P2.5 PRINCIPAL COMPONENT ANALYSIS OF MONTH-TO-MONTH PRECIPITATION VARIABILITY FOR DOWNTOWN LOS ANGELES (1877-8 THROUGH 2003-4 SEASONS)

Charles J. Fisk *
NAWCWPNS-Point Mugu, CA

1. INTRODUCTION

The Downtown Los Angeles monthly precipitation history, commencing with July 1877, is one of the lengthiest records of its kind for a single locality in the western United States. As of late 2004 there were 127 full seasons of individual monthly and aggregate rainfall statistics available to the researcher (in this context, given Los Angeles' propensity for a winter rainfall maximum and summer drought, an individual "full" season or "year" is a July-June series rather than the more traditional January-December one).

For a single station with such a rainfall regime, an obvious first step in assessing a given year's precipitation character is 1) comparing the July-June aggregate total with the long-term station average and/or with totals from other years. Another step might be 2) examination of the year's month-to-month anomaly patterns, especially for those months that encompass the height of the rain season, climatologically. Comparing these patterns on a year-to-year basis, it may appear that certain contiguous anomaly sequences or "modes" seem to recur in subtle but recognizable ways. Objective statistical confirmation of these would be a likely useful result.

In a previous study, Fisk (2002), using Principal Components Analysis, investigated for the seven NCDC California climatic divisions, the existence of month-to-month precipitation variability modes. The original data were areal monthly averages, by year, for the period 1895-2001, and the analysis made use of 1) a merging of monthly categories that encompassed the drier months of the year, 2) statistical transformation to normality, and 3) standardization.

With some methodological changes, the analyses are repeated here for Downtown Los Angeles. One modification is reduction of the monthly selection from July-June to October-June. The rationale behind this is that, generally speaking, precipitation over the excluded July-September months is very light, and of different origin (e.g., low stratus drizzle, monsoonal influxes and tropical cyclone remnants). The Los Angeles record is also about 20 percent longer than the Climate Divisions', due mostly to its' earlier start date (1877-78 season). While there have been some local station moves over the long history, it is believed they would not compromise results of a study of this kind.

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

2. METHODS AND PROCEDURES

Much of the theory that supports the use of Principal Component Analysis assumes multivariate normal data. Monthly precipitation data in raw form are notoriously non-normal, frequently exhibiting positive skewness and/or "trace" or zero observations on the bounded left side. To allow the use of measurable and non-measurable data alike and to bring them into approximate univariate-normal conformance, the Box-Cox power transformation procedure (Johnson and Wichern, 1982) was applied to the individual monthly series, preceded by addition of an .01 constant or "start" (Mosteller and Tukey, 1977) to those that had least one zero or 'trace' observation (this was the case for all the months considered). Results were assessed by inspection of normal probability plots. Power transformations on combined October-November and April-June categories had to be adapted because individual-month conformance to normality was poor even after Box-Cox transformation, attributable to a large number of very dry or rainless months. The transformed series were also standardized. Thus, the monthly series were analyzed in terms of power transformed relative anomalies (z-scores) in place of arithmetic anomalies.

Interpreted, a given transformed distribution's z-score represented a relative deviation from a level slightly less (usually) than its' parent distribution's median. To illustrate, for the six periods considered (October-November, December, January, February, March, and April-June), parent-series precipitation magnitudes corresponding to zero z-scores in their transformed distributions ranged from equivalency to 9% lower than their actual medians, but were always at least 28% lower than their means. Overall average of the six periods' zero z-score "magnitudes" (1.63") was about five percent less than the average of their corresponding medians (1.72").

With the precipitation data transformed and standardized, the PCA was next performed. Eigenvalue magnitudes of greater than or equal to 1.0 were used as a cutoff criterion, resulting in retention of three components. To sharpen eigenvector coefficient contrasts in these three, a varimax rotation was then performed. As a tradeoff for this eigenvector coefficient contrast enhancement, eigenvalue magnitudes contrasts became slightly less.

