周期時系列の統計解析

(11)コサイン関数モデルによる主成分分析の検証

nino

2019 年 3 月 23 日

主成分分析はわかりにくい？ため，その結果の解釈が難しいのではないだろうか？ここでは，周期時系列としてコサイン関数モデルを用い，それに相関係数行列による主成分分析を適用し検討した．

主成分分析ソフト

主成分分析ソフトに関しては，様々な市販ソフトやフリーソフトがあるが，後述するモデル等を使って計算結果を確認しておくことをお勧めする．また，Excelを用いた主成分分析の計算方法を紹介・解説しているサイト（参考文献１）もある．

コサイン関数モデル

コサイン関数モデルとして，次の変数u_1,u_2を用いた．u_1,u_2は，振幅と周期がともに１で，位相がそれぞれα，βの最も簡単なモデルである．ここで，$\alpha > \beta \geqq 0$とした．

$$\begin{aligned} u_1 &= \cos(\theta+\alpha) \\ u_2 &= \cos(\theta+\beta) \end{aligned} \tag{1}$$

相関係数行列による主成分分析を適用する場合は，u_1,u_2を基準化し，平均が0で，分散が１にしておく必要がある．u_1,u_2の平均は0で，分散は等しいので，サンプル数をn，不偏標準偏差をσ_{n-1}とすると，基準化した変数x_1，x_2は次式で与えられる．

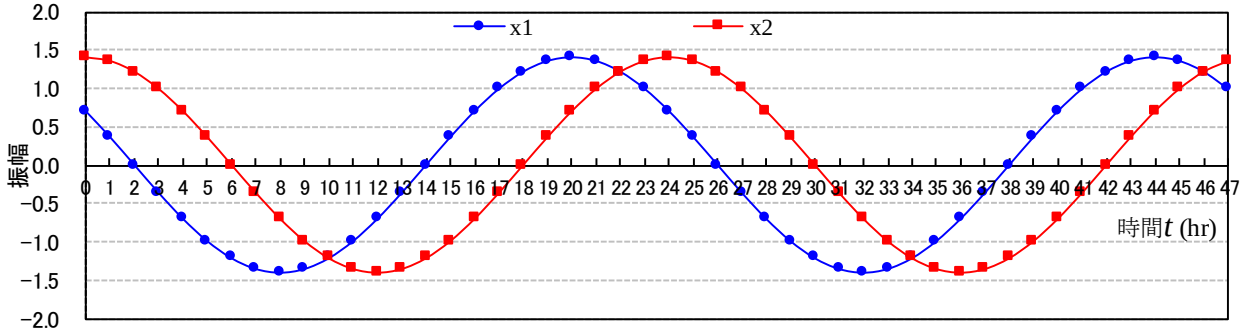
$$\begin{aligned} x_1 &= u_1/\sigma_{n-1} = \cos(\theta+\alpha)/\sigma_{n-1} \\ x_2 &= u_2/\sigma_{n-1} = \cos(\theta+\beta)/\sigma_{n-1} \end{aligned} \tag{2}$$

具体例でx_1，x_2の時系列を説明する．

式(2)において，離散変数を$\theta_t = 360^\circ\ t/n$とし，サンプリング間隔＝15°で２周期分，そして$\alpha$＝60°，$\beta$＝0°の場合について調べる（$n$＝48，$t$＝0, 1, 2, ・・・, 47）．この場合は，１周期＝360°/15°=24，であるから，例えば，気象観測値の日変化を想定すると，1hr間隔で，２日間（n＝2×24＝48hr）の時系列を表し，x_1はx_2より（$\alpha - \beta$）/15°＝60°/15°＝4hr進みとなる．

図１にx_1,x_2の時系列を示した．この場合，n=48なので，$\sigma_{47} \fallingdotseq 0.7146$が得られ，振幅は$1/0.7146 \fallingdotseq 1.399$となる．また，$x_1$と$x_2$の位相は，例えば，24hr前後の最大振幅についてみるとそれぞれ20hrと24hrであり，位相差は4hr（＝60°）となる．なお，x_1とx_2の相関係数r_{x1x2}はそれらの位相差のコサインで表されるので，$r_{x1x2} = \cos(\alpha - \beta) = \cos(60^\circ) = 0.5$が得られる（参考文献２）．

