前に述べた「分布を数量的に扱う」ことができるようになるのである。

分布図の訂正要求で、「この点を消せ」という場合、指示した人は一つの産地の一つのレコードを想定したのだろうが、実際にはそうでない場合の方が多かった。同じ位置座標をもつ20件ものレコード(それも異なる産地の)を消したこともある。

一方、特定の標本について、「これは誤りだから削除せよ」と指示しても、分布点が期待通りに消えない場合が少なくなかった。複数のレコードが関係しているからである。標本をみんな削除しても、野帳のデータが残ることもあつた。だから、分布図上で気に入らない点を削除するためには、元になったレコードをたどった上で判断する必要があるのだが、この手続きを十分に行うだけのヒマはなかった。

水平分布図の点が垂直分布図に反映されていない、という指摘があつたが、元のレコードに高度が入っていなければ、これは当然起こることなのである。こういうケースはずい分多かった。

4. 削除されたレコードはどうするか?

「削除せよ」と指示されても、私の方ではこれ迄述べたいろいろな理由があるので、そ

れを素直に実行するわけにはゆかない。削除やむなしとしても、「分布点の削除＝レコードの消去」でよいのかどうかを考えねばならない。データベースの削除は書き物とちがって、赤線を引くわけではない。消去してしまえば、そのレコード、ひいてはその元になった標本・資料は存在しなかったことになるからである。

削除の理由は大きくみて、「分布位置が違う」場合と「同定が違う」場合がある。前者なら、位置を正せばよい。だから、消去する理由にはならない。後者ならば植物名を正せばよい。結局どちらにせよ、レコードを消去する理由にはならないのである。しかし「削除」の指示はただそれだけであって、「こう正せ」という指示はほとんどの場合伴わなかった。本当は、同定も位置も「正しい」のに、「そこにはない筈だ」と削除されたものもかなりあるようだ。こういう資料は訂正のしようがないから、永久に浮かばれない「さまよえるオランダ人」になるだろう。しかし植物地理学的には、こういうレコードこそ再検討に値するものと思う。

一番困ったのは、垂直図上の点の削除を求められた場合である。指示した人はその高度の産地を否定したいわけだが、東西図と南北図の対応する点が指示されていれば、水平図上の点を特定でき、その高度をもつレコードを削除することができる。しかし垂直図の一方だけで指定されたり、二つの垂直図の指示が水平図の一点を指示していなければ、多くの場合、その高度を持つ複数の位置座標の点が削除の該当者になってしまう。そうすると垂直図の訂正だけをしたつもりなのに、水平分布図に予期しない変化が起こってしまうのである。指示する側から見ると、違っていることはわかっても、元のレコードなり資料なりを見ていないから、「こう正せ」と言えないわけである。もとの資料にたどりついたにしても、それが野帳の記録であれば、それ以上追求できない。文献資料にしても同じである。

だから訂正は「保留」せざるを得ない。

標本についてはもっと深刻である。いくら削除といわれても、標本は実在し、「誤った」名前なり位置座標がついているのだから、データベース上で「消去」しても、標本・資料の存在は抹殺できない。それどころか、いつか「こういう標本があるのに、見落としか？」と言われるだろう。標本の名前を付け替えてもらわねばならないのだが、差し当たりそこまでは無理である。つまりこれも「保留」ということになる。

結局、削除を指示されたレコードは、その植物名を取り外して保存し、将来の訂正を待つというのが、一番無難であるということになった。名前を取り外すといっても、真っ白にしまうより、以前はどんな名前だったかがわかる方が、後でそういうレコードを見つけるのに都合がよいだろう、と思った。そこで和名の先頭を一字空白にした。科名、学名は元のままである。こうしておけば植物名でまとめるときに不便はないし、こうして「隠居」させたレコードも、元の名前が大体わかる。大事なことは、科名や属名が保存されているため、近縁の植物のレコードと離れないでいてくれることである。削除ということで全部を消してしまえば、折角入力したレコードが無駄になる。誤同定だからと和名学名を空白にすれば無縁仏になってしまい、そのレコードをあとで訂正しようとしても、見つけ出すことはほとんどできないだろう。