3. RESULTS

3.1 – Downtown Los Angeles

Table 1 displays the summary PCA statistics for the Downtown Los Angeles station. The primary mode ("#1" column) contrasts comparatively high eigenvector coefficients for January, February, and March (+.399, +.567, and +.531, respectively) with negligible values for October-November, December, and April-June (-0.190, +0.017, and +0.018, respectively). Rotated loadings for January, February, and March (+.501, +.752, and +.701, respectively) are also much higher than the rule-of-thumb absolute value 0.300 statistic for variable retention, so interpreted in this light, mode one can be considered as a precipitation pattern that emphasizes January, February, and March relative anomaly character. The secondary mode ("#2" column), only about 7% less in eigenvalue magnitude, contrasts large eigenvector coefficients for October-November and December (+.618 and +.598, respectively); and a somewhat lesser figure for January (+.283), with undistinguished magnitudes for the remaining three. The October-November through January periods' loading magnitudes each exceed +.300, so again by the rule-of-thumb criterion, mode two can be considered one that emphasizes October-November, December, and January relative anomaly character. Mode three contrasts an overwhelmingly high coefficient for April-June (+.922) with insignificant values for the other periods. Alto-gether the three components explain a somewhat modest 59.5% of the variance, compared to the expected 50%.

The existence of primary and secondary modes that each cover contiguous months seems to at least suggest that Los Angeles' seasonal rainfall character may have a subtle tendency to adhere to one of two alternative regimes ("early" vs. "later"?). For the balance of the paper, the focus will be on these two.

3.2 – California Division #6 ("South Coast Drainage")

Comparing the Downtown Los Angeles results with that of Calfornia Division #6 [Fisk, 2002], within which Los Angeles is located (see Figure 1 on the next page for a map of the California Climate Divisions), shows very similar results, not altogether unexpected. Inspecting the most prominent loadings and eigenvector coefficient magnitudes, it is clear that the primary, secondary, and tertiary modes are the "same" ones in each Table. About the only noticeable contrast is that for mode # 2, which finds the January rotated loadings' magnitude for the Climate Division #6 (+0.166) appreciably lower than that for Los Angeles (+0.322) and well below the arbitrary 0.300 retention threshold.

Table 1 - Statistics for First Three Components of Transformed Precipitation Anomalies - Downtown Los Angeles (1877-78 to 2003-04 data)

ROTATED LOADINGS			
PERIOD	#1	#2	#3
OCNV	-0.075	**+0.751**	-0.189
DEC	-0.026	**+0.729**	0.169
JAN	**+0.501**	**+0.322**	0.192
FEB	**+0.752**	-0.152	0.039
MAR	**+0.701**	-0.077	-0.166
APJU	+0.026	+0.005	**+0.900**
EIGENVALUES			
MODE	#1	#2	#3
VALUE	1.314	1.228	1.030
CUMULATIVE PCT VARIANCE			
MODE	#1	#2	#3
VALUE	21.91	42.35	59.54
EIGENVECTOR COEFFICIENTS			
PERIOD	#1	#2	#3
OCNV	-0.190	**+0.618**	+0.211
DEC	+0.017	**+0.598**	-0.139
JAN	**+0.399**	**+0.283**	-0.176
FEB	**+0.567**	-0.088	-0.045
MAR	**+0.531**	-0.022	+0.157
APJU	+0.018	0.039	**+0.922**

Table 2 - Statistics for First Three Components of Transformed Precipitation Anomalies - California Division #6 –-"South Coast Drainage" (1895-96 to 2000-2001 data) – [after Fisk, 2002]

ROTATED LOADINGS			
PERIOD	#1	#2	#3
JLNV	-0.259	**+0.654**	+0.012
DEC	+0.171	**+0.801**	+0.022
JAN	**+0.587**	+0.166	-0.031
FEB	**+0.658**	-0.141	-0.006
MAR	**+0.701**	-0.107	+0.086
APJU	+0.029	+0.031	**+0.996**
EIGENVALUES			
MODE	#1	#2	#3
VALUE	1.365	1.128	1.001
CUMULATIVE PCT VARIANCE			
MODE	#1	#2	#3
VALUE	22.754	41.660	58.346
EIGENVECTOR COEFFICIENTS			
PERIOD	#1	#2	#3
JLNV	-0.147	**+0.566**	-0.002
DEC	+0.181	**+0.727**	-0.021
JAN	**+0.447**	+0.190	-0.069
FEB	**+0.477**	-0.080	-0.035
MAR	**+0.507**	-0.051	+0.054
APJU	-0.029	-0.013	**+0.997**

Figure 1. NCDC California Climate Divisions

3.3 First Component Scores' Time Series

With the principal eigenvector functions identified for Downtown Los Angeles, time series plots of the primary and secondary mode standardized scores were next constructed. From online sources, The NCEP Climate Prediction Center's list of "Previous ENSO Events (1877-present)" along with Quinn and Neal's list of "El Nino Occurrences Over the Past Four and a Half Centuries" (warm events only) were also utilized to relate the scores in a general way to warm, cold, and neutral episodes. Figures 2 and 3 are time series plots of the first and second component scores, respectively, for Downtown Los Angeles.

