P1.3 IDENTIFICATION OF DIURNAL WIND PATTERN MODES, OPTIMALLY NUMBERED, FOR FOUR COASTAL SOUTHERN CALIFORNIA STATIONS UTILIZING THE V-FOLD CROSS-VALIDATION ALGORITHM APPLIED TO K-MEANS CLUSTERING ANALYSIS

Charles J. Fisk *
Naval Base Ventura County, Point Mugu, CA

1. INTRODUCTION

Climatological wind variability is a meteorological element frequently evaluated in planning and decision-making activities in which weather conditions are of some importance. Wind rose diagrams, for example, depicting the most favored compass directions/ associated speeds, are routinely utilized for such purposes. Resultant wind statistics, providing distilled statistics like mean vector wind directions, speeds, and constancies, are informative also.

Potentially valuable as well should be information on the most prominent contiguous hourly wind patterns that occur climatologically over particular months or seasons. Such information should complement the more conventional, individual-hour focused approaches. In the same manner as there are favored individual hourly directions and related speeds, there should be preferred hour-to-hour patterns or "modes". Resolving modes of this kind could be considered a clustering problem, and K-means clustering analysis is a frequently used technique. One complication to K-means is that the number of clusters has to be guessed in advance, the ultimate choice of how many there "are" requiring a combination of trial-and error iterations and subjective judgment. Recent developments in the Data Mining field, however, have resulted in adaptation of the V-fold Cross-Validation algorithm, which incorporated into K-means allows for a less subjective determination of the likely number of patterns.

Utilizing this hybrid tool, identification of the "optimal" number of hour-to-hour diurnal wind patterns for the summer season June-August is performed for four Southern California coastal stations: North Island Naval Base (near San Diego), Long Beach FAA, Los Angeles International Airport (LAX), and Santa Barbara . Periods of record cover the late 1940's to 2010, inclusive. Data input consists of hourly derived u/v wind components and relative humidities, creating a potential clustering problem in 72-dimensional space. Relative humidity data is hypothesized to separate out patterns associated with fair or cloudy conditions. Upon generation of the clusters, the respective mean hourly u's and v's are recombined into hourly resultant wind statistics (direction, speed, and derived constancy values) to go with the associated mean relative humidities. From these, wind mode analyses for the four stations are conducted.

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

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).

As applied to K-means clustering, 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 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 from their cluster centers ("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 Southern California stations' data were normalized (an automatic software feature) to reduce them to a common scale and lessen the influence of outliers. For example, the scale of humidities (0 to 100 percent) was much larger than those of the u and v components (approximately an order of magnitude so), and failure to adjust for this would have produced results highly biased towards the humidities.

Also, the distance improvement cutoff threshold was reduced from the default 5 percent to 3 percent, promoting a larger final K. In this regard, generation of the "optimal" number of clusters is not completely automatic – the user can influence this by changing the threshold. Indeed, if there was no improvement increment cutoff at all, results could conceivably produce as many clusters as there are observations (assuming every individual observation was unique). Nature being as complex as it is, knowledge of the "real" number of clusters in a given application is probably an unattainable result; nonetheless, the V-fold cross-validation algorithm enhances the methodological objectivity of a clustering technique like K-means.

The Euclidean statistical distance option was selected to produce the clusters.

3. DATA AND PROCEDURES

The raw hourly data were downloaded, decoded, and processed from the National Climatic Data Center online site ("NCDC-online"). Only those individual daily series that had complete hour-to-hour sets of wind observations/ relative humidities were retained. Non-retentions were due to such factors as missing observations, or periods of record over the station history in which it was not operational the entire day.

With a large number of complete hourly data sets of winds and humidities in hand for the four stations, it was deemed preferable upon further consideration to "weight" the analysis to the "daytime" observations. Nighttime winds are mostly light and variable during summer over coastal Southern California, humidities generally high over a small range, and since daytime patterns are more variable and likely of more interest operationally, the contiguous-hour selection was arbitrarily pared down to 0600 LST to 1800 LST, leaving a 36 rather than 72 dimensional analysis to be performed. Resultant wind, mean scalar wind speed, and relative humidity statistics for the excluded hours would still be presented on a cluster-by-cluster basis alongside those for the daytime hours that were utilized to produce the clusters.

Figure 1 below is a map of the coastal Southern California area with the locations of the four stations included in this analysis labeled.

Figure 1. Geographic locations of Coastal Southern California Stations (North Island, Long Beach, Los Angeles International Airport or "LAX", and Santa Barbara) whose Summer (June-August) diurnal wind pattern modes are investigated in this study.

3. RESULTS

.

3.1. – *North Island Naval Base*

North Island Naval base (32.7 N, 117.2 W: elev: 26 ft), the southernmost station to be analyzed, is located

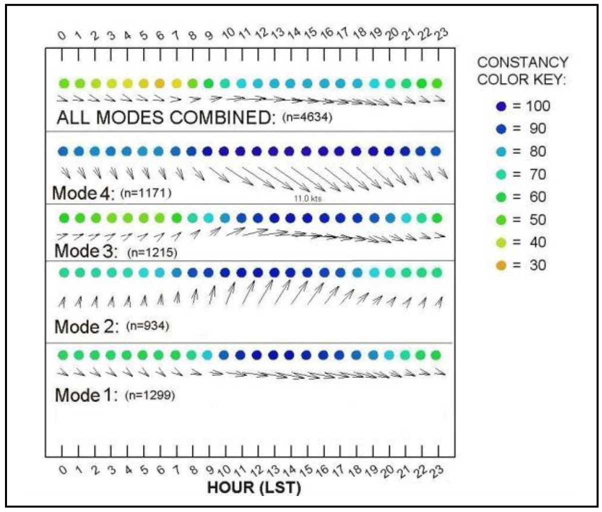


Figure 2 – Hourly Mean Vector Winds at North Island, June-August, for the four derived clusters and overall climatology

at the north end of the Coronado Peninsula, a few miles directly south of San Diego Lindbergh Field.

