

可視化関数と学生の興味

日本大学・生物資源科学部 五十嵐 正夫 (Masao IGARASHI)*
College of Bioresource Sciences, Nihon University*

1 はじめに

計算機に初めて手を触れたのは、今から 45 年以上前の学生時代である。授業科目は「数値解析及び演習」で、4 単位選択必修科目であった。宇野利雄先生が講義、永坂秀子先生が演習を担当された。連立 1 次方程式の解法やニュートン法などを習った。メモリーの節約方法や理論と実際との違いが、面白かった。使用した計算機の名前は FACOM128B である。メモリーの考え方にもよるが 128 は 128 ワード程度である。演算素子はリレーであるため、計算が止まると、そのリレーを引き抜き、接点を磨く必要があった。磨き直して、再度差し込むと足踏みしていたリレーの音がリズムカルに鳴り出した。今回の研究集会のプログラムで、同僚の鈴木潔光先生が FACOM128B の計算精度等の仕様書等について話されたのを拝聴し、大変感慨深く、私的なことで恐縮ではあるが記した。

学生時代に計算機に触れたことが縁で、農学・生命系学部的一般教育、専門教育、大学院教育のなかで情報処理関係の講義や指導に当たることになった。計算機に対する学生の興味は低く、教室の設備は今日とは比較にならないほど貧弱であった。初期の段階でのプログラム言語は BASIC と FORTRAN のみ、8 インチの FD を入れ替え差し替えし、教える内容は現行の高校数学 B の「統計とコンピュータ」程度であった。

それが、今から 25 年前ぐらいから、学生一人ひとりにパソコンがいき渡るようになった。使いやすさ、計算速度、メモリー、アプリケーションソフト数等が飛躍的に増加、進化し、一昔前のスーパーコンピュータが各家庭に設置され、しかもネットワーク化されたような時代がやってきた。

そのような急激な変化の中で、情報処理や数式処理、あるいは数値計算を専門としない学部における「可視化による情報教育」についての現状と課題について考察する。

2 教室環境と授業科目

所属学部は、農学・生命系で 11 学科、入学定員 1410 名、1 年生のほとんどが情報科学関係の授業科目を履修している。コンピュータ教室は、大教室 (150 名) 1 つ、中教室 (70 名) 2 つ、小教室 (20 名) 1 つである。従って、一度に 300 名程度の受講ができるようになっている。授業で利用できる情報科学関係のアプリケーションソフトには、語学系アプリは別として、Office, Mathematica, Minitab, Adobe, 一太郎などがある。教室内設備は、かなり整っている。通常 3 名程度の TA (大学院院生) が教員のアシスタントをする体制になっている。

授業科目は、情報科学、情報処理論、数理情報科学、情報処理演習などである。コンピュータ教室は統計学関係の授業にもよく利用される。内容の概略は次の通りで、何れも半期 15 回授業を原則としている。

- (1) 情報科学：学部固有の情報リテラシー教育
- (2) 情報処理論：学科固有のデータ処理法教育
- (3) 数理情報科学：学科固有の工学系解析
- (4) 情報処理演習：学科固有のデータ処理演習

情報リテラシー教育は、学部のネットワークシステムや図書館にログインできて、セキュリティーを学び、Word でレポートを書き、Excel で図表を作成し、Power Point でプレゼンテーションができれば、ほぼ終わりとなる。

情報処理論は、間口がたいへん広い。情報を処理する道具が大変使いやすくなり操作手間が省け、機器の立ち上がりが早くなったことは、ありがたいことである。しかしながら、例えばプログラミングを教える手間が省けるようになったことは「学びたいことだけ学習する」には効率良くなったが、知識の奥行きと幅の狭い、いわゆる「議論でかつち的」な学生を世に送り出しているのではないかと反省することもある。

数理情報科学では、生物成長曲線や熱伝導方程式などを例に取り、微分方程式の数値解法の講義などを行っている。Mathematica が利用できるため、プログラムもすっきりし、また、可視化も可能となり、学生の理解も進んでいるようである。