図 1



固有値と固有ベクトル

x_1,x_2の第 1 主成分z_1と第 2 主成分z_2は次式で表される．

$$z_1 = l_1 x_1 + l_2 x_2$$
$$z_2 = m_1 x_1 + m_2 x_2 \qquad (3)$$

相関係数をrとすると，次の主成分分析結果が得られる（参考文献３）．
固有値λは，

$$\lambda = 1 + r \quad \text{または，} \quad 1 - r \qquad (4)$$

であり，固有ベクトル$(l_1,\ l_2)$，$(m_1,\ m_2)$は，
$\lambda = 1 + r$（$r > 0$）のとき，

$$(l_1,\ l_2) = (1/\sqrt{2},\ 1/\sqrt{2}) \quad \text{あるいは} \quad (-1/\sqrt{2},\ -1/\sqrt{2})$$
$$(m_1,\ m_2) = (1/\sqrt{2},\ -1/\sqrt{2}) \quad \text{あるいは} \quad (-1/\sqrt{2},\ 1/\sqrt{2}) \qquad (5)$$

$\lambda = 1 - r$（$r < 0$）のとき，

$$(l_1,\ l_2) = (1/\sqrt{2},\ -1/\sqrt{2}) \quad \text{あるいは} \quad (-1/\sqrt{2},\ 1/\sqrt{2})$$
$$(m_1,\ m_2) = (1/\sqrt{2},\ 1/\sqrt{2}) \quad \text{あるいは} \quad (-1/\sqrt{2},\ -1/\sqrt{2}) \qquad (6)$$

となる．

この結果だけでは，良くわからない．そこで，先のコサイン関数モデルの具体例（$r = 0.5 > 0$）に複数のソフトにより主成分分析を行った結果，次の固有ベクトルが得られた．

$$(l_1,\ l_2) = (0.7071,\ 0.7071)$$
$$(m_1,\ m_2) = (-0.7071,\ 0.7071),\ \text{または，}\ (0.7071,\ -0.7071) \qquad (7)$$

式(6)から，固有ベクトル$(m_1,\ m_2)$については，ソフトによって異なる２つの結果が得られた．この理由はわからないが，後述するように，主成分分析結果を解釈する上で大きな問題はないと思われる．また，式(6)では，固有ベクトルの値は0.7071で示されている．これは数値計算によって算出された値であり，解析的には式(4)と式(5)で記されているように $1/\sqrt{2}$（≒0.7071）である．

主成分の導出

コサイン関数モデルを用いて，式(6)の３つの固有ベクトルに対応した第１主成分z_1，第

２主成分z_2，およびもう一つの第２主成分z'_2を求めた．

z_1は，$(l_1,\ l_2)=(1/\sqrt{2},\ 1/\sqrt{2})$より，

$$
\begin{aligned}
z_1 &= \frac{1}{\sqrt{2}}(x_1+x_2) \\
&= \frac{1}{\sqrt{2}\sigma_{n-1}}\{\cos(\theta+\alpha)+\cos(\theta+\beta)\} && \text{和積公式より} \\
&= \frac{2}{\sqrt{2}\sigma_{n-1}}\cos\left(\theta+\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right) && \text{半角公式より} \\
&= \frac{\sqrt{2}}{\sigma_{n-1}}\sqrt{\frac{1+\cos(\alpha-\beta)}{2}}\cos\left(\theta+\frac{\alpha+\beta}{2}\right) && \text{相関係数:}\cos(\alpha-\beta)=r\text{より} \\
&= \frac{\sqrt{1+r}}{\sigma_{n-1}}\cos\left(\theta+\frac{\alpha+\beta}{2}\right) && 1+r=\lambda_1\text{より} \\
&= \frac{\sqrt{\lambda_1}}{\sigma_{n-1}}\cos\left(\theta+\frac{\alpha+\beta}{2}\right) && (8)
\end{aligned}
$$

