

ニューラルネットワークによるサクラ開花日の 学習・予測実験 (山形・新潟を例に)

高野 哲夫*

1. はじめに

近年、人工知能技術の発展が目覚ましい。そのアルゴリズムには様々なものが発表されているが、最も基本的なものの一つとして「ニューラルネットワーク」が挙げられる。ニューラルネットワークに関しては、これまで「天気」誌上でも、例えば小泉 (1997) にて解説されており、その応用例については柳野 (1998a), 柳野 (1998b) などにて解説されている。

筆者は Iwata and Matubara (1996) および熊沢 (1998) に基づき、ニューラルネットワークを用いた数値モデル (以下、ニューロモデル) を構築し、過去の気象データの学習および予測実験を試みた。

今回、題材に取り上げたのは、サクラ (ソメイヨシノ) の開花日の予測である。このような予測は、温度変換日数を用いた統計的手法 (例えば、青野・小元 1990 など) が多く用いられている。このことから、開花時期までの気温変動のシナリオが、開花日を決定する上で重要な影響を与えていることが理解できる。

従って、開花時期までの気温変動のシナリオを入力変数、開花日を出力変数とするニューロモデルを構築し、過去の観測データを学習させた上で、サクラの開花日の予測に応用することを提案した。

本研究では、山形・新潟を例に挙げ、ニューロモデルを独自に開発し、予測実験を行うことにより、サクラの開花日予測への応用の可能性を検討した。

2. 実験の概要

山形・新潟におけるサクラ (ソメイヨシノ) の開花日の平年値は各々 4月15日、9日である。従って、本

研究では、2011年～2015年の各年において、当年の3月11日、21日、4月1日の3回に渡って開花日の予測を実施する事案を想定し、実験を行った。

まず、2月上旬から4月上旬までの各々の旬平均気温の平年偏差を入力変数、開花日の平年偏差を出力変数とし、ニューロモデルに前年までの入出力関係を学習させ、これを用いて当年の開花日の予測を実施した。この詳細を2.2および2.3にて述べる。

次に、2月上旬から4月上旬までの各々の旬平均気温の平年偏差を説明変数、開花日の平年偏差を目的変数とする重回帰分析を行い、得られた重回帰式 (以下、重回帰モデル) を用いて、当年の開花日の予測を実施した。この詳細を2.4にて述べる。

そして、上記のニューロモデルと重回帰モデルによる予測結果、および実況値の比較により、ニューロモデルの予測結果の評価を行った。

2.1 使用した観測データ

使用した観測データは、山形は1953年～2015年、新潟は1961年～2015年の各2月上旬から4月上旬までの旬平均気温、開花日および各々の平年値である。

2.2 ニューロモデルによる予測実験

今回は、2011年～2015年の各年において、予め前年までの観測データをニューロモデルに学習させた上で、当年の3月11日、21日、4月1日の3回に渡って予測を実施する事案を想定し、実験を行った。

例えば2015年の開花日を予測する場合は、予め前年の2014年までの観測データを学習させる。すなわち、新潟は1961年～2014年、山形は1953年～2014年の観測データを学習する。

学習に際しては、前年までの各シナリオについて、入力値 (旬平均気温の平年偏差) に応じた出力値と、正解に相当する実況値 (開花日の平年偏差) を比較して、その誤差を低減するようモデルの最適化を図るも

* Tetsuo TAKANO, 株式会社 SnowCast.

takano@snowcast-web.com

© 2015 日本気象学会

のとし、このプロセスを10万回反復するものとした。

また、例えば3月11日の時点での予測の際には、2月上旬から3月上旬までの入力値は既知となるが、予測実施日（3月11日）以降の3月中旬から4月上旬までの入力値は未知となる。

従って、既知の入力値はそのまま使用する一方、未知の入力値については $-3.0\sim+3.0^{\circ}\text{C}$ の範囲で 0.1°C ずつ順次変化させて適用するものとする。すなわち、未知の1旬当たり60通りの入力値を適用することになり、この条件において想定し得る全てのシナリオを計算するものである。

具体的には、3月11日の時点の予測では以降3旬が未知となるので $60^3=216000$ 通り、以下同様に3月21日では $60^2=3600$ 通り、4月1日では $60^1=60$ 通りのシナリオについて計算を実施するものである。

