JP1.8 IDENTIFICATION OF INTRA-MONTH DAILY MEAN TEMPERATURE MODES UTILIZING PRINCIPAL COMPONENTS ANALYSIS

Charles J. Fisk *
NAVAIR-Point Mugu, CA

1. INTRODUCTION

A commonly used statistic for summarizing day-to-day temperatures is the “daily mean”, usually the arithmetic average of the extreme maximum and minimum readings over a 24-hour period (generally midnight-to-midnight for first-order stations). Long-term smoothed climatological averages of same-day means serve as “normals”, enabling calculation of “departures from normal” and determining anomaly character.

Climatological daily mean temperatures are, of course, statistical idealizations, and it goes without saying that considered over a calendar month interval, for example, their smoothed, linear-like forms frequently do not typlify “normal” day-to-day *patterns* in mean temperature that actually occur from one year to the next. These are more likely to be a quasi-sinusoidal mix. Given this fact, it might be a useful exercise from a descriptive climatological point of view to explore statistical idealizations of these patterns as well, and the degree to which they replicate actual historical ones.

To this end, the 85-year Downtown Los Angeles daily mean temperature record (1921-2005) is analyzed, by calendar month, for the existence of day-to-day mean temperature modes, utilizing Linear Principal Components Analysis. The observational data reflect the usual inhomogeneities associated with a lengthy history (several instrument relocations, evolving heat-island effects, etc.,) but since evaluation of a methodology is the primary objective, and a long period of record is desirable, this is not considered critical.

In a previous PCA study on Downtown Los Angeles temperatures, the patterns of daily maximum and minimum temperatures taken as a unit were explored [Fisk, 2004]. From the results, first component percents of variance explained were so predominant, and related scores and loadings so highly correlated with certain climatological parameters (e.g., long-term average max/min’s, average daily ranges, and standard deviations of maxima/minima taken collectively) that it was possible to use this first component PCA information as an objective means of ranking and identifying the extreme-most individual years’ max/min patterns in terms of “shape” and “spread”. The methodology was applied on time units ranging from calendar month(s) to a full year.

In this study, the use of daily means confined to a calendar month period only, largely eliminates this first component “overall-climatology” effect, allowing resolution of modes that would otherwise be overwhelmed and obscured statistically.

** Corresponding author address:* Charles J. Fisk, NAVAIR, Point Mugu, CA. 93042: e-mail: charles.fisk@navy.mil

2. METHODS AND PROCEDURES

Generation of the PCA daily mean temperature pattern statistics was accomplished as follows. First, each individual year’s daily means for a calendar month of interest were assembled into an N-row by Y-column matrix, where N was the number of days in the particular month, and Y the number of years. For example, since there were 85 years’ Downtown Los Angeles data, a matrix of February daily means would have 28 rows and 85 columns of daily means (Feb. 29 leap year data excluded to preserve matrix rectangularity). A correlation PCA was then performed (O-mode decomposition with no rotation), the loadings’ statistics calculated by year and ordered by absolute magnitude. Loading statistics equal in magnitude but opposite in sign would essentially reflect patterns “inverted” relative to one another in appearance.

The modes were quantified by daily standardized scores, and their correlations (the above “loadings”) with the various years’ and months’ actual daily means represented the degree of conformance to a particular mode.

To provide a visual feel for the range of components’ patterns, the first four were graphed for each of the calendar months (twelve graphs with overlays of four different scores’ patterns per graph). A succeeding series of plots then depicted actual daily mean patterns for selected months that exhibited particular conformance to the four modes, or some other feature(s) of interest.

To evaluate the degree to which the components could replicate actual historical patterns on a *multi-modal* basis, a sample analysis was performed on the 85 years’ of January data, incorporating each of the four components into multiple regression prediction models. In addition to evaluation of the multiple correlation results, the prediction models were inspected, year-by-year, for statistically significant coefficient magnitudes, the affiliated components identified and utilized to build a frequency “typology” of important multi-modal subset combinations.

3. RESULTS

3.1 - Correlation PCA Tabular Statistics

The correlation PCA resolved 14 modes with eigenvalues >= 1.0 for eleven of the twelve months.

MONTH	CORR 1ST COM-PONENT %-VAR (1)	CORR 2ND COM-PONENT %-VAR (2)	CORR 3RD COM-PONENT %-VAR (3)	CORR 4TH COM-PONENT %-VAR (4)
JAN	20.81% (30) (10)	16.30% (20) (10)	11.34% (13) (2)	9.03% (6) (1)
FEB	20.81% (29) (9)	15.65% (23) (4)	12.66% (15) (0)	8.02% (4) (2)
MAR	17.02% (22) (8)	12.66% (15) (2)	10.73% (11) (2)	9.37% (12) (0)
APR	**16.13%** (22) (10)	**10.54%** (15) (0)	10.44% (11) (2)	9.20% (12) (2)
MAY	17.39% (24) (5)	14.09% (19) (1)	11.86% (15) (2)	8.90% (8) (0)
JUN	**30.77%** (45) (23)	12.30% (13) (2)	**10.11%** (9) (2)	**7.53%** (9) (1)
JUL	23.54% (36) (13)	15.35% (19) (3)	12.90% (14) (4)	9.63% (10) (2)
AUG	20.67% (27) (10)	14.36% (18) (7)	**13.04%** (14) (3)	**9.96%** (14) (1)
SEP	21.07% (32) (12)	15.41% (19) (5)	11.59% (15) (1)	9.46% (8) (2)
OCT	20.92% (28) (9)	11.94% (12) (3)	11.26% (11) (3)	9.18% (9) (0)
NOV	21.96% (31) (13)	14.30% (20) (9)	11.81% (13) (1)	9.76% (11) (2)
DEC	19.86% (28) (11)	**17.36%** (25) (9)	12.46% (16) (2)	9.94% (11) (1)

Table 1 - Percents of Variance Explained, by Month, for first Four Components, and Number of Cases with Absolute Loadings' figures >=.50 and >=.70 (in parentheses) - Downtown Los Angeles Daily Mean Temperature Time Series, (1921-2005 data)

The only exception, September, had 13 such modes. As will be seen in the graphs to follow, eigenvector one's day-to-day standardized scores suggested propagation of long-waves and, occasionally, seasonal trend as the most important intra-month daily mean temperature pattern. The other modes, in descending rank order, seemed to portray progressively higher frequency quasi-sinusoidal patterns with varying phase shifts, suggestive of a gradation in relative importance from long-waves to synoptic or more random influences.

Table 1 lists the percents-of-variance-explained statistics, by calendar month, for the four most important modes. Included also, in parentheses below the percent figures, are the counts of cases in which the absolute loadings' magnitudes were >=.50 or >=.70 (the latter in red).

In general, the percents-explained figures are relatively uniform across the months, by component. The only major exception is the significantly higher first component magnitude for June (30.77%); this compares with the minimum first component figure (16.13%) for April. Second component magnitudes range from 17.36% (December) to 10.54% (April), third component statistics from 13.04% (August) to 10.11% (June), and fourth component values from 9.96% (also August) to 7.53% (also June).

Total percents of variance explained for the first four modes range from 46.3% (April) to 61.4% (July).