同様にして，z_2は，$(m_1,\ m_2)=(-1/\sqrt{2},\ 1/\sqrt{2})$より，

$$
\begin{aligned}
z_2 &= \frac{1}{\sqrt{2}}(-x_1+x_2) \\
&= \frac{\sqrt{1-r}}{\sigma_{n-1}}\sin\left(\theta+\frac{\alpha+\beta}{2}\right) \\
&= \frac{\sqrt{\lambda_2}}{\sigma_{n-1}}\sin\left(\theta+\frac{\alpha+\beta}{2}\right) && (9)
\end{aligned}
$$

z'_2は，$(m_1,\ m_2)=(1/\sqrt{2},\ -1/\sqrt{2})$より，

$$
\begin{aligned}
z'_2 &= \frac{1}{\sqrt{2}}(x_1-x_2) \\
&= \frac{-\sqrt{\lambda_2}}{\sigma_{n-1}}\sin\left(\theta+\frac{\alpha+\beta}{2}\right) \\
&= -z_2 && (10)
\end{aligned}
$$

これらの式によると，振幅はz_1で$\sqrt{\lambda_1}/\sigma_{n-1}$，$z_2$と$z'_2$では$\pm\sqrt{\lambda_2}/\sigma_{n-1}$が得られた．位相は$z_1$，$z_2$とも$x_1$と$x_2$の平均位相$(\alpha+\beta)/2$で同じだが，$z_1$はコサイン関数であるのに対して，$z_2$および$z'_2$はサイン関数であった．$z_2$，$z'_2$はコサイン関数で表すと$\cos\{\theta+(\alpha+\beta)/2\mp 90^\circ\}$であるから，$z_1$との相対的な位相$(\mp 90^\circ)$より求められる．

通常，固有値が１より小さい主成分は元の変数よりも情報量が少ないので固有値が１以上の主成分までを使うのを目安としている（参考文献１，３）．これまでの結果より，固有値が１以上の値をもつz_1が求まれば，１以下の値をもつz_2，z'_2はz_1との相対的な位相関係で決まると考えれば理解できる．また，主成分分析は多変量データを要約する（低い次

元に落とす）ための強力なツールと言われている．これは，同一周期を有する多変数が固有値が１以上の値をもつ一つの主成分との相対的な位相関係で表わされるためと考えられる．このように，周期時系列では，変数相互の位相関係により主成分分析結果を解釈できた．

一方，位相の$(\alpha-\beta)/2$に着目して式(8)を変形すると，次式が得られる．

$$z_1=\frac{\sqrt{2}}{\sigma_{n-1}}\cos\left\{\frac{\alpha-\beta}{2}\right\}\cos\left\{(\theta+\alpha)-\frac{\alpha-\beta}{2}\right\} \qquad (8)'$$

$$z_1=\frac{\sqrt{2}}{\sigma_{n-1}}\cos\left\{\frac{\alpha-\beta}{2}\right\}\cos\left\{(\theta+\beta)+\frac{\alpha-\beta}{2}\right\} \qquad (8)''$$

同様に，式(9)については，次式が得られる．

$$z_2=\frac{\sqrt{2}}{\sigma_{n-1}}\sin\left\{\frac{\alpha-\beta}{2}\right\}\cos\left\{(\theta+\alpha)-\left[\frac{\alpha-\beta}{2}+\frac{\pi}{2}\right]\right\} \qquad (9)'$$

$$z_2=\frac{\sqrt{2}}{\sigma_{n-1}}\sin\left\{\frac{\alpha-\beta}{2}\right\}\cos\left\{(\theta+\beta)+\left[\frac{\alpha-\beta}{2}-\frac{\pi}{2}\right]\right\} \qquad (9)''$$

これらの式の意味するところは，z_1，z_2のx_1，x_2との位相関係を表していることである．例えば，式(8)’はz_1がx_1より$(\alpha-\beta)/2$遅れ，そして，式(8)”はz_1がx_2より$(\alpha-\beta)/2$進みであることを示している．z_2とx_1，x_2との位相関係についても同様である．このように，z_1，z_2の位相はx_1とx_2との位相差で表すことができた．