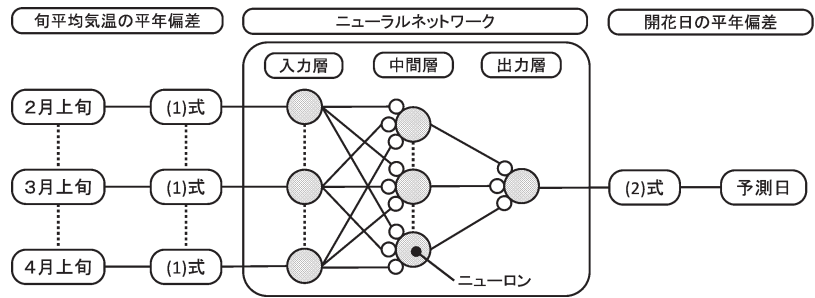
但し、予測時においては未知の入力値（1旬当たり60通り）の出現し得る確率分布を知り得ないため、本研究では $-3.0\sim+3.0^{\circ}\text{C}$ の範囲において一様な分布を仮定した。

以上の計算結果を基に、平年値の18日前～18日後の範囲で、各対象日が当年の開花日となる確率分布を求める。さらに、この確率分布を基に開花日の期待値を求め、これをニューロモデルによる予測開花日とするものである。

2.3 ニューロモデルの構造

第1図には、本研究で用いるニューロモデルの構造を示す。本ニューロモデルは入力層-中間層-出力層の3層からなるフィードフォワード型ニューラルネットワークを中枢としており、学習方法には誤差逆伝搬法（バックプロパゲーション法）を用いている。

入力層は、旬別に対応する入力変数を割り当てるため、2月上旬から4月上旬



第1図 ニューロモデルの構造。

までの7旬に各々対応する7個のニューロンを設けた。出力層は、開花日の平年偏差のみに対応するため、1個のニューロンを設けた。また、中間層には閾値の異なる29個のニューロンを設けた。

各層のニューロンの出力値はシグモイド関数を経て出力されるため、出力層ニューロンの出力値は0から1までの実数値（アナログ値）となる。本モデルではこれに合わせて、入力層ニューロンに与える入力値も同様に0から1までの実数値（アナログ値）を与えるものとした。

いま、入力値である旬平均気温の平年偏差を x 、 x が取り得る範囲を $x_{\min} \leq x \leq x_{\max}$ とする。この時、次式(1)で定義される x_* を入力層ニューロンに入力する。

$$x_* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

第1表 予測年別の重回帰式の係数および決定係数（山形）。

予測年	切片	2月上旬	2月中旬	2月下旬	3月上旬	3月中旬	3月下旬	4月上旬	R^2
2011	-0.1612	-0.3223	-0.3309	-0.2894	-0.2226	-0.8869	-0.9039	-1.0522	0.8270
2012	-0.1705	-0.3311	-0.3275	-0.3014	-0.2096	-0.8860	-0.8915	-1.0531	0.8270
2013	-0.1626	-0.3349	-0.3432	-0.3033	-0.1946	-0.8946	-0.8772	-1.0765	0.8323
2014	-0.1685	-0.3342	-0.3436	-0.2969	-0.2004	-0.8960	-0.8748	-1.0788	0.8324
2015	-0.1612	-0.3363	-0.3501	-0.2903	-0.2122	-0.8994	-0.8527	-1.0792	0.8327

第2表 予測年別の重回帰式の係数および決定係数（新潟）。

予測年	切片	2月上旬	2月中旬	2月下旬	3月上旬	3月中旬	3月下旬	4月上旬	R^2
2011	-0.0227	-0.7210	-0.5574	-0.5466	-0.4149	-0.8768	-1.0746	-0.7667	0.9365
2012	-0.0065	-0.6611	-0.5835	-0.4944	-0.4842	-0.8685	-1.1310	-0.7752	0.9349
2013	-0.0063	-0.6612	-0.5838	-0.4944	-0.4838	-0.8687	-1.1308	-0.7755	0.9362
2014	-0.0832	-0.6630	-0.5781	-0.4129	-0.5797	-0.8986	-1.0805	-0.8057	0.9270
2015	-0.1055	-0.6671	-0.5555	-0.4274	-0.5589	-0.8927	-1.1291	-0.7955	0.9268