Using overall averages of the month-to-month count figures, 354 of the 1020 individual months, or 34.7%, "conformed" to their respective months' first component modes at the .50 loading level or higher, 13.0% at the .70 level. For the lesser modes, the figures were, for the .50 level: 21.4% for mode 2, 15.4% for mode 3, and 11.2% for mode 4; at the .70 level: 5.4% for mode 2, 2.4% for mode 3, and 1.4% for mode 4.

Summing the .50-level counts across the four modes for all months combined (excluding same case multiple-conformance "extra" counts), 72.8 % of the monthly series' patterns conformed to at least one of the four primary modes. For the .70 threshold, (no multiple conformances observed at this level), the figure was 23.5 %. Beyond the fourth component, another 3 cases (or 0.3%) absolute loadings' figures greater than equal to absolute value .70 were encountered, all for the fifth component. About 10 % of the individual months loaded higher than absolute .50 on two modes simultaneously; one case had two loadings at +.633 and -.640, respectively.

This somewhat modest percentage for the .70 loading level indicates that, occasionally (about one-fourth of the time), actual patterns resembled a given idealized one quite well (a .70 loading or correlation, represents 49 % of the variance explained), but more often they did so only in a general sort of way (a .50 loading, of course, indicates 25 % variance explained), or less favorably.

Highest absolute loadings figures for the four modes were as follows: (+).920 for mode 1, (+).889 for mode 2, (-).889 for mode 3, and (-).838 for mode 4.

3.2 Standardized Scores Plots, by Month, for the First Four Modes

3.2.1 - First Component Plots

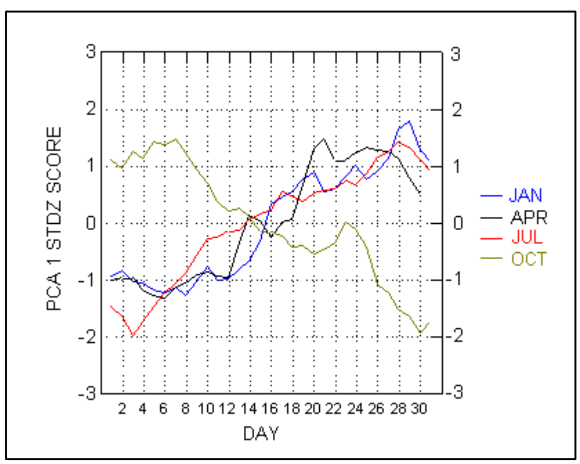

Figure 1. First Component Correlation PCA Scores for January, April, July, & October - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

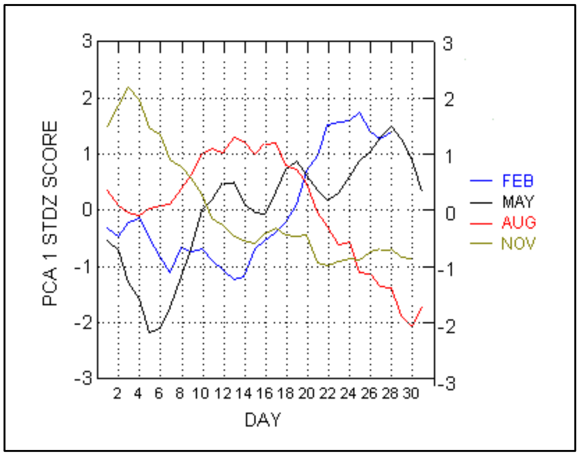

Figure 2. First Component Correlation PCA Scores for February, May, August, & November - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

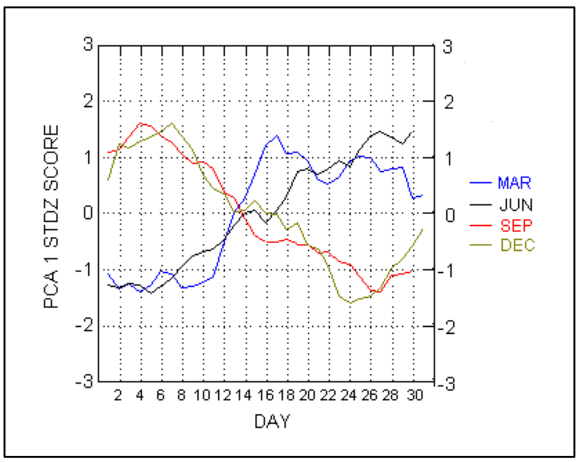

Figure 3. First Component Correlation PCA Scores for March, June, September, & December - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

Figures 1 through 3 are plots of the various first component day-to-day scores', by month. Emphasizing that the patterns are first and foremost synthetic statistical idealizations, a few remarks are offered about their features and the physical processes they might reflect.

The majority seem to exhibit steady upturns or downturns over most of their days, suggestive, at least, of long-wave propagation or seasonal trends. January, March and April all show increases in scores after about days 8-12 (March, however, leveling off after mid-month). June commences a rise earlier, from about day 5. February starts a rise about day 14 but levels off at day 22. The September through December plots, in contrast, show general declines, especially after the first week, December also exhibiting an increase for the last week, suggestive of a two-week idealized "half-cycle" in daily mean temperature progressions. May shows a "wavy" upward trend, and August presents a mid-month peak, more characteristic of second-component patterns, as the next series of graphs will show.

3.2.2 Second Component Plots

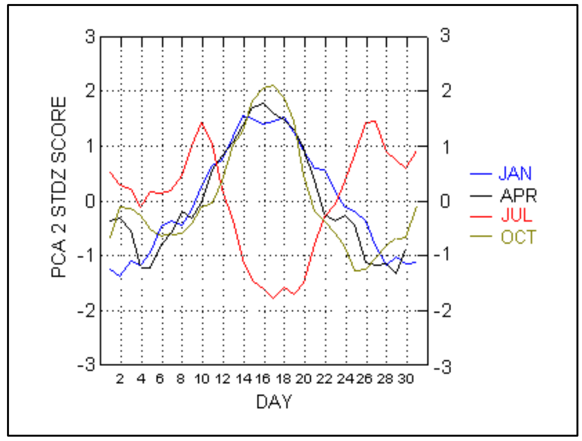

Figure 4. Second Component Correlation PCA Scores for January, April, July, & October - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

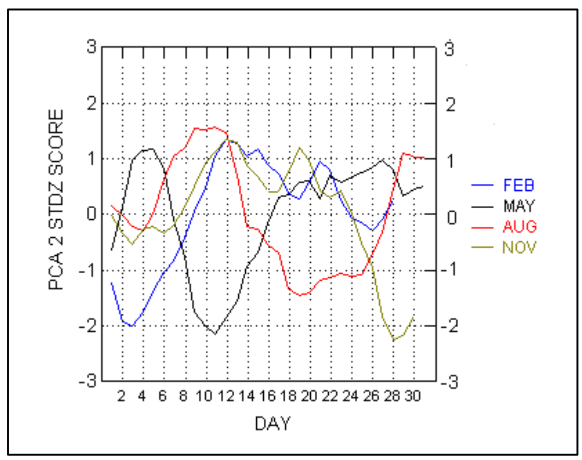

Figure 5. Second Component Correlation PCA Scores for February, May, August, & November - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

Figures 4 through 6 are plots of the monthly second component day-to-day scores. Compared to the first component, the scores' patterns show lesser half-periods (approximately 7-10 days) when there are sinusoidal indications. Also, the peaks (troughs) are frequently concentrated around the middle of the month. These properties are especially true for the months represented In Figures 4 and 6, but in Figure 5 the