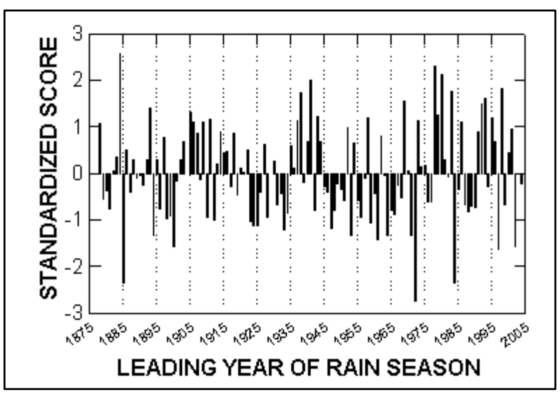

Figure 2. Time Series Plot of Downtown Los Angeles First Component Scores for 1877-78 to 2003-04 seasons

Primary mode for Los Angeles, as was indicated by Table 1, consisted of relatively significant positive coefficients for January (+.399), February (+.567), and March (+.531). Positive (negative) scores would thus be particularly enhanced by power transformed relative anomalies for those months that were also strongly positive (negative).

Inspecting Figure 2, component scores at or above an arbitrary +1.00 level seem to be clustered over four multi-season periods, covering just a small portion of the total record. For example, four are noted over the seven-season period 1904-05 to 1910-11, inclusive; four more over the seven-season interval 1936-37 to 1942-43, inclusive; five additional ones encompassing the nine-season interval 1977-78 to 1985-86, inclusive; and four additional cases over the seven-season period 1991-92 to 1997-98, inclusive. Thus, 17 of the 23 +1.00 or greater scores over the 127-year history (74%) are concentrated over a span of just 30 seasons (24% of the total record). Another noticeable feature, from the early 1970's on, is the presence of considerably more inter-season variability. For example, the two most negative scores (-2.75 for 1971-72 and -2.36 for 1983-84, respectively), and the second and third most positive (+2.31 for 1977-78 and +2.11 for 1979-80, respectively), occur over just thirteen seasons.

The 1883-84 absolute maximum score (+2.57) reflects to a large extent the excessively heavy falls for February (z=+1.95) and March (z=+2.39) matching up well with the significantly positive first component eigenvector coefficients for those two months (January's z-score was a more modestly positive +0.37). This season was classified as a "strong-plus" ENSO by Quinn and Neal, but is excluded from the NCEP list (the latter also omits the 2002-3 El Nino). The 1971-72 extreme negative statistic (-2.75), conversely, is markedly influenced by the heavy falls of December (z=1.40), followed by marked deficits for January (z=-2.19), February (z=-1.42), and March (z=-2.21). In absolute hydrological terms, January thru March's meager 0.13" total is by far the least ever recorded for that interval in Downtown Los Angeles history.

Referencing the Climate Prediction Center list (but including 1883-84 as an El Nino), fourteen of the scores greater than or equal to +1.00 were associated with El Nino events (including eight "Strong" events), nine with neutrals, and none with La Nina's. Also, nine of the ten most positive scores (z=+1.48 or higher) were associated with El Nino's. Of the eighteen scores less than or equal to –1.00, seven were related to El Nino's (including the strong event of 1911-12), eight to neutrals, and three to La Nina's (all the La Nina's weak). The 1971-72 season was a neutral, 1977-78 and 1979-80 each weak El Nino's, and 1983-84 a weak La Nina.

It should be mentioned that of the 127 total seasons analyzed, 29 % were classified as "El Nino", 50 % "Neutral", and 21% "La Nina".