For the analysis, 1949-2010 data were examined, and 4634 intact observational sets were retained, or 81.2% of possible. The K-Means/V-Fold methodology produced four clusters (or "modes"), three of which seem to have a relatively straightforward physical interpretations.

Figure 2 above shows the clusters' hour-to-hour mean vector wind series, along with that for climatology ("ALL MODES COMBINED" section). All the vectors, section-by-section, are drawn to the same scale, the 11.0 knot labeling for 1300 LST in the Mode 4 section being the highest mean vector speed statistic produced. Constancy values are depicted by the color-coded circles directly above their respective vectors – blue representing the highest constancies or most persistent orientations. The constancies are also presented as line plots in Figure 4.

Mode 1 (lowermost section) has the highest case membership (n=1299 or 28.0% of the total). It exhibits relatively light magnitude northwesterly vectors for the early morning hours, shifting to more westerly or west-northwesterly ones for the afternoons. The mean vectors here also resemble climatology the most closely of the four clusters. The consistently onshore orientations, night and day, is apparently characteristic of summertime flow right at the coast in summer, the land-breeze essentially weak or absent. Mode 1 may represent a diurnal flow pattern of very local character, more or less free of synoptic influences.

Second most frequent pattern is Mode 3 (n=1215 or 26.2 % of the total), the most difficult of the four to interpret. The most prominent feature is not the winds, but the consistently low mean relative humidities compared to the other clusters and climatology (see orange trace in Figure 5). Vector wind directions for the late morning hours are slightly southwesterly at light magnitudes, although they turn westerly over the afternoon hours. Constancy values are comparable to those of Mode 1 at those times. It's possible that the low relative humidities, not the u and v properties, were the primary influence in producing this cluster,

Interestingly, if the cluster membership frequencies are evaluated on a month-by-month basis (See Table 1), Mode 1 shows a 30.8 % increase from June to August (n=364 to n=476), Mode 3 a 23.7% *decrease* (n=447 to n=341).

In third place is Mode 4 (n=1171 or 25.3 %), the vector directions and high magnitudes suggestive of Post-Frontal influences (or alternatively, unusually enhanced west-to-east surface pressure gradients). Orientations are exclusively northwesterly and sometimes north-northwesterly, with hour-to-hour constancies and mean scalar speeds (blue traces in Figures 3 and 4, respectively), the highest of the four clusters. The mean scalar speed disparity is especially so for the late-morning and early afternoon hours when a local sea-breeze enhancement occurs. Membership frequencies of this cluster increase 44.2 % over June to August (from n=317 to n=457).

Ranking fourth is Mode 2 (n=984 or 21.2%), incorporating wind observations from the Catalina Eddy phenomenon. According to the NOAA National Weather Service Glossary: "a Catalina Eddy (coastal eddy) forms when upper level large-scale flow off Point Conception interacts with the complex topography of the Southern California coastline. As a result, a counter clockwise circulating low pressure area forms with its center in the vicinity of Catalina Island. This formation is accompanied by a southerly shift in coastal winds, a rapid increase in the depth of the marine layer, and a thickening of the coastal stratus. Predominately these eddies occur between April and September with a peak in June".

From Figure 2, the North Island vectors are clearly oriented in this direction (southerly to south-southwesterly) at relatively high magnitudes and constancies (see red trace in Figure 4). The trademark presence of coastal stratus is indirectly reflected also by the comparatively high average hourly humidities (see red trace in Figure 5).

Catalina eddy vortexes are frequently dissipated in the afternoon hours by the more normal westerly or northwesterly onshore flow patterns, this reflected in Figure 4 by the decreased constancy levels at those hours (although the mean vector directions are still southerly – see Figure 1). The fact that Catalina eddies can be a predominantly morning feature might also be deduced from Figure 3 (see red trace), in which the mean scalar speed (5 knots) at 0800 LST is nearly an overall maximum for that hour; by mid-afternoon, however, the mean speeds are the lowest of the clusters. Mode 2 monthly frequencies decline 26 % from June to August (n=373 to n=276), consistent with the known falloff in frequencies from Spring into Summer.

As a last point, the pronounced increases and decreases in relative cluster frequencies over June to August seen in Table 1 was a somewhat surprisingly result, the mix of summer months' circulation patterns over time evidently less homogeneous than supposed. Indeed, however, a chi-square test of frequency uni-formities in Table 1 rejects the null hypothesis at the .001 level (chi-square= 72.635; 6 d. f.)

