76. CLUSTER ANALYSIS OF PREFERRED MONTH-TO-MONTH PRECIPITATION ANOMALY PATTERNS FOR LOS ANGLES/SAN DIEGO AND SAN FRANCISCO WITH BAYESIAN ANALYSES OF THEIR OCCURRENCE PROBABILITIES REALTIVE TO EL NINO, NEUTRAL, OR LA NINA EPISODES

Charles J. Fisk *
NAVAIR-Point Mugu, CA

1. INTRODUCTION

Long-term monthly averages are a traditional means of characterizing climatological precipitation variability for given weather stations over the course of a rain year. Frequently based on the 30-year period of record, they serve as monthly precipitation "normals" which are a basis for departure or "anomaly" calculations.

Such normals, of course, are only statistical idealizations, and actual individual years' month-to-month rainfall patterns invariably depart from mean climatology in some fashion, not necessarily randomly. Inherent tendencies may exist, for example, for occasional clustering of wet or dry anomalies over multi-month sequences, or alternatively, gradual progressions from wet (dry) to dry (wet) regimes, reflecting mean trough (ridge) to ridge (trough) propagations. Perhaps also there may be distinctive ENSO phase (El Nino, Neutral, or La Nina) influences on the nature and frequency of the patterns. Information on such tendencies, to whatever extent they are real, would represent a useful complement to the more conventional climatological characterizations.

To explore these possibilities, the following study investigates the existence and relative frequencies of month-to-month precipitation anomaly patterns for three California localities with lengthy periods of record: the downtown stations of Los Angeles, San Diego, and San Francisco, CA. The K-means clustering analysis methodology integrated with the V-Fold Cross Validation Algorithm is applied. The latter, an automated software-based training and testing sample type data mining procedure, tends to optimize the number of K clusters identified, subject to user selected preliminary settings in such elements as distance metric type, percent improvement cutoff threshold, seed, number of folds, and others.

In this study, the Squared Euclidean distance metric is utilized, the other settings left to their default values. The final selection of K clusters is based on Scree-plot inspections of the Squared Euclidean results, focusing on inflection points. The nature and frequencies of the patterns are described, and then, utilizing Bayesian analyses, the conditional occurrence probabilities for each of the patterns by station, relative to El Nino, Neutral, and La Nina episodes are calculated.

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

Periods of record examined for all three stations are the1877-78 thru 2013-14 (July-June) rain seasons. Given the winter rainfall maximum and summer drought character of California coastal stations, the calendar period selection includes October-November, December, January, February, March, and April-May. Also, given the close proximity of Los Angeles and San Diego (just 120 miles apart) and their similar rainfall climatologies, data for the two stations are merged into a single data base. Thus, the Los Angeles/San Diego cluster analysis is 12-dimensional, the San Francisco one, six.

2. THE K-MEANS AND V-FOLD CROSS VALIDATION METHODOLOGIES

The original K-means methodology was introduced by Hartigan (1975), and the basic methodology consists of assigning observations to a designated number of K clusters such that the multivariate means across the clusters are as different as possible. The differences can be measured in terms of Euclidean, Squared Euclidean, City-Block, and Chebychev statistical distances (Nisbet, et. al., 2009).

The V-fold cross-validation scheme, as applied to K-means clustering involves dividing the overall data sample into V "folds", or randomly selected subsamples. K-means analyses are then successively applied to the observations belonging to the V-1 folds (training sample), and the results of the analyses are applied to sample V that was not used in estimating the parameters (the testing sample) to assess the predictive validity or the average distances of the training sample arrays from their cluster center centroids. The procedure is repeated for cluster sizes K+1. K+2, …, etc., until the incremental improvement in the average distances is less than some threshold, at which time the "optimal" cluster size is considered attained (Nisbet, et. al., 2009).

The STATISTICA Data Miner Clustering module was utilized to employ this technique. Preliminary to the analyses, the period-to-period precipitation data were normalized, an internal automatic software feature, to reduce them to a common scale (between 0.0 and 1.0) and lessen the influence of outliers. The various cluster results' reports would, however, be presented in the units of the pre-normalized data.

Since as already stated, the distance threshold and other settings can be changed, generation of the "optimal" number of clusters is not completely automatic. Nonetheless, the V-fold cross-validation algorithm enhances the methodological objectivity of a clustering technique like K-means. And, the graphical (scree-plot) option is available as an additional decision aid for deciding on the “ideal” number.

3. BAYESIAN ANALYSIS

From *Wikipedia*, Bayesian inference is a method of which Bayes’ rule is used to update the probability estimate for a hypothesis as additional evidence is acquired. In the context of this study, the initial hypothesis would be a probabilistic belief, or “Prior Probability”, that a given anomaly pattern (cluster) would occur unconditionally (historical percent frequency of the pattern), updated by a processing of evidence that relates the occurrence of the pattern to ENSO phase. The latter could be referred to as “accounting for evidence” and the result, or “impact”, multiplied by the “Prior Probability” would produce a “Posterior Probability” that incorporates this new conditional information (the ENSO phase influence) into a revised probabilistic belief that the given pattern will result. A desirable outcome would be a marked contrast in magnitudes between the Posterior and Prior probabilities which would indicate that knowledge about the conditional variable “matters”. The actual Bayesian expression will appear in a later section in which a case example is demonstrated on the Los Angeles/ San Diego precipitation data.

