3.2 OBJECTIVE IDENTIFICATION OF EXTREME-MOST ANOMALOUS DAILY MAX/MIN HISTORICAL TEMPERATURE PATTERNS USING PRINCIPAL COMPONENTS ANALYSIS

Charles J. Fisk *
NAWCWPNS-Point Mugu, CA

1. INTRODUCTION

A common means of depicting day-to-day temperature data is the floating-bar or "hi-lo" chart (also used extensively for presenting daily stock price data). As utilized, a given bar's top boundary represents the daily maximum temperature, usually from midnight-to-midnight local time, the bottom boundary the daily minimum; the length of the bar indicates the daily temperature range. Depending of the time interval shown (usually a calendar month or complete calendar year), the day-to-day progression of bars reflects the varied diurnal, synoptic, long-wave, and seasonal influences on temperature over time. In viewing an individual chart, the question may arise on how typical or unusual the pattern is, especially if it is irregular or in some other way appears anomalous. In this regard, it would be perhaps be interesting for its' own sake and useful to have a pattern-oriented measure or measures that can characterize a given period's maximum and minimum daily temperature "configuration" relative to those of other years.

Utilizing daily maximum and minimum temperature data from a first-order station with a lengthy history (Los Angeles Civic Center, 1921-2002), the purpose of the following is to explore Principal Components Analysis (PCA) as a tool for objectively identifying extreme-most maximum/minimum daily time series patterns in temperature for a hierarchy of time intervals (annual, seasonal, and monthly – seventeen intervals in all). As it was considered desirable to have as long a record as possible to work with, and the primary objective was exploratory, the issue of data set inhomogeneity was not considered critical.

In a previous study [Fisk, 1995], average correlations and average covariances of individual-year maximum/minimum temperature time-series vs. other years' same-chronology series were used to identify most "anomalous" patterns, as indicated by highest or lowest mean correlation or covariance rankings. The analyses were done for a single station by calendar month. Among the results, highest average correlations were associated with day-to-day maximum/minimum patterns that closely resembled climatological maxima and mimima profile appearances, irregardless of overall departure from normal. Average covariances were affected by unseasonable temperature progressions too, and the most extreme values also seemed to reflect peculiar patterns in average daily ranges (average daily maximum less average daily minimum), be it magnitude or trend.

The initial objective of this study was to use average correlation statistics to identify the most anomalous complete calendar-year patterns for a collection of stations, but along the way it was determined that if a correlation matrix PCA was done instead, the first component loadings were almost perfectly correlated on a year-by-year basis with their counterpart average correlations. Moreover, correlation PCA standardized first component *scores* were also almost perfectly correlated with climatological mean daily maxima and minima statistics (See Figure 1 below). This same almost perfect association was seen for first component covariance loadings versus counterpart average covariances. Thus, the decision was made to go with the more conventional PCA approach, confine the analysis to a single station, and add meteorological seasons and calendar months to the time-interval selection. An additional advantage of the PCA method, especially the correlation matrix option, is that it provides relatively interpretable information on other components or "modes".

Yarnal [1993] compared and contrasted the use of correlation and covariance PCA's in the context of synoptic climatology pattern-recognition and classification, the broad concepts of which are applicable here. In brief, the correlation coefficient is strongly influenced by feature or "shape" similarities, absolute magnitude contrasts or offsets across maps of otherwise identical patterns producing identical correlations. This property is not unlike the high average correlations, described above, that were generated between monthly maxima/minima temperature series that conformed closely to climatological profiles, irrespective of departures from the mean. Covariance, the formula of which is identical to that of the correlation except that the x and y standard deviation product term is excluded from the denominator, incorporates the absolute magnitude differences or anomaly ("intensities") that the correlation is transparent to, although it can be biased "by large variances related to the seasonal cycle".

2. METHODS AND PROCEDURES

Generation of pattern statistics for a specified calendar period of interest was accomplished as follows. First, each individual year's maxima and minima were merged into single column arrays of length N times 2, where N was the number of days in the interval. For example, since there were 82 years' Los Angeles Civic Center data, an "annual" array of 365 pairs of daily

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

maxima and minima (Feb. 29 leap year data excluded) would have 730 rows and 82 columns. Correlation and covariance PCA's were then performed (O-mode decomposition with no rotation), the loadings' statistics calculated by year and ordered by magnitude. Graphs of the original daily maxima/minima for some of the more interesting and extreme anomalous patterns, as identified by both maximum and minimum correlation and covariance loadings statistics, were also made. In addition, using joint correlation and covariance loadings information, sample confidence ellipses were constructed to identify the most extreme patterns in the bivariate sense.

3. RESULTS

3.1 – Correlation Matrix PCA

Table 1 lists various first component statistics connected with the Los Angeles Civic Center Correlation Matrix PCA for the 17 individual calendar intervals. For each, the first or "reference climatology" component (see column 1) was by far the most important, explaining between 69.3% of variance for December-February ("WNTR") to 87.9% for July. These percent-of-variance figures were almost identical, period-by-period, to those exhibited by the covariance matrix first component (not shown). In general, the proportions were highest for summer, lowest for the winter and spring. Highest second component figure was just 5.5%, for January, the lowest a meager 1.0% for the complete calendar year period (see column 4). The other columns (2,3,5, and 6) identify the highest and lowest first component correlation and covariance loadings statistics, respectively, and the years for which they were calculated.