情報処理演習は、統計的な計算が多くなる。Excel の利用頻度が高くなるが、分布関数のや累積分布関数の形状、上側確率と言ったことを、図的に理解することはなかなか難しいようである。特に 2007 と 2010 の統計関数の違いは、学生のみならず、教師も混乱することがある。表や図を用いず、単に計算だけで解答を得ようとする、失敗することがある。もちろん Mathematica でも正規分布以外の t -分布や χ^2 分布を学生に理解させながら、図示することは面倒なことである。特に、SHOW で 2 つのグラフを重ね合わせるとき、2 番目のグラフが優先し、1 番目のグラフの一部が切れてしまう場合がある。

3 可視化の具体例

コンピュータ教室では、学生が授業中にどんなサイトに接続しているか、教師や TA は端末から常時見ることができる。どこの大学でも見られるように、授業に興味を持てなければ、学生は思い思いのサイトで検索という名の遊学に出てしまう。それを防止するには、興味ある例題を選び、その結果が自ら考える「素材」、特に競争意識を醸し出す素材となることが大事であると思える。

農学・生命系学部であるため「気温」を題材として選び、気温のデータが非線形であるため、非線形フィッティング問題をテーマとした^[1]。

3.1 目的

- (1) 温暖化を自分なりに考える素材を見つける。
- (2) 線形回帰と非線形フィッティングの違いを理解する。
- (3) R^2 の値がなぜ、非線形フィッティングに関しての当てはまり具合に準用されるのかを理解する。

3.2 材料

気象庁の公開している「気象統計情報」のなかの「過去の気象データ検索」項から、都道府県の「観測開始からの月ごとの値」を選択し、データを得る。ここでは、南極昭和基地の1967年からのかく月の最高気温と最低気温と、日本の平均気温を決める17観測所を例に挙げて説明する。

3.3 方法

- (1) 気温データに適合するモデルを決定する。
- (2) 気温は年毎に線形的上昇しているので線形項を入れる。
- (3) 気温は月毎に周期性を持つため周期項を入れる。
- (4) 計測誤差項は無視する。
- (5) カタストロフィー的な項は無視する。
- (6) モデルのパラメータ係数決定には最小2乗法とニュートン法を用いる。
- (7) 簡単のため Mathematica の `FindFit` 関数を利用する。
- (8) モデルの当てはまり具合は、データとそれに対応する数値解の相対誤差の算術平均を用いる^[2]。

4 結果

次の形のモデルを学生に提案する。教員は試行錯誤的にモデルを選ぶわけであるが、なぜ「そのようなモデルを選んだか」については詳しくは説明しない。ただ、水文学等の分野で、それに近いモデルのあることは説明する。

気温を時間 t の関数 $x(t)$ とする。データが月単位で与えられるため、 t は月を表すことを強調する。

$$x(t) = \underbrace{a_1 + a_2 t}_{\text{トレンド (年)}} + \underbrace{a_3 \sin(a_4 + \frac{\pi}{n} t)}_{\text{周期 (月)}} \quad (1)$$

ここで a_i , $i = 1, 2, 3, 4$ は未知のパラメータ, n は $1 \leq n \leq 100$ の範囲の整数とする. $\sin(a_4 + \pi t/n)$ の基本周期は $2n$ であるため, $n = 6$ が最適解を与えれば, 12 か月で気温が周期的に変化することになり, モデルはデータに良く当てはまっている可能性がある.

実際, n を 1 から 100 まで反復し, それぞれの n に対して, データと数値解の相対誤差の算術平均を求めると $n=6$ の時が最小で 0.08 程度となり, データと数値解は平均して 10 進で 1 桁以上一致していることが言える.

昭和基地の気温の例

気象庁は 1957 年から南極昭和基地の月平均気温, 月最高気温, 月最低気温などをネット上で公開している. ここでは, データの欠落ない 1970 年 1 月から 2009 年の 12 月までのデータを利用する. 暫定値気温はそのまま利用した. データ数は最高気温, 最低気温とも 480 個である.

Mathematica の FindFit 関数を用いると次のような最高, 最低気温に関する非線形フィッティング関数が得られる. ここで n はデータとその予測値の相対誤差の算術平均が最も小さくなるように決めた.

$$x(t) = -7.68598 + 0.000342789t + 8.39044 \sin(0.909812 + \pi t/6) \quad (2)$$

$$x(t) = -14.0793 + 0.00103592t + 9.25668 \sin(0.845499 + \pi t/6) \quad (3)$$

480 個のデータとそれに対応する数値解 (予測値) との相対誤差の算術平均は 0.08 と 0.07 であった. この結果から, データと数値解は 1 桁以上一致していると言える.