4. DATA AND PROCEDURES

Data for the three stations were secured from various online sites, including those of the National Climatic Data Center and National Weather Service. The precipitation histories of each station include a number of station moves locally, but for the purposes of this analysis, the moves are assumed to have negligible influence on results’ outcomes.

Also, identification of ENSO episodes is a not completely objective or definitive process, different researchers have composed different lists, and there is likely more uncertainty with years further back than closer to the present. For the purpose of this research, the lists utilized are those formulated by the NOAA Climate Prediction Center. The first covers the years 1877-2001, the second 1950-2013. Those years that overlap (1950-2013) are given the designations assigned by the latter.

5. RESULTS

5.1 - Downtown Los Angeles and Downtown San Diego Combined Results

Figures 1 and 2 are bar graphs depicting mean overall Los Angeles and San Francisco precipitation figures, respectively, for the six calendar periods under consideration.

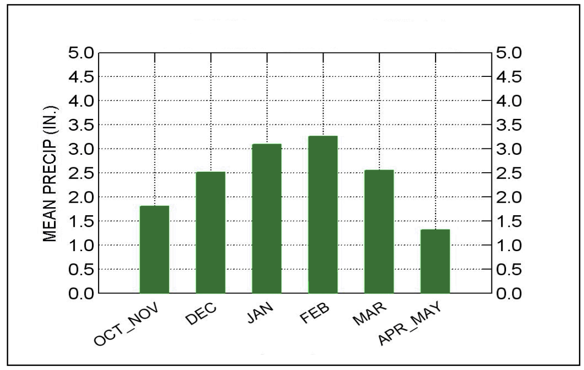

Figure 1. Mean Downtown Los Angeles Precipitation for October-November, December, January, February, March, and April-May calendar periods, 1877-78 to 2013-14 Period of Record.

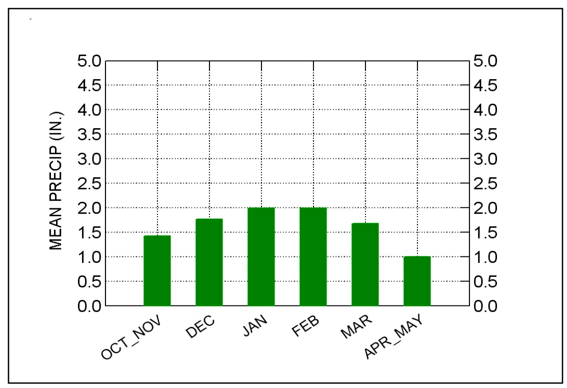

Figure 2. Mean Downtown San Diego Precipitation for October-November, December, January, February, March, and April-May calendar periods, 1877-78 to 2013-14 Period of Record.

__5.1.1__ – Scree Plot

Combined use of the Squared Euclidean Distance Metric and inspection of the associated Scree Plot results brought about "resolution" of four clusters for Los Angeles/San Diego. From Figure 3, the magnitude of the curve at K=5 was scarcely less than that at K=4; hence a suitable stopping or "cutoff" point was at the latter.

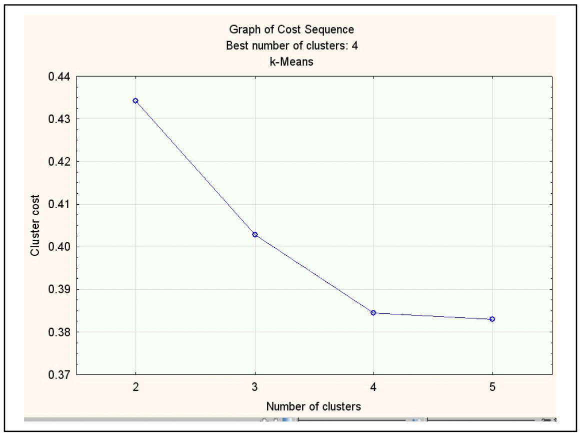

Figure 3 – Scree Plot of K-Means/V-Fold Cross Validation Algorithm Analysis of Los Angeles/San Diego October-November, December, January, February, March, and April-May Precipitation Anomalies.

5.__1.2__ - Downtown Los Angeles and Downtown San Diego Combined Results

Figures 4 thru 7 present the mean anomalies for each of the four patterns, in descending order of importance. Except for Figure 4a, the period-to-period anomalies for a given pattern are the deviations from the overall 137-year mean figures. Annotations above or below the individual bars are the actual cluster means for the period concerned.