Table 1 - Selected First Component Correlation and Covariance Matrix PCA Statistics for Los Angeles Civic Center Daily Maximum/Minimum Temperature Time Series, by Period (1921-2002 data)

PERIOD	CORR FIRST COM-PONENT %-VAR (1)	CORR MATRIX MAXIMUM LOADING & YEAR (2)	CORR MATRIX MINIMUM LOADING & YEAR (3)
YEAR	78.4	.915 ('48)	.852 ('71)
WNTR	69.3	.896 ('93-4)	.730 ('22-3)
SPR	71.7	.902 ('02)	.780 ('67)
SUM	84.9	.961 ('48)	.837 ('79)
FALL	75.3	.923 ('47)	.741 ('33)
JAN	70.5	.922 ('66)	.664 ('71)
FEB	71.7	.945 ('33)	.672 ('96)
MAR	73.0	.954 ('49)	.745 ('86)
APR	71.8	.941 ('44)	.667 ('89)
MAY	75.4	.963 ('52)	.765 ('67)
JUN	84.1	.982 ('58)	.766 ('79)
JUL	87.9	.984 ('63)	.873 ('85)
AUG	87.0	.973 ('23)	.855 ('35)
SEP	77.8	.965 ('32)	.698 ('39)
OCT	73.9	.950 ('44)	.733 ('21)
NOV	75.7	.941 ('99)	.756 ('24)
DEC	73.4	.946 ('76)	.735 ('55)
PERIOD	CORR SECOND COM-PONENT %-VAR (4)	COV MATRIX MAXIMUM LOADING & YEAR (5)	COV MATRIX MINIMUM LOADING & YEAR (6)
YEAR	1.0	12.567 ('49)	9.662 ('28)
WNTR	2.5	11.085 ('75-6)	7.319 ('26-7)
SPR	2.3	10.191 ('87)	7.078 ('83)
SUM	1.5	11.002 ('52)	8.234 ('93)
FALL	2.3	12.576 ('48)	8.430 ('32)
JAN	5.5	13.394 ('76)	6.489 ('95)
FEB	5.3	12.235 ('02)	5.923 ('62)
MAR	3.8	11.093 ('88)	6.102 ('83)
APR	5.0	11.443 ('87)	6.702 ('83)
MAY	3.7	10.586 ('78)	6.034 ('82)
JUN	2.6	11.516 ('76)	5.807 ('69)
JUL	2.0	11.938 ('47)	7.049 ('93)
AUG	2.5	11.823 ('48)	7.938 ('79)
SEP	4.2	12.854 ('48)	6.770 ('76)
OCT	3.6	12.694 ('99)	7.257 ('74)
NOV	3.8	13.142 ('48)	7.307 ('82)
DEC	4.3	12.364 ('76)	6.247 ('55)

Figure 1 is a plot of first component day-to-day standardized scores for the full calendar year. Since, as stated above, the scores were almost perfectly correlated with the 82-year statistical mean daily maxima and minima (shown in Figure 2) the profiles are virtually identical in appearance or ("shape").

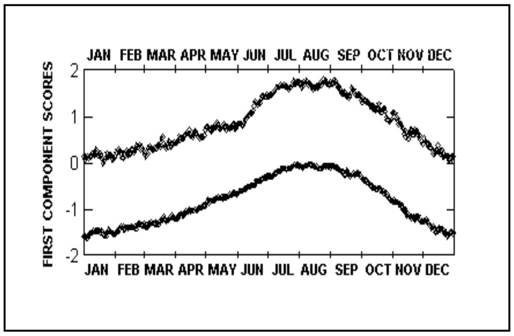

Figure 1. Calendar Year Plot of First Component Standardized Scores from Correlation Matrix PCA of Los Angeles Civic Center Daily Maximum/Minimum Temperatures 1921-2002)

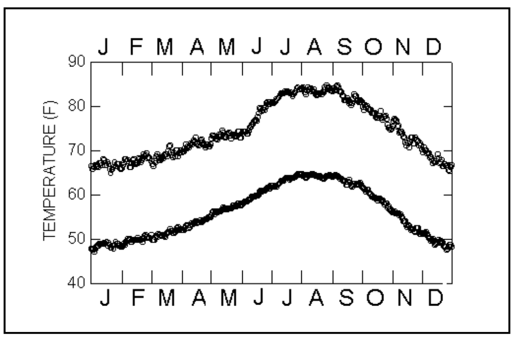

Figure 2. Calendar Year Plot of Average Daily Maximum and Minimum Temperatures for Los Angeles Civic Center (1921-2002)

***3.1.1** – Some Individual Period results*

Next, a few graphs are presented for those years that produced the most extreme correlation matrix loadings statistics. The selection includes the complete calendar year ("YEAR"), a four-month segment ("FALL"+ December), and two individual calendar months (June and September).

Figures 3 and 4 are floating-bar graphs of the daily maximum and minimum temperatures for 1948 and 1971, respectively, the years with the most extreme correlation first component loadings (+.915 and +.852, respectively).