データ (ドット) と式 (2) を Plot を用いて図示したのが図 1, データ (ドット) と式 (3) を Plot を用いて図示したのが図 2 である.

応用

次にデータを次のように 4 区分して, 各区分ごとのフィッティング関数を計算する.

- (1) 1970 年 1 月から 2009 年 12 月
- (2) 1980 年 1 月から 2009 年 12 月
- (3) 1990 年 1 月から 2009 年 12 月
- (4) 2000 年 1 月から 2009 年 12 月

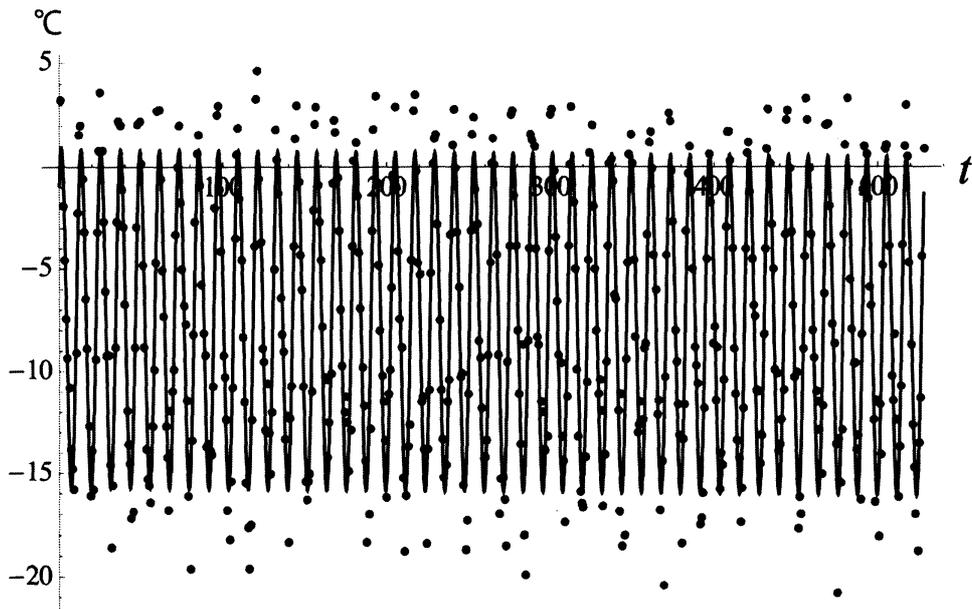


図 1: 南極昭和基地の最高気温のデータ (ドット) と得られた曲線

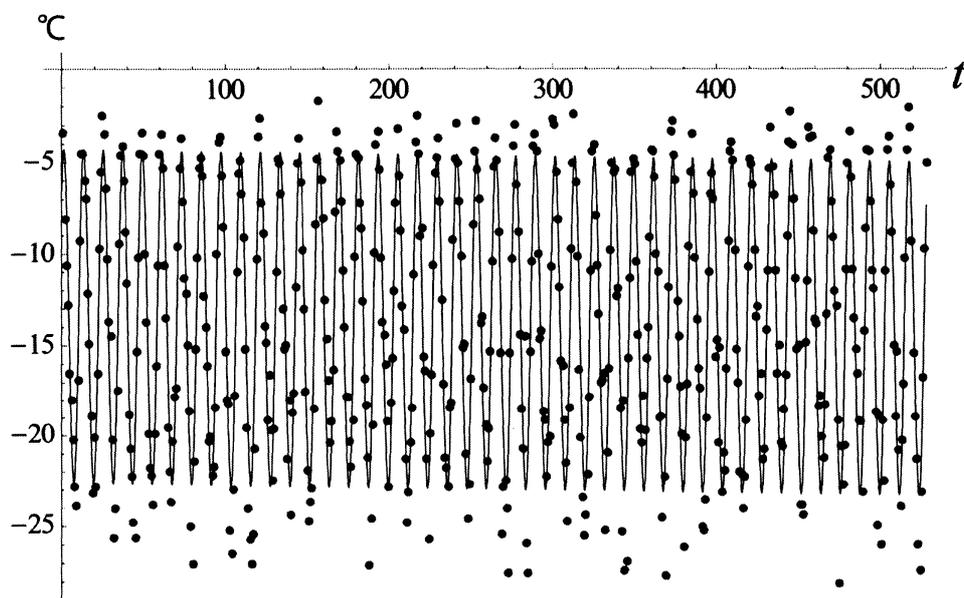


図 2: 南極昭和基地の最低気温のデータ (ドット) と得られた曲線

図2は得られたフィッティング関数の1次項 t の係数から、100年間で気温が何度上昇するかを示したものである。

例えば1970年のMAX上の0.41は上の1区分のデータを用いると100年後には1970年の最高気温よりも0.41℃上昇、同様に最低気温では1.24℃上昇するという意味である。