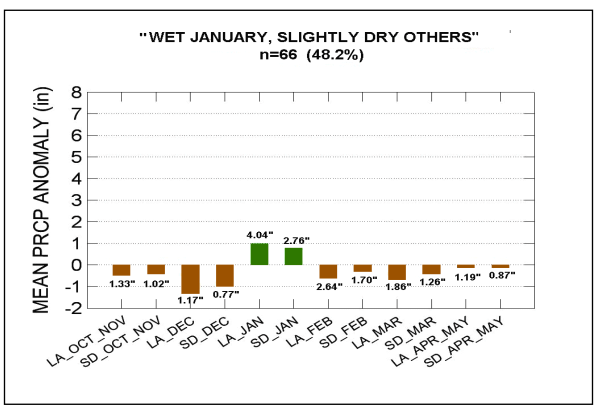

Figure 4 – Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the "Wet January, Slightly Dry Others" Pattern.

The "Wet January, Slightly Dry Others" Pattern (Figure 4) is by far the most frequent of the four, some 66 seasons or 48.2% falling into this cluster. In general, both stations show mean anomalies very much alike, period-to-period, a sensible result for two stations so close in distance, with a known similar rainfall climate. The Los Angeles departures (both positives and negatives) are a bit more pronounced on an individual basis than San Diego's, this to be expected as Los Angeles is wetter climatologically. Both the January mean anomalies for Los Angeles and San Diego are positive (wet), the only departures of that sign; the December negatives (dry) for each Los Angeles and San Diego are the most pronounced of that sign for any period.

Bearing in mind that the pattern in Figure 4 is an idealized one, and the anomaly configurations of the 66 different seasons contained within the group would show individual variations, a few generalizations are possible. First it appears that there may be a natural tendency over a given season for dry to wet successions between December and January, with essentially unremarkable, slightly negative ones for the other periods. It is also true, however, that most individual anomalies relative to long-term means are likely negative because of the frequently positive skews of precipitation distributions, creating "inflated" mean figures (i.e., the means higher than the medians). Since the clustering algorithm worked with the original data in normalized form, with eliminated or reduced skewnesses, conceivably it produced clusters with centroids that were more affiliated with medians rather than means. If this is a reasonable assumption, aside from December and January in this case, it could be generalized that roughly half of the individual seasons associated with this mode for Los Angeles/San Diego were relatively non-descript anomaly-wise, period-to-period.

Figure 4a depicts the period-to-period cluster anomalies relative to the long-term overall *medians*. Supporting the above argument to some extent, the anomalies seem to become less pronounced in absolute magnitude, save for those of January, and April-May. Mean departures for all the12 periods are -0.306 in Figure 4 and +0.302 in Figure 4a, but excluding January and April-May, the comparative figures are -0.649 and -0.044, respectively.

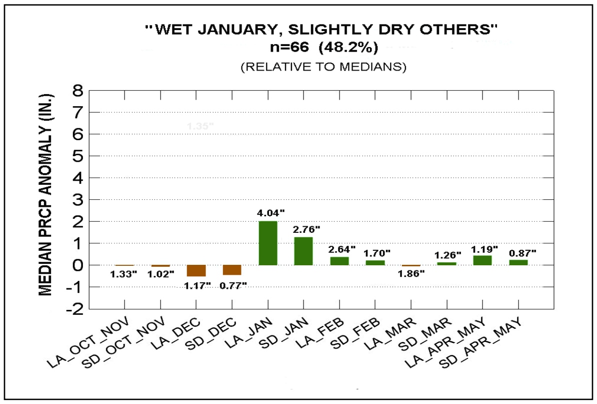

Figure 4a – Mean Period-to-Period Anomalies Relative to Long-Term Overall Medians for the "Wet January, Slightly Dry Others" Pattern.

Ranking second in relative frequency (n=31, 22.6% frequency), was the "Wet December, Dry January & February" pattern (See Figure 5). This displays a clear drop-off from wet Decembers for both Los Angeles and San Diego to noticeably dry Januarys and Februarys for each. Mean anomalies for the remaining periods, October-November, March, and April-May were slightly negative relative to their long-term means.

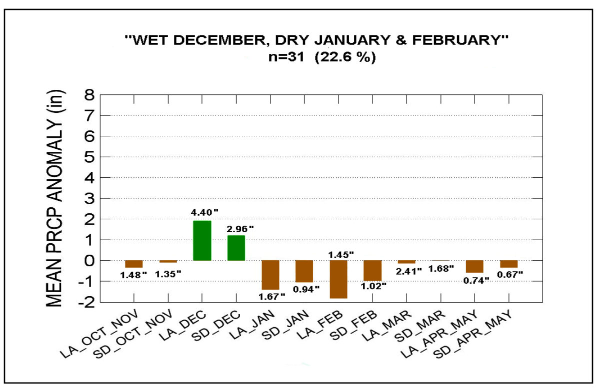

Figure 5 – Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the "Wet December, Dry January & February" Pattern.

Third in importance was the "Very Wet February & March pattern (n=25, 18.2 % frequency – see Figure 6), This displays a striking contrast in anomalies between February and March vs. the other four periods. Except for January, which displays negative departures for both stations, the only other negative anomaly is seen for October-November for Los Angeles. The 8.32" mean Los Angeles figure for February is 255% of average (3.26"), that for San Diego (4.29") 216% of the norm (1.99"). The corresponding mean cluster statistics for March are 4.81" in Los Angeles (189% of average) and 3.10" for San Diego (186% of average).