因子負荷量の導出

因子負荷量は，主成分と変数の相関係数で表され，主成分と強い関係を持つ因子を選択するのに使われる（参考文献３）．そこで，z_1，z_2とx_1，x_2との因子負荷量を求めた．
先述したように，コサイン関数モデルにおける相関係数は位相差のコサインで表されることを利用すると，簡単に求まる．z_1とx_1の相関係数r_{z1x1}は，

$$\begin{aligned} r_{z_1x_1} &= \cos\left\{\left(\frac{\alpha+\beta}{2}\right)-\alpha\right\} \\ &= \cos\left\{-\left(\frac{\alpha-\beta}{2}\right)\right\} \\ &= \sqrt{\frac{1+\cos(\alpha-\beta)}{2}} \\ &= \sqrt{\frac{1+r}{2}} \\ &= \sqrt{\frac{\lambda_1}{2}} \end{aligned} \qquad (11)$$

となり，z_1とx_2の相関係数r_{z1x2}は次式が得られた．

$$r_{z_1x_2} = \cos\left\{\left(\frac{\alpha+\beta}{2}\right)-\beta\right\}$$
$$= \cos\left(\frac{\alpha-\beta}{2}\right)$$
$$= \sqrt{\frac{\lambda_1}{2}} \qquad (12)$$

同様にして，z_2とx_1およびx_2の相関係数r_{z2x1}およびr_{z2x2}については，次式が得られた．

$$r_{z_2x_1} = -\sin\left(\frac{\alpha-\beta}{2}\right) = -\sqrt{\frac{\lambda_2}{2}} \qquad (13)$$

$$r_{z_2x_2} = \sin\left(\frac{\alpha-\beta}{2}\right) = \sqrt{\frac{\lambda_2}{2}} \qquad (14)$$

式(11)～式(14)によると，それぞれの因子負荷量 $r_{zk,xi}$は固有値のルート（$\sqrt{\lambda_k}$）と固有ベクトルとの積（$\sqrt{\lambda_k}\,l_{ki}$）で表される（参考文献３）．また，周期時系列の因子負荷量は先の相関係数の定義から主成分と変数の位相差が寄与しており，例えば，z_1とx_1, x_2の因子負荷量はそれらの位相差（時間のズレ）が小さいほど１に近い値となり，いわゆる「強い関係」にあることを示している．

