IDENTIFICATION OF CALIFORNIA, OREGON, AND WASHINGTON COASTAL AREA CLIMATE DIVISIONS' RELATIVE ANOMALY MODES IN TOTAL SEASONAL PRECIPITATION WITH CHARACTERIZATION OF THEIR OCUURRENCE PROBABILITIES RELATIVE TO EL NINO, NEUTRAL, AND LA NINA EPISODES

Charles J. Fisk

1. INTRODUCTION

In a recent study (Fisk, 2015), the existence and character of idealized relative statistical anomaly patterns or "modes" in July-June total precipitation were investigated for the seven NCDC California Climate Divisions, utilizing the 1895-96 thru 2013-14 period of record, K-means clustering analysis, and calculations of Bayesian probabilities of the patterns' likelihoods, given the particular ENSO phase ("La Nina", "Neutral", and "El Nino") in place. Results resolved six modes, some of which for the two southernmost divisions displayed noticeable contrasts in relative anomaly character compared to those of the other five. The ENSO phases also showed different affinities for different clusters. Extending the scope to include the Pacific states collectively, Oregon and Washington as well as California, and refining the ENSO phase designations by adding the "Strong La Nina", "Weak El Nino", "Moderate El Nino", and "Strong El Nino" subcategories, the following study repeats the objective of the previous study, incorporating in the process an additional season's data (2014-15).

Twenty-six total climate divisions comprise the three States and to parse down the number of divisions and at the same time keep a near-coastal focus, the selection is reduced to include those closest to the Pacific (and in the case of Oregon and Washington, essentially those west of the Cascade Mountains). This leaves twelve divisions, three in California, four in Oregon, and five in Washington. The K-Means Clustering methodology is integrated with the V-Fold Cross Validation Algorithm, an iterative training sample type procedure that optimizes the number of clusters created, depending on the choice of statistical distance metric (Euclidean, Squared Euclidean, etc.), percent improvement cutoff threshold (e.g., 5 percent), and other settings. In this study the K-means approach utilizes the Squared Euclidean metric combined with the 5 percent distance improvement cutoff threshold; also the precipitation data are normalized in advance, by division. Seven clusters are resolved.

Then, through referencing and adaptation of bi-monthly ranked statistics from the MEI ENSO Index Data base back to 1895-96, a Bayesian statistical analysis is done which addresses the following questions: given the presence of a "Strong El Nino", "Moderate El Nino", "Weak El Nino", "Neutral", "Ordinary La Nina" , or "Strong La Nina" episode, what are conditional probabilities that each of the seven idealized anomaly patterns would be expressed for a given July to June rain season. Results are described and interpreted, including both descriptions and discussions of the patterns along with their Bayesian probabilities.

. ---

* *Corresponding author address:* Charles J. Fisk, e-mail: cjfisk@att.net

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).

Applied to K-Means, the V-fold cross-validation scheme 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 the 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 Climate Division 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. Cluster results would be presented in pre-normalized data form.

Since the percent improvement threshold default setting (5 percent) can be changed, potentially resulting in a different "best" cluster size, an alternative graphical tool is available that can provide a different selection option. This tool, the Scree Plot, traces the actual (usually decreasing) mean training sample statistical distances over a range of increasing K's. Inflection points on the Scree Plot can be interpreted as "natural" cutoff points, the "best" cluster size corresponding to the inflection point's Kth position on the graph. The percent improvement cutoff K may differ (the iterations having stopped at K+1), so, alternatively, if one opts to choose the inflection point as the "right" K and it is different than the percent improvement threshold K, the program can be rerun, "forcing" the "optimal" cluster size and accompanying analysis and information to correspond to that at K, and K only. If one is interested in a more exhaustive analysis, the forcing could be done at a Scree K-value that exhibits essentially zero change in mean training sample statistical distance from the preceding K-1 level. At even higher K levels, of course, the statistical distance curve might trend back upward, reflecting over-fitting.

In this study, the 5 percent default distance improvement cutoff threshold was utilized along the Squared Euclidean distance metric (default: Euclidean), together with Scree Plot inspection.

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) into a revised probabilistic belief that the given pattern will occur. 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 Climate Division precipitation data.