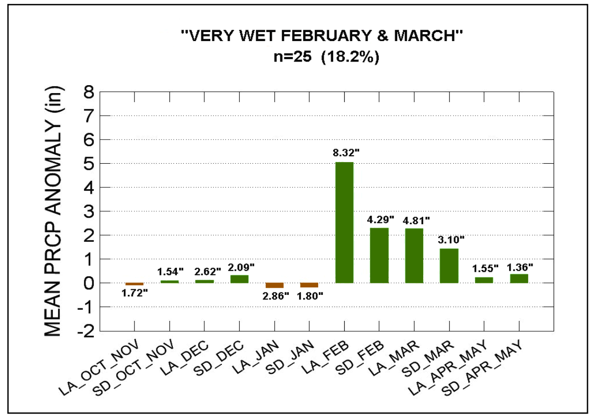

Figure 6 – Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the "Very Wet February & March" Pattern.

Finally, in fourth place was the "Very Wet October-December, Dry January-March pattern (n=15, 11.0% frequency- see Figure 7). This seems to be an expanded version, forward and backward, of the "Wet December, Dry January & February pattern" (see Figure 5), the "Wet" in this case extending into the earlier portion of the rain season (October-November), the "Dry" into the later part of the season (March). Also, in contrast with Figure 5, the anomalies for April-May are both positive for Los Angeles and San Diego.

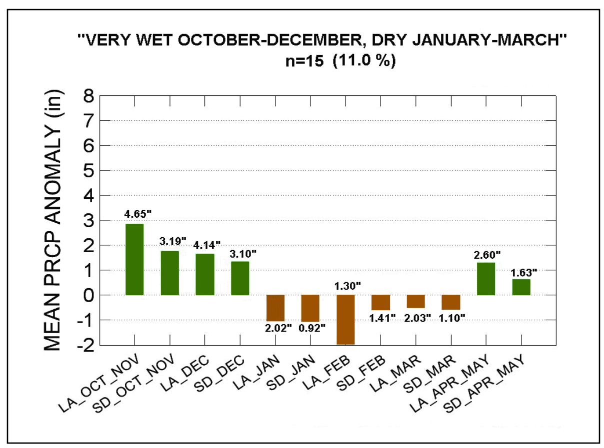

Figure 7 – Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the "Very Wet October-December, Dry January-March" Pattern.

5.1.3. *– Pattern Probabilities Conditioned on El Nino, Neutral, or La Nina occurrences – Bayesian Determinations*

While the percent frequencies of the above four patterns may be considered as probabilities that they may occur individually for a given rain year, there are other climatic indicators that should provide additional, more refined probabilistic information on occurrence likelihoods. ENSO phase (“El Nino, “Neutral”, or” La Nina”) is one indicator known to influence California rainfall patterns, so the next step was to investigate the possible modifying influences of these three episode types on the “baseline” prior probabilities above of the four patterns. This was a conditional probability exercise, and the method of choice, already introduced, would be Bayesian Analysis.

After ENSO episode classification for each of the 137 seasons, the Bayesian conditional probabilities were calculated. Since there were four patterns and three different ENSO phases, 12 separate calculations were performed. Figure 8 shows the Bayesian theorem along with the steps of a sample computation, that for the probability of the “Very Wet February & March” pattern (see Figure 6) given an El Nino episode.

- BAYES THEOREM -

POSTERIOR PROB

PRIOR PROB

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A_c)P(A_c)}.$$

$$P\left(\text{"VERY WET FEBRUARY AND MARCH PATTERN"} \middle| \text{"EL NINO"}\right) = \frac{P\left(\text{"EL NINO"} \middle| \text{"VERY WET FEBRUARY AND MARCH PATTERN"}\right) P\left(\text{"VERY WET FEBRUARY AND MARCH PATTERN"}\right)}{P\left(\text{"EL NINO"} \middle| \text{"VERY WET FEBRUARY AND MARCH PATTERN"}\right) P\left(\text{"VERY WET FEBRUARY AND MARCH PATTERN"}\right) + P\left(\text{"EL NINO"} \middle| \text{"OTHER THREE PATTERNS"}\right) P\left(\text{"OTHER THRI PATTERNS"}\right)}$$

$$P\left(\text{"VERY WET FEBRUARY AND MARCH PATTERN"} \middle| \text{"EL NINO"}\right) = \frac{(12/25) \times (25/137)}{[(12/25) \times (25/137)] + [(28/112) \times (112/137)]}$$

$$P\left(\text{"VERY WET FEBRUARY AND MARCH PATTERN"} \middle| \text{"EL NINO"}\right) = \frac{.0876}{.0876 + .2044} \text{ or } .30$$

Figure 8 – Bayes Theorem *(*from *Wikipedia)* and a Sample Calculation of the Conditional Probability of the “Very Wet February & March” Pattern for Los Angeles/San Diego being realized, given an El Nino episode.