The slight range in these values indicates that the annual patterns are certainly more alike than different, but differences nonetheless exist that are visually apparent. While 1948 exhibits its' share of day- to-day variability, overall conformance to the first-component

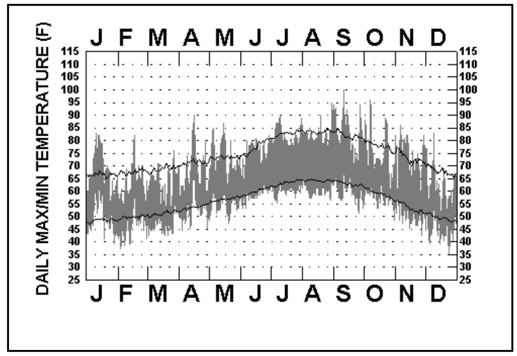

Figure 3. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1948 (Maximum First Component Correlation Loading Statistic)

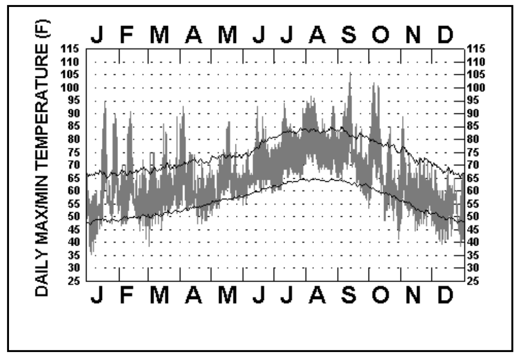

Figure 4. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1971 (Minimum First Component Correlation Loading Statistic)

absolute scores' profile or "shape" is such so as to create the maximum statistic. Further exemplifying the nature of this "shape" property, if a constant is added to each of 1948's maxima and minima and the PCA redone, the exact same loading magnitude and rank would result. Also, as already alluded to, if the daily max/min temperatures are correlated with the climato-logical max/min's (depicted by the upper and lower line traces in the figure) the statistic would be virtually identical. While 1948's pattern is "anomalous" in its own right, 1971's extreme minimum figure identifies a much more visually obvious extreme pattern, the salient property of the correlation loadings minimum statistic. Among 1971's main features are marked short-term variability, numerous "spikes" of much above normal temperatures spells lasting several days ("offshore" or Santa Ana episodes), and a number of days with reduced daily ranges.

An issue concerning the correlation coefficient is that it can sometimes be unduly influenced by extreme values in a data set. With a complete calendar year analysis of this kind, the seasonal effect (higher temperatures typically in summer, lower in winter) might create such a bias. To visually assess this possibility with the Los Angeles Civic Center data, Figure 5 is a correlation influence scatterplot of the 1971 daily maximum and minimum temperatures vs. first component scores. There is only a slight change in the sizes of the symbols corresponding to the low and high temperature (x-axis) regions, indicating that a significant bias does not likely exist, probably at least in part due to the large number of data points overall (n=730). The plot for 1948 is similar in this regard, and from these two cases it might be concluded that the loading (correlation) statistics are essentially unbiased representations of "shape" anomalies for the Los Angeles Civic Center. One might also make the case that since perception of a chart's pattern is probably influenced to a large degree by the most extreme temperatures depicted and their relative calendar-day position, perhaps they "should" be more important in producing an overall "shape" statistic.

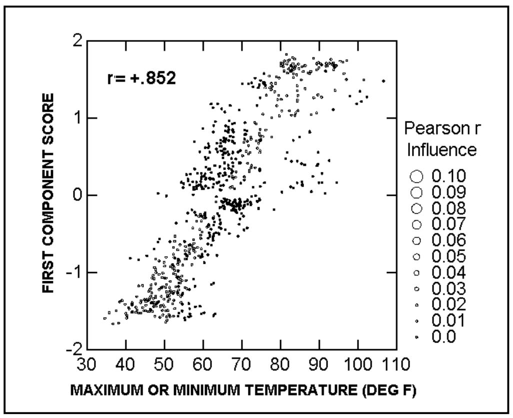

Figure 5. Correlation Influence Plot of Los Angeles Civic Center 1971 Daily Maximum and Minimum Temperatures vs. First Component Correlation PCA Scores

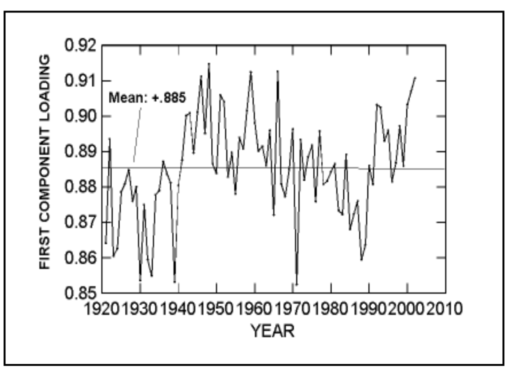

Figure 6. Time-Series Plot of First Component Correlation Loadings for Individual Calendar Years, Los Angeles Civic Center (1921-2002)

In addition to identifying the extreme-most patterns, one could also do a time-series plot of the loadings year-by-year to visualize their variation over the station history. Figure 6 is a time-series plot of the correlation loadings for the Civic Center "annual " series, 1921-2002. Allowing for station/instrumentation changes and evolving microclimatological influences over time, most of the years prior to 1940 had low magnitudes or comparatively irregular, "non-climatological-like" profiles, in the mold of 1971. In contrast, since about 1990, most of the years have had relatively high magnitudes or more regular, "linear" profiles, similar to 1948.