4. THE DATA

The raw data were downloaded via an NCDC online link which has the complete history for the July 1895 to June 2015 period of interest. Figures 1 through 3 are maps of the California, Oregon, and Washington Climate Divisions, respectively, included in this study. For California (Figure 1), the three divisions are 1). "South Coast", 2.) "Central Coast", and 3.) "North Coast". For Oregon (Figure 2), they are 1.) "Coastal Area" , 2.) "Southwestern Valleys", 3.) "Williamette Valley", and 4.) "Northern Cascades". Finally, for Washington (Figure 3) they are 1.) "West Olympic Coastal", 2.) "East Olympic Cascade Foothills", 3.) "Puget Sound Lowlands", 4.) "Cascade Mountains West", and 5.) "NE Olympic San Juan". For presentation purposes, all of these 12 titles, by necessity, appear in abbreviated form.

Also, Figure 4 is a bar chart of the 120-year mean July-June precipitation figures, by division, and Figure 5 a similar type bar chart of the standard deviation statistics, by division. From Figure 4, there is a wide range of mean statistics, from 108.66" in the Washington ""West Olympic Coastal" Division, to 17.38" in the California "South Coast" Division. The standard deviation statistics in Figure 3 range from 16.21" for the Washington "West Olympic Coastal" to 4.01" for the Washington "NE Olympic San Juan" Division.

With such a wide division-to-division range in overall mean precipitation and variability across the State, it makes sense from an interpretation standpoint to express the individual cluster results, division-by-division, in terms of relative or standardized deviations from the overall averages in Figure 4, based on the overall standard deviation statistics depicted in Figure 5.

Figure 1 – Map of California Climate Divisions included in this study – from NCDC.

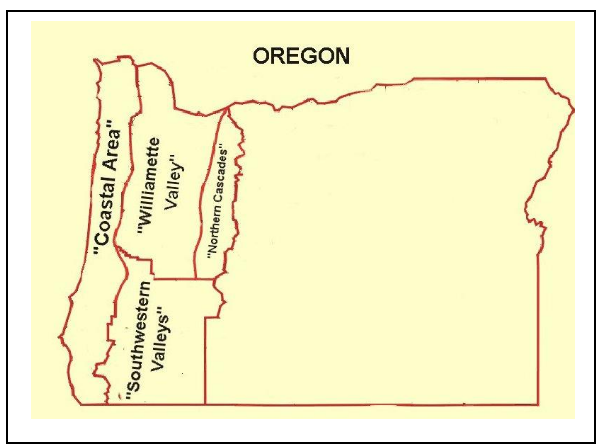

Figure 2 – Map of Oregon Climate Divisions included in this study – from NCDC.

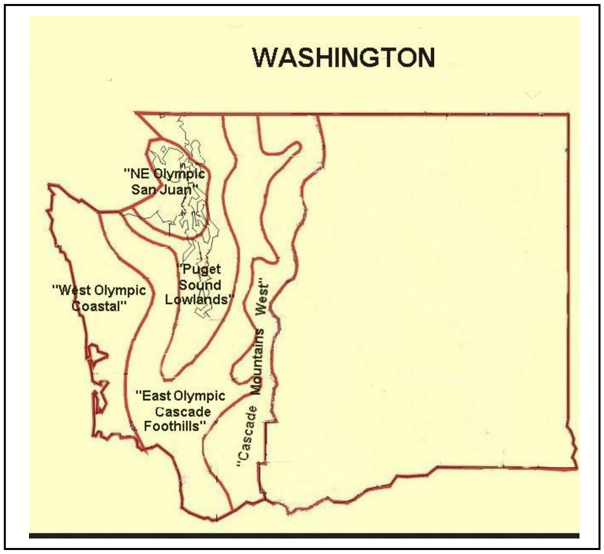

Figure 3 – Map of Washington Climate Divisions included in this study – from NCDC.

