IDENTIFICATION OF MIDNIGHT-TO-MIDNIGHT HOURLY CLIMATOLOGICAL WIND PATTERN MODES 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 an important element in planning and decision-making activities in which wind conditions are crucial, in addition to being an interesting topic for its own sake. Wind rose diagrams, for example, depicting the most favored compass directions and associated speeds for an individual or combined group of hours, allow for quick visual interpretations of climatological wind character. Resultant wind calculations, producing single-value statistics (mean vector wind direction and speed) distilled from many separate individual observations, are useful as summarization metrics. These can be depicted graphically as well.

Also interesting and potentially insightful should be information on the most prominent *contiguous* hour-to-hour wind patterns. Such data should be a logical extension to the wind rose and resultant wind methods, which by their individual period focus, are not intended in a direct way towards characterizing adjacent-hours' variability. In the same manner as there are favored individual hourly directions and related speeds, there should be preferred collective hourly patterns, or "modes". Resolving patterns of this kind could be considered a clustering problem, and K-means clustering analysis is frequently utilized for such problems. One drawback associated with traditional K-means is that one has to guess in advance the number of clusters, the ultimate choice of how many there "are" requiring trial-and-error iterations combined with subjective judgment. Recent advances, however, have resulted in adaptation of the V-fold Cross-Validation Algorithm, a "training-sample" type methodology, which incorporated into K-means allows for a more objective determination of the optimal number of clusters.

In a previous study (Fisk, 2012) using the K-Means/V-Fold methodology, the "optimal" number of diurnal wind pattern modes were resolved for four Southern California coastal stations (North Island, Long Beach, Los Angeles International Airport, and Santa Barbara), using ~ 60 years' data for the summer months (June, July, and August combined) and the hours 0800 to 1800 LST, inclusive. In most cases, results identified patterns (three or four modes) that were interpretable in a physically unambiguous way. Among these was the "Catalina Eddy" feature.

The present study, utilizing data for a single station (La Guardia Airport, New York) expands the scope to every hour of the day, utilizing data for the months January, April, July, October; and all months combined. In addition, for the latter (all months) group, utilizing ranked statistical distances from cluster centroids, a few of the most extreme individual 24-hour wind observation days in the historical record are presented.

The raw input wind data consist of hour-to-hour derived u/v wind components, creating a clustering problem in 48-dimensional space. Upon generation of the clusters, the respective mean hourly u's and v's (coordinates of the 48-D centroids) are recombined into hourly resultant wind statistics (direction, speed, and constancy values, the latter defined as mean vector wind speed divided by mean scalar wind speed times 100). Mean hourly temperature, relative humidity and scalar wind speed data, by hour/cluster accompany the mean vector wind output, serving as interpretative aids. These are not incorporated into the original clustering process.

The Southern California study's results showed that aside from those hours that had naturally light and variable winds (e.g., post-midnight nocturnal), or were preferred diurnal changeover times (e.g., land breeze to sea-breeze and vice-versa), the K-means/V-Fold method produced highly "distilled" mean vector wind directions, by hour/ cluster (i.e., directions with high constancy values – many figures in the 80's or higher), which in the practical interpretive sense rendered them as mean *scalar* wind directions/speeds.

A similar analysis on a more wind-variable station like LaGuardia, highly susceptible to synoptic-scale influences over most of the year with a wide range of wind directions around the compass (not to mention a sea-breeze), should conceivably be able disentangle its own mix of observations into clusters clearly differentiated in terms of hourly resultant wind statistics.

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

----

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

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 La Guardia Airport were normalized (an automatic software feature) to reduce them to a common scale and lessen the influence of outliers.

As the distance threshold can be changed, generation of the “optimal” number of clusters is not completely automatic. 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 is probably an unattainable result in most all applications; nonetheless, the V-fold cross-validation algorithm enhances the methodological objectivity of a clustering technique like K-means.

In the present study, the 5 percent default distance improvement cutoff threshold is retained with the Squared Euclidean distance metric chosen (default: Euclidean).