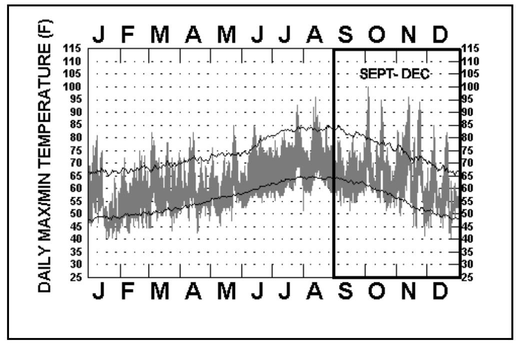

Figure 7. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1933 (Minimum September-December Correlation Loading Statistic)

Focusing on the minimum correlation loading statistic and reducing the time-interval from a full year to four months, Figure 7 shows the highly irregular pattern for September-December 1933. This covers the "conventional" Fall meteorological season (September-November) plus December. Superimposed on a seasonally normal downward trend in temperature are six brief episodes of far above normal temperatures

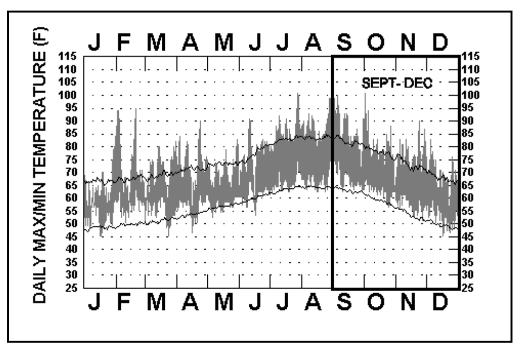

Figure 8. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1995 (Maximum September-December Correlation Loading Statistic).

and low daily ranges. The correlation loading statistic for September-December 1933 (+.776) is 3.8 standard deviations below the mean, second most extreme statistic of its kind calculated in the study.

In comparison, Figure 8 shows the chart for 1995, the year with the highest September-December correlation loading statistic (+.931). Evident is a much more “linear” floating-bar pattern. Nearly all the daily minima and most of the maxima are above seasonal norms, but the general profile “shape” conforms well to climatology, which is the governing factor for a high statistic.

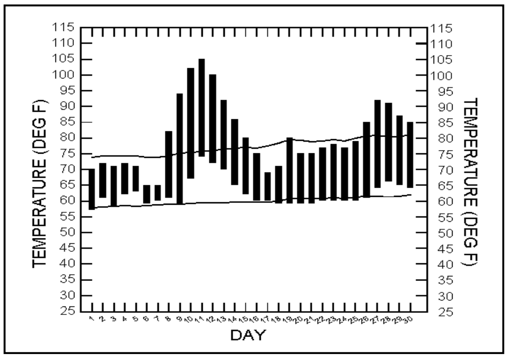

Figure 9. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for June 1979 (Minimum June Correlation Loading Statistic)

Further reducing the time scale to the individual calendar month level, Figures 9 and 10 show sample most extreme patterns (minimum loading statistics only) for the months of June (1979) and September (1939), respectively.

June 1979 (Figure 9) exhibited an exceptionally late and several-day long offshore flow event during the second week, with daily maxima as high as 105 F (30 F above average) and several days’ minima in the 70’s. Daily mean temperature on 11 June (88.5 F) is 5.1 standard deviations above the 1921-2002 average for the day.

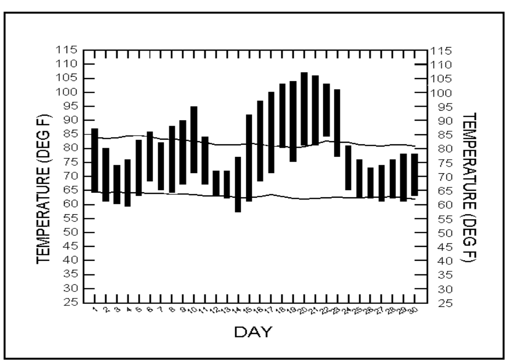

Figure 10. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for September 1939 (Minimum September Correlation Loading Statistic)

The correlation loading statistic for the month’s pattern (+.766 – see Table 1) is itself 4.2 standard deviations below the mean, the most extreme statistic of its kind generated in the study, and further exemplifying the spell’s uncommon character.

September 1939 (Figure 10) experienced a remarkably intense and protracted heat wave late in the month, with seven successive daily maxima at 100 F or higher and five of seven days with minima in the 80’s. Daily mean temperature for the 20th is 4.6 standard deviations above average for the day. This record heat spell was likely associated at least in part with the divergent outflow, subsidence, and offshore flow resulting from the trajectory of an approaching Eastern Pacific hurricane, the remnants of which abruptly terminated the heat wave when it made landfall as a tropical storm at Long Beach. September 1939’s loading figure (+.698 – see Table 1) is 3.2 standard deviations below the mean.