From Figure 8, the top expression shows the general Bayes Theorem, that immediately below the expression adapted to the variables of the sample exercise. In the numerator on the right side of the equation, “P(A)” is the Prior Probability of the “Very Wet February & March” pattern, simply the original proportion of the137 seasons that were so classified by the K-Means/V-Fold algorithm (25/137 or .182, or 18.2%). P(B|A) is the proportion of “Very Wet February & March” cases that were associated with El Nino episodes (in this case, 12/25 or .480, a high relative figure). P(A) and P(B|A) are then multiplied together, yielding .0876, this result also copied into the denominator, to be added to the product of the proportional incidence of El Nino’s in the other non-“Very Wet February & March” pattern cases (28/112 or .250) times the converse of the Prior Probability (.818). This yields .0876+.2044, the final quotient (.0876/(.0876+.2044) giving 30.0% as the Posterior Probabilty, P(A|B): the likelihood that the “Very Wet February & March” pattern will be realized, given an impending El Nino. The Posterior Probability in this example is nearly 2/3rds higher than that of the Prior, indicating that evidence of an El Nino episode does “matter”, in this instance, increasing the odds significantly that the “Very Wet February & March” pattern” will be expressed for the given rain season.

Table 1 lists the Posterior Probability results for the 12 combinations of 3 ENSO types (columns) and 4 Patterns (rows).

Pattern #	Name	Posterior P(A|B) El Nino	Posterior P(A|B) Neutral	Posterior P(A|B) La Nina	Prior P(
1	Wet January, Slightly Dry Others	45.0%	50.8%	47.1%	48.2
2	Wet December, Dry January & February	12.5%	20.6%	38.2%	22.6
3	Very Wet February & March	30.0%	19.0%	2.9%	18.2
4	Very Wet October to December, Dry January-March	12.5%	9.5%	11.8%	10.9
	Total:	**100.0%**	**100.0%**	**100.0%**	**100.(**

Table 1 – Posterior Probability Results for all combinations of ENSO Type vs. Los Angeles/San Diego Pattern.

To interpret, for example, the El Nino Posterior Probability column (third from the left), reading down, lists the conditional probabilities that each of the four patterns will happen, given an El Nino episode. The 30.0% figure for the “Very Wet February & March” pattern, just calculated, is seen in row three, but the most frequent pattern for the El Nino’s (indeed for the Neutrals and La Nina’s as well) is the “Wet January, Slightly Dry Others” pattern with a 45.0 % Posterior. For the La Nina’s, the “Wet January, Slightly Dry Others” pattern has a 47.2% Posterior, but not far behind is the 38.2% figure for the “Wet December, Dry January & February” mode. For the Neutrals, aside from the 50.8 % Posterior for the “Wet January, Slightly Dry Others” pattern, there are no marked Posterior versus Prior contrasts for the other three modes.

So in summary, comparing the El Nino and La Nina results in Table 1, most affected seasons in the Los Angeles and San Diego areas would seem to exhibit non-descript anomaly patterns period-to-period. But, a sizeable minority of secondary modes also exhibit marked and contrasting configurations, in particular an El Nino tendency to gravitate around heavy rains in the late season (February & March), and a La Nina proclivity for wet conditions relatively early in the season (December) followed by dry ones over January and February.

5.2 - Downtown San Francisco Results

Figure 9 shows bar graphs depicting mean overall San Francisco precipitation figures, respectively, for the six calendar periods under consideration. Mean seasonal San Francisco precipitation is significantly higher than that for Los Angeles and San Diego, about 40% so more than Los Angles and more than double that of San Diego.

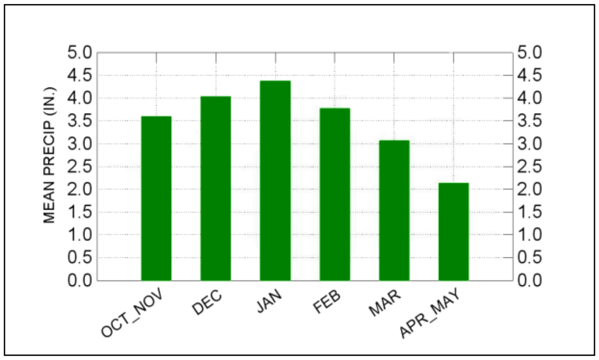

Figure 9. Mean Downtown San Francisco Precipitation for October-November, December, January, February, March, and April-May calendar periods, 1877-78 to 2013-14 Period of Record.

***5.2.1** – Scree Plot*

Combined use of the Squared Euclidean Distance Metric and inspection of the associated Scree Plot results brought about "resolution" of six clusters for San Francisco. From Figure 10, the magnitude of the curve at K=7 was about the same as for K=6, the "cutoff" point thus being at K=6.