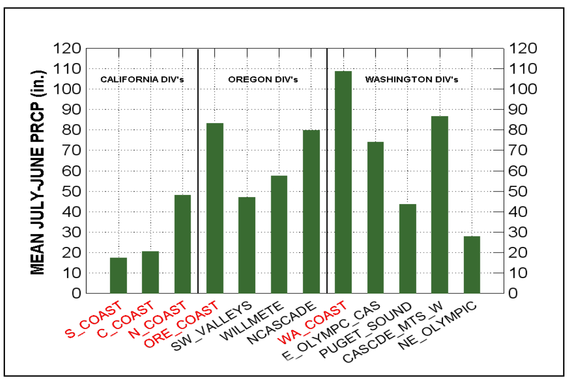

Figure 4 – Mean Seasonal (July-June) Precipitation (In.) For 12 NCDC Near-Coastal California, Oregon, and Washington Climate Divisions (1895-96 thru 2014-15 Period of Record

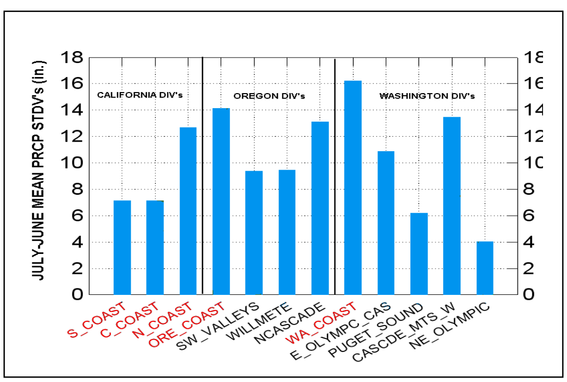

Figure 5 – Seasonal (July-June) Precipitation Series' Standard Deviations (In.) – 12 NCDC Near-Coastal California, Oregon, and Washington Climate Divisions (1895-96 thru 2014-15 Period of Record)

It will be noted that in both charts, five of labels below the horizontal axis are shaded red; these are interpretative aids, identifying those (elongated and narrow) divisions that border directly on the Pacific Ocean, All three of the California divisions fit this category along with one each for Oregon/Washington.

5. RESULTS

The K-Means/V-Fold algorithm produced seven clusters, ranging in percent frequency from 23.3% to 4.2%.

5.1. *– Scree Plot*

Figure 6 is a Scree Plot of the iterative results. The "Best K", determined by the 5% default improvement cutoff level is at K=7; this is somewhat arbitrary as the inflection point between K=7 and K=8 is not a particularly dramatic one, a somewhat lesser percent improvement figure would have likely allowed the process to stop at K=8.

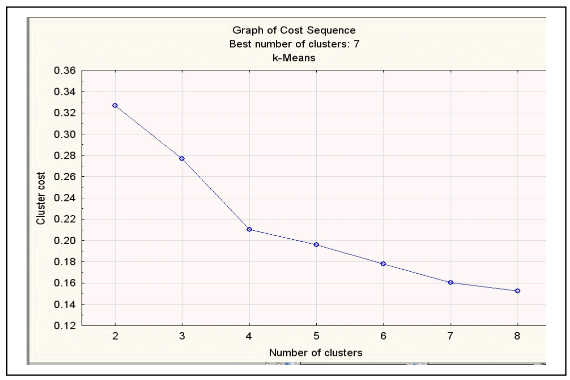

Figure 6 – Scree Plot of K-Means/V-Fold Cross Validation Algorithm Analysis of NCDC Near-Coastal California, Oregon, and Washington Climate Divisions Seasonal (July-June Total) Precipitation Anomalies.

__5.2.__ – Standardized Mean Anomaly Charts for the Individual Patterns.

Figures 7 thru 13 present the division-by-division standardized mean anomalies for each of the seven patterns, in descending order of importance.

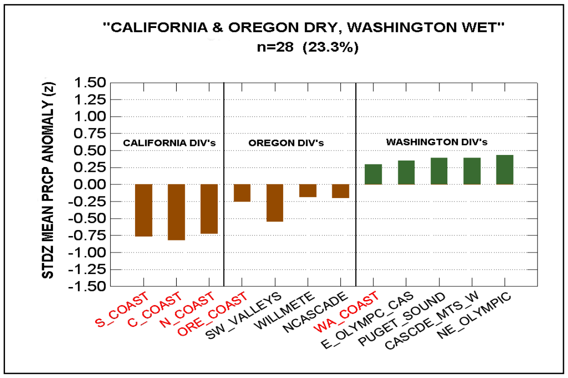

Figure 7 – Standardized Mean Division-by-Division Anomalies for the "California & Oregon Dry, Washington Wet" Pattern - Ranking Mode #1.