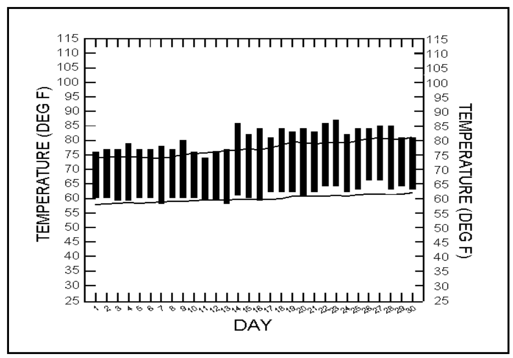

Figure 11. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for June 1958 (Maximum June Correlation Loading Statistic)

Finally, Figure 11 shows the maxima/minima for June 1958, that with the highest correlation loading statistic (+.982) of any June, and second highest figure encountered for any individual period in the study. Exhibited is an atypically uniform, linear, and thus, ”anomalous” pattern.

3.1.2 – Second Principal Components

While second components never explained more than 5.5% of the variance in a correlation matrix (January) nor had eigenvalues greater than 4.5 (also for January), it is of interest to visualize what sort of “idealized” pattern they describe. Using January’s results as an example, Figure 12 is a plot of its’ second component scores. For a given day of the month, the score for the daily maximum comes first (labled with an “X”), followed by that for the minimum (labeled with an “N”), plotted a half day to the right; the points are also connected with lines. The pattern shows exclusively positive scores over the first half of the month with a steady downward trend thereafter almost to the close. For most days after mid-month, the maxima scores are actually less than those of the minima, not a physical sensible result, but not inconceivable considering that the daily maxima and minima were included in the same data set. This, of course, was not an issue with the first component scores plotted in Figure 1 (not line-connected).

January 1933's second component loading magnitude (+.526) was one of the highest such figures exhibited for a non-first component in this correlation matrix PCA application. Figure 13 shows the month's daily maxima and minima plotted in the same format as Figure 12, Figure 14 is the floating bar version. From inspection, the standardized scores' "shape" pattern for the second component resembles the daily maxima and minima of January 1933 in a modest sort of way. The major contrasts are the low daily "ranges" of scores in Figure 12, and of course, the cases of "daily minima" scores being higher than "daily maxima" ones.

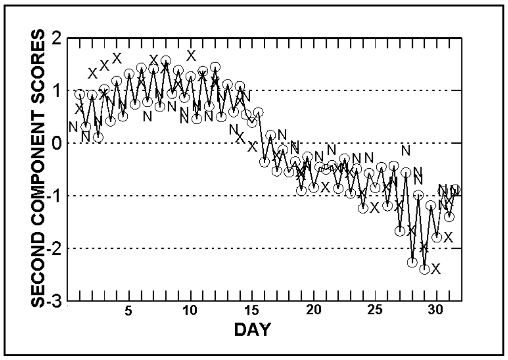

Figure 12 - Plot of January Second Component Standardized Scores from Correlation Matrix PCA of Los Angeles Civic Center Daily Maximum/Minimum Temperatures 1921-2002

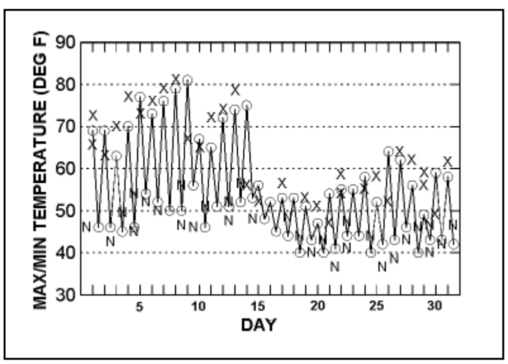

Figure 13 - Plot of January 1933 Los Angeles Civic Center Daily Maximum/Minimum Temperatures (same format as Figure 12)

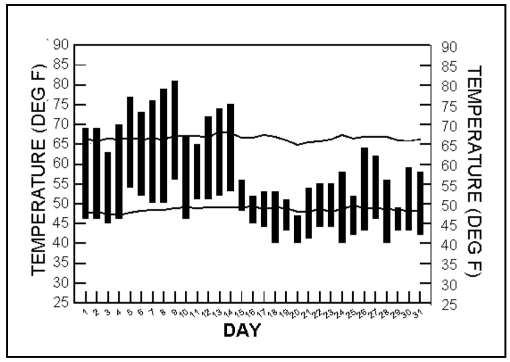

Figure 14 - Floating Bar Plot of January 1933 Los Angeles Civic Center Daily Maximum/Minimum Temperatures

3.2 – Covariance Matrix PCA

First component correlation loadings statistics, as evidenced by the graphs, seemed to discriminate well between Civic Center daily max/min pattern extremes, and the component's standardized daily scores were almost perfectly correlated with a simple "proxy" variable, the climatological maxima/minima. Additional information was yet to be had by examining first component covariance loadings figures.

While the correlation loadings statistic was predominately a "shape" measure, the first component covariance loadings figure captured "spread". In this floating-bar temperature application, the spread had two intertwined components, a diurnal one, since both maxima and minima were being analyzed from the same data set, and a seasonal one, analogous to the "seasonal cycle" influence described by Yarnal.

Table 2 includes some selected statistics that quantify, by period, the statistical associations between first component covariance loadings and four other variables of interest. From column 1 (relationships with first component correlation loadings), with the exception of April (r= .000) and June (r=-.096), the figures are all modestly positive, most so for December (+.508), January (+.479), and "WNTR" (+.456); the calendar year's figure is +.387. While the two measures are thus not independent or orthogonal (April the lone exception), the associated correlation magnitudes are still low enough to reject the notion of collinearity or redundancy of information. Average figure for the seventeen correlations in column 1 is just +.256.