3.4 Second Components Scores' Time Series

Figure 3 is a time-series plot of the Second Component's scores. From Table 1's information, this mode consisted of relatively significant positive coefficients for the contiguous periods' October-November, December, and January. In analogous fashion to that of the first component, positive (negative) scores would be

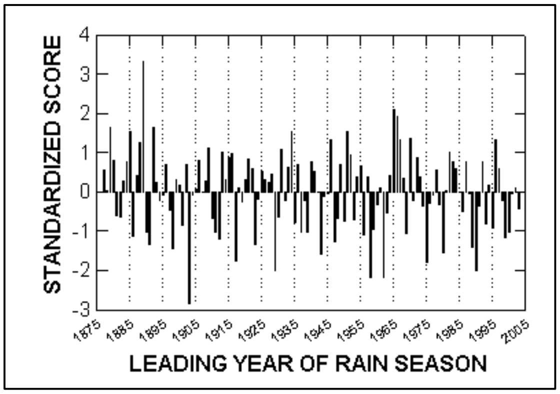

Figure 3. Time Series Plot of Downtown Los Angeles Second Component Scores for 1877-78 to 2003-04 seasons

enhanced by relative anomalies for these periods that were likewise consistently positive (or negative). Also, by definition, the 127 Second Component scores were uncorrelated with those of the First.

Inspecting Figure 3, second component scores at or above +1.00 seem to be more randomly distributed over the record than those in Figure 2, although four occurrences in six seasons are noted between 1965-66 and 1970-71. Most positive figure (the extraordinarily high +3.33 statistic for 1889-90 – a strong La Nina) reflects the heavy falls of October-November (z=+2.23), December (z=+2.63), and January (+1.40). In absolute hydrological terms, some 31.94 inches of rain fell over October-January 1889-90, including 6.96" in October (a record maximum for that calendar month) and 15.80" in December (a record maximum for any calendar month). The second and third most positive scores (+2.10 and +1.93) are noted for the successive seasons (1965-66 and 1966-67, respectively). The former, a moderate El Nino, had very heavy falls in both November and December (z=+2.48 and z=+2.36 values for the two months, respectively) but a light amount for January (z=-.37). The latter 1966-67 season, a "Neutral", had consistently heavy falls for all three periods (z=+1.26 for October-November, z=+1.14 for December, and z=+1.05 for January, respectively). The very next season, 1967-68, also had a markedly positive score +1.34) as did 1970-71(+1.35), just three years hence; these two seasons were classified as "Neutral" and "moderate La Nina", respectively, by the Climate Prediction Center. Altogether in the 127 year series, six scores greater than or equal to +1.00 were associated with El Nino events, nine with neutrals, and two with La Nina's.

Most negative figure (-2.82 for 1903-04 – also a strong La Nina) reflects to a considerable extent the non-measurable falls for October-November (z=-2.08) and December (z=-2.37) along with the light 0.14" fall for January (z=-1.52). Of the twenty-four second component scores less than or equal to –1.00, six were associated with El Ninos, twelve with neutrals, and six with La Nina's.

In summary, except for some occasional periods of like-signed clustering and increased variability, there were no obvious long-period time-series non-uniformities for the either component. For mode one (essentially January-February-March), the most noteworthy feature was the association of high positive scores (greater than equal +1.0) with El Nino events, and to a lesser extent with "Neutrals" (La Nina's were completely unrepresented at these high positive levels). At even more positive levels (scores greater than equal to +1.5), the seasons in question were almost universally El Nino's.

For mode two (essentially October/November-December-January), although the individual most positive and negative scores were each associated with strong La Nina events, on a more general basis, there were no clear-cut disproportionate adherences to El Nino's, Neutrals, or La Nina's at the high absolute score levels (either greater than or equal to +1.0, or less than or equal to –1.0).

3.5 Two-Dimensional Treatment

Finally, on an all-inclusive, two dimensional-basis, Figure 4 is a scatterplot of the first component and second component scores for each of the 127 years. The points are designated as El Nino's (red "W"), Neutrals (black "N"), and La Nina's (blue "C").

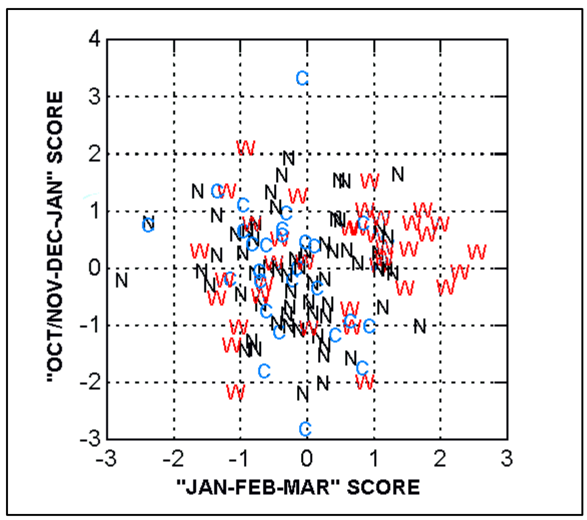

Figure 4. Scatterplot of First Component ("Jan-Feb-Mar") and Second Component ("Oct/Nov-Dec-Jan") scores for Downtown Los Angeles, with ENSO designations.