具体例を用いた主成分分析

これまでに得られた結果を確認するため，先の具体例（α＝60°＝4hr，β＝0°＝0hr，r＝0.5)に主成分分析を適用し検証した．

主成分分析結果の固有値，寄与率，累積寄与率を表１に，因子負荷量を表２に示した．

表１

	z_1	z_2
固有値	1.50	0.50
寄与率	0.75	0.25
累積寄与率	0.75	1.00

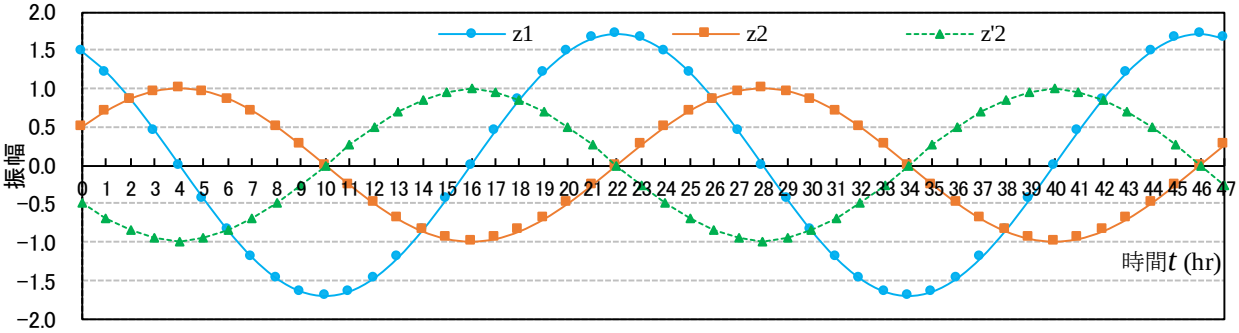
表２

因子負荷量	z_1	z_2
x_1	0.866	−0.500
x_2	0.866	0.500

表１，表２より，式(8)～式(14)の理論通りの結果が得られた．

次に，各主成分得点z_1，z_2，z'_2の時系列（図２）について調べた．

図２



まず，振幅についてみると，z_1は$\sqrt{\lambda_1}/\sigma_{47} = \sqrt{1.5}/0.7146 \fallingdotseq 1.714$，$z_2$は$\sqrt{\lambda_2}/\sigma_{47} = \sqrt{0.5}/0.7146 \fallingdotseq 0.9895$となった．また，$z'_2$の振幅は$z_2$の振幅と等しいが，逆位相（$-z_2$）であった．
次に，位相については，24hr前後の最大振幅を示す時間帯において，x_1，x_2の時系列（図１）と比較検討した．最大振幅を示す時間はx_1で20hr，x_2で24hrであるが，z_1では（20＋24）/2＝22hrを示した．また，z_2はz_1に対して－90°/15°＝－6hrすなわち6hr遅れの22＋6＝28hrを示し，z'_2はz_1より+90°/15°＝+6hrすなわち6hr進みの22－6＝16hrを示した．
なお，位相と時間の関係については，参考文献４などを参照されたい．

相関係数が負の場合の主成分分析

これまでは相関係数 r＞0の場合について検討したが，ここでは，相関係数 r＜0の場合における主成分分析（主成分z_1，z_2）の結果の概要をまとめた．

r＜0の場合における固有値は$\lambda_1 = 1 - r$，$\lambda_2 = 1 + r$，固有ベクトルは$(l_1,\ l_2) = (1/\sqrt{2},\ -1/\sqrt{2})$，$(m_1,\ m_2) = (1/\sqrt{2},\ 1/\sqrt{2})$となり，次の$z_1$，$z_2$が導出された．

$$z_1 = \frac{1}{\sqrt{2}}(x_1 - x_2) = \frac{-\sqrt{\lambda_1}}{\sigma_{n-1}}\sin\left(\theta + \frac{\alpha+\beta}{2}\right) = \frac{\sqrt{\lambda_1}}{\sigma_{n-1}}\cos\left(\theta + \frac{\alpha+\beta}{2} + \frac{\pi}{2}\right) \tag{15}$$

$$z_2 = \frac{1}{\sqrt{2}}(x_1 + x_2) = \frac{\sqrt{\lambda_2}}{\sigma_{n-1}}\cos\left(\theta + \frac{\alpha+\beta}{2}\right) \tag{16}$$

主成分と変数の因子負荷量については，

$$r_{z_1x_1} = \sqrt{\lambda_1}(1/\sqrt{2}) = \sqrt{\lambda_1}l_1 \tag{17}$$
$$r_{z_1x_2} = \sqrt{\lambda_1}(-1/\sqrt{2}) = \sqrt{\lambda_1}l_2 \tag{18}$$
$$r_{z_2x_1} = \sqrt{\lambda_2}(1/\sqrt{2}) = \sqrt{\lambda_2}m_1 \tag{19}$$
$$r_{z_2x_2} = \sqrt{\lambda_2}(1/\sqrt{2}) = \sqrt{\lambda_2}m_2 \tag{20}$$

が得られた．

具体例として，α＝120°（8hr），β＝0°（0hr），相関係数r＝cos120°＝－0.5＜0の場合について調べた．
主成分分析結果（固有値，寄与率，累積寄与率）を表３に，因子負荷量を表４に示した．