Figure 7 shows the most frequent mode (23.3% incidence), dubbed the "California & Oregon Dry, Washington Wet" one. A generalized dry to wet trend, north to south, is evident between States. For Coastal California, the standardized anomalies are consistently and significantly negative (from -0.72 z to -0.81 z) for all three divisions. Utilizing the divisional standard deviation statistics (depicted in Figure 5), these correspond to absolute July-June precipitation mean deficits of 5.4" and 5.8" for the South and Central Coasts, respectively, and 9.1" for the North Coast. For Oregon, the anomalies are more modestly negative, and in a "wet" shift, all five divisions for Washington exhibit positive departures, between + 0.30 z and +0.44 z, corresponding, for example, to excesses for the naturally rainy "Wa_Coast" and "Cascade_Mts_W" regions of 4.8" and 5.3", respectively.

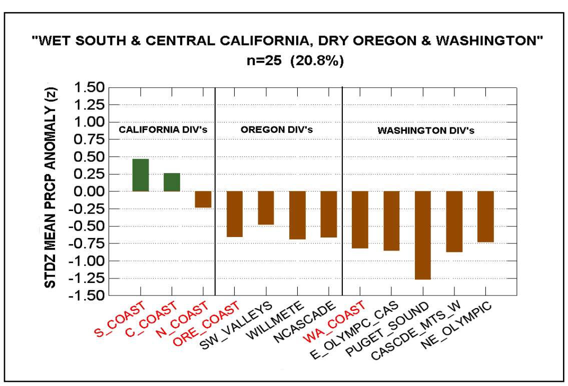

Figure 8 – Standardized Mean Division-by-Division Anomalies for the "Wet South & Central California, Dry Oregon & Washington " Pattern –Ranking Mode #2.

Ranking second is the "Wet South & Central California, Dry Oregon & Washington" pattern (20.8% incidence – see Figure 8). This displays pronounced negative mean anomalies for Oregon and Washington, the "Puget_Sound", Washington division, for example, showing a marked -1.27 z mean departure, denoting a 7.9" deficit. In striking contrast, however, those of the California South Coast are positive at +0.47 z (equivalent to a 3.3" excess), the Central Coast's modestly so (+0.26 z), and the North Coast modestly negative (-0.23 z). Such dichotomies between California and Oregon/Washington will be seen in other charts to follow, along with occasional inter-California ones between the South Coast/Central Coasts and the North Coast.

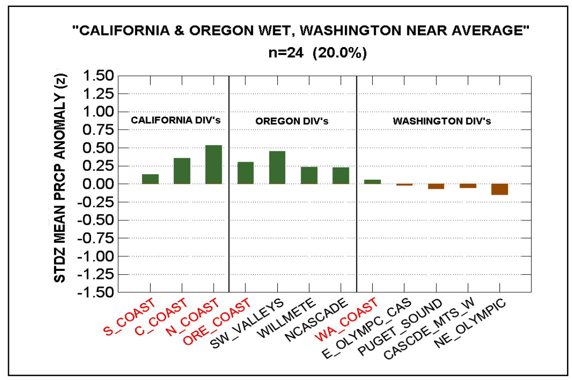

Figure 9 – Standardized Mean Division-by-Division Anomalies for the "California & Oregon Wet, Washington Near Average " Anomaly Pattern – Ranking Mode # 3.

Third most frequent is the "California & Oregon Wet, Washington Near Average" Pattern (20.0% incidence - see Figure 9). In this case, California/Oregon contrast with Washington, the former two states exhibiting a mix of modestly to significantly wet seasons, compared to essentially near average conditions for all of Washington's five divisions. The California North Coast's mean departure statistic is slightly greater than +0.54 z, translated into a 6.8" excess.

The highest three ranking modes discussed so far make up nearly two-thirds or 64.1 % of the total cases.