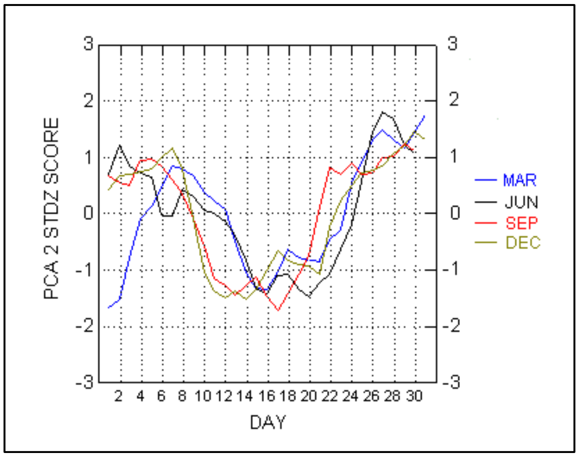

Figure 6. Second Component Correlation PCA Scores for March, June, September, & December - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

patterns for February and November, in particular, are noticeably dissimilar from this “norm”.

3.2.3 Third Component Plots

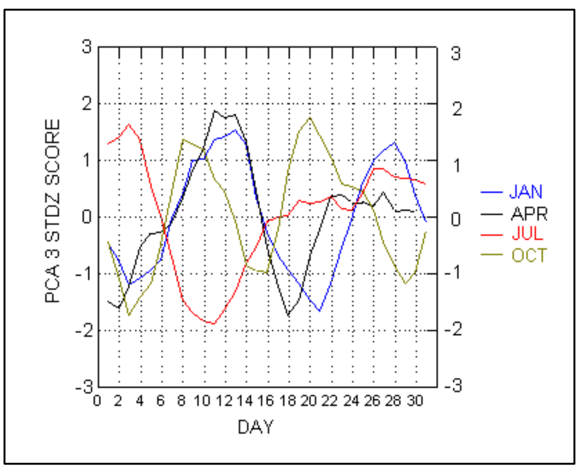

Figure 7. Third Component Correlation PCA Scores for January, April, July, & October - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

Figures 7 through 9 are plots of the monthly third component day-to-day scores. Relative to the second-component displays, the patterns seem more diverse with shorter half-periods, sometimes less than seven days. Centers of the most prominent peaks and troughs are positioned at scattered non-mid-month locations. The September pattern (Figure 9) is a noticeable one with a particularly high amplitude score on the 23rd, approaching +3.00.

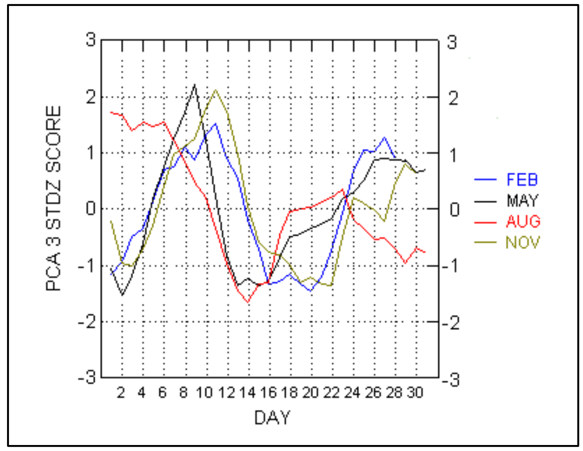

Figure 8. Third Component Correlation PCA Scores for February, May, August, & November - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

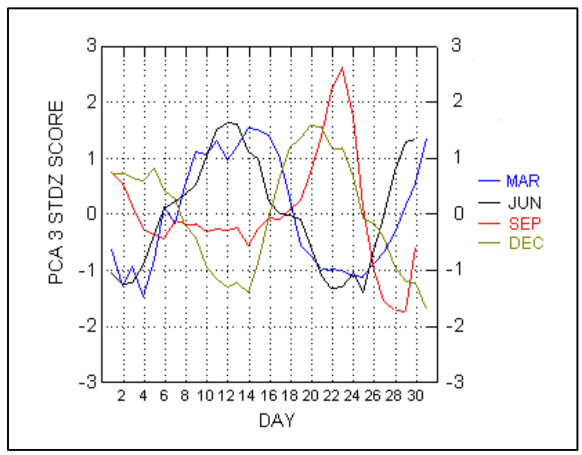

Figure 9. Third Component Correlation PCA Scores for March, June, September, & December - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

3.2.4 Fourth Component Plots

Finally, Figures 10 through 12 show the synthetic patterns for the fourth component. They seem even more varied than those in Figures 7 through 9.

This non-coherence/noisiness of the month-to-month patterns becomes even more pronounced for the fifth components and beyond, the number of relatively high-level individual loadings’ statistics also decreasing.

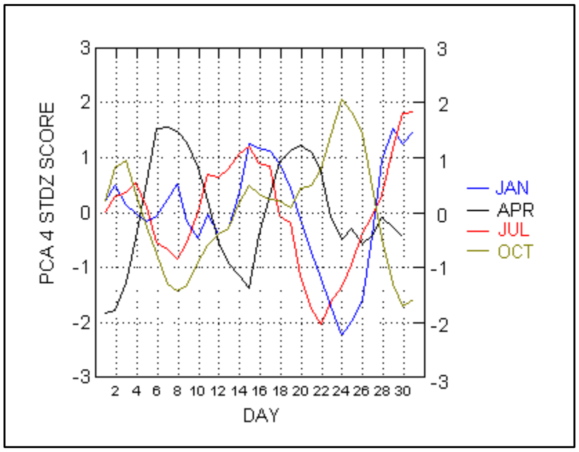

Figure 10. Fourth Component Correlation PCA Scores for January, April, July, & October - Downtown Los Angeles Daily Mean Temperatures (1921-2005)

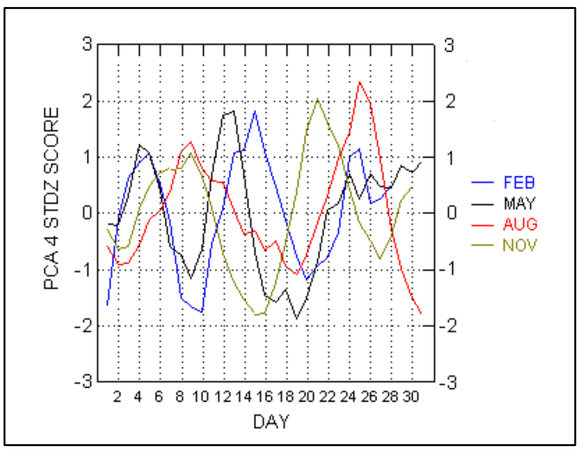

Figure 11. Fourth Component Correlation PCA Scores for February, May, August, & November - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

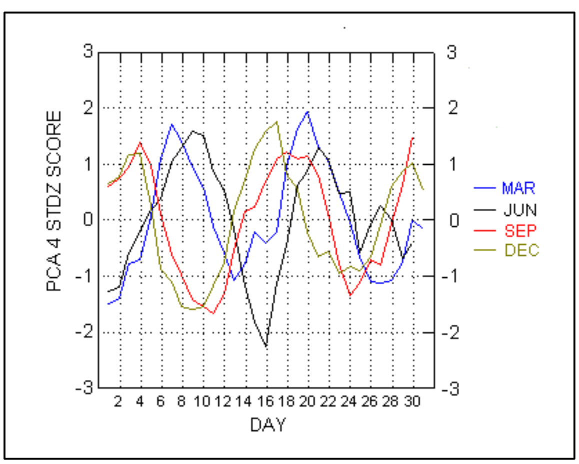

Figure 12. Fourth Component Correlation PCA Scores for March, June, September, & December - Downtown Los Angeles Daily Mean Temperatures 1921-2005)

3.3 Selected Actual Months Patterns and Their Conformances to Modes 1 through 4

The first four modes' patterns having been visualized/characterized, Figures 13 through 25 present actual daily mean patterns for selected months over the history that exhibited high conformances to one, and occasionally, combinations of the four.