表３

	z_1	z_2
固有値	1.500	0.500
寄与率	0.750	0.250
累積寄与率	0.750	1.000

表４

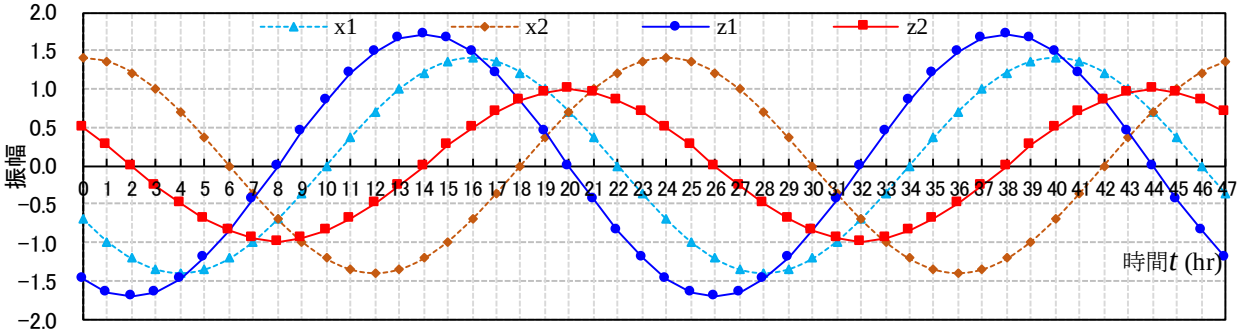
因子負荷量	z_1	z_2
x_1	0.866	0.500
x_2	−0.866	0.500

表３，４によると，固有値はz_1，z_2ともに相関係数r＝0.5の場合と等しかった．しかし，因子負荷量については式(17)～(20)で示されるように符号が一部で変化した．

次に，x_1，x_2とz_1，z_2の主成分得点の時系列を図３に示した．

図３から，最大振幅を示す時間はx_1で16hr，x_2で24hrであるので，z_1では式(15)より（16＋24）/2＝20hrより6hr進み（+90°）の20－6＝14hrを示した．z_2は式(16)より（16＋24）/2＝20hrを示した．このように，相関係数が負になると主成分の位相が変化するので，注意する必要がある．

図３

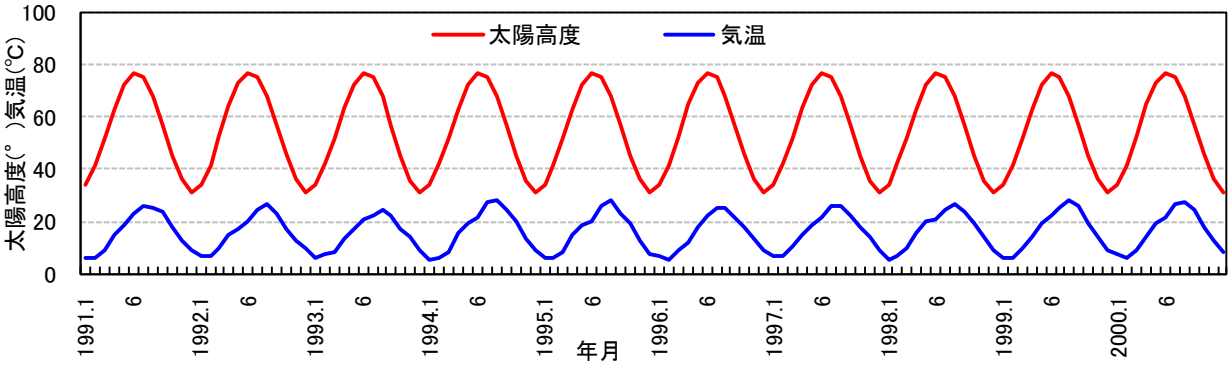


観測値を用いた主成分分析

通常，２変数の主成分分析を行うことはないが，これまでの結果を確認するために，２つの観測値に主成分分析を適用し検討した．

用いた観測値は，太陽高度と月平均気温(横浜)であり，対象期間は1991年1月から2000年12月までの10年間である（参考文献２）．図４にそれらの時系列を示した．

図４



太陽光はまず地面を暖め，その熱が空気を暖めるため，太陽高度と気温のピークに時間のズレ（位相差）が生じる．１年間で太陽高度が最も高くなるのは夏至（6月）だが，最も暑いのは8月頃であり，気温の最高値は太陽高度よりも時間遅れで現れる．