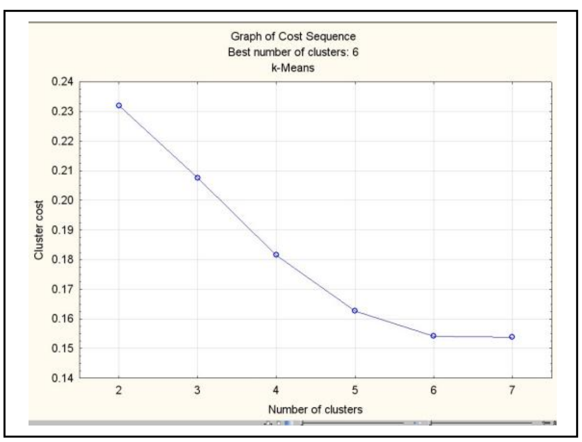

Figure 10 – Scree Plot of K-Means/V-Fold Cross Validation Algorithm Analysis of San Francisco October-November, December, January, February, March, and April-May Precipitation Anomalies.

5.2.2 - Downtown San Francisco Results

Figures 11 thru 16 depict the mean anomalies for each of the six patterns, like Los Angeles/San Diego, in descending order of importance. Also, in analogous fashion, except for Figure 11a, the period-to-period anomalies for a given pattern are the deviations from the overall 137-year mean statistics. As previously, annotations above or below the individual bars for all the charts are the actual cluster means for the period concerned.

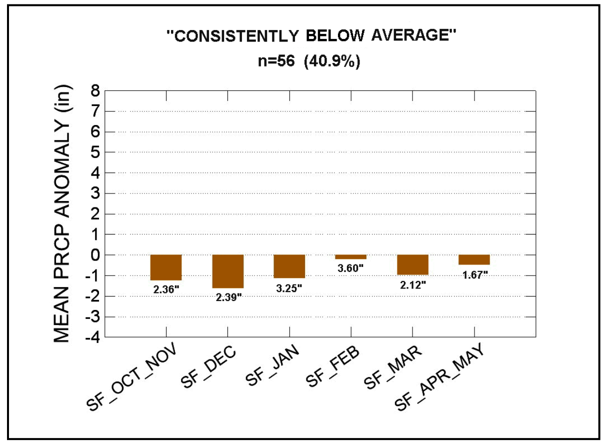

Figure 11 – Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the "Consistently Below Average" Pattern.

The "Consistently Below Average" pattern (Figure 11) is easily the most frequent of the six, some 56 seasons or 40.9% affiliated with this group. Like that for Los Angeles/San Diego, the configuration of the bars likely reflects the clustering algorithm's data normalization, and to investigate this possibility once again, Figure 11a shows the departures relative to the medians.

Similar to those in its counterpart chart of Figure 4a, most of the individual bars in Figure 11a are positioned nearer to zero, the February and April-May ones, formerly negative, now positive. Mean anomaly magnitude for the six periods in Figure 11 is -0.922, that in Figure 11a, just -0.297.

Thus, similar to that for Los Angeles/San Diego, the generalization could be made, allowing for intra-cluster individual season variation, that a sizeable proportion of seasons for San Francisco (41%) display relatively unexceptional period-to-period anomaly configurations.

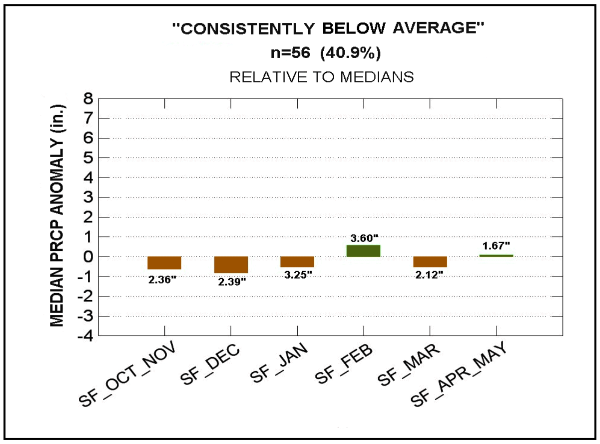

Figure 11a – Mean Period-to-Period Anomalies Relative to Long-Term Overall Medians for the “Consistently Below Average” Pattern.

Second in importance was the “Wet December, Dry February and March” pattern (n=23, frequency: 16.8% - See Figure 12). This displays a transition from an exceptionally wet December regime (mean anomaly approaching +5”, to a relatively dry pattern covering both February and March, each about 1” below average. Physically, this is suggestive of a mean trough to ridge “propagation” between the December and February-March. Such early to later seasonal transitions are also displayed in similar but not identical form in Figures 5 and 7 for Los Angeles/San Diego.

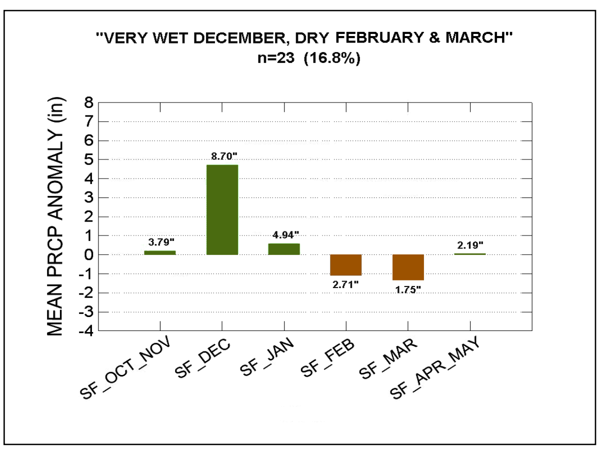

Figure 12 - Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the “Very Wet December, Dry February & March” Pattern.