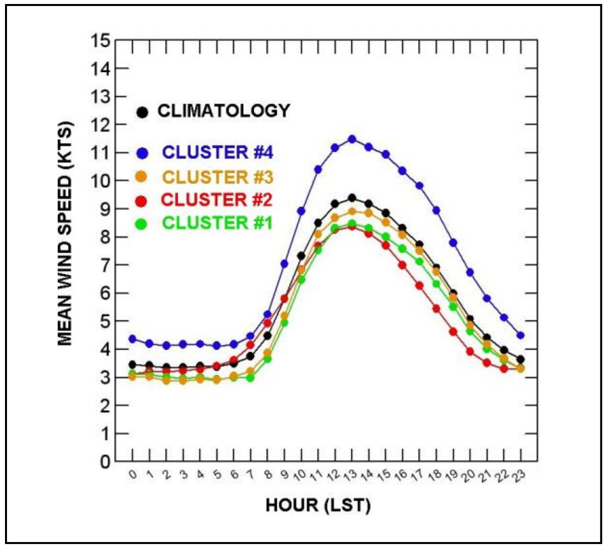


Figure 3 – Mean Hourly Scalar Wind Speeds for North Island Clusters, and Climatology (June-August)

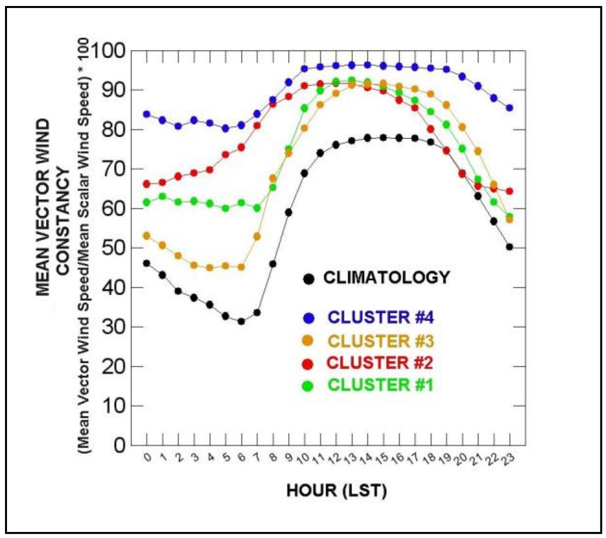


Figure 4 – Mean Vector Wind Constancies for North Island Clusters, and Climatology (June-August)

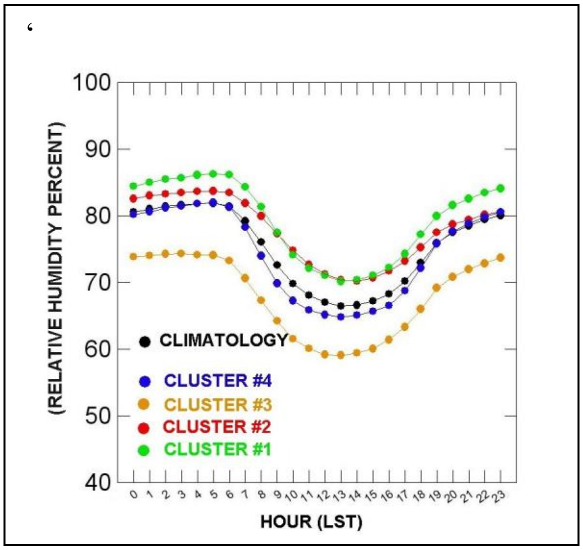


Figure 5 – Mean Hourly Relative Humidities for North Island Clusters and Climatology (June-August)

Frequencies

CLUSTER (rows) by MONTH (columns)

	JUNE	JULY	AUGUST	Total
CLUSTER #4	317	397	457	1171
CLUSTER #3	447	427	341	1215
CLUSTER #2	373	300	276	949
CLUSTER #1	364	459	476	1299
Total	1501	1583	1550	4634

Table 1 – Cross-Tabulation of North Island Cluster Frequencies, by Month

3.2. – *Long Beach FAA Airport*

Long Beach FAA Airport (33.8 N, 118.2 W – elev. 66 ft), located several miles from the Pacific Ocean, has its diurnal wind regime influenced significantly by the local topography such as the Palos Verdes Hills to the west/southwest (elevations approaching 1500 feet). Prevailing wind directions in summer exhibit a kind of morning to afternoon dichotomy – a local southerly sea-breeze seen frequently for the pre-noon to early afternoon hours, succeeded by a more westerly one (more regionally influenced?) later on.

Period of record for Long Beach was also 1949-2010, but due to more missing observations and the fact that it has been a 24-hour station less often than North Island, the number of complete, intact 24-hour wind and humidity cases was just 2373 or 41.6 % of possible.

The K-Means/V-Fold methodology again produced four clusters.

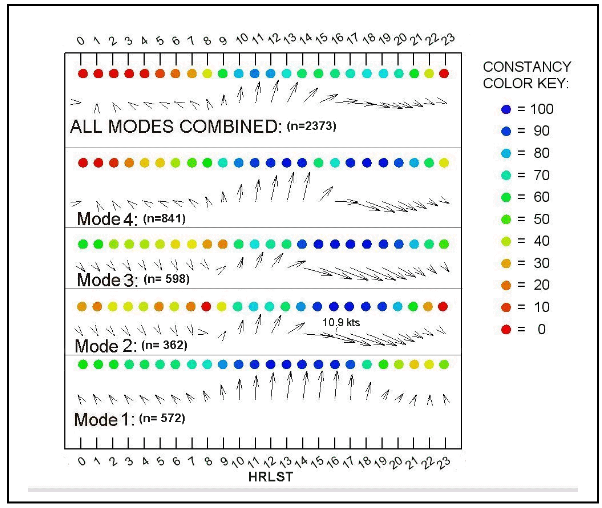


Figure 6. Hourly Mean Vector Winds at Long Beach , June-August, for the four derived clusters and overall climatology

Using the same format as North Island's, Figure 6 shows the Long Beach clusters' hour-to-hour mean vector wind series, along with that for climatology. Highest magnitude vector (10.9 knots) is the west-northwesterly one depicted in Mode 2 for 1600 LST.