Columns 2 and 3 list the correlations with average daily ranges (the "diurnal" component), and the standard deviation of maxima and minima as a unit (capturing the "seasonal" component if the period is long enough), respectively. Both variables are highly correlated with the covariance loadings, especially the average daily ranges. For the individual calendar months, the latter all have correlations of at least +.953, the corresponding figure +.842 for the standard deviations. This suggests that the easily calculated average daily range statistic itself could serve as a "proxy" variable for the first component covariance loading statistic for periods in which there was no significant climatological trend or cyclical variation. With the Civic Center data, this seems to be the case for all the periods except “YEAR” and “FALL”. The high correlation with the standard deviation statistic exhibited by “YEAR”, +.952, indicates that it too could serve as an analogous “proxy” variable for the covariance loading. So also for “WNTR”, the corresponding figure +.989, although its’ correlation with average daily range is a more respectable +.873 (that for “YEAR” is +.630).

Finally, column 4 lists the correlations of linear models that regress both average daily range and standard deviation on the covariance loading statistic. In nearly all cases, the resulting correlation is only marginally better than that for the average daily range, the only major exceptions again being “YEAR” and “FALL”, where seasonal variation is significantly more important than diurnal. This mix, of course, would undoubtedly be different for other stations, particularly more continental and/or poleward ones.

***3.2.1** – Some Individual Period results*

As was done with the correlation loadings application, the following shows some sample graphs for those years that produced the most extreme covariance matrix loadings statistics. Included are those for the complete calendar year (“YEAR”), and a calendar month (January).

Figures 15 and 16 are floating-bar graphs of the daily maximum and minimum temperatures for 1949 and 1928, respectively, years with the most extreme covariance PCA loading statistics (12.567 and 9.662, respectively).

The year 1949 exhibited both a pronounced seasonality (primarily attributable to January, which produced the coldest recorded calendar month in 1921-2002 Civic Center history), and high average daily ranges. Standard deviation statistic for the daily maxima and minima (14.16 F), is the highest for any year, and the average daily range (19.9 F), is the third highest.

In contrast, the year 1928’s extreme minimum figure is due almost exclusively to its low seasonal variation, apparent from the chart, and more in line with the relative importance of seasonal influences on covariance loadings magnitudes for “YEAR”. The year 1928’s standard deviation figure (11.05) is third lowest, but the average daily range (18.2 F) is very average, ranking as 39th lowest.

Table 2 - Correlation Statistics for First Component Covariance Loadings vs. Selected other Variables, by Period

PERIOD	FIRST COMP COV LOAD VS. FIRST COMP CORR LOAD (1)	FIRST COMP COV LOAD VS. AVG DAILY RNGE (2)	FIRST COMP COV LOAD VS. STDV OF DAILY MAX/ MIN’S (3)	FIRST COMP COV LOAD VS STDV & AVG. DAILY RNGE (4)
YEAR	+.387	+.630	+.952	+.955
WNTR	+.456	+.873	+.989	+.989
SPR	+.352	+.920	+.898	+.947
SUM	+.314	+.889	+.899	+.936
FALL	+.406	+.603	+.866	+.876
JAN	+.479	+.996	+.873	+.996
FEB	+.337	+.993	+.884	+.994
MAR	+.261	+.985	+.906	+.987
APR	0	+.986	+.843	+.986
MAY	+.009	+.996	+.917	+.996
JUN	-.096	+.964	+.947	+.976
JUL	+.148	+.987	+.954	+.989
AUG	+.210	+.998	+.941	+.998
SEP	+.219	+.986	+.842	+.987
OCT	+.108	+.953	+.858	+.959
NOV	+.251	+.953	+.922	+.970
DEC	+.508	+.978	+.907	+.981

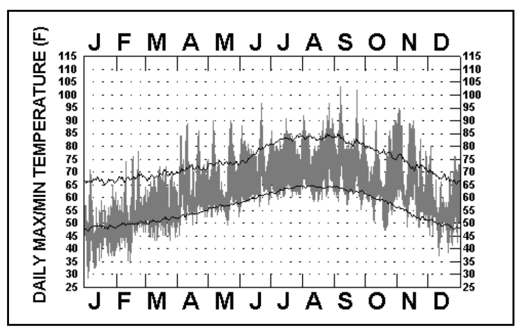

Figure 15. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1949 (Maximum First Component Covariance Loading Statistic)

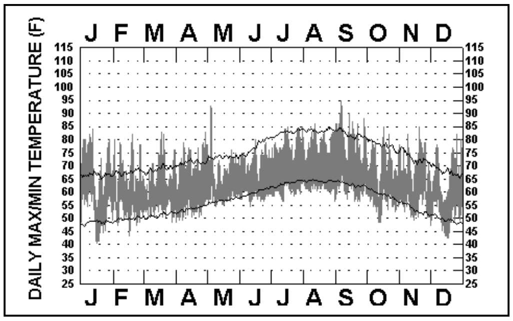

Figure 16. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for 1928 (Minimum First Component Covariance Loading Statistic)

At the calendar month level, Figure 17 and 18 show the floating bar patterns for January 1976 (covariance loading maximum: 13.394) and January 1995 (covariance loading minimum: 6.489), respectively; the former statistic is also 3.3 standard deviations above the 82-year average.