図４によると，太陽高度と気温はともにコサイン関数モデルに類似した周期変化を示すが，気温は気象変化などの影響を受け年毎に少し変化していた．また，気温の最高値は太陽高度よりも1～2か月程度の遅れを生じていた．両者はほぼコサイン関数モデルで近似できるので，相関係数rが0.763であるから，位相差は約40°となり，月換算では約1.3月に相当すると考えられた（参考文献２）．

主成分分析の結果を表５，６に，主成分得点の時系列を図５に示した．

表５

	z_1	z_2
固有値	1.763	0.237
寄与率	0.881	0.119
累積寄与率	0.881	1.000

表６

因子負荷量	z_1	z_2
太陽高度	0.939	−0.344
気温	0.939	0.344

図５

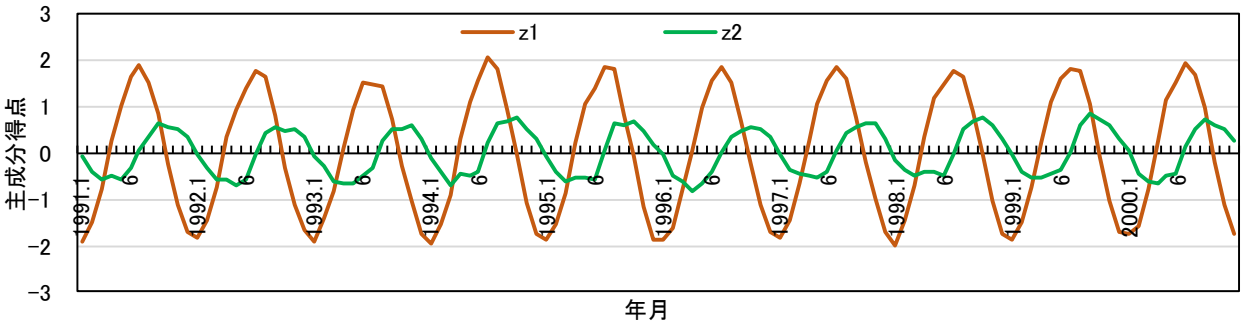


表５によると，z_1の固有値は1.763（＝1+r)を示した．太陽高度と気温の位相差が約40°（r＝0.763)であるので，z_1の固有値が比較的大きくなったといえる．また，表６から，z_1と太陽高度および気温の因子負荷量はともに0.939となり，位相差としては約20°（約0.67月)であった．これは，式(11)および式(12)より，z_1と太陽高度および気温の位相差|($\alpha-\beta$)/2|＝|40°/2|＝20°に一致している．z_2も同様に式(13)と式(14)を満たしている．
一方，図５から，最高値はz_1が7月頃にあり，夏至(6月)と最高気温の8月のほぼ平均値と一致した．また，z_2が10月頃にあり，z_1とは約3か月（90°）のズレが認められた．

このように，コサイン関数モデルの主成分分析結果を良く説明できた．これは，太陽高度と気温がともにコサイン関数モデルに類似した変化パターンを示すからである．他の観測値では変化パターンがそうならなかったり誤差が大きかったりする場合が多いと思われる．その場合は，相関係数がモデルに比べて小さくなり，位相差は大きくなるので（参考文献５)，解釈に注意が必要である．変数と主成分の時系列を比較して検討することが肝要であると考えられる．

参考文献

１．高校数学の基本問題，Excelを用いた主成分分析
http://www.geisya.or.jp/~mwm48961/statistics/syuseibun1.htm
２．物理のかぎしっぽ，周期時系列の統計解析(1)相関係数と位相差
http://hooktail.sub.jp/contributions/shuki01.pdf
３．多変量解析法，奥野忠一ら，日科技連出版社(1984)
４．電磁気学の基本問題，交流の基本
http://www.geisya.or.jp/~mwm48961/electro/alternate1.htm
５．物理のかぎしっぽ，楕円の相関係数(その10)
http://hooktail.sub.jp/contributions/ellipse_cor_10v1.pdf