For this demonstration, the emphasis is arbitrarily on January, although a selection of charts for a few other calendar months that displayed some particular feature(s) of interest is also included.

3.3.1 January First Component Plots

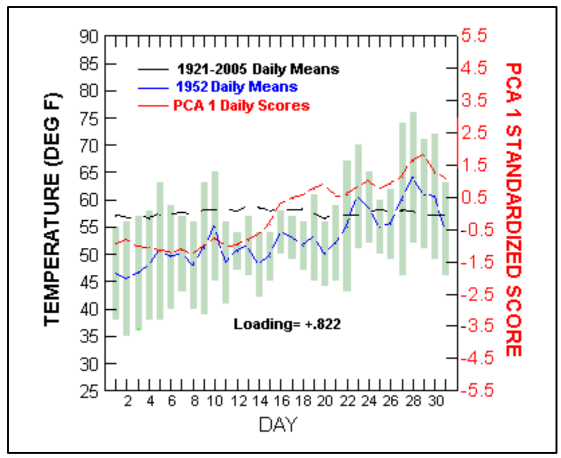

Figure 13. Downtown Los Angeles Daily Temperature Chart for January 1952 (High Positive Conformance of Daily Means to Mode 1)

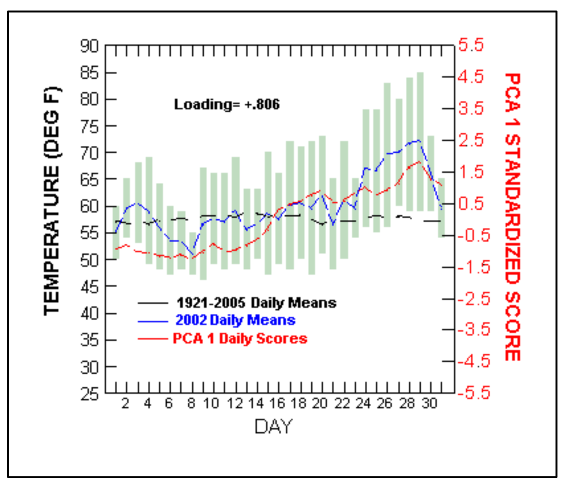

Figure 14. Downtown Los Angeles Daily Temperature Chart for January 1931 (High Positive Conformance of Daily Means to Mode 1)

Figures 13 and 14 are plots of the actual day-to-day temperature patterns experienced in Downtown Los Angeles for the Januarys of 1952 and 1931, respectively, months whose daily means showed high positive conformance to the scores of the January first mode (see Figure 1). In each graph, the blue dashed traces connect the actual daily means (arithmetic averages of the daily maxima/minima, depicted by the light green floating bars in the background). The black-dashed trace represents the 1921-2005 climatological daily means for January (exhibiting an almost trendless day-to-day pattern). All of these statistics reference the graph's left-axis scale.

In addition, red-dashed lines trace the first mode's standardized scores ("PCA 1") for January; these relate to the red-lettered scale on the right-axis. The range of values on the right-axis is arbitrarily stretched from that of its Figure 1 counterpart graph to facilitate a more intuitive comparison between the red-traced PCA 1 scores and the blue-traced daily means. In any case, the correlation coefficient (or loading) between the two, a measure of shape (a+bX form) similarity [Yarnal, 1993], objectively determines the degree of statistical conformance.

Both months exhibited an upward day-to-day trend in daily means, January 1952 (loading: +.822) essentially from the start of the month, January 1931 (loading +.806) from the second week on.

Based on the daily mean progressions, it at least appears that both months' patterns may have reflected the gradual progression from an upper-air trough regime (i.e., cooler daily means accompanied by occasional rains) to a ridge pattern (i.e., warmer daily means accompanied by drier conditions). This, however, was only partially true, especially in 1952. In that January, nearly 10 inches' rain fell over the 12th-18th, this especially wet (and presumably trough-influenced) period's cloudiness-induced decrease in the daily maxima more than offset by the increase in the daily minima, the net result being a continued upward trend in the daily means.

This early month vs. late month trough to ridge scenario was more valid in 1931, the month's coolest daily means, over the 7th-8th, being associated with heavy rains (nearly 2") on those two days. The ensuing three weeks saw a steady, progressive warming trend, with less than ½ inch additional rain received.

The loading statistics, of course, could be negative as well as positive, highly negative ones reflecting to some extent an idealized progression from a ridge to trough regime over the course of the month.

In this regard, the patterns for January 2002 (Figure 15) and January 1933 (Figure 16) exemplify highly *negative* loadings (correlations) with first mode scores, or alternatively, highly positive loadings with sign-reversed scores (multiplied by -1). Parameter-wise, the 2002 and 1933 graphs are identical to those of 1952 and 1931 except that the red PCA 1 traces reflect these sign-reversed scores.

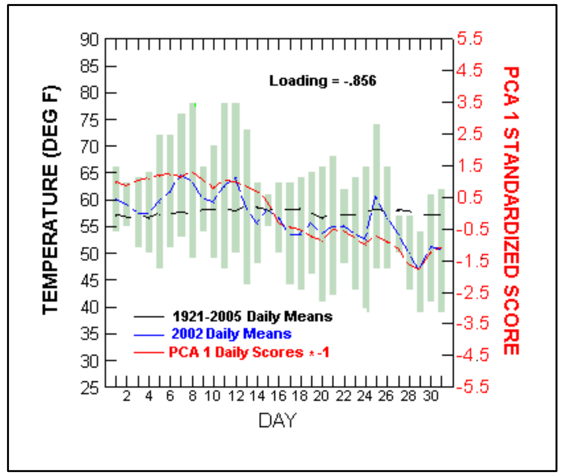

Figure 15. Downtown Los Angeles Daily Temperature Chart for January 2002 (High Positive Conformance of Daily Means to Mode 1 times -1)

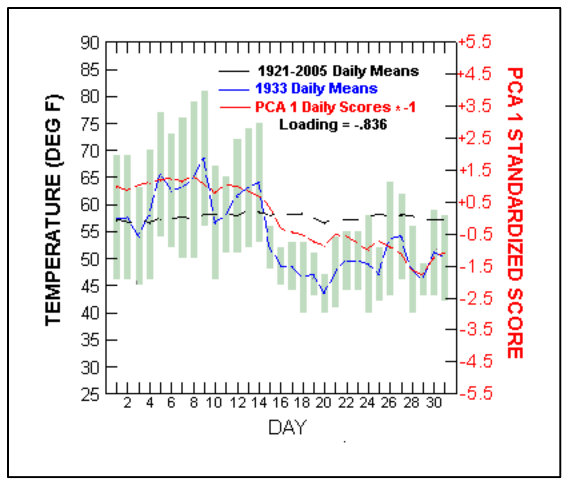

Figure 16. Downtown Los Angeles Daily Temperature Chart for January 1933 (High Positive Conformance of Daily Means to Mode 1 times -1)

January 2002 (loading: -.856) had the most extreme first mode loading magnitude of any of the 85 Januarys. Aside from a few individual-day "spikes", a clear and steady downward trend in daily means (blue trace) is evident along with an obviously strong correlation with the sign-reversed PCA 1 red trace. Only four days during the month had measurable rain, almost all of it (0.68" of the 0.80" monthly total) coming over the 27th-28th.