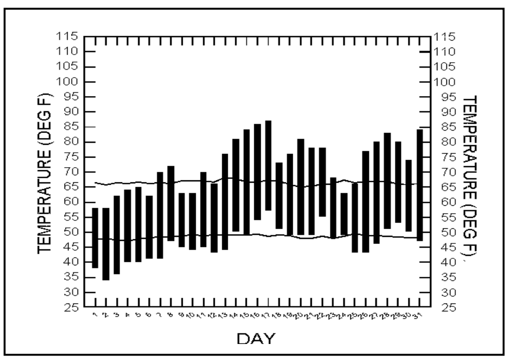

Figure 17. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for January 1976 (Maximum January Covariance Loading Statistic)

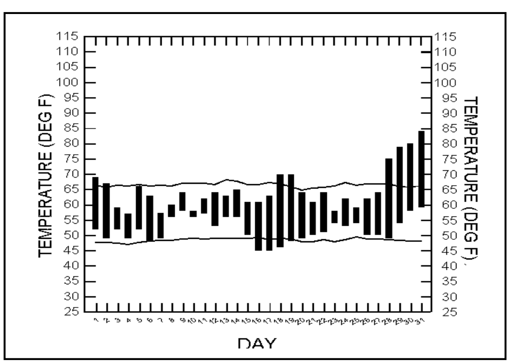

Figure 18. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for January 1995 (Minimum January Covariance Loading Statistic)

Consistent with the relationship shown in Table 2, January 1976 also has the highest average daily range statistic of any January in the record (26.2 F), and January 1995 the lowest (13.0 F). January 1976 also has the highest standard deviation figure (15.13 F), and January 1995 the second lowest (8.53 F). The former had no measurable rainfall during the month, the latter nearly 13 inches.

3.3 – Joint First Component Correlation ("shape") and Covariance Loading ("spread") Information

With two sets of statistics now available to characterize the floating-bar patterns, the logical next step was to combine them into a two-dimensional measure of pattern anomaly. Two sample bivariate analyses are provided below, one for the complete calendar year and the other for January.

Figure 19 is a scatterplot of the first component covariance vs. correlation loadings statistics for "YEAR", fitted to a 99% bivariate normal confidence ellipsoid. Each of the covariance and correlation loadings arrays in this instance are univariately normal. Points for those years that exhibited the correlation and covariance loadings extremes, the day-to-day temperature patterns of which have already been plotted above, are identified.

While the difference is slight, 1971's position is just outside the 99% confidence interval contour, 1949's just within. Thus, 1971 is identified as having the most extreme pattern from the combined correlation and covariance loadings statistics

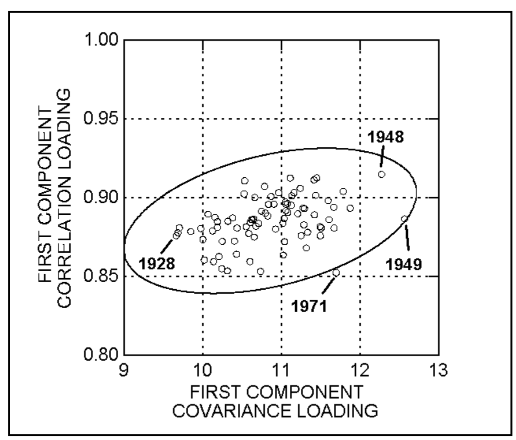

Figure 19. Scatterplot and 99% Confidence Ellipsoid of First Component Covariance vs. Correlation Loadings Statistics for Los Angeles Civic Center Complete Calendar Year Daily Maximum/Minimum Temperatures (1921-2002)

The individual calendar month (January) application was more complicated. The original correlation loadings were not normally distributed and there was more variance. A nominal 99% confidence ellipsoid would have a portion of its boundary extend beyond the 1.00 correlation loadings level, an impossible value statistically. A data transformation was thus in order, and application of the logit (or "folded-log") transformation (log(1/1-x)), where x is the original variable magnitude, rendered the correlation loading data univariately normal (the covariance loading data were already so), and permitted a theoretically reasonable confidence ellipsoid graph to be plotted.

Figure 20 is a scatterplot of the January first component covariance loadings vs. corresponding logit transformed first component correlation loadings, fitted to a 99.5% bivariate normal confidence ellipsoid. The years 1976 and 1995, which exhibited the covariance loading maximum and minimum, respectively, are identified, along with 1971, which had an extreme pattern of its own (see Figure 21).

While January 1971's pattern is clearly more irregular than 1976's, Figure 20's bivariate plot indicates the latter year as positioned just outside the contour and the former just inside. Thus, on this abstract basis, 1976 is identified as having the most extreme January pattern.