3. DATA AND PROCEDURES

Period of consideration was January 1949 thru December 2011, inclusive. The raw hourly data were downloaded, decoded, and processed from the Integrated Surface Hourly (“ISH”) data base, accessible 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, temperature, and relative humidities were retained (69%-76% of possible for the periods considered). Most of the incomplete digitized daily sets came from 1973 on.

3.1 – *The New York LaGuardia Station*

New York LaGuardia Airport (with equal applicability to New York City proper) is described in its 2011 LCD Annual Summary as being “close to the path of most storm and frontal systems which move across the North American Continent”. Continental influences “predominate” (most weather systems come from the west) but oceanic influences are “by no means absent” with local sea-breezes in summer (and presumably other seasons). Coastal storms “occurring most often in the Fall and Winter months” can produce bouts of heavy precipitation.

Figure 1 is a map of the New York City area with three major airports labeled:

Figure 1 – Map of New York City with Three Major Airports Numbered. LaGuardia is (2). *Source: Wikipedia*

4. RESULTS

4.1 – *January*

The ISH data base produced 1487 observation-days for January (76.1% of possible) with complete, intact 24-hour observations of winds, temperature, and humidity.

Figures 2 through 6 depict the mean vector wind results along with the supporting graphical analyses of hourly mean temperatures, humidities, wind speeds, and constancies.

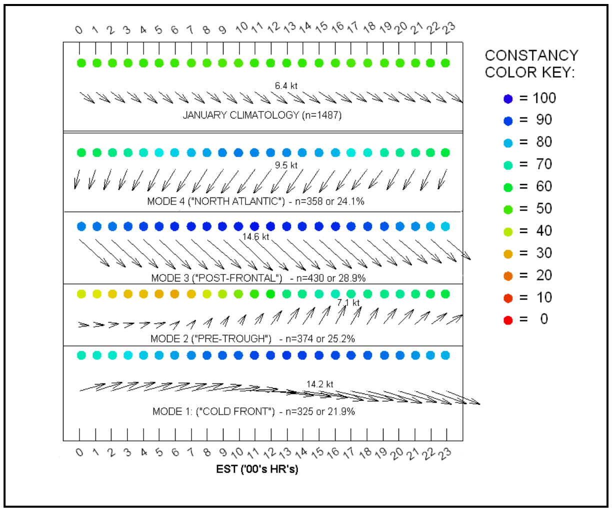

Figure 2 - January Contiguous Hourly Mean Vector Wind Modes (plus Climatology) for LaGuardia

Four modes were resolved by the K-means/V-Fold algorithm, and given subjective titles, in order of importance: "Post-Frontal" (28.9% incidence), "Pre-Trough" (25.4%), "North Atlantic" (24.1%), and "Cold Front" (21.9%). The contrasts in magnitudes are comparatively slight, not unexpected considering LaGuardia's proximity to fronts in January, and the fact that three of the four (with the possible exception of "North Atlantic") could be considered part of a synoptic event sequence.

Figure 2 depicts the hour-to-hour mean vectors for January Climatology (top panel) followed by those of the four individual clusters. The vector magnitudes are drawn to a common scale, this particular scaling to be retained for all the other vector charts presented in the following sections. In each panel, the hourly vector with the highest (longest) magnitude is labeled. Constancy statistics are represented above each vector as circles, and color coded according to the legend on the right. The mode numbers are those originally assigned by the software output.

The "Climatology" vectors (top panel) exhibit a persistent northwesterly orientation for all hours of the day, reflecting the net predominance of continental polar air masses at La Guardia during January. Constancy values are consistently in the 50's (see also black trace in Figure 6), reflecting the fact that mean vector wind speeds are about half as great as the mean scalar wind speeds of (typically about 11-12 knots – see black trace in Figure 5). The magnitudes, while modest, are actually the highest on a collective hour-to-hour basis for any individual calendar month at La Guardia (see Figure 34).