第2区分の1980年から2010年までのデータを用いると、最高気温では0.55℃、最低気温では2.23℃のマイナスとなる。

4区分をまとめたのが図3である。このデータからは、南極では10年ごとに、推定気温がプラス、マイナスと交互に変化していることが理解できる。

実際のデータを基に、簡単ではあるが「温暖化」の指標となるデータ区分の大切さ、すなわち利用するデータ区分によっては、恣意的な結論を導き出せる可能性がある、ことを理解させることにしている。

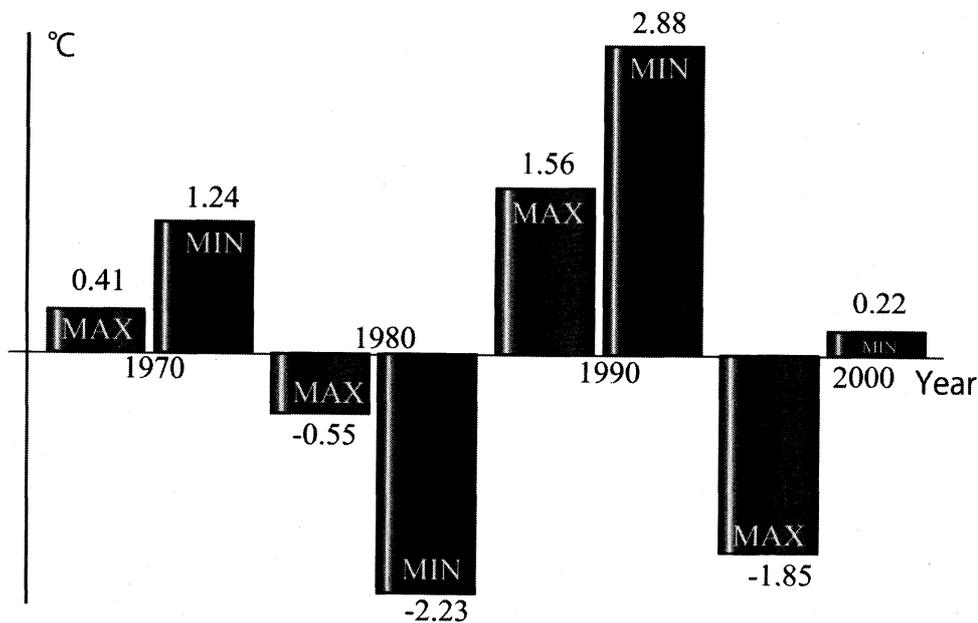


図3: 最高気温と最低気温の100年間の推定気温

発展

気象庁の公開データには日本各地の観測所の気温も公開されている。海洋観測所を含めると主なものでも57か所になる。例えば東京では大手町が測定場所として選ばれている。すぐ分かるように、様々な人為的影響を受けやすい所である。それを考慮して気象庁は日本の平均的な気温を定めるための次の17か所を選んでいる。

網走、根室、寿都、山形、石巻、伏木、長野、水戸、飯田、銚子、境、浜田、彦根、宮崎、多度津、名瀬、石垣島

これらの観測所の平均気温から、10年区分ごとのデータを用いての気温変化を推定すると図4になる。この結果は次のように読み取れる。

- (1) 1880年から2010年までのデータとモデル(1)を用いると100年間で0.60℃上昇。
- (2) 1890年から2010年までのデータとモデル(1)を用いると100年間で0.89℃上昇。
- (3) 1900年から2010年までのデータとモデル(1)を用いると100年間で1.08℃上昇。
.....
- (4) 1970年から2010年までのデータとモデル(1)を用いると100年間で2.63℃上昇。
- (5) 1980年から2010年までのデータとモデル(1)を用いると100年間で3.09℃上昇。
- (6) 1990年から2010年までのデータとモデル(1)を用いると100年間で0.2.00℃上昇。
- (7) 2000年から2010年までのデータとモデル(1)を用いると100年間で1.09℃上昇。