Clearly evident is the cluster of El Nino designated seasons at the most positive score levels on the "Jan-Feb-Mar" or first component axis. A number of these points are positive on the "Oct/Nov-Dec-Jan" or second-component axis as well, although the scores are more modestly positive. Neutrals seem to be uniformly spread throughout the four quadrants. Only a few La Nina's have positive scores on the "Jan-Feb-Mar" axis but they are well-distributed in both positive and negative space relative to the "Oct/Nov-Dec-Jan" axis. Further emphasizing this El Nino/La Nina contrast, average "Jan-Feb-Mar" score for the 37 El Nino's was +.43; that for the 27 La Nina's was -.31. A t-test for the difference was significant at the .01 level (t=2.807; d.f.= 62). Interpreted, La Nina's tend to be of significantly opposite character (drier) for this mode than El Nino's. Also, the fairly appreciable negative magnitude (-.31) suggests that they are dry in an absolute hydrological sense as well. Average "Oct/Nov-Dec-Jan" score for El Nino's was +.13, that for La Nina's -.02; the difference was not statistically significant. Corresponding mean values for Neutrals were -.12 for the "Jan-Feb-Mar" mode and –.07 for "Oct/Nov-Dec-Jan".

4. SUMMARY AND CONCLUSIONS

Using normally transformed, standardized precipitation data, Principal Components Analysis (PCA), and varimax rotation, the existence and character of month-to-month modes in October-June precipitation variability were investigated for Downtown Los Angeles monthly rainfall history for the 1877-78 through 2003-2004 seasons. Results resolved three components that explained the variance of a full-fledged original variable and together they explained a modest 59% of the variance (expected: 50 %).

Mode one described a statistically idealized rainfall regime with significant eigenvector coefficients (anomalies) of the same sign for January, February, and March. Mode 2, covering a generally earlier sequence of contiguous months, did the same for the combined October/November period, December, and January. Mode three strongly emphasized anomaly character for April. While the modes' individual eigenvalue magnitudes were not especially pronounced, a few useful conclusions are possible when the results are stratified according to ENSO designations. First, consistent heavy rains (that is, falls above the median) during the January-February-March period can occur during both Neutral and El Nino years, but the exceptionally heavy falls are historically associated with El Nino's. Conversely, during La Nina episodes, consistently above median rainfalls during January, February, and March are unlikely, in fact, the months might be quite dry.

5. ACKNOWLEDGEMENTS

The California Divisional Data and Figure 1's map were downloaded by accessing the NOAA National Climatic Data Center link:
http://www.ncdc.noaa.gov/onlineprod/drought/main.html

The NCEP Climate Prediction Center page with the list of ENSO events is accessed from the link:
http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears_1877-present.html

The Quinn and Neal data were accessed from the following site:
http://www.wxresearch.com/outlook/eos98.htm

The Historical Downtown Los Angeles Monthly Precipitation Series was accessed from the Oxnard/Los Angeles NWS site:
http://www.nwsla.noaa.gov/climate/cvc_rainfall.html

6. REFERENCES

Fisk, C.J., 2002; "Principal Component Analysis of Month-to-Month Precipitation Variability for NCDC California Climatic Divisions (1895-6 through 2000-1 Seasons)" . *Proceedings of the 16th Conference on Probability and Statistics in the Atmospheric Sciences, American Meteorological Society, Orlando*.

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

Manly, B., 1994; *Multivariate Statistical Methods - A Primer, 2nd edition,* Chapman & Hall, 215 pp.

Mosteller, F., and Tukey, J.W., 1977; *Data Analysis and Regression – A Second Course in Statistics,* Addison-Wesley Publishing Company, 588 pp.

Wilkinson, Leland, 1990: *SYSTAT: The System for Statistics,* SYSTAT, Inc., 677 pp.

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