和名の前を空白にすると、和名ソートのとき隠居レコードが先頭にかたまって面白くない。こういうヤクザなレコードはうしろに引っ込んでいてもらいたいのである。だから空白の代わりに、片仮名よりコードの大きい文字をもってくる必要がある。隠居させたレコードは1054件ある。これらはいずれ、指示した方に訂正をお願いすることになるだろう。さもないと、折角作ったデータが役立たずのまま眠ってしまうのである。これらは分布図にかかわるものだけだから、実際には隠居さ

せるべきレコードはずい分多いにちがいない。

5. 分布図に点が追加された場合

水平分布図上で「ここへ点を打て」という指示が少なくなかった。わざわざ地図上の位置を割り出して指示しなくても、その点に該当する資料の産地、高度、採集者、日付、標本番号などの詳細が付記されていれば、新しいレコードを作ることができておたがいに楽なのだが、しかし互いにまだこういうことに慣れていないものだから、手間のかかるやり方になり勝ちである。多くの場合は、点の付加の指示だけだった。その点をつけ加えないと、分布図に不都合が生ずるのだから、とにかくその位置座標をもつレコードを追加した。このレコードは位置座標と資料種別と植物名以外のデータを持たない。これでは分布図のカッコウはついて、資料リストでは空白となり、数は少ないがたいへんみっともないものになる。しかも資料種別は、ほとんどの場合「標本」である。これもいずれは指示した方に、データの詳細をお伺いせねばならないのである。片方の垂直図上だけでの追加指示は、位置座標さえ決められないので、追加のしようがなかった。

6. データベースの追加訂正と分布図

前述のとおり、データベースには様々なエラーが含まれている。エラーが発生したことに文句をつけるつもりはない。人の手でデータベースを作るのだから当然の結果で、コンピュータで処理しているからこそ、エラーの発見も訂正もやり易いのである。そこで訂正できるものはたえず訂正してゆく。また新しいレコードも、受け取り次第追加される。だから、執筆者の要求に従って分布図を訂正しても、再出力するときには前とは異なるレコードが含まれている可能性がある。そうすると「前回の分布図と違う」とか、削除したあとに追加データが顔を出して「指示通りに訂正されていない」ということが起こる。これはデータベースが絶えず更新されてゆく結果で、辛抱強くつきあっていただく他はない。植物誌データベースには「これで終わり」ということはないのである。こうやって直して行くと、データベースと標本ラベルとに違いが出てくるので、そちらの方の手当ても考えねばならない。

引用文献

環境庁 1985. 生きもの地図ができました. 96pp.
国土地理院 1995. 日本の市町村位置情報要覧. 358pp.

〔植物ニュース〕

ヒロハヌマガヤ *Diarrhena fauriei* (Hackel)
Ohwi

日本のイネ科では希品中の希品（日本イネ科植物図譜 長田武正 1989による）とされるヒロハヌマガヤは、長野県の下伊那郡大鹿村と南佐久郡南相木村、北相木村だけで確認されている貴重種である。

昨年9月24日、上田市誌自然編の植物調査の際に、上田市上室賀の大沢池（標高600m）の道端の林下で1か所に直立した茎が十数本見られた。周辺のくわしい調査はしていないので、ほかにもあるかどうかは不明であるが新産地として報告する。（篠原 修・宮下昇三郎・池田登志男）

〔植物ニュース〕

シナノアキギリ *Salvia koyamae* Makino

シナノアキギリは南佐久郡小海町松原湖畔で小山海太郎（旧制野沢中学校－現野沢北高校教師）によって発見され、長い間、佐久特産とされていたが、その後群馬県側でも発見されている。現在佐久では10か所ほどの自生地が確認されているが、県下では佐久以外からは報告されていない。

昨年5月3日・9月28日の上田市誌自然編の植物調査によって、独鈷山（上田市東塩田）の平井寺コースの標高1000～1100mあたりの数か所に大群落を確認した。佐久では、これほどの群生地は見えていない。（篠原 修・宮下昇三郎・池田登志男）