Ranking third was the “Very Wet March” pattern (n=22, frequency 16.1% - See Figure 13), the most notable feature of the graph being the highly positive mean anomaly for March (approaching +4”). The anomalies for the contiguous periods October-November, December, and January are all negative, those for February, March, and April-May all positive.

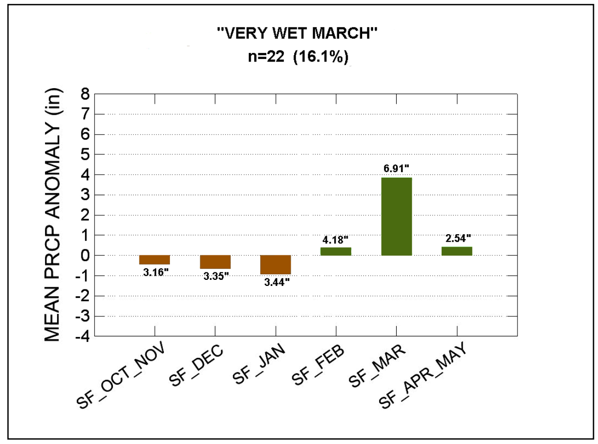

Figure 13- Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the “Very Wet March” Pattern.

In fourth place was the “Very Wet October-November/Dry February” pattern (n=20, frequency 14.6% - See Figure 14). This exhibits a very pronounced positive anomaly for October-November (departure in excess of 4”) followed by a noticeably negative one (approaching -2”) in February, suggestive physically of another kind of early-season to later season transition from a predominant trough regime to a more ridge-influenced one. Compared to Figure 12, however, Figure 14’s is less “rapid”, three periods separating October-November and February compared to two for the former (i.e., December vs. February).

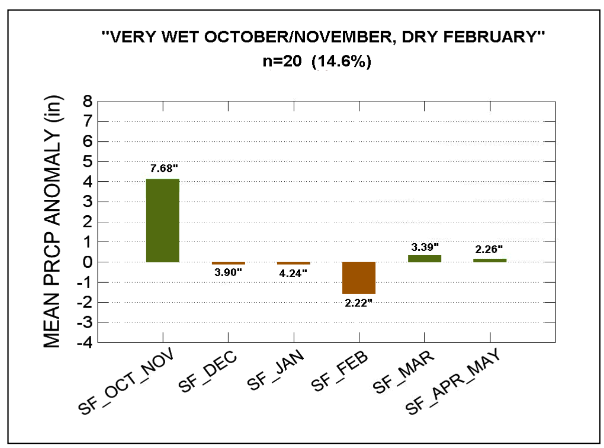

Figure 14 - Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the “Very Wet October/November, Dry February” Pattern.

Ranking fifth was the “Wet January & February” pattern (n=14, frequency 10.4% - see Figure 15). This is the only San Francisco pattern with significant positive anomalies shown for contiguous periods; such was a routine feature for Los Angeles/San Diego. Mean positive anomaly for January approaches 5 ½”, that for February nearly 4”. The other four periods show mostly modestly negative departures.

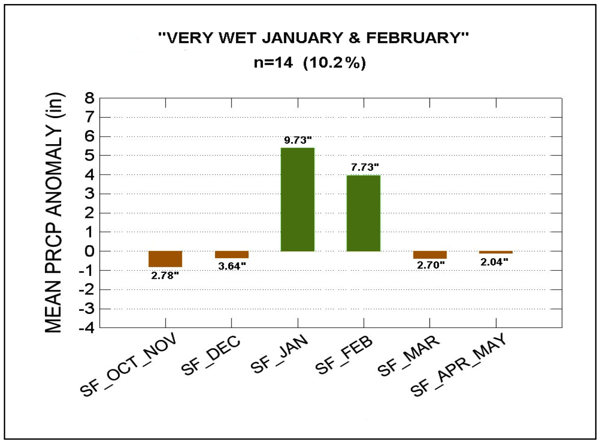

Figure 15 - Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the “Very Wet October/November, Dry February” Pattern.

Ranking last was the “Very Wet April-May” pattern (n=2, frequency: 1.5% - see Figure 16). With a sample size of just 2, it’s obviously not very significant in climatological scheme of things. Its major feature is the very large positive mean anomaly shown for April-May, approaching +7”, and reflecting an average precipitation amount in excess of 400% of average.