図からは1980年まで平均気温が上昇傾向、その後下降傾向となっていることが読みとれる。

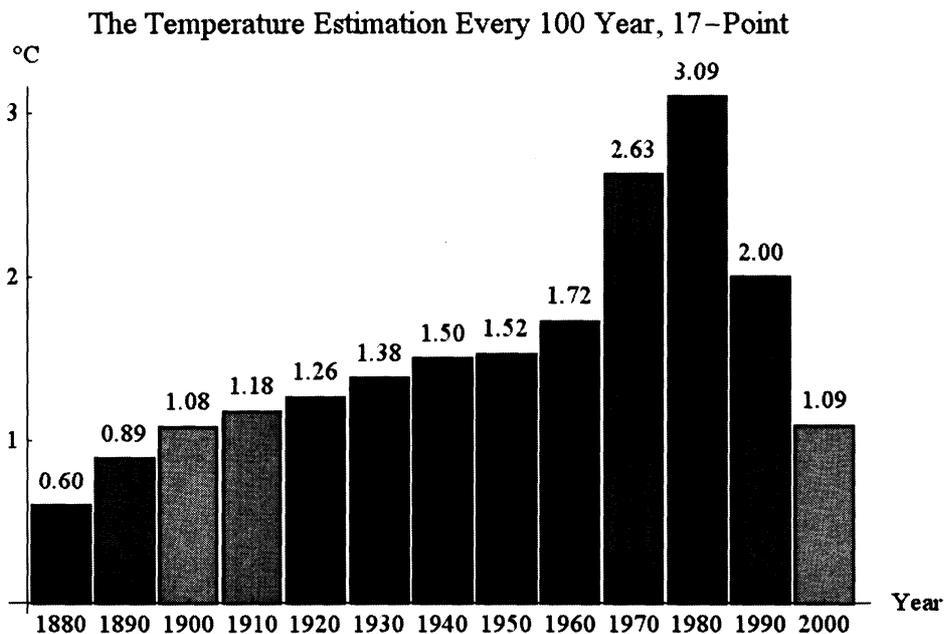


図4: 日本の平均的な気温変化推定

5 おわりに

本稿では、学生の興味と可視化の観点から気温を選んで、「数学ソフトウェアと教育」を考えてみた。

農学・生命系の学部では、非線形なデータに対してそれに良く適合する曲線が必要とされる場合が多い。例えば植物や動物の生長や、陸地、海面の気温データなどは、多くの例において非線形な曲線となる。この場合、次の3つの問題点がある。

- (1) 農学・生命系の学生は、プログラムを書くことに慣れてないために、データ解析入り口でギブアップすることがある。
- (2) 誤差や精度に関する知識が少ない。この原因はおそらくそのようなことを学習する機会がなかったためと思われる。
- (3) 実験やデータの計測・収集に追われている。この傾向は学部では3年生頃から、大学院ではその傾向に拍車がかかっている。

このような点を解決するのに「数学ソフトウェア」は大変便利であり、十分実用的である。巨大な潜在能力を持った「赤ちゃん」が、大活躍し始めた感じがする^[3]。

その一方で、理論ばかりの「頭でっかち」の学生さんから、自己主張の強い「議論でっかち」の学生さんが増えてきたことには注意を払う必要がある。分けも分からずにアプリケーションソフトを利用し、その結果を鵜呑みにするからである。特に統計ソフトを利用する場合は「仮定があつて定理がある」ことや「標本がその仮定をほぼ満足していることを検証したか」と言ったことを注意することにしていく。上滑り的な知識のみの「アプリでっかち」的な学生さんにしないような教授上の方策を常に考案し、その方策は毎年更新することが必要であると考えている。

参考文献

- [1] 五十嵐正夫, 他, Mathematica を用いての NDVI に対する非線形カーブフィッティング, RIMS, 1735, pp.191-197(2011).
- [2] S D コンテ, C. ドボアー著, 吉澤正訳, 電子計算機による数値解析と算法入門, ブレイン図書出版, 1980.
- [3] 竹之内 脩 編著, コンピュータと数学教育, 日本評論社, 1985.