また、開花日の年平均偏差を y とし、 y が取り得る範囲を $y_{\min} \leq y \leq y_{\max}$ 、出力層ニューロンの出力値を y^* とすると、次式(2)の関係が成立する。

$$y = y_{\min} + y^*(y_{\max} - y_{\min}) \quad (2)$$

すなわち、出力層の出力値に式(2)を適用することで、開花日の年平均偏差(予測値)に変換するものである。

本モデルでは、旬平均気温の年平均偏差の取り得る範囲を「 -5.0°C 以上、 $+5.0^{\circ}\text{C}$ 以下」と設定し、上記の式(1)において $x_{\min} = -5.0$ ($^{\circ}\text{C}$)、 $x_{\max} = +5.0$ ($^{\circ}\text{C}$) を適用した。

また、開花日の年平均偏差の取り得る範囲を「 -18 日以上、 $+18$ 日以下」と設定し、上記の式(2)において $y_{\min} = -18$ (日)、 $y_{\max} = +18$ (日) を適用した。

2.4 重回帰モデルによる予測実験

2月上旬を第1旬目、4月上旬を第7旬目とするとき、第 i 旬目の旬平均気温の年平均偏差を T_i 、係数を a_i とし、切片を a_0 、さらに $T_0 = 1$ とおくと、この場合の開花日の年平均偏差の予測値 y は次式(3)で表される。

$$y = \sum_{i=0}^7 a_i T_i \quad (3)$$

本研究では、2011年～2015年の各年において、予め前年までのデータを基に重回帰分析を行い、得られた重回帰式を、当年の予測のための重回帰モデルとした。第1表および第2表には予測年別の重回帰式の係数および決定係数 (R^2) を地点別に示す。

また、上記2.2と同様に、平年の18日前～18日後の範囲で、各日が当年の開花日となる確率分布、さらに開花日の期待値を求め、これを重回帰モデルによる予測開花日とした。

3. 実験結果

本研究では、ニューロモデルと重回帰モデルによるサクラ開花日の予測実験を行った。これらの実験結果を以下に示す。

3.1 ニューロモデルによる確率分布の結果

第2図(次ページ)には、2011年～2015年の各年3月11日、21日、4月1日の各回

におけるニューロモデルの予測結果を地点別に示す。ここで、横軸は開花日の平年値を基準 (± 0 日) とする対象日、縦軸はその対象日が開花日となり得る確率密度であり、矢印は実際の開花日を示している。

全体を通して、早い段階の3月11日の予測ではグラフの形状も扁平でほぼ水平状となる一方、後の段階の4月1日の予測ではグラフの形状も極大域がより明瞭になっている。これに伴って予測される開花の時期も明瞭になってくる。

これは、未知の入力値が一様分布になると言う仮定の影響が現れたものと考えられる。さらに、母集団の要素に当たるシナリオの総数が3月11日の予測の場合は216000と非常に膨大であるのに対し、4月1日の予測では60と大幅に絞り込まれるためと考えられる。