Mode 4 (second section from the top) has the highest case membership (n=841 or 35.4% of the total). It shows a striking resemblance to climatology ("ALL MODES COMBINED" section). The mean vectors are south-southwesterly at increasing magnitudes for the early morning hours (showing blue or high constancy magnitudes as late as 1400 LST). Then as previously described, a changeover occurs to westerly, then west-northwesterly by late afternoon/early evening. During the peak transition hours (1500 to 1600 LST), constancy values drop to light blue levels, then recover again to dark blue as the westerly flow pattern becomes predominant. Inspecting the comparative mean wind speed and relative humidity charts (Figures 7 and 9, respectively), Mode 4's statistics (orange traces) match climatology's (black traces) very closely for most hours of the day, and examining the cross-tabulation of frequencies, by month (Table 2), Mode 4 displays no apparent trend.

Second most frequent pattern is Mode 3 (n=598 or 25.2 % of the total), its mean vector wind series showing some contrasts with Mode 4. The late-morning/early afternoon southerly-oriented vectors are less prominent (light blue rather than blue constancy colorings) and the shift to the westerlies occurs about two hours earlier; at their late afternoon peak magnitudes the vectors are also of higher magnitudes. Mean scalar speeds at 1500 and 1600 LST are about 2 knots higher as well (blue trace in Figure 7). Mode 3's frequencies exhibit a 45.2 % increase in frequencies over June to August (n=166 to n=241), suggesting perhaps that this may be a more seasonally influenced pattern.

Third in rank is Mode 1 (n=572 or 24.1 % of the total).

This appears to represent Long Beach's version of the Catalina Eddy. Mean vector orientations are strongly southerly from mid-morning through late-afternoon at high constancy (blue) magnitudes – no indications of any westerly changeovers in the mean. Further signature "tip-offs" come from the mean scalar wind speed chart (Figure 7), Mode 1's mean speeds (red trace) the highest of the clusters through mid-morning (1000 LST), but the lowest in the late afternoons/early evenings. From Figure 8, constancy magnitudes (red trace) are the highest of any of the clusters for every hour, inclusive, from 0100 to 1200 LST, and, from Figure 9, mean relative humidity values (red traces) are at cluster maximum values without exception from 0900 to 1800 LST, indirect indication of cloudier conditions (and lower temperatures). Finally, like that for North Island, occurrence frequencies (see Table 2) decrease from June through August, in this case 29.7% (from n=222 to n=156).

In fourth place is Mode 2 (n=362 or 15.3 % of the total). This cluster's distinctiveness stems from its much lower than average hourly humidities (green trace in Figure 9), average percentage figures in the 40's and high 30's for the late mornings and afternoons, compared to climatology's low 60's and 50's. Mode 2's mean vectors (Figure 6) resemble those of Mode 3 to a considerable extent (i.e. orientation and strength of vectors, time of changeover) so this seems to suggest that this may reflect episodes of clear skies and high temperatures (strong ridging episodes?). Also, mean wind speeds are at inter-cluster maxima levels for the hours 1400 to 1600 LST (Figure 7 – green trace) further suggestive of a stronger west-east pressure gradient. Mode 2 exhibits no apparent trends in frequencies from June to August (n=112 for June and n=114 for August in Table 2.

Overall, a chi-square test of frequency uniformities in Table 2 again rejects the null hypothesis at the .001 level (chi-square= 29.333; 6 d. f.)

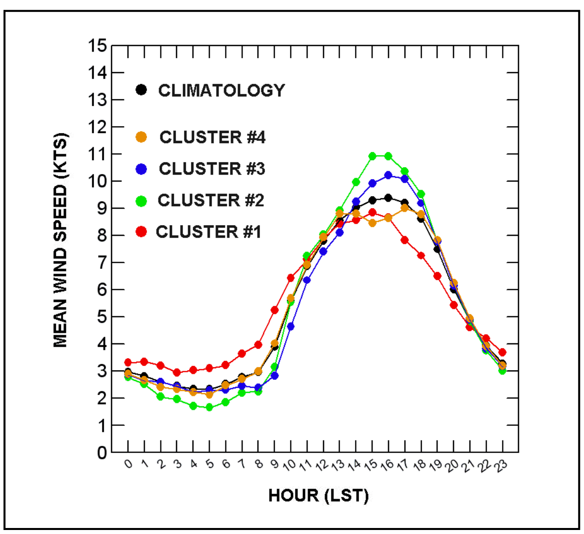


Figure 7. - Mean Hourly Scalar Wind Speeds for Long Beach Clusters, and Climatology (June-August)

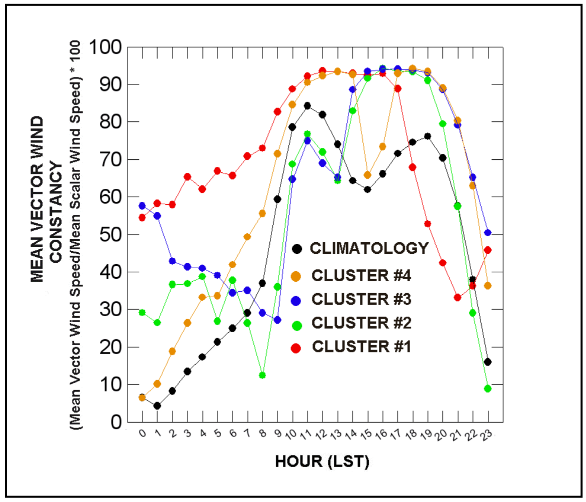


Figure 8 – Mean Vector Wind Constancies for Long Island Clusters, and Climatology (June-August)