The most frequent pattern, Mode 3 or ("Post Frontal": 28.9% frequency) displays strong northwesterly vectors for all hours of the day. Constancy levels are at 85-95 levels almost throughout (blue trace in Figure 6), and mean scalar wind speeds (blue trace in Figure 5), through 1300 LST, are also the highest of the clusters, between 15-16 knots, surpassed only thereafter by the mean figures of "Cold Front". Mean hourly temperatures (blue trace in Figure 3) are easily the coldest, subfreezing throughout the day with a mean range of about 26 F to 30 F. Reflecting the dry character of polar air masses associated with this pattern, in spite of the low mean temperatures, mean hourly humidities (blue trace in Figure 2) are consistently the lowest as well, except for the early evening hours through 1100 LST, when they are roughly equivalent to those of "Cold Front".

Next in importance (Mode 2 or "Pre-Trough": frequency 25.4%) exhibits much lighter magnitude, southwesterly vectors, maximum magnitudes reached for the afternoon hours associated with constancies in the high 60's to low 70's (brown trace in Figure 5). Mean hourly temperatures (brown trace in Figure 2) are distinctly the warmest of the clusters, approaching the mid 40's F in the early afternoon, relative humidities (brown trace in Figure 3) noticeably higher in the mean compared to all except "North Atlantic". Mean hourly wind speeds (brown trace in Figure 4) are consistently the lightest.

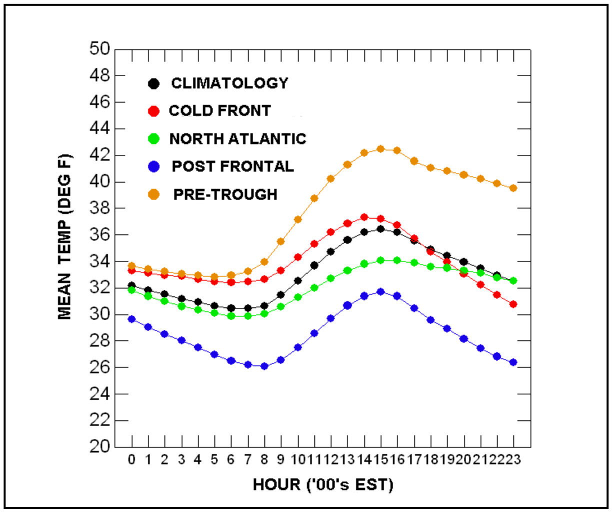

Figure 3 – Mean Hourly Temperatures, by Cluster and for Climatology, for LaGuardia - January

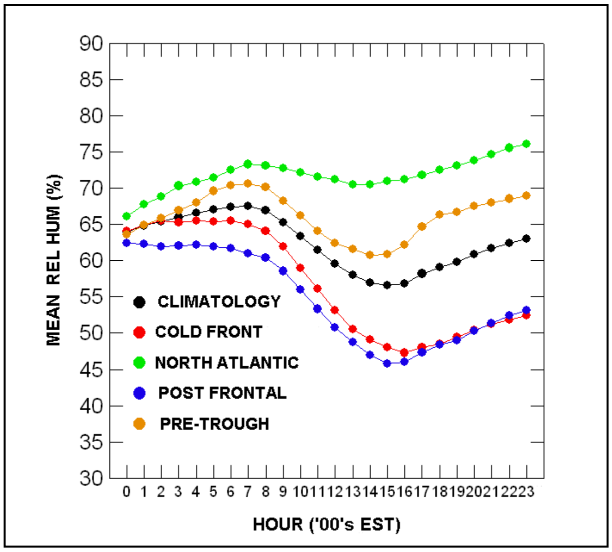

Figure 4 – Mean Hourly Relative Humidities, by Cluster and for Climatology, for LaGuardia - January

Third in ranking is Mode 4 or "North Atlantic" (frequency 24.1%), displaying northeasterly vectors throughout the day at constancies as high as the mid 80's (green trace in Figure 6) for the mid-morning hours. Mean temperatures (green trace in Figure 3) show the least diurnal range of the clusters (about 30 F to 34 F), coupled with the highest humidities (green trace in Figure 4). Mean wind speeds are below climatology for all hours of the day (green trace in Figure 5), but higher than those of "Pre-Trough". Northeasterly wind episodes are of course well-known in the New York and surrounding areas for the cool, damp, and cloudy conditions associated with "Backdoor cold fronts", and occasional storms, such as "Nor'easters".