January 1933 (loading: -.836) exhibited a more abrupt and step-like downward progression in daily means. A ridge to trough scenario in this case was almost unmistakable – no measurable rain was recorded over the first fourteen days, more than 8 ½" on thirteen days over the last eighteen.

3.3.2 January Second Component Plots

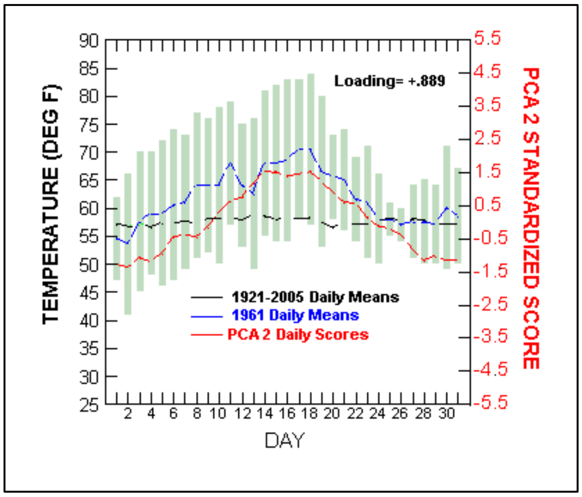

Figure 17. Downtown Los Angeles Daily Temperature Chart for January 1961 (High Positive Conformance of Daily Means to Mode 2)

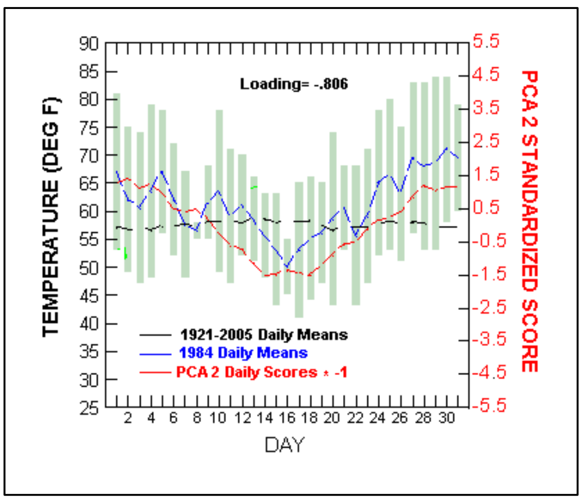

Figure 18. Downtown Los Angeles Daily Temperature Chart for January 1984 (High Positive Conformance of Daily Means to Mode 2 times -1)

Moving on to a demonstration of noteworthy mode two patterns, Figures 17 and 18 are plots of the day-to-day temperature patterns experienced in Downtown Los Angeles for the Januarys of 1961 and 1984, respectively, months whose daily means showed, in succession, highly positive and negative loadings relative to second mode scores. A hallmark feature of mode two were peaks (troughs) in scores, concentrated around mid-month.

January mode two scores, of course, were also uncorrelated with those of mode one, the percent of variance explained by the former, 16.30%, about 22 % less than the latter's (20.81 % - see Table 1).

Both of these two months were very rain-infrequent. January 1961 (loading: +.889) had only two measurable rain-days,1 ¼" recorded over the 25th-26th, associated with the low daily ranges depicted for those two days by the floating bars. January 1984 (loading: -.806) also had only two such rain-days, .16" recorded on the 16th, corresponding to the extreme low daily mean for the month, and .01" on the 21st.

Physically, each of these months were obvious examples of ridge-dominated droughty regimes interrupted by a one or two brief wet spells.

3.3.3 January Third Component Plot

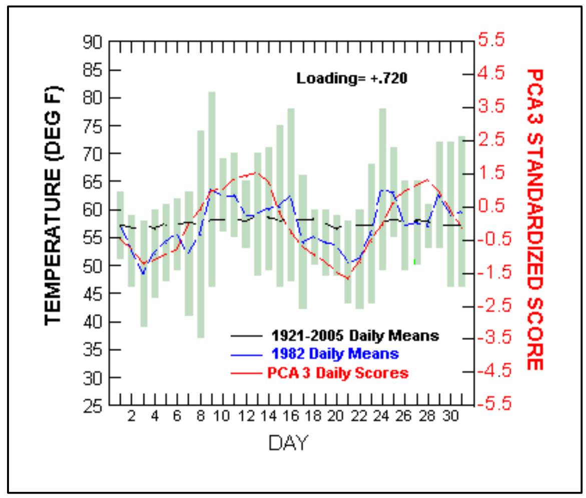

Figure 19. Downtown Los Angeles Daily Temperature Chart for January 1982 (High Positive Conformance of Daily Means to Mode 3)

Figure 19 is a plot of the day-to-day temperature pattern for January 1982 (loading: +.720), a month that exhibited a highly positive correlation with the January third mode. From Table 1, this mode explained 11.34 % of the variance, about 30 % less than that of Mode 2.

As previously described, the lesser ranking modes seemed to idealize progressively higher frequency quasi-sinusoidal patterns, and in Figure 19, the red trace depicts a daily scores' configuration with two relative minima – on the 3rd and 21st, the same dates of January 1982's coolest daily means. The highest scores also match up well with the month's highest daily means, the former appearing for the 13th and 28th.

The first five days of January 1982 were rainy, as were the 19th-21st.