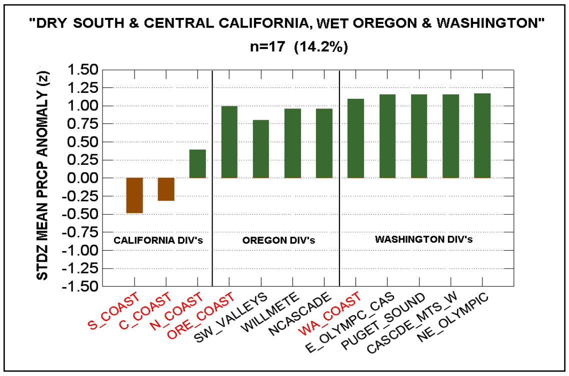

Figure 10 – Standardized Mean Division-by-Division Anomalies for the “Dry South & Central California, Wet Oregon & Washington Pattern – Ranking Mode # 4.

In fourth place is the “Dry South & Central California, Wet Oregon & Washington” Mode (14.2 % incidence – see Figure 10). This is essentially a mirror image (oppositely signed anomaly configurations) of Figure 8, Oregon and Washington all very wet, slightly more so in absolute anomaly terms than they were dry in Figure 8 In contrary fashion once again, the California South and Central Coast divisions show *negative* departures (the former -0.48 z, the latter -0.31 z); the North Coast shows a positive +0.39 z figure, but this is significantly less than those of Oregon/Washington, which range uniformly between +1.1 z and +1.2 z.

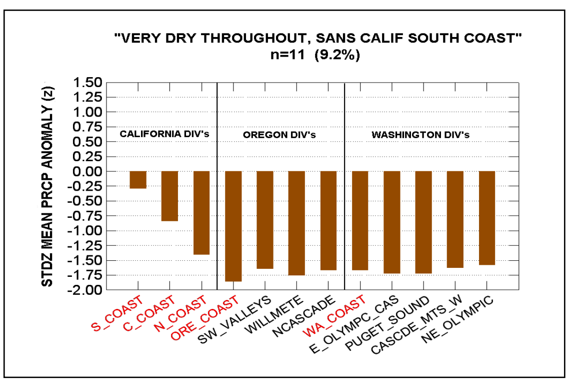

Figure 11 – Standardized Mean Division-by-Division Anomalies for the “Very Dry Throughout, Sans Calif South Coast” Relative Anomalies” Pattern – Ranking Mode # 5.

Ranking fifth is the “Very Dry Throughout, Sans Calif South Coast” Pattern” (9.2 % incidence – see Figure 11). This displays a decidedly negative series of departures, especially for Oregon and Washington, and to a somewhat lesser extent, but still at significant levels, the California North and Central Coasts. The Oregon and Washington deviations range from -1.58 z to -1.85 z.

While California’s departures are also negative, once more the relative character disparity with Oregon’s and Washington’s gets greater as one goes farther south. The North Coast’s anomaly is roughly comparable (-1.40 z), the Central Coast’s -0.83 z, but the South Coast’s is just -0.29 z.

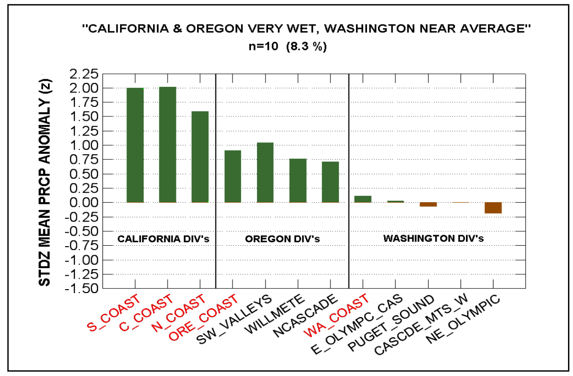

Figure 12 – Standardized Mean Division-by-Division Anomalies for the “California & Oregon Very Wet, Washington Near Average” Pattern – Ranking Mode #6.

Sixth in relative importance is the “California & Oregon Very Wet, Washington Near Average” Pattern (8.3% incidence – See Figure 12). This configuration is not unlike Figure 7’s, with essentially near average departures for Washington, contrasted with wet ones for California and Oregon, except in this case, California’s are *very* wet, those of the South Coast and Central Coast reaching +2.00 z and +2.01 z, respectively; the North Coast’s is +1.59 z. Those of Oregon are at lesser but still quite wet-indicative levels (between +0.71 z to +1.05 z.

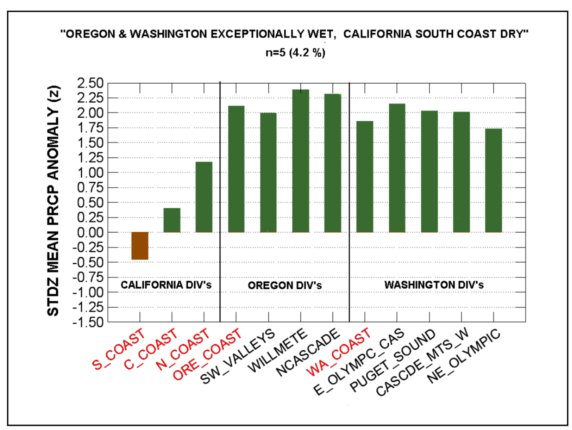

Figure 13 - Standardized Mean Division-by-Division Anomalies for the “Oregon & Washington Exceptionally Wet, California South Coast Dry Pattern – Ranking Mode #7.

Ranking seventh (and last) is the “Oregon & Washington Exceptionally Wet, California South Coast Dry” pattern (Incidence: 4.2 % - See Figure 13), comprising only 5 cases. This is not unlike Figure 8’s configuration, except for the fact that the Oregon and Washington mean anomalies are amplified to exceptionally high levels, most at +2.00 z or higher, the "WILLMETE" division (relating to the Willamette Valley) reaching +2.39 z. Once more, there are disparities relative to California, especially pertaining to the Central and South Coast – the latter showing a significant - 0.45 z departure, indicative of a mean 3.2" deficit. In contrast, the Oregon "NCASCADE" (Northern Cascades) division exhibits a greater than 30" excess.

Reviewing the seven patterns configurations collectively, and comparing them across States, the previously described non-conformities of the California mean relative departures relative to those of Oregon and Washington are present in four of the seven, constituting 71.8 % of the total cases. Also, as previously described, these dissimilarities get more pronounced as one moves from the California North Coast to the South Coast.

5.3. *– Pattern Probabilities Conditioned on El Nino, Neutral, or LaNina occurrences – Bayesian Determinations*

While the percent frequencies of the above seven patterns may be considered as probabilities that they may occur individually for a given July-June 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 West Coast rainfall patterns, so the next step is to investigate the possible modifying influences of these three episode types on the "baseline" Prior probabilities above of the six patterns. This would be a conditional probability exercise, and the method of choice, already introduced, would be Bayesian Analysis.

First, the 120 seasons are assigned ENSO episode classification. To begin this task the online MEI ("Multivariate ENSO index") web site is referenced (Wolter, 2015). The most current data base consists of standardized index values encompassing two-month running periods, going back to Dec 1949/Jan 1950. A separate legacy data base has index values covering Dec 1870/Jan 1871 through Nov/Dec 2005. As the NCDC Climate Divisions' history goes back to 1895, both the legacy and current versions of Wolter's data base are utilized.

A strategy is adopted to merge the two, joining the legacy data set up through 1949 (the1950-2005 portion excluded) with the 1950-present data set, the combined data sets' two-month running periods then re-standardized as a single unit. Next, the re-standardized data set is restructured into a July-June format (a "season" covering the JunJul to MayJun periods). Following an approach that Wolter uses, each of the two-month moving periods are assigned a rank; the individual ranks then added and averaged to create a seasonal average rank (based on the 1871-2014 period). La Nina episodes, for example, which typically display negative indices on a period to period basis, have "lower" numerical ranks (the strongest La Nina with a rank: 1).

Since West-Coast precipitation tends to decrease markedly late in the season and El Nino episodes, in particular, have a tendency also to decay in strength in Spring, it was of preliminary interest to evaluate the overall relationships of the individual two-month ranks with the overall average ranks (i.e., determine how "representative" they were relative to the overall picture).

Correlation coefficient calculations revealed that all the individual periods had correlations greater than +.910 except JunJul (+.834), AprMay (+.786), and MayJun (+.604), so in another arbitrary step, the AprMay and MayJun variables were dropped from further consideration. The average overall ranks were then recalculated using the remaining ten periods.