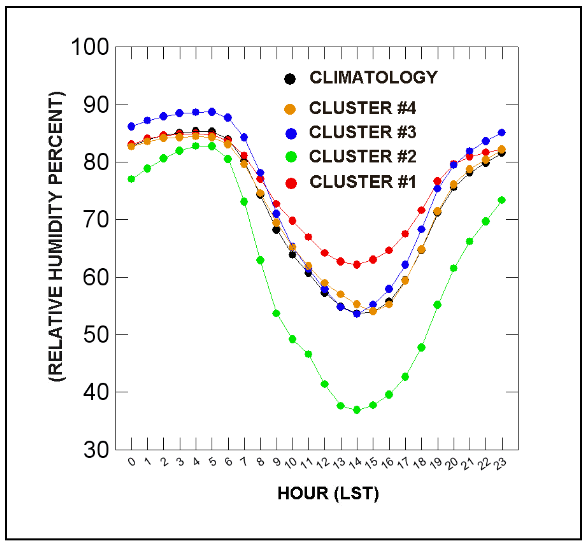


Figure 9 – Mean Hourly Relative Humidities for Long Beach Clusters, and Climatology (June-August)

Frequencies

CLUSTER (rows) by MONTH (columns)

	JUNE	JULY	AUGUST	Total
CLUSTER #1	222	194	156	572
CLUSTER #2	112	136	114	362
CLUSTER #3	166	191	241	598
CLUSTER #4	263	289	289	841
Total	763	810	800	2373

Table 2 – Cross-Tabulation of Long Beach Cluster Frequencies, by Month

.

3.3. – *Los Angeles International Airport (LAX)*

Los Angeles International Airport or "'LAX" (33.9 N, 118.4 W; elev: 66 ft) located at the westward edge of the Los Angeles Basin, about three miles east of the Pacific Ocean. Its location and the coastline orientation result in a predominately southwesterly onshore wind regime in daytime.

An hourly data set covering the period 1947-2010 was utilized for the analysis, and 4903 intact 24-hour observations of winds and humidities were extracted (83.7 % of possible).

The K-Means/V-Fold methodology produced three clusters (or "modes"), with mostly subtle distinctions.

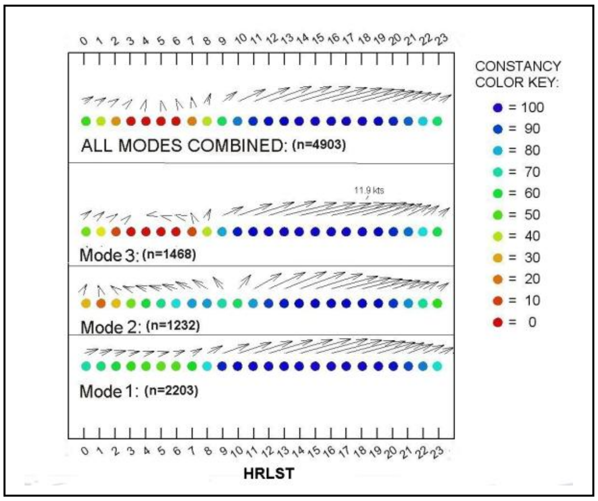


Figure 10. Hourly Mean Vector Winds at LAX, June-August, for the three derived clusters, and overall climatology

Mode 1 has the highest case membership (n=2203 or 44.9% of the total). It exhibits southwesterly oriented vectors for the whole 24 hours, light magnitudes prior to 0800 LST (still at green or blue constancy levels though), but relatively strong ones from late morning to late afternoon at dark blue levels (in the high 90's). From Figure 13, its mean hourly relative humidities (red trace) are the highest of the clusters for each hour, although mean wind speeds (Figure 11) are very close to climatology.

In second place is Mode 3 (n=1468 or 29.9% of the total). Its early morning vectors (0700 LST and before) are mostly light and variable, but from 1000 LST on they are very much like mode 1's in orientation, magnitude, and constancy; indeed, the magnitude of the 1500 LST vector (11.9 knots), is the highest individual statistic of the three clusters. Mode 3's mean scalar speeds (blue trace in Figure 11) are a bit less than Mode 1's from 0000 through 0900 LST (a knot lower at the greatest), higher from 1000 through 1800 LST (a knot higher at the greatest) but very close in mean magnitude thereafter. Relative humidities are significantly lower than the other two clusters', this parameter obviously being the one that was the greatest separating influence. Similar to Long Beach cluster #2, this may reflect high temperature episodes at LAX; the higher mean wind speeds in Figure 11 (blue trace) likewise indicative of a stronger west to east pressure gradient (see Figure 7 for Long Beach's mean wind speeds).

Third in rank is Mode 2 (n=1232 or 25.1 %). This cluster likely includes the Catalina Eddy cases, although given the orientation of the LAX area coast and its tendency to promote southerly-component flow in general, not exclusive to the Eddies. The hourly mean vector wind series in Figure 10 shows southeasterly orientations of not insignificant magnitudes for the hours 0600 to 0900 LST with the constancies values at moderate blue levels for 0700 to 0800 LST, atypically high for the time of day. Then, over 0900 and 1000 LST, the shadings curiously dip down to green levels as the vectors rotate to the southwest; Figure 12 (green trace) also depicts this constancy decrease, from the low 80's at 0700 LST to the low 60's by 0900 LST. This could be inferred as a statistical transition time between regimes of different wind character (e.g., Catalina Eddy to the more normal southwesterly onshore/sea-breeze flow), although Eddies doubtlessly persist in some form beyond 1000 LST in some cases.

The mean scalar wind speed chart (Figure 11) also shows a contrasting pattern compared to the other clusters – the green trace showing a curious bump at 0700 LST to ~ 5 knots, then actually decreasing slightly by 0900 LST, and remaining below the levels of the two other modes for the rest of the day. This probably also reflects the incidence of Catalina Eddies, in addition to those cases for whatever reason had reduced sea-breeze intensities (e.g., cloudy skies or reduced west to east regional pressure gradients).

While there are no discernable July-August trends present in Table 3, The chi-square test of frequency uniformities still rejects the null hypothesis at the .002 level (chi-square= 16.871; 6 d. f.). Non-uniformities, of course, could arise from features other than upward or downward June-August trends.

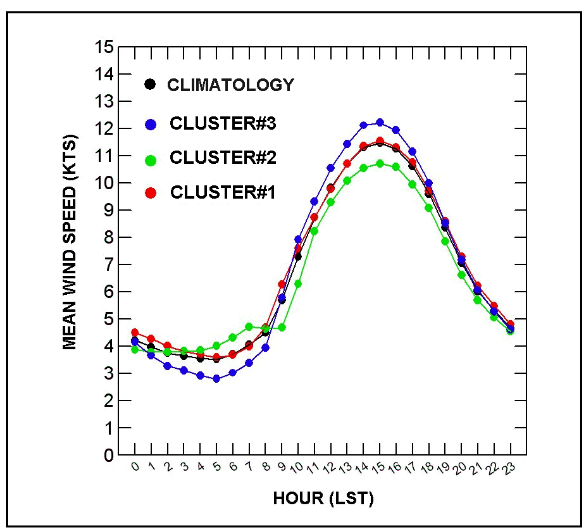


Figure 11. - Mean Hourly Scalar Wind Speeds for LAX Clusters, and Climatology (June-August)

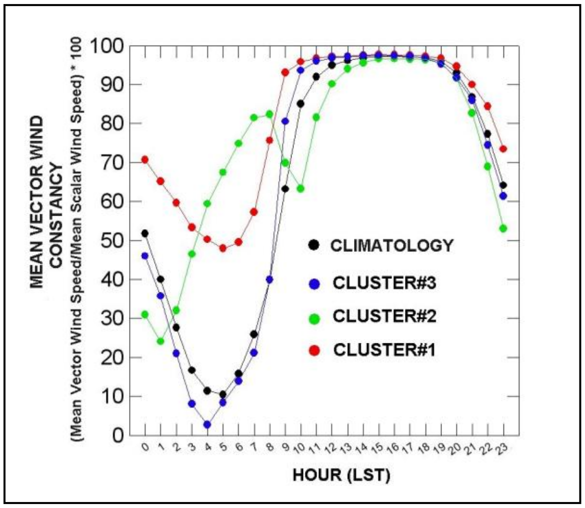
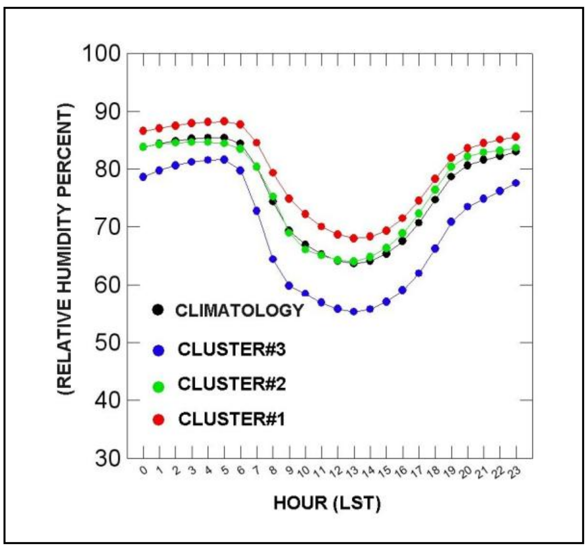


Figure 12 – Mean Vector Wind Constancies for LAX Clusters, and Climatology (June-August)



.
Figure 13 – Mean Hourly Relative Humidities for LAX Clusters, and Climatology (June-August)

Frequencies

CLUSTER (rows) by MONTH (columns)

	JUNE	JULY	AUGUST	
CLUSTER #1	670	785	748	2203
CLUSTER #2	489	472	507	1468
CLUSTER #3	453	386	393	1232
Total	1612	1643	1648	4903

Table 3 – Cross-Tabulation of LAX Cluster Frequencies, by Month

3.4. – *Santa Barbara WSO*

Santa Barbara WSO (34.4 N, 119.6 W; elev:7 ft), the northernmost station analyzed, is located on a coastline facing directly south; thus its typical sea-breezes normally have a strong southerly component.

An hourly data set covering the period 1950-2010 was utilized for the analysis, and 2739 intact 24-hour observations of winds and humidities were extracted (48.8 % of possible). Like Long Beach, it has been a non 24-hour station through much of its history.

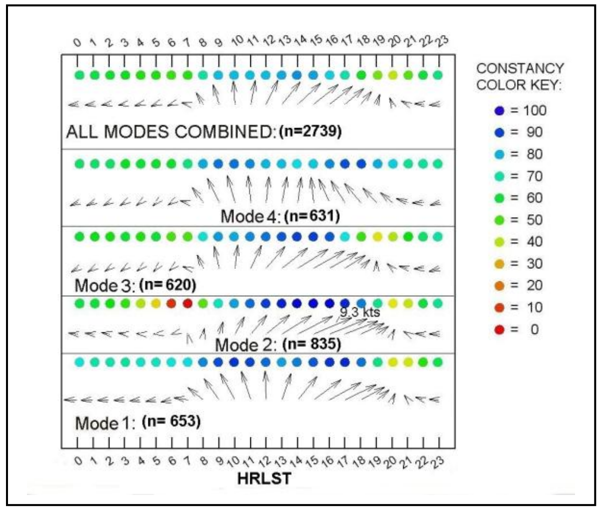


Figure 14.- Hourly Mean Vector Winds at Santa Barbara, June-August, for the four derived clusters and overall climatology

The clustering technique generated four clusters, and Mode 2 (n= 835 or 30.5% of the total) was the most prominent. The Figure 14 mean vector wind chart shows light easterly vectors through 0700 LST, with a subsequent turning to south and southwesterly at progressively higher magnitudes through mid-afternoon. The highest individual magnitude vector of all the clusters (9.3 knots) is observed for 1500 LST; constancy shadings are at dark blue levels from Noon through 1700 LST. Mean scalar wind speeds (green trace in Figure 15) show a relatively pronounced diurnal amplitude - mean speeds lower than the other clusters from 0800 to 1100 LST, inclusive, but higher for 1400 to 1600 LST. Average relative humidities (green trace Figure 17) are relatively high throughout the 24 hours, either matching those of Mode 1 (red trace), or slightly less.

Second in frequency is Mode 1 (n=653 or 23.8 % of the total). Its vectors start to pick up in strength a little earlier in the day than Mode 2, and are not nearly as southwesterly as the latter's. Mode 1's are southeasterly oriented from 0800 to 1100 LST, southerly or south-southwesterly through about 1400 LST, and not really full southwesterly until 1500 through 1800 LST. Also, Mode 1's constancies (red trace in Figure 16) take a dip from around 90 at 0900 LST (southeasterly at that time) to 80 at Noon, recovering to 90 again by 1500 LST (southwesterly at that time). This at least hints at some kind of wind pattern transition (Catalina Eddy to Sea-breeze?). Mode 1's mean scalar wind speeds (red trace in Figure 15) are generally higher than the other modes' figures through about 1000 LST, and its relative humidities (Figure 17), as already mentioned, vie with Mode 2 consistently for highest hour-to-hour cluster means.

Ranking third was Mode 4 (n=631 or 23.0% of the total). It's mean vector wind orientations (Figure 14) are either southeasterly or southerly, the former from about 0800 to 1000 LST, the latter from 1100 to about 1500 LST, and the former again from 1600 LST into the early evening hours when a dissipation to light easterlies occurs. Mode 4 also displays a dip in constancy values (orange trace in Figure 16), from the mid 80's at 0900 LST (southeasterly) to the high 70's at 1500 LST (southerly), recovering to around 90 for 1600 and 1700 LST (southeasterly). Mean scalar wind speeds for the mid afternoons are noticeably lighter (~ 2 knots lower – see orange trace in Figure 15) with relative humidities roughly matching or a bit lower than climatology for all hours of the day (orange trace in Figure 17). Mode 4 also seems to hint at some possible Catalina Eddy signatures, although the southerly orientation of the Santa Barbara coastline complicates this interpretation. The clustering algorithm, of course, cannot distinguish Catalina Eddy southerly-component flow from normal southerly-component sea-breeze flow, unless there are other clear physical interpretable distinctions - in this particular analysis, atypical diurnal variations in surface mean scalar wind speeds, constancies, and relative humidities.

In fourth place was Mode 3 (n=620 or 22.6 % of the total. Curiously, this mode's mean vector orientations resemble climatology's the most closely (see Figure 14). There is no indication of a wind regime transition (i.e., pronounced mid-day dips in constancy magnitudes), and the orientations are southwesterly with dark blue shadings from about Noon to 1600 LST. Mean scalar wind speeds (blue trace in Figure 15) match climatology's quite well throughout the whole day, but relative humidities (blue trace in Figure 17) are consistently below the other curves; hence, the likely origin of this cluster.

Finally, inspecting Table 4's frequency tabulations, Mode #2 exhibits a 61.9 % increase from June to August (n=210 to n=340), and Mode #4 a 29.7 % decrease (from n=256 to n=180). If the latter cluster is indeed the one that includes most of the Catalina Eddy cases, this would be consistent with the fact that the Eddies, as previously mentioned, decrease in frequency from the spring through the summer months. The chi-square test of frequency uniformities rejects the null hypothesis at the .001 level (chi-square= 54.953; 6 d. f.)

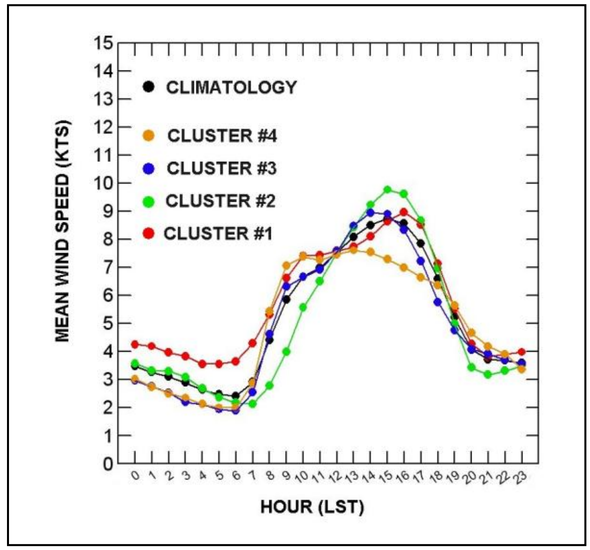


Figure 15. - Mean Hourly Scalar Wind Speeds for Santa Barbara Clusters, and Climatology (June-August)

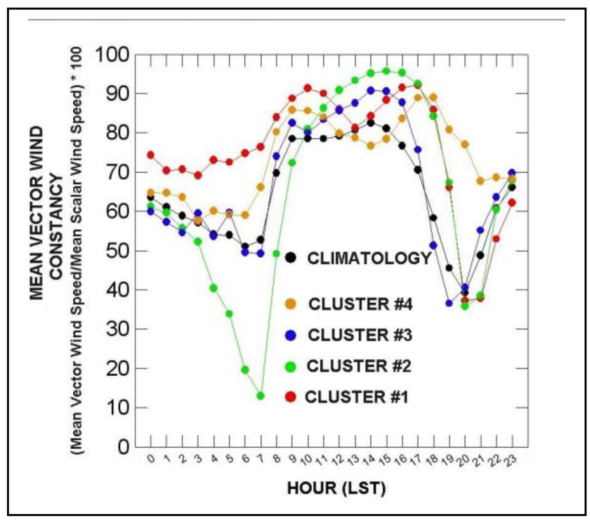


Figure 16 – Mean Vector Wind Constancies for Santa Barbara Clusters, and Climatology (June-August)

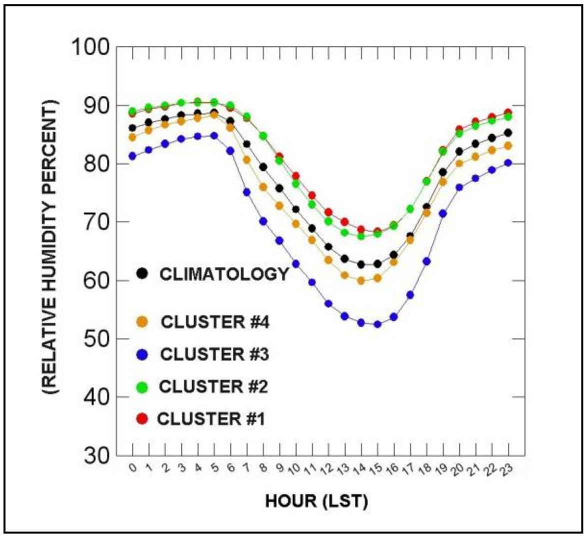


Figure 17 – Mean Hourly Relative Humidities for Santa Barbara Clusters, and Climatology (June-August)

Frequencies

CLUSTER (rows) by MONTH (columns)

	JUNE	JULY	AUGUST	Total
CLUSTER #1	201	240	212	653
CLUSTER #2	210	285	340	835
CLUSTER #3	234	191	195	620
CLUSTER #4	256	195	180	631
Total	901	911	927	2739

Table 4 – Cross-Tabulation of Santa Barbara Cluster Frequencies, by Month

4. SUMMARY AND CONCLUSION

Utilizing the data mining clustering tool K-Means, accompanied by the V-fold cross validation algorithm. The existence and identification of contiguous hourly wind patterns was explored four Southern California coastal stations (North Island, Long Beach, Los Angeles International Airport or LAX, and Santa Barbara) using approximately 60 years' data for months June to August. Input was normalized u/v wind components and relative humidities.

Results resolved four clusters (or "modes") for North Island, Long Beach, and Santa Barbara; and three for LAX. Likely signatures of the most well-known "alternative" wind mode, the Catalina Eddy, were seen for all four stations, interpretation more problematic, however, for the more southward facing ones in which normal sea-breeze (southerly) onshore flow is not that dissimilar from that of the Eddies. It should be emphasized that statistical modes are not necessarily physical modes.

In addition to station-by-station charts depicting mean vector winds, by hour of the day, by cluster, additional (line) graphs were presented portraying mean scalar wind speeds, vector wind constancies, and relative humidities. The line graphs, to some extent, provided additional interpretative insights.

The original intent of incorporating humidity data was to tease out "subtler" wind pattern modes, and it did so in some cases, namely the higher amplitude mean scalar wind speed curves for Long Beach and LAX associated with low humidities (mean vector wind direction contrasts in these cases, however, essentially negligible). These cases, however, reflected differences within the same physical type (diurnal sea-breeze flow).

Considering the restricted time of year (summer), the physical settings of the stations (coastal with a generally invariant sea-breeze/land breeze flow regime at this season), and small number of variables, as a first trial, the technique performed quite satisfactorily.

California has a number of different climatological wind regimes which should produce more interesting results if the scope is expanded to non-coastal stations for additional seasons with possibly more surface variables.

5. REFERENCES

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

www.weather.gov. *National Weather Service Glossary.* NOAA National Weather Service