なお、重回帰モデルの場合でも同様の傾向が見られたが、ここでは省略する。

3.2 ニューロモデルと重回帰モデルの予測結果

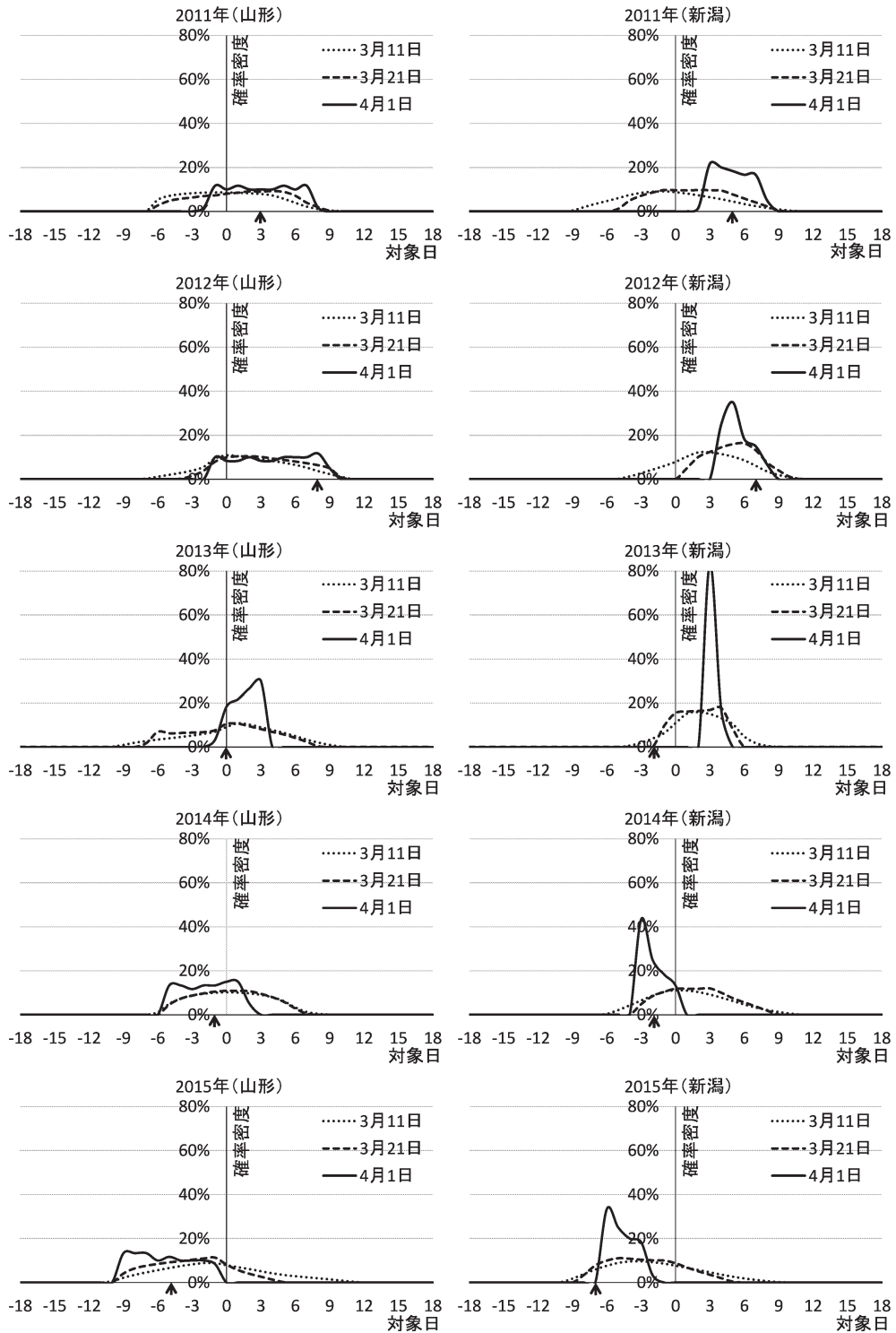
第3表および第4表には、ニューロモデルと重回帰

第3表 予測値と実況値の比較 (山形)。平年値を基準 (± 0 日) とする相対日数で表示。網掛けは、ニューロモデルと重回帰モデルによる予測開花日と比較し、より実況値に近い方を示す。太字は予測値と実況値が一致した箇所を示す。

年	予測実施日						開花日
	3月11日		3月21日		4月1日		
	ニューロ	重回帰	ニューロ	重回帰	ニューロ	重回帰	
2011	0	0	1	1	3	3	3
2012	2	2	3	3	4	3	8
2013	1	1	0	1	2	1	0
2014	0	1	0	1	-2	-1	-1
2015	-1	-1	-3	-2	-5	-4	-5

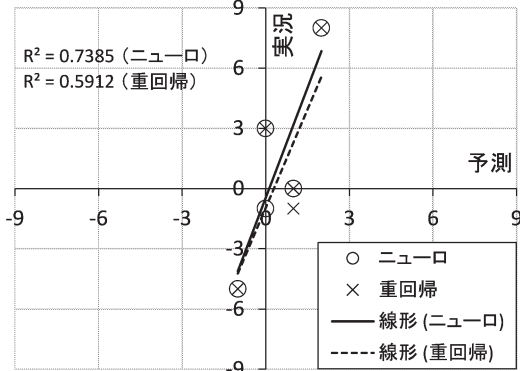
第4表 予測値と実況値の比較 (新潟)。平年値を基準 (± 0 日) とする相対日数で表示。網掛けは、ニューロモデルと重回帰モデルによる予測開花日と比較し、より実況値に近い方を示す。太字は予測値と実況値が一致した箇所を示す。

年	予測実施日						開花日
	3月11日		3月21日		4月1日		
	ニューロ	重回帰	ニューロ	重回帰	ニューロ	重回帰	
2011	0	-1	1	0	5	2	5
2012	3	3	5	4	5	5	7
2013	2	2	2	1	3	3	-2
2014	1	2	2	2	-2	-1	-2
2015	-2	-2	-3	-3	-5	-4	-7

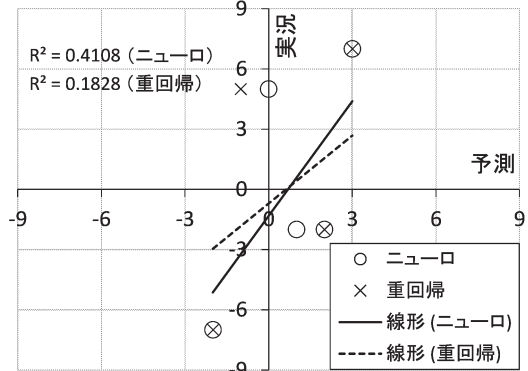


第2図 ニューロモデルにより予測された開花日の確率分布 (左：山形, 右：新潟)。横軸は開花日の平年値を基準 (± 0 日) とする対象日, 縦軸はその対象日が開花日となり得る確率密度。横軸直下の矢印は実際の開花日を示す。

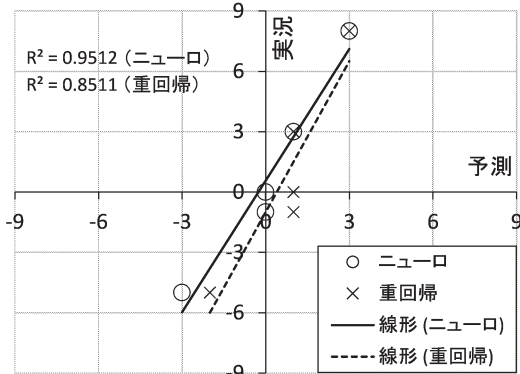
03月11日の予測(山形)



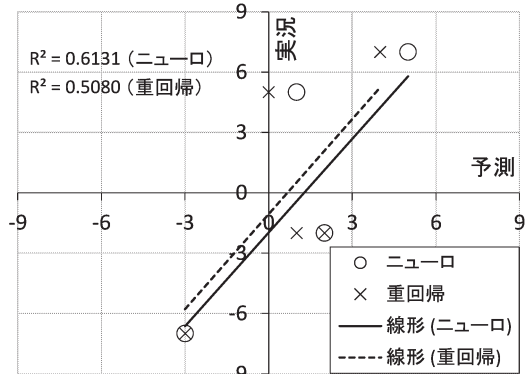
03月11日の予測(新潟)



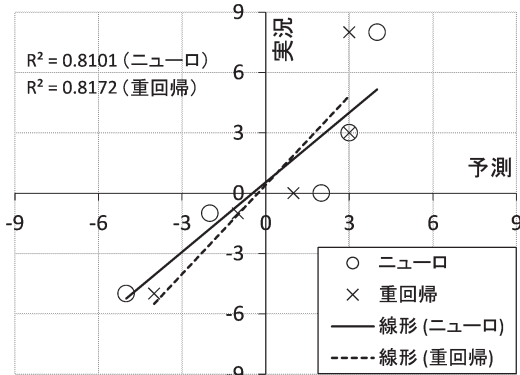
03月21日の予測(山形)



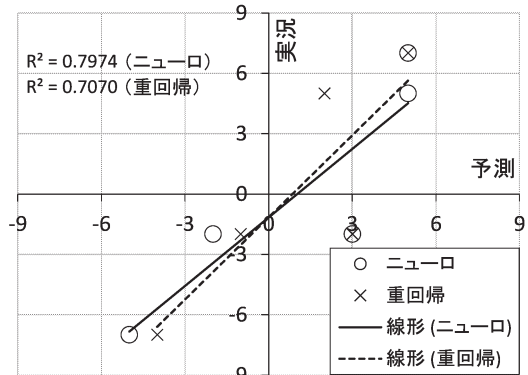
03月21日の予測(新潟)



04月01日の予測(山形)



04月01日の予測(新潟)



第3図 予測値と実況値の相関関係(山形)。横軸は予想値、縦軸は実況値であり、いずれも平年値を基準(±0日)とする相対日数で表示。

第4図 予測値と実況値の相関関係(新潟)。横軸は予想値、縦軸は実況値であり、いずれも平年値を基準(±0日)とする相対日数で表示。

モデルによる予測開花日と各年の実況値の比較を示す。いずれも平年値を基準（±0日）とした相対日数で表示している。

予測開花日は、3月11日、21日、4月1日の各回においてニューロモデルと重回帰モデルによる予測結果を比較し、より実況値に近い方に（より誤差の少ない方に）網掛けを施してある。但し、両モデルの予測開花日が等しい場合は網掛けを行っていない。

ニューロモデルと重回帰モデルの予測開花日は概ね同様の結果となっているが、網掛けの数の内訳を見ると、重回帰モデルが3ヶ所であるのに対し、ニューロモデルは13ヶ所であり、重回帰モデルに比べてやや改善されていることが判る。

また、全体を通して、早い段階の予測開花日より後の段階の予測開花日の方が、より実況値に近づいており、精度が向上する傾向が見られた。

続いて、第3図および第4図は、山形および新潟の3月11日、21日、4月1日の各回におけるニューロモデルと重回帰モデルによる予測開花日、および各年の実況値の相関関係を散布図に表したものである。

横軸は予想値、縦軸は実況値であり、いずれも平年値を基準（±0日）とした相対日数で表示している。ここでは、相関の強さを表す指標として決定係数（ R^2 値）を付記している。

山形・新潟共に、後の段階の予測になるにつれて、予測開花日と実況値の相関も強まる傾向が見られた。また、重回帰モデルに比べてニューロモデルの方が、概ね相関がより強い傾向にあることが判る。

4. おわりに

本研究では、独自のニューロモデルを開発し、予測実験を行うことで、サクラの開花日予測への応用の可能性を検討した。この結果、ニューロモデルをサクラの開花日予測へ応用できる可能性は示されたものと考えられる。

その一方で、手法の詳細についてはまだ検討の余地はある。例えば、ニューロモデルの中間層ニューロンの最適数や内部パラメータのチューニングも引き続き検討する余地がある。

また、2.2にて述べたように、予測時点では既知の入力値と未知の入力値が存在する。本研究では、未知の入力値については想定し得る全てのシナリオを考慮して確率分布を求める手法を用いた。

ここで例えば、週間天気予報の気温予報値を用いることにより、予測時点の直後の旬の入力値をある程度推定することができる。この推定が適切に行われ、確率分布を求める際のシナリオを絞り込むことができれば、さらなる精度の向上も期待できる。

また、本研究では入力条件として2月上旬から4月上旬までを考慮したが、さらに遡って考慮対象の期間を広げることについても検討の余地がある。

謝辞

研究や教育を本務としない環境にあつて本研究を進めるに当たり、株式会社 SnowCast 代表取締役・杉浦 聡気象予報士より御支援を頂きました。

また、本調査ノートの投稿に際しては、担当編集委員の藤部文昭様、並びに匿名の査読者より有益なアドバイスを頂きました。

地上および桜の開花日に関する観測データは気象庁ホームページにて公開されたデータを使用しました。

以上、各関係者の皆様に心より感謝を申し上げます。

参考文献

- 青野靖之、小元敬男、1990：チルユニットを用いた温度変換日数によるソメイヨシノの開花日の推定。農業気象、45、243-249。
- Iwata, A. and T. Matubara, 1996：ニューラルネットワーク入門。http://www-ailab.elcom.nitech.ac.jp/lecture/neuro/menu.html (2010.4.16閲覧)。
- 小泉 耕、1997：ニューラルネットワーク。天気、44、51-55。
- 熊沢逸夫、1998：電子情報通信工学シリーズ 学習とニューラルネットワーク。森北出版、182pp。
- 柳野 健、1998a：気象の予測と研究のためのニューラルネットワーク I。天気、45、865-872。
- 柳野 健、1998b：気象の予測と研究のためのニューラルネットワーク II。天気、45、907-913。