Next, data for the seasons 1871-72 to 1894-95 were dropped, a new data set created including 1895-96 to 2014-15 statistics only. The seasonal ranking averages of these were then sorted with following ENSO classification scheme applied: Rankings 1 to 12: "Strong La Nina", rankings 13-40: "Other La Nina", rankings 41-80: "Neutral", rankings 81-94: "Weak El Nino":, rankings 95-108 : " Moderate El Nino" and finally, rankings 109-120: "Strong El Nino". Thus, there would be 12 "Strong" La Ninas, 28 "Other" La Ninas, 40 "Neutrals", 14 "Weak" El Ninos, 14 "Moderate" El Ninos, and 12 "Strong" El Ninos.

The seasons designated as "Strong" El Ninos, for example, in descending order of mean rank magnitude were 1.) 1997-98, 2.) 1982-83, 3.) 1930-31, 4.) 1972-73, 5.) 1941-42, 6.) 1991-92, 7.) 1902-03, 8.) 1896-97, 9.) 1918-19, 10.) 1940-41, 11.) 1957-58, and 12.) 1965-66.

Next, the Bayesian conditional probabilities were calculated. Since there were seven patterns and six different ENSO phases, there would be 42 separate calculations. Figure 14 shows the Bayesian theorem along with the steps of a sample calculation, that for the conditional probability of Pattern #2 ("California & Oregon Very Wet, Washington Average" – see Figure 12) as associated with an imminent "Strong" El Nino episode.

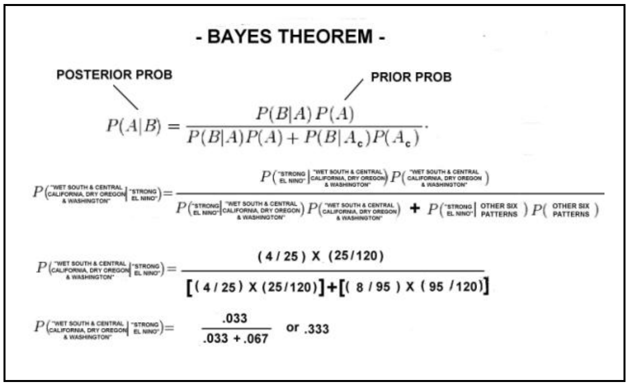

Figure 14 – Bayes Theorem *(from Wikipedia)* and a Sample Calculation of the Conditional Probability of the "Wet South & Central California, Dry Oregon & Washington Pattern (# 2) being realized, given an impending or ongoing "Strong" El Nino.

From Figure 12, the top expression shows the general Bayes Theorem, that immediately below it 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 "Wet South & Central California, Dry Oregon & Washington" Pattern, simply the original proportion of the 120 seasons that were so classified by the K-Means/V-Fold algorithm (25/120 or .208, or 20.8%). P(B|A) is the proportion of the pattern's cases that were associated with Strong El Nino episodes (in this case, 4/25 or .160), P(A) and P(B|A) are then multiplied together, yielding .033, this result also copied into the denominator, to be added to the product of the proportional incidence of Strong El Ninos in the other six patterns collectively (8/95 or .084 times the converse of the Prior Probability (.792), yielding +.067. The final quotient (.033/(.033+.067) or 33.3% is the Posterior Probabilty, P(A|B): the likelihood that the "Wet South & Central California, Dry Oregon & Washington" pattern will be ultimately be realized, given an impending Strong El Nino. The Posterior Probability in this example is about 2/3rds higher than that of the Prior, indicating that a Strong El Nino episode does boost the odds noticeably that this pattern will be expressed for a July-June rain season on the West Coast

Table 1 lists the Posterior Probability results for all the 42 combinations of 6 ENSO types (columns) and 7 Patterns (rows). Posteriors of particular interest are shaded in red.

Pattern #	Name	Posterior P(A|B) "Strong El Nino"	Posterior P(A|B) "Moderate El Nino"	Posterior P(A|B) "Weak El Nino"	Posterior P(A|B) "Neutral"	Posterior P(A|B) "Other La Nina"	Posterior P(A|B) "Strong" La Nina	Prior P(A)
1	"California & Oregon Dry, Washington Wet"	8.3%	0	28.6%	30.0%	28.6%	25.0%	23.3%
2	"Wet South & Central, California, Dry Oregon & Washington"	33.3%	42.9%	14.3%	15.0%	17.9%	16.7%	20.8%
3	"California & Oregon Wet, Washington Near Average"	25.0%	14.3%	21.4%	22.5%	21.4%	8.3%	20.0%
4	"Dry South & Central California, Wet Oregon & Washington"	0.0%	7.1%	0.0%	20.0%	14.3%	33.3%	14.2%
5	"Very Dry Throughout, Sans Calif South Coast"	8.3%	14.3%	28.6%	5.0%	7.1%	0.0%	9.2%
6	"California & Oregon Very Wet, Washington Near Average"	25.0%	21.4%	7.1%	5.0%	3.6%	0.0%	8.3%
7	"Oregon & Washington Exceptionally Wet, California South Coast Dry"	0.0%	0.0%	0.0%	2.5%	7.1%	16.7%	4.2%

Table 1 – Posterior Probability Results for all combinations of ENSO Type vs. Pattern

To interpret, for example, the "Strong El Nino" Posterior Probability column (third from the left), reading down, lists the conditional probabilities that each of the seven patterns will be realized, given a Strong El Nino episode. The 33.3% figure, shaded red for the "Wet South & Central California, Dry Oregon & Washington" pattern (shown in Figure 8 and having already served as the Bayesian computation example above) is the pattern most likely to happen of the seven. As already discussed, the 33.3 % figure is significantly higher than the pattern's 20.8% Prior shown in Column 9. Interestingly, the Posterior shown for this pattern relative to a "Moderate" El Nino is an even higher 42.9 %, more than double the Prior. Both Strong and Moderate El Nino's thus seem to prefer this "Wet South & Central California, Dry Oregon & Washington" configuration. They also prefer as a pair the "California and Oregon Very Wet, Washington Near Average" pattern (Figure 12), the posteriors (25.0% and 21.4%, respectively) in this case, far higher than the Prior (8.3 %).

In contrast, the most favored pattern for "Strong La Ninas" (33.3% Posterior Probability) is the "Dry South and Central California, Wet Oregon & Washington" pattern, shown in Figure 10; this is markedly higher than the Prior (14.2%).

Also curiously, the "Very Dry Throughout, Sans Calif South Coast" pattern (Figure 11), seems to be preferentially associated with Weak El Nino's (the 28.6 % Posterior much higher than the Prior, 9.2%).

There are also a number of other interesting Prior vs. Posterior contrasts, so, in conclusion, conditioning the occurrence probabilities of the seven patterns on ENSO phases did provide more refined insights on their likelihoods. The range of their Priors was 4.2% to 23.3%, that for the Posteriors 0.0% to 42.7%.

6. SUMMARY

Utilizing the clustering tool K-Means, integrated with the V-fold cross validation algorithm, the existence and character of seasonal (July-June total) precipitation modes were explored, collectively, for the seven NCDC California climate divisions, accessing the 1895-96 to

2013-14 period of record. Inputs were normalized, areal-averaged total precipitation statistics season-by-season, and division-by division.

Results resolved seven clusters (also “patterns” or “modes”), characterizing a variety of anomaly configurations across the three States. Individual pattern frequencies (“Prior probabilities) ranged from 4.2% to 23.3%. Then, using Bayesian statistical methodology, conditional probability estimates (Posterior probabilities) were made of the occurrence likelihoods of the seven patterns, given Strong El Nino, Moderate El Nino, Weak El Nino, Neutral, Other La Nina, or Strong La Nina episodes imminent or already in place. In many of the 42 Posterior Probability calculations, the Posterior magnitudes differed markedly from the Priors, indicative that El Nino type was a useful predictive indicator. These figures ranged from 0.0% to 42.7%.

A combined Clustering/Bayesian analysis of this kind might prove similarly useful in other climatological-related applications.

6. REFERENCES

Fisk, C., 2015, Identification of California Climate Division Rain Year Precipitation Anomaly Patterns (1895-96 to 2013-14 Seasons) with Bayesian Analyses of Occurrence Probabilities Relative to El Nino, Neutral, or La Nina Episodes: AMS 13th Conference on Artificial Intelligence, Phoenix, AZ, 6 January, 2015.
https://ams.confex.com/ams/95Annual/webprogram/Paper260219.html

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

Wolter K., (2015) - (MEI Index Web-page): http://www.esrl.noaa.gov/psd/enso/mei/