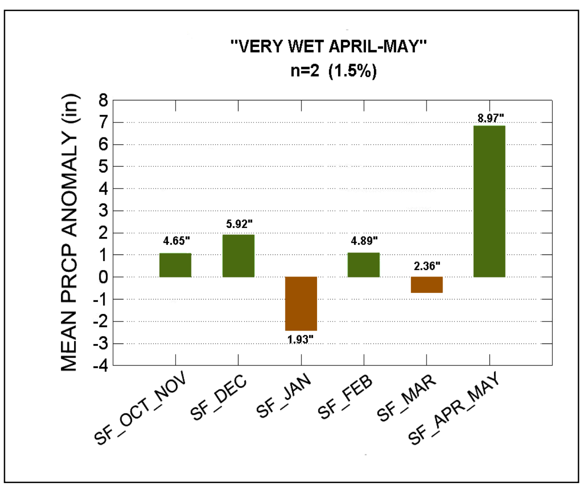

Figure 16 - Mean Period-to-Period Anomalies Relative to Long-Term Overall Means for the “Very Wet April-May” Pattern.

5.2.3. *– Pattern Probabilities Conditioned on El Nino, Neutral, or La Nina occurrences – Bayesian Determinations for San Francisco*

Pattern #	Name	Posterior P(A|B) El Nino	Posterior P(A|B) Neutral	Posterior P(A|B) La Nina	Prior P(A)
1	Consistently Below Average	45.0%	39.7%	38.2%	40.9%
2	Very Wet December, Dry February & March	15.0%	20.6%	11.8%	16.8%
3	Very Wet March	15.0%	14.3%	20.6%	16.1%
4	Very Wet October-November, Dry February	10.0%	14.3%	20.6%	14.6%
5	Very Wet January & February	15.0%	9.5%	5.9%	10.2%
6	Very Wet April-May	0.0%	1.6%	2.9%	1.5%
	Total:	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Table 2 – Posterior Probability Results for all combinations of ENSO Types vs. San Francisco Patterns

Repeating the Bayesian Analyses as performed for Los Angeles/San Diego, Table 2 shows the Posterior Probability results for San Francisco**.**

Comparing the Posterior Probability figures, the “Consistently Below Average” pattern has the highest posterior figures for El Nino (45.0%), Neutral (39.7%), and La Nina (38.2%), and in contrast with Los Angeles/San Diego (See Table 1), there are no close second place competitors across any of the three episode types. For the El Nino’s, three patterns tie for second place (15.0 % posteriors each), and for the La Nina’s, two tie for second (20.6% posteriors each).

The most conspicuous contrasts between El Nino and La Nina posteriors are seen for the “Very Wet January & February” and “Very Wet October-November, Dry February” patterns. In the former, the El Nino figure exceeds that of the La Nina 15.0% to 5.9%, and in the latter the La Nina posterior exceeds that of El Nino 20.6% to 10.0%. These are not nearly as distinct as those analogous comparisons shown in Table 1. For example, the posterior for the “Very Wet February-March” pattern had a 30.0% to 2.9% contrast between El Nino and La Nina, that for the “Wet December, Dry January & February pattern” a 38.2% to12.5% disparity between La Nina and El Nino.

In sum, ENSO type does not seem to affect the San Francisco patterns’ Bayesian posteriors like it does the Los Angeles/San Diego ones. This is likely due to the more northerly location of San Francisco and its greater susceptibility to precipitation producing influences other than those related to ENSO phase.

6. SUMMARY

Utilizing the K-Means/V-Fold Cross-Validation clustering methodology, the foregoing investigated the existence and relative frequencies of month-to-month precipitation anomaly modes for the Downtown Los Angeles and Downtown San Diego stations (analyzed as a unit), and that for Downtown San Francisco, each covering the 1877-78 through 2013-14 periods of record. The Squared Euclidean distance metric combined with Scree Plot inspections were utilized to generate and decide upon the number of clusters (or “patterns”). The calendar units of analysis were

October-November, December, January, February, March, and April-May.

Overall, four such patterns were resolved for Los Angeles/San Diego, six for San Francisco. Primary modes for each (48% and 41% incidences, respectively) were those that displayed, with a few exceptions, very modest mean precipitation deviations period-to-period, relative to overall climatology. This was interpreted to mean that the most frequent anomaly patterns expected in these cases should be relatively unexceptional ones, period-to-period (the ultimate "unexceptional" pattern, of course, would be period-to-period precipitation totals that matched climatology exactly). Some of the lesser, secondary patterns included "standalone", single period highly positive (wet) anomalies, and progressions that suggested trough (wet) to ridge (dry) transitions over differing time scales. Los Angeles/San Diego seemed to be more prone to significant anomaly patterns that extended across contiguous periods.

Following the pattern resolutions and evaluations, Bayesian analyses were performed to investigate the influences of ENSO phases ("El Nino, "Neutral", and "La Nina") on occurrence likelihoods ("Posterior Probabilities"). Strong contrasts in posterior magnitudes for El Nino's versus La Nina's were seen for two of the four Los Angeles/San Diego patterns, but no similarly pronounced differences were noted for any of San Francisco's six modes.

7. REFERENCES

Nisbet, R., Elder, J., and Miner, G., 2009: Handbook of Statistical Analysis & Data Mining Applications Elsevier, 824 pp.

http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears_1877-present.shtml

http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml