

周期時系列の統計解析 (2)最小二乗法による近似 v2

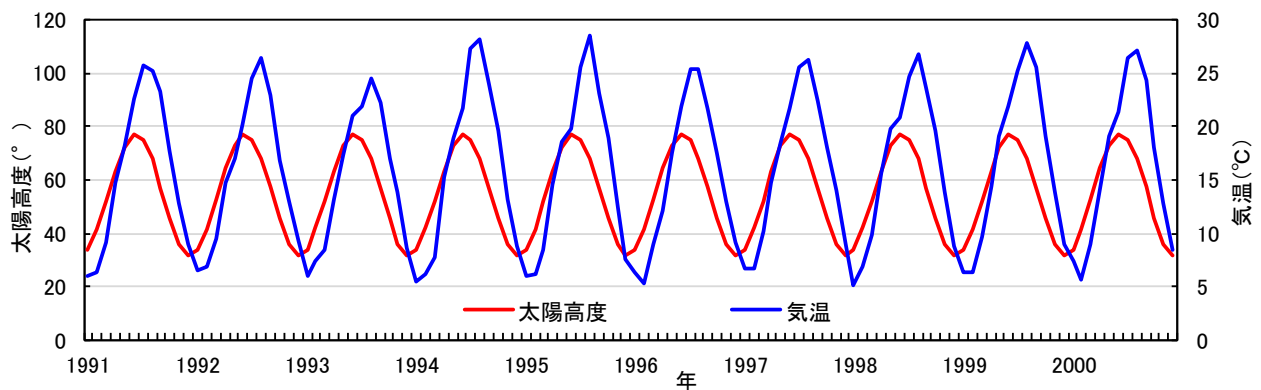
nino

2017年11月23日

最小二乗法は、観測データに最も当てはまりの良いモデル関数を求める手法の一つである。ここでは、前報で取り上げた太陽高度と気温を例として、これら周期時系列をコサイン関数を含むモデル関数で近似し、最小二乗法を用いてパラメータ（時間のズレなど）を推定するとともに、モデル間の比較検討を行った。

太陽高度と気温の観測値

太陽高度と気温の観測値は、前報と同様に、Casioのkeisanサービスの太陽高度（同時刻の年間変化）と気象庁の月平均気温（横浜）を用いた。ただし、太陽高度は横浜における時刻が12時の5日間隔のデータから月平均値に変換したものである。期間は1991年1月から2000年12月までの10年間とした（時系列図）。なお、1990年代は近年のなかでも比較的気温上昇率の高い期間に相当する。



時系列図によると、太陽高度は約 50° を平均として振幅が 20° 程度のコサイン関数に類似した周期変動を示した。気温も平均約 15°C で振幅が 10°C 程度の同様な周期変動を示すが、年毎に高低がみられ、梅雨期ごろに少し低くなる傾向を示した。また、気温のピークは太陽高度よりも時間遅れが生じている。

モデル関数と最小二乗法

太陽高度と気温の時系列は周期変動や傾向変動（トレンド）を含むとともに、両者には時間のズレ（位相差）が生じている。したがって、モデル関数のパラメータとしては、コサイン関数に関するもの以外に、トレンドや位相のパラメータが必要である。そこで、太陽高度と気温の時間 t におけるモデル関数 $f_x(t)$ と $f_y(t)$ を、それぞれ次式で定義した。

$$f_x(t) = H_x \cos(\theta_t + \varphi_x) + M_x + tK_x$$

$$f_y(t) = H_y \cos(\theta_t + \varphi_y) + M_y + tK_y$$

ここで、パラメータは、振幅； H_x, H_y , 平均； M_x, M_y , トレンド； K_x, K_y , 位相； ϕ_x, ϕ_y (rad)であり、周期を T とすると $\theta_t = 2\pi(t/T)$ で表される。また、 ϕ_x と ϕ_y の単位を rad から月に換算する式は $[\phi \times T / (2\pi)]$ である。

太陽高度と気温の観測値 $X(t)$ と $Y(t)$ を用いて、次の残差二乗和を最小にするパラメータを求める。

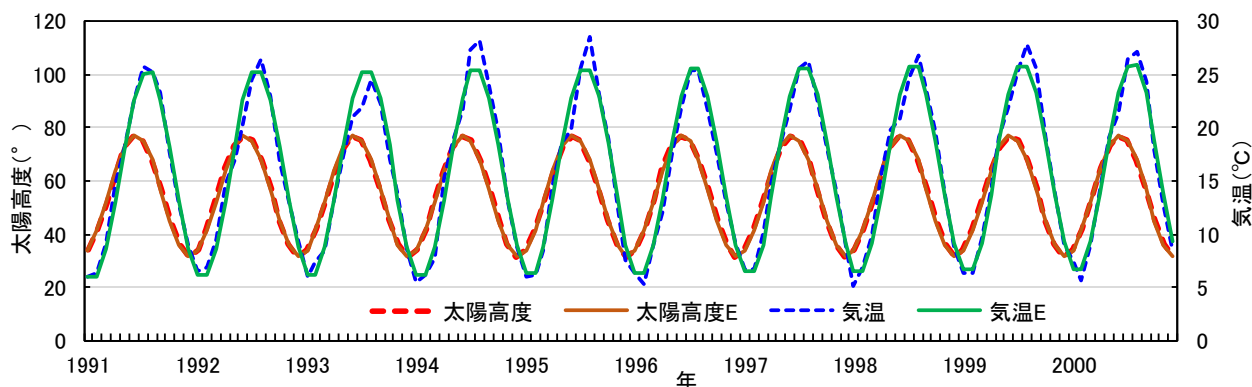
$$\sum_{t=0}^{n-1} \{X(t) - f_x(t)\}^2 + \sum_{t=0}^{n-1} \{Y(t) - f_y(t)\}^2$$

なお、最小二乗法でパラメータを求めるためにExcelのソルバーを用いたが、そのやり方は様々なサイトで紹介されている。その数例を文末に記したので参照されたい。

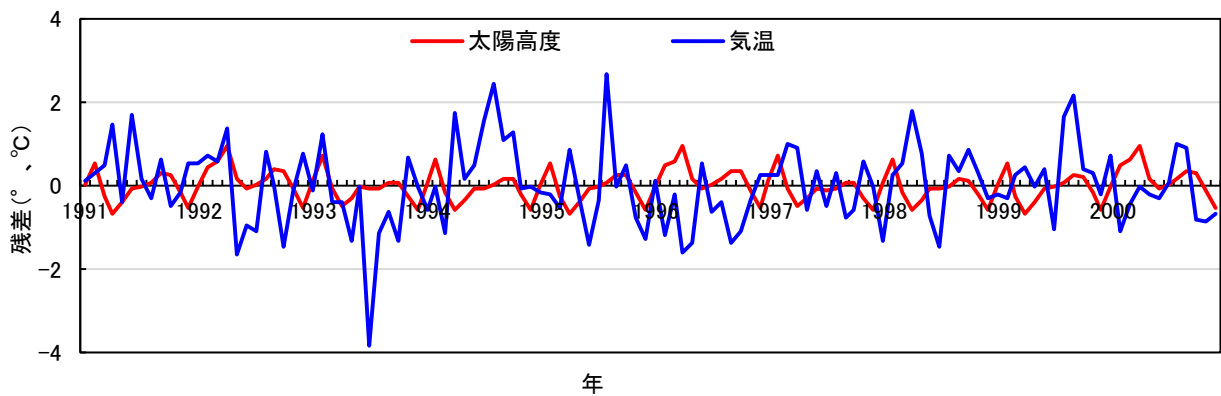
次表は、最小二乗法により求めたモデル関数（太陽高度Eと気温E）のパラメータである。太陽高度Eと気温Eの位相はそれぞれ5.80か月、4.50か月であり、位相差は-1.30か月となった。すなわち、気温Eのピークは太陽高度Eのそれより約1.3か月遅れで現れる。トレンドは、太陽高度Eでほぼ0°/月、気温Eで0.0064°C/月であった。また、最小二乗和は、太陽高度が気温の約1/9程度であり、前者はモデル関数と当てはまりが良かった。

	太陽高度E	気温E
振幅 $H(^{\circ};^{\circ}C)$	22.7	9.9
平均 $M(^{\circ};^{\circ}C)$	54.5	15.5
位相 ϕ (月)	5.80	4.50
トレンド $K(^{\circ};^{\circ}C/月)$	0.0001	0.0064
最小二乗和 Σ	16	142

観測値（太陽高度と気温）と予測値（太陽高度Eと気温E）を比較してみると（時系列図）、太陽高度Eは10年間を通じて太陽高度とほとんど重なっていた。しかし、気温Eと気温は年毎や月毎に少し差がみられ、また、梅雨期ごろに気温の低下が認められた。地上気温は気象変化の影響を受けるためである。

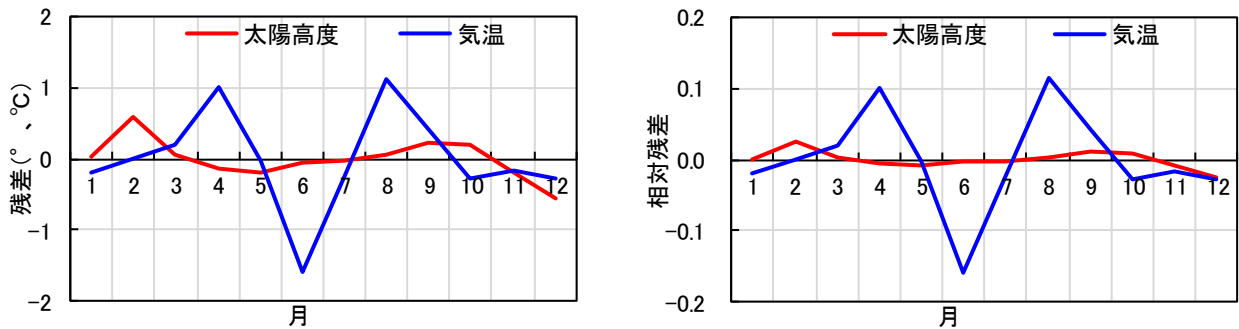


観測値と予測値の違いを調べるため、それらの残差（太陽高度，気温）を求めた。



残差時系列によると、太陽高度resはほぼ一様に変化しているが、気温resは年毎の変化が大きく、また、6月ごろに低くなる傾向を示した。

残差の季節変化の特徴を調べるため、10年間の月別平均の残差を求めた（左図）。また、単位が異なる太陽高度と気温の残差を比較検討するために、残差を振幅Hで除した値（無次元値で、以下、相対残差という）についても同様に求めた（右図）。



残差については、気温は6月に -1.5°C を下回り、4月と8月には $+1^{\circ}\text{C}$ 程度を示した。一方、相対残差については、太陽高度は $-0.025\sim 0.025$ の範囲、気温では $-0.16\sim 0.11$ の範囲にあり、太陽高度より気温のほうが5倍程度大きい変動を示した。

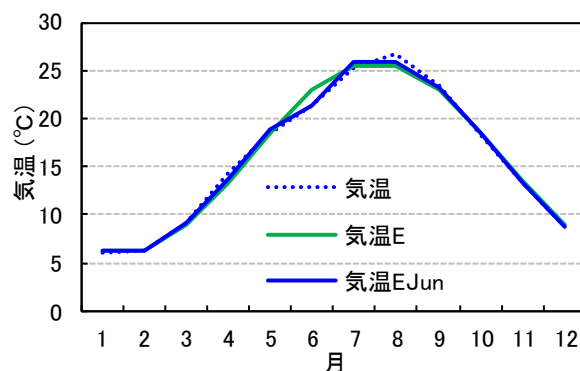
気温の残差は特に梅雨期の6月に低い値を示すので、観測値の周期時系列は6月（ピークの左側）が少しつぶれた波形となる。そのため、観測値をコサイン関数を前提とした波形モデルで近似した場合、近似波形のピークは本来より時間遅れの方向に偏る、言い換えると、位相の予測値は本来の値より低くなると推察される。

この6月の影響を確認するため、気温のモデル関数において、6月以外の月のモデル式は変えず、6月のモデル式だけ次のように6月のパラメータ（6月効果：Junとする）を一つ追加して、同様に最小二乗法によりパラメータを求めた。

$$f_Y(t) = H_Y \cos(\theta_t + \varphi) + M_Y + tK_Y + Jun$$

表に、6月効果を含むモデル関数（気温EJun）の結果と、6月効果の無いモデル関数の結果（気温E）を再掲して示した。また、季節変化の時系列図には、気温の観測値（気温）と気温Eおよび気温EJunを示した。

	気温E	気温EJun
振幅 H(°C)	9.9	10.1
平均 M(°C)	15.5	15.7
位相 φ(月)	4.50	4.54
トレンド K(°C/月)	0.0064	0.0065
6月効果(°C)		-2.1
最小二乗和 Σ	142	109



表によると、気温EJunの6月効果として -2.1°C が得られた。気温EJunの位相は4.54か月となり、気温Eの位相はそれより0.04か月低い4.50か月であった。言い換えると、気温Eの位相は気温EJunのそれより0.04か月遅れであった。これは、先述したように、波形の違いが影響したためと考えられるが、その影響の程度は小さかった。気温EJunと太陽高度の位相差は $4.54 - (5.80) = -1.26$ か月となり、小数点2桁目を四捨五入すると1.3か月が得られた。さらに、モデル関数に4、8月のプラス効果も含めて検討したところ、位相差は同じく約1.3か月となった。このように、いずれのモデルでも太陽高度と気温の位相差は小数点第一位までを考えると(0.1か月=約3日)、1.3か月で一致した。最小二乗和については、気温Eの142に対して気温EJunは109となり、6月効果を含むモデルは6月効果の無いモデルより気温観測値との当てはまりが良いことが示唆された。

季節変化の時系列図では、気温EJunは気温Eに比べてより気温に近い変動を示しており、気温観測値との当てはまりが良かった。最小二乗和も考慮すると、気温Eよりも気温EJunのモデルのほうが妥当だと考えられた。

ただし、モデルのパラメータ数が多くなるほど最小二乗和は小さくなるので、最小二乗和の大小だけで、どのモデルが妥当かを判断すべきではない。パラメータの数を考慮したモデル選択の統計検定にはAIC(赤池の情報量基準)を用いる方法があるが、どのモデルも同じ結果となったので、ここでは言及しなかった。

参考文献

Excelソルバーを用いた最小二乗法の解説サイト例

<http://tsuyu.cocolog-nifty.com/blog/2010/07/excel-24f0.html>

<http://blog.cfms.jp/wordpress/?p=1273>

http://godfoot.world.coocan.jp/square_solver.htm

Casioのkeisanサービス <http://keisan.casio.jp/exec/system/1185782617>

気象庁 <http://blog.cfms.jp/wordpress/?p=1273>