3.3.4 January Fourth Component Plot

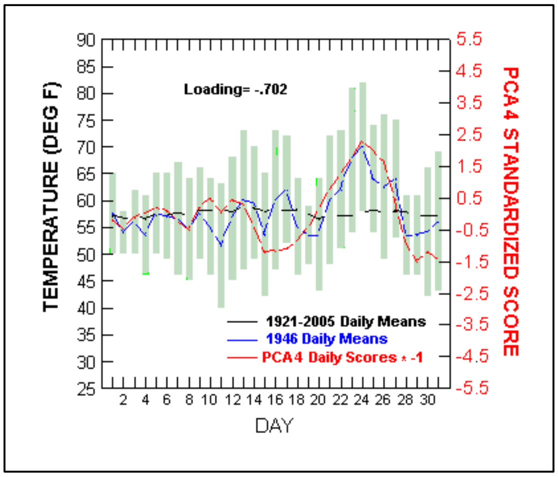

Figure 20. Downtown Los Angeles Daily Temperature Chart for January 1946 (High Positive Conformance of Daily Means to Mode 4 times -1)

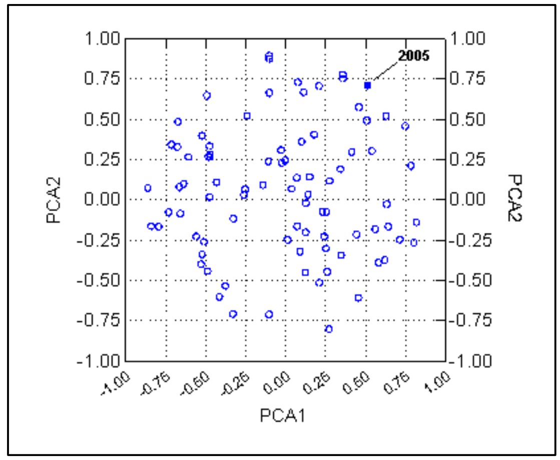

Figure 21. Scatterplot of January PCA 1 vs. PCA 2 Loadings - Downtown Los Angeles Daily Temperatures (1921-2005 Data)

Lastly, Figure 20 plots the day-to-day temperature pattern for January 1946 (loading: -.702), a month that exhibited a relatively high positive correlation on the sign-reversed scores of the January fourth mode. For the most part, the daily means correspond well to the scores (originals multiplied by -1) throughout the month, the maxima for each coming on the 24th.

January 1946 was the only such month that loaded above absolute value .70 on the fourth mode.

3.3.5 Multiple Conformances at Moderate Loadings Levels.

While the idealized component scores were uncorrelated in the aggregate, none of the individual monthly patterns, of course, were ever represented purely by one mode at the expense of all the others. The patterns loaded on all the components to at least some degree, a result of the varying mixes of long-wave, synoptic, and quasi-random influences on daily mean temperature progressions. As mentioned previously, about 10% of the months in the Downtown Los Angeles history loaded at or above absolute value 0.50 on more than one component simultaneously.

Figure 21 is a scatterplot of the January PCA 1 vs. PCA 2 loadings. While the scatter indicates no overall dependence of PCA 2 on PCA 1 (and vice-versa), there are a number of years that loaded relatively significantly on both modes in the absolute magnitude sense.

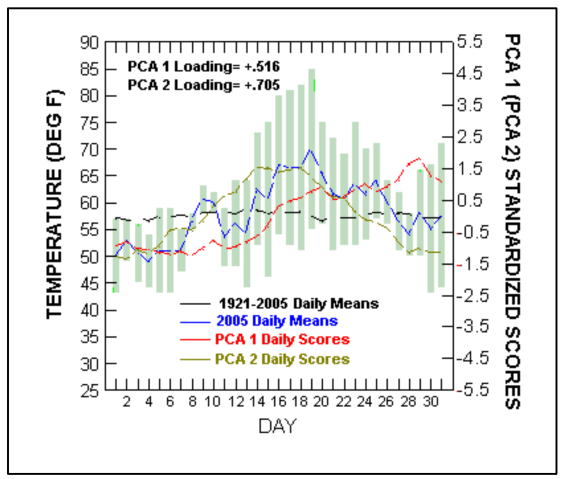

Figure 22. Downtown Los Angeles Daily Temperature Chart for January 2005 (Moderate Positive Conformance of Daily Means to Modes 1 AND 2

The day-to-day pattern for January 2005 (Figure 22) is a case-in-point (its loadings' point representation also shaded in blue in Figure 21), the month displaying both the signatures of an upward trend (idealized by the mode 1 red trace) and a mid-month peak (idealized by the mode 2 brown trace, which incidentally shows no net linear trend from the beginning to end of the month). The January 2005 daily means seemed to exhibit a moderate correspondence with the PCA 1 scores through day 25, departing noticeably thereafter - the overall +.516 loading figure reflects this. The more prominent signature, though, is the middle-of-month

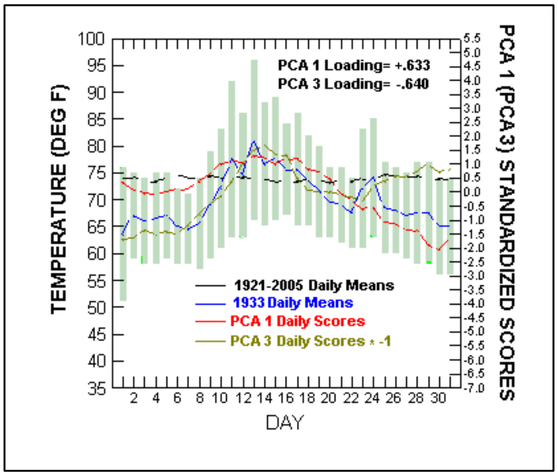

Figure 22. Downtown Los Angeles Daily Temperature Chart for August 1933 (Moderate Positive Conformance of Daily Means to Modes 1 AND Mode 3 Times -1

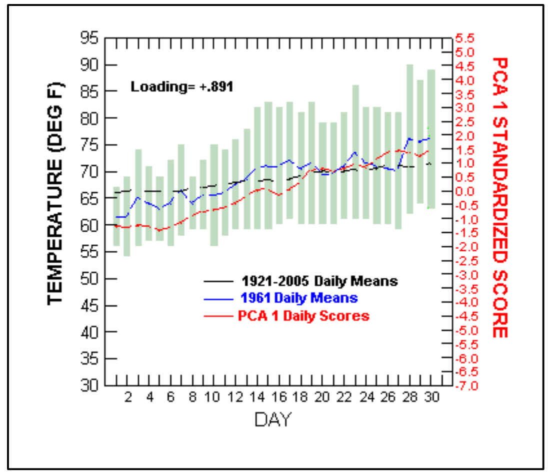

Figure 23. Downtown Los Angeles Daily Temperature Chart for June 1961 (High Positive Conformance of Daily Means to Mode 1)

peak, and thus January 2005's correlation with the PCA 2 scores was a significantly higher +.705.

In another example, Figure 22 shows the pattern for August 1933, the only month in the entire history that loaded higher than absolute value .60 on more than one component simultaneously.

Both modes (PCA 1and sign-reversed PCA 3) captured the relative peak feature, but sign-reversed PCA 3 (brown trace) better "handled" the relatively low daily means over the first week or so, PCA 1 (red trace) the relatively low ones over the last week. August 1933's daily means loaded on PCA 1 at magnitude +.633 and on PCA 3 at magnitude -.640.

3.3.6 A Few Other Charts of Interest.

As described previously, the month-to-month percents of variance explained for Downtown Los Angeles in Table 1 seemed to be more similar than different in magnitudes, mode-by-mode. The major exception was the first component percentage for June, some 30.77% of the variance explained, more than 7% higher than that for the next higher figure, that for the July first mode (23.54%).

Figure 23 is the pattern for June 1961, that which loaded to the highest degree (+.891) on mode 1. The red trace depicts a steadily upward trend in scores, similar in pattern to the black-dashed climatological daily means, but steeper. Of the 43 years that loaded at or above 0.50 absolute magnitude on this mode, all of their signs were positive, reflecting the upward progression in day-to-day mean temperatures that is the rule at this time of year. Only six loadings were negative, the most pronounced a modest -.229.

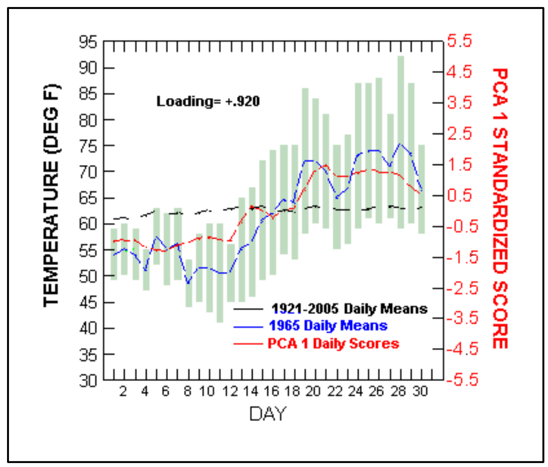

Figure 24. Downtown Los Angeles Daily Temperature Chart for April 1965 (Highest Loading Statistic, +.920, of any in the 1921-2005 history)

The highest individual loading statistic for any month and mode (+.920), was April 1965's loading on its first component scores. Figure 24 depicts the day-to-day pattern.

The red-trace depicts a basically trendless first twelve days in scores, followed by a steep rise through the 21st, succeeded by a leveling off through month-end. This matches very well with the actual progressions in daily means that were recorded in April 1965.

Finally, Figure 25 shows the pattern for January 1941 (Figure 25) the January with the least average absolute daily departure (2.1 F) from climatology (black traces) in the 85-year history. Inspecting the blue traces, depicting the actual daily means, it is evident that even that year displayed a more quasi-sinusoidal than linear pattern, with three separate slight troughs and peaks in the daily means.

The pattern conforms to a moderate degree (loading +.475) to PCA 3 (red trace), a slight phase shift discrepancy partially responsible for the somewhat reduced positive loading figure.

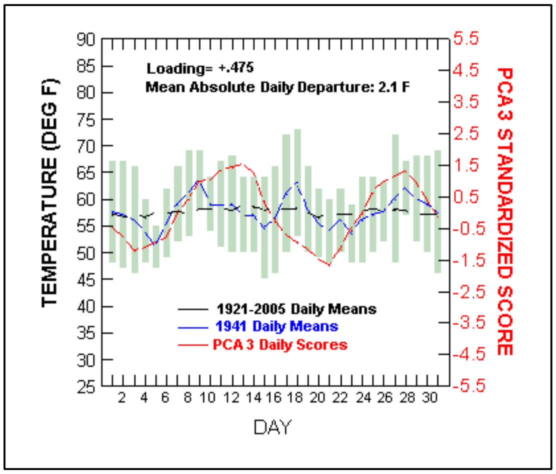

Figure 25. Downtown Los Angeles Daily Temperature Chart for January 1941 (Least Average Absolute Departure from Daily Mean Climatology of any January in record – 1921-2005.

3.4 Analyzing the Components Collectively

The above has shown how a Linear Correlation PCA decomposed and ordered a hierarchy of day-to-day mean temperature variability modes for Downtown Los Angeles.

Some visual examples, the most highly correlated cases to these idealized patterns, were also displayed for each of the first four modes

For the most part, however, the percents of variances explained, even for the most important components, were relatively modest. Also, the long-wave and more synoptic-scale influences were essentially portrayed as separate, uncorrelated components, which departs from physical reality (i.e., the two almost always influence day-to-day mean temperature variability during the course of a given month in tandem, synoptic scale features being embedded in the long-waves.)

For the above PCA to convey more useful and realistic information (for example, creating a pattern "typology"), the components should be considered collectively in some fashion.

3.4.1 Multiple Regression Treatment of the Component Scores Information.

One method that could be used to analyze the components' information as a group is multiple regression analysis. For a calendar month of choice, each of the individual years' arrays of daily means could be fit to a multiple regression model, incorporating a range of components (i.e., their scores) as independent variables. The resulting regression coefficients could then be evaluated for statistical significance individually at some critical value, those that rejected the null hypothesis designated as a significant mode(s) of variability. Since the original component scores were, by mathematical design, uncorrelated, there would be no multicollinearity issues.

Selecting January again as the illustration month, a multiple regression model incorporating the first four modes was fitted to each of the 85 years' day-to-day daily means' patterns. The component scores did not have to be signed-reversed, as the resulting coefficients could be either positive or negative, and the statistical significance test was two-tail at the .05 level. From the results, a crude typology of pattern mixes was thus possible. Table 2 lists the 15 possible combinations of modes, one to four, and the number of Januarys whose coefficient magnitudes were statistically significant.

Possible Combinations of Modes 1 to 4, taken once, twice, thrice at a time, or all at once (1)	# of Januarys with Statistically Significant Regression Coef-ficients on given combination in (1) (2)
1,2,3	15
1,2,3,4	3
1	5
2	2
3	3
4	6
1,2	8
1,3	5
1,4	3
2,3	3
2,4	5
3,4	4
1,2,4	7
1,3,4	6
2,3,4	2
None	8

Table 2 - Frequency Tabulation of Number of Januarys with Significantly Significant Regression Model Coefficients at .05 level (two-tail test), by Modal Combination - Downtown Los Angeles Daily Mean Temperature Time Series, (1921-2005 data)

From Table 2, sixty-one of the Januarys had significant coefficient magnitudes on at least two modes, with only eight of no exceedances (the "none" category). These eight had their significant magnitudes on modes 5 and higher. This confirmed the rather obvious notion that Downtown Los Angeles January day-to-day patterns in mean temperature are multi-modal in character, at least those defined by the Linear Correlation PCA.

The most frequently represented category with statistically significant regression coefficients on all specified modes was the "1, 2, 3" combination. Some 15 Januarys had statistically significant coefficient magnitudes associated with each of these three. This implied an "ensemble" pattern of long-wave (ridge to trough or trough to ridge), mid-month peak (or trough), and a double trough-to-peak (or peak-to-trough) couplet in daily means.

Multiple correlation values (analogous to loadings) for the 15 Januarys ranged from +.714 to +.926. Figure 26 shows the day-to-day pattern for January 2001, the most highly correlated pattern.

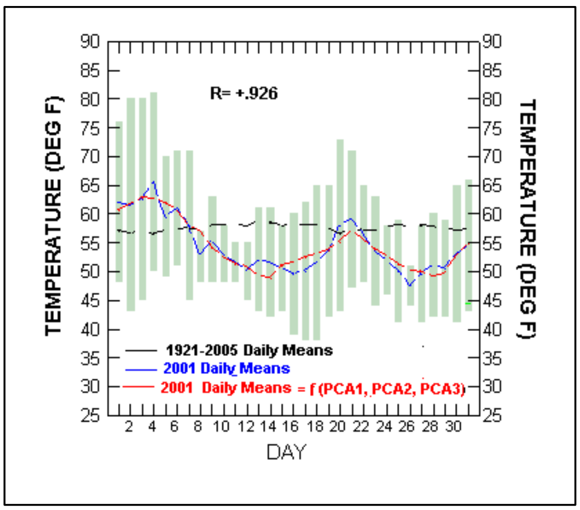

Figure 26. Downtown Los Angeles Daily Temperature Chart for January 2001 (Highest Multiple Correlation Statistic, +.926, on Multiple Regression Model incorporating January Modes 1 through 3 as independent variables)

The coefficients in the model were all negative, the exact expression being:

Y= 54.387 –2.269*PCA1 –2.100*PCA2 – 3.054*PCA3.

From the chart, a general downward trend in means is evident, expressing (sign-reversed) mode 1 and the mid-month trough (sign-reversed mode 2). The more negatively weighted coefficient for mode 3, again reflecting a sign-reversed pattern, is responsible for the secondary peak around day 20. During the month, more than 4 inches' rain was received over the 10th-12th, and about 1 inch over the 24th-26th.

In addition from Table 2, three cases had statistically significantly magnitudes on all four modes (the "1,2,3,4" category). Reverse-signed Mode 4, it can be seen from the Figure 20 red trace, portrays a sinusoidal pattern of progressively increasing amplitudes which displays a pronounced peak late in the month, on the 24th.

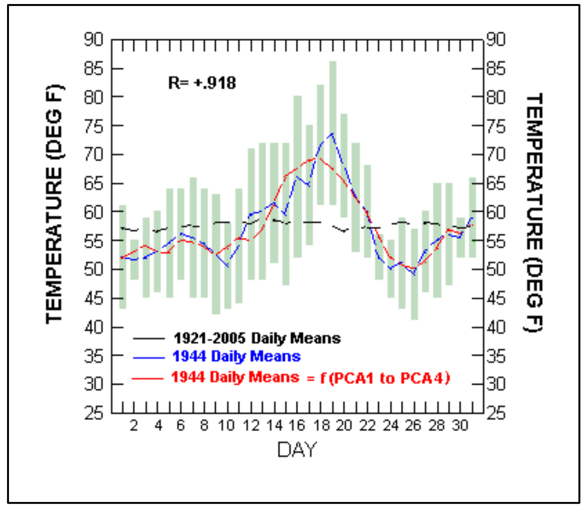

Figure 27. Downtown Los Angeles Daily Temperature Chart for January 1944 (Highest Multiple Correlation Statistic, +.918, on Multiple Regression Model incorporating January Modes 1 through 4 as independent variables)

Figure 27 shows the day-to-day pattern for January 1944, the most highly correlated pattern (+.918) of those Januarys that had statistically significant coefficients on each of the first four modes (the other two such years, 1922 and 1995, had +.826 and +.706 correlation figures, respectively). All but Mode 3's coefficients are positive, the multiple regression expression being:

Y= 57.194 +1.35*PCA1+ 4.444*PCA2 -2.277*PCA3 +2.66*PCA4

Evident from Figure 27 is a slight overall upward trend (mode 1), superimposed on a mid-month peak (mode 2), superimposed further on a double peak/trough in reverse sign form (mode 3). As in Figure 26, the coefficient weighted mode 3 scores helped produce the peak around the 20th. Mode 4 helped reinforce the post mid-month peak somewhat, but its coefficient-weighted scores were much more influential in producing the dip around the 24th.

A second variation on the multiple regression approach would be to fit the monthly patterns as before to a pre-selected range of component scores, except that the additional step of refitting the data using components attached only to statistically significant coefficient magnitudes (as was done for Figure 26) would be omitted.

From Table 1, ten of the Januarys, or 11.8%, loaded on the first component at the absolute value of .70 or higher. If this second variation multiple regression approach is applied to components 1 and 2 only, the count of model fitting correlations at +.70 or higher (multiple correlations are always positive) increases to 30 of the 85 Januarys. Fitting the first through third modes enhances the tally to 46, and for the first through fourth components, the frequency increases to 60. Incorporating the fifth component scores into the model raises the count to 69.

Viewing the fifth component results another way, a model with just five synthetic independent variables (or component scores), which numbered one-seventeenth the total number of years in the history, could fit greater than 80% of the individual years' daily mean temperature patterns to a 0.70 or better correlation.

Possibly PCA information of this kind could be used to create realistic simulations of daily mean temperature patterns. One complicating factor that would likely have to be addressed, however, is the existence of intercorrelations between the component *coefficients*, and with the multiple regression constant. The 85 four-component January models, at least, had slight (less then absolute value .10) intercorrelations between coefficients, but PCA 3 had a +.176 correlation of its coefficients with the corresponding regression constants. Other issues to be considered in this regard would be the relative normalities or non-normalities of the regression coefficients' distributions (as well as that of the regression constant). Resampling techniques would likely prove useful for these tasks. Preceding all of this, of course, the number of components to be actually included in the model would have to be decided upon.

.

4. SUMMARY

Utilizing Linear Principal Components Analysis, the foregoing analyzed and visualized, by calendar month, the most prominent patterns of day-to-day mean temperature variability for Downtown Los Angeles, based on the 1921-2005 period of record. Results focused on the first four statistical components or "modes", quantified by component scores.

Percents of variance explained by the first component were 21% on average (highest: 31% for June, lowest: 16% for April), and 56% percent on average for the first four (highest: 61% for July, lowest: 46% for April).

Mode 1 seemed to reflect propagation of long-waves and, occasionally, seasonal trend as the most important influence on day-to-day mean temperature variability over the course of a month. Lesser modes seemed to portray progressively higher frequency quasi-sinusoidal patterns with varying phase shifts from month-to-month, suggestive of a gradation in relative importance to synoptic-scale influences and more random influences peculiar to a few years in the sample history.

A large proportion of the individual monthly patterns exceeded the 0.50 loading figure on at least one component, sometimes two at a time (about ten percent of the cases). Only about one-fourth did so on the 0.70 benchmark, never more than one component at a time at this level.

Focusing primarily on the month of January for illustration purposes, a number of actual years' patterns were presented that conformed particularly well on modes 1 through 4 individually.

To demonstrate the usefulness of the PCA information as a foundation for further evaluating combinations of modes (more representative of physical reality), multiple regression prediction models were generated utilizing the first few modes' scores as independent variables. More than 70% of the 1921-2005 January daily mean patterns, for example, could be fit at a multiple correlation level of .70 or higher using the first four components' scores alone as independent variables. A few actual years' patterns were visualized that had been precisely replicated by these multi-modal models.

It would interesting to compare these results with those of other first-order stations with lengthy histories, although with long-wave and synoptic-scale influences on day-to-day temperature variability so inherent in the middle and higher latitudes, they would undoubtedly be similar. Stations with more seasonal contrasts in temperature than Downtown Los Angeles, however, might show more variances explained on the first component, especially for those calendar months which typically exhibit pronounced warming or cooling trends over their course.

5. REFERENCES

Johnson, R. A., and Wichern, D.W., 1982*:* *Applied Multivariate Statistical Analysis Prentice Hall, Inc.,* 594pp.

Wilkinson, L., Blank, G., and Gruber, C.,1996; *Desktop Data Analysis with SYSTAT,* Prentice-Hall Inc., 798 pp.

Yarnal, Brent, 1993: *Synoptic Climatology in Environmental Analysis, A Primer.* Bellhaven Press, 195 pp.

Fisk, C., 2004: "Objective Identification of Extreme-Most Anomalous Daily Max/Min Historical Temperature Patterns Using Principal Components Analysis". American Meteorological Society 17th Conference on Probability and Statistics in the Atmospheric Sciences. Seattle, 2004. **ams.confex.com/ams/pdfpapers/69198.pdf**