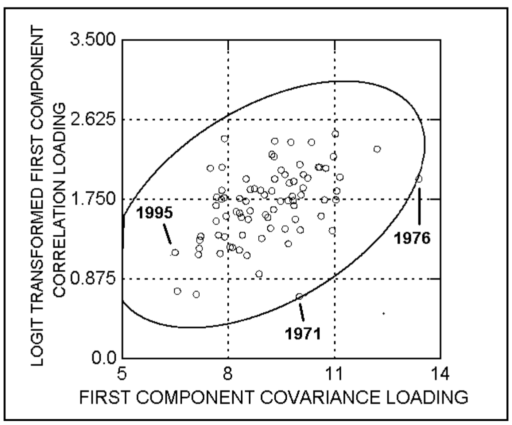

Figure 20. Scatterplot and 99.5% Confidence Ellipsoid of First Component Covariance vs. Logit Transformed Correlation Loadings Statistics for Los Angeles Civic Center January Daily Maximum/Minimum Temperatures (1921-2002)

4. SUMMARY

Using information provided by Principal Components Analysis correlation and covariance matrix loadings statistics, the purpose of this study was to identify 82-year Los Angeles Civic Center extreme-most daily maximum/minimum temperature time series patterns for a hierarchy of time periods (yearly, seasonal, and monthly), and in the process evaluate the technique.

Results had the first component, closely related to reference climatology, explaining the great portion

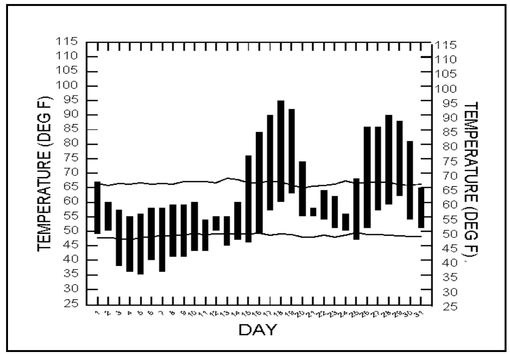

Figure 21. Los Angeles Civic Center Daily Maximum and Minimum Temperatures for January 1971

(between 69% and 88%) of the variance for both the correlation and covariance matrix options, the exact preponderance varying by period. This effective reduction of matrix data to a single, physically meaningful dimension (climatology) simplified the identification and interpretation process significantly for the Civic Center data, and it seems probable that this outcome would be repeated for other stations, although the percents of variance explained might be lower for more interior-type locations in some instances. For example, applying this technique for two calendar segments with Minneapolis-St. Paul, MN data for 1921-2002 resulted in the correlation matrix first component explaining 84.4% of the variance for "YEAR" (higher than the Los Angeles Civic Center's figure: 78.4%), but 35.4% for January (compared to the Civic Center's 70.4%).

Correlation loading information essentially measured a given floating bar temperature pattern's conformance to the day-to-day climatological profile or "shape". Covariance loading information reflected diurnal and seasonal "spread". For the calendar periods under consideration, these two "dimensions" were either orthogonal (April, and effectively May), or quasi-independent with weak to modest correlations (all the others), the degree of the latter depending on the period.

In addition to pattern anomaly evaluation, the analysis identified variables that that could serve as substitute or "proxy" statistics, highly correlated with the loadings, and thus useful if the capability to do a PCA was not available. For example, the "shape" measure (first component correlation loading statistic) could be virtually duplicated by calculating the correlation coefficient between the climatological max/min's and the observed max/min's for the period in question. The "spread" measure (first component covariance loading statistic) could be replicated approximately by average daily range statistics for periods with no appreciable climatological trend. Alternatively, for periods in which there was a significant trend or a seasonal cycle (e.g., some meteorological seasons or the calendar year), standard deviations of the maxima and minima as one set of data could serve as a substitute.

A principal component analysis like the foregoing, of course, was not the only way to evaluate floating-bar temperature patterns. In addition to non-linear techniques, another linear-based option would be Canonical Correlation Analysis, which allows multiple x and y variables simultaneously. Daily maximum and minimum temperature data would readily conform to this kind of treatment, although it is uncertain if the mode(s) would have the same relatively straightforward interpretation that a linear PCA provides.

In this study the correlation loading minima (capturing highly irregular, nonlinear profiles or “shapes”), and to a lesser extent the covariance loading minima (capturing unusually low seasonality or diurnal variability), seemed to identify more visually intuitive extreme patterns than those represented by the corresponding maxima (especially the correlation maxima), the latter or “anomalously linear” patterns requiring a more abstract interpretation.

Also, the departure from normal, or “offset” dimension’s effect on pattern was not evaluated directly, the “shape” property, among other variables, being completely transparent to this. For example, on a complete calendar year basis, “YEAR” departure statistics from mean 1921-2002 temperature had +.00 and +.09 correlations, respectively, with the first component correlation and covariance loadings statistics, indicative that both were more or less independent to “offset”.

Finally, statistical “objectivity” is not a be-all and end-all. A number of Civic Center charts for “YEAR”, while not loading extremely high or low on first component correlation and covariance scales, had patterns that were subjectively “extreme” in appearance and interesting in their own right.

5. REFERENCES

Fisk, C.J., 1995: “An Objective Technique for Identifying Extreme-most Pattern Anomalies in Daily Maximum/Minimum Temperature Time Series”, *Proceedings of the 9th Conference on Applied Climatology, American Meteorological Society, Dallas.*

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

Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division. 2000: *Existing Climate Data Sources and Their Use in Heat Island Research – Historical Analysis of the Urban Climate of Los Angeles* http://eetd.lbl.gov/EA/Reports/41973/41973-2c-part1.html

Mosteler, F., and Tukey, J. W., 1977: *Data Analysis and Regression - A second course in Statistics :* Addison Wesley Publishing Company, Inc., 588 pp.

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.