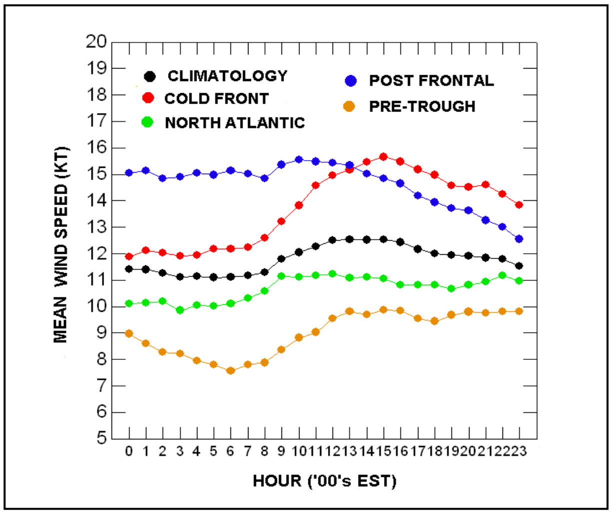

Figure 5 – Mean Hourly Scalar Wind Speeds, by Cluster and for Climatology, for La Guardia - January

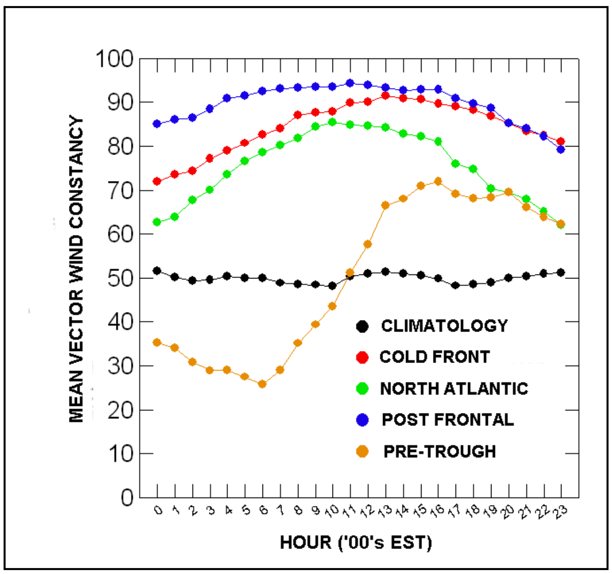

Figure 6 – Mean Vector Wind Constancies, by Cluster and for Climatology, for La Guardia - January

Fourth is Mode 1 or "Cold Front" (frequency: 21.9%). This pattern's vector orientations consist of a progressive turning from west-southwest to westerly and then to strongly northwesterly, an idealized cold-frontal passage. Constancies (red trace in Figure 6) increase from the 70's to as high as near 90, matching and even slightly exceeding those of "Post Frontal" for the hours just prior to midnight. Mean temperatures (red trace in Figure 3) are slightly higher than Climatology (black trace in Figure 3) into the early afternoon, but as the front "passes", show a relatively sharp drop by late evening. Reflecting the influx of drier, colder air, relative humidities (red trace in Figure 4) then fall significantly to roughly "Post Frontal" levels by evening. Characteristic of cold frontal passages, mean wind speeds pick up markedly after Noon LST (red trace in Figure 5) approaching 16 knots at 1400 LST.

4.2 – *An Exploratory Cluster Analysis of a Cluster*

At the outset it was explained that a hypothesized advantage of utilizing a clustering analysis on u and v components was that enough of a mean vector "distillation" should take place such that results could be idealized/described, essentially, as a 24-hour array of mean scalar wind directions/speeds. Inspecting Figure 2, this seems to be largely the case, the individual clusters showing frequently blue (high) constancy magnitudes, especially in the afternoons and evenings, certainly better than overall Climatology. On that note, the contrasts with Climatology will become even more apparent for the other calendar periods considered.

One noticeable exception for the January modes was the very low constancies (orange and yellow shadings) shown for "Pre-Trough", especially for the pre-sunrise hours. As troughs themselves are not diurnal features, perhaps the cluster so-named "Pre-Trough" was in reality a mixture of other more infrequent or indistinct patterns not isolated by the K-Means/V-Fold method, given the settings in place.

To further investigate, one could rerun the cluster analysis with a lower cutoff threshold (encouraging more clusters) or alternatively, perform a cluster analysis of the "Pre-Trough" cluster itself. Since the other three clusters showed satisfactory results, the second alternative is selected as an exploratory exercise in "clustering a cluster".

Figure 7 below shows the results. Four sub-clusters were generated, all with green to blue constancy shadings, those for mode 2b and mode 2d with equal incidence frequencies (7.7% relative to the entire January data set). Mode 2b displays southwesterly vectors throughout with lessening magnitudes, however, from the early afternoon on. For perhaps lack of a better title, it is given another "Pre-Trough" designation. Mode 2c exhibits a diurnal changeover in vector orientations from northwesterly to southwesterly, hence the designation: "Warm front". Mode 2d (overall incidence: 5.9%), shows a late-day shift of orientations to westerly/northwesterly with increased constancies, hence the designation: "Cold Front". Finally, Mode 2a (frequency 3.9%) is designated "Sea-Breeze", showing light and variable vectors for the morning hours (but with constancy levels at modest green levels), followed by a shift to southerlies at blue, sometimes dark blue, constancies for the afternoon hours. This signature will be seen at more prominent frequency levels for the later calendar periods analyzed. The meager "Sea-Breeze" frequency for January could be attributable to the shorter days, low sun angle, and the stronger synoptic circulation patterns of this time of year which would tend to obliterate it.

Returning to Mode 2b again, inspection of the Figure 1 map shows somewhat of a southwest to northeast

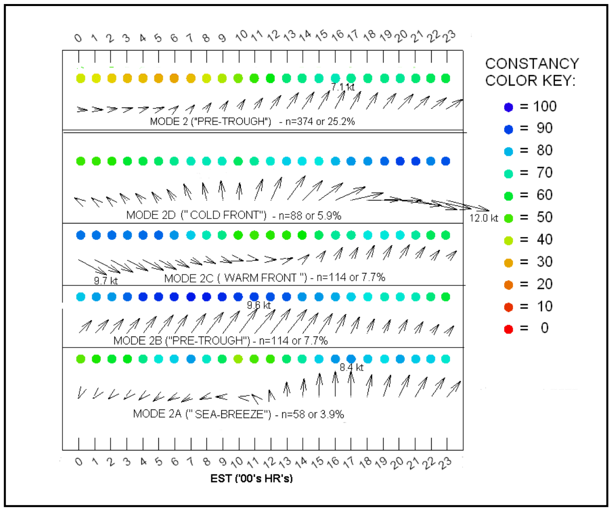

Figure 7 - January "Pre-Trough" Category Sub-clusters for LaGuardia

orientation of the area topography. Possibly this might play a subtle role in enhancing southwesterlies (channeling?), when larger scale synoptic influences are similarly and favorably directed.

4.3 – *April*

The ISH data base produced 1355 observation-days for April (71.7% of possible) with complete, intact 24-hour observations for the parameters of interest. Five modes were resolved and given the designations: "Sea-Breeze" (25.1% incidence), "Post-Frontal" (23.8%), "Pre-Trough" (17.6%), "North Atlantic" (17.0%), and "Cold Front" (16.5%). Except for the new "Sea-Breeze" cluster, all the designations are repeats from January.

Figure 8 depicts the mean vectors for April Climatology (top panel) followed by those for the five modes. Reflecting the highly variable wind character for April, the hourly Climatology vectors are very weak northwesterly for the morning thru early afternoon hours, weakly westerly through early evening, and then weakly northwesterly again for the balance of hours. Constancy values are very low (black trace in Figure 12); frequently less than 20, and for the evening hours near 10.

The "Sea-Breeze" pattern (Mode 5), discernable only (and just barely) from the January sub-cluster analysis, is the most "prominent" April mode, still encompassing, however, just one-fourth of the total observations. Mean vectors are light northerly at constancies generally in the 60's through late morning (yellow trace in Figure 12), but showing a major shift in direction over 1200 to 1300 LST (Figure 8), accompanied by a ~10 constancy statistic for 1300 LST. Thereafter, they become south-southeasterly at increasing magnitudes through 1800 LST, at which time the constancy attains an 82 value (yellow trace in Figure 12) The relatively late time of day at which the mean sea-breeze predominates statistically could probably be attributed to the intervening topography of Long Island, of which the flow has to traverse.

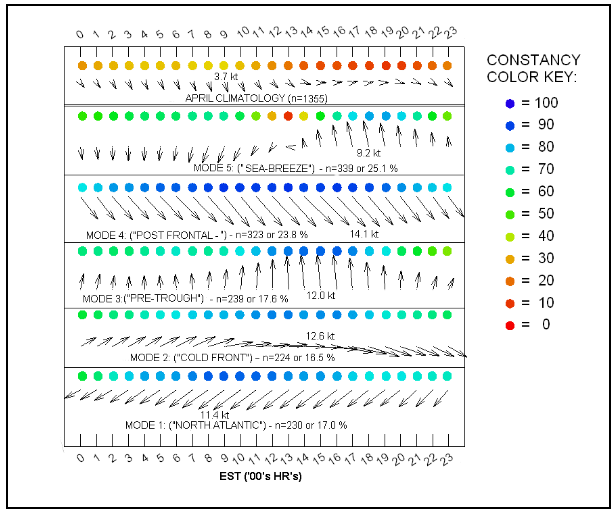

Figure 8 - April Contiguous Hourly Mean Vector Wind Modes (plus Climatology) for LaGuardia

The mean hourly temperature curve (yellow trace in Figure 9) shows a fairly sharp drop relative to Climatology after 1400 LST, having reached its peak about the same time as the latter, however. Relative humidities (yellow trace in Figure 10) also show a sharp increase after about 1400 LST, and mean hourly wind speeds (yellow trace in Figure 11), easily the lightest of the clusters over the mid-morning to late afternoon hours, exhibit a very late maximum (~1800 LST).

Second most important pattern, Mode 4 or "Post-Frontal" (incidence: 23.8% - down from January's 28.9% figure) again exhibits northwesterly vectors of high magnitudes and constancies for all hours. Mean hourly temperatures (blue trace in Figure 9) are the coldest of the clusters for the morning hours, higher only those of damp and cloudy "North Atlantic" in the afternoon. Mean relative humidities (blue trace in Figure 10), however, are at minimum levels for all hours, again reflecting the dry character of the cold air masses having just passed through. Mean wind speeds and constancies (blue traces in Figures 11 and 12, respectively) are also at cluster maxima levels for all hours of the day.

In third place is Mode 3 or "Pre-Trough" (incidence: 17.6% -- down from January's 25.2%). The mean vectors, in particular, show a south to slightly south-southeasterly bent at high magnitudes and constancies during the afternoon hours. Mean hourly temperatures (brown trace in Figure 9) are the highest of the clusters from about 0700 to 1300 LST, roughly matching that of "Cold Front" into early evening when the latter's shows a faster decline. Mean relative humidities (brown trace in Figure 10) show a steep rise in the afternoons, consistent with advection of moister air but without an accompanying decline in temperatures that would accompany a sea-breeze (and would be visible in Figure 9). Mean late morning to mid-afternoon mean scalar wind speeds (brown trace in Figure 11) are slightly higher than that of Climatology, but less than those of "Post Frontal" and "Cold Front". Vector wind constancies (brown trace in Figure 12) for the hours 1300 to 1500 LST approach the 90 mark, just short of the maximum levels for "Post Frontal".

Fourth in importance is Mode 1 or "North Atlantic" (incidence: 17.0% - down from January's 24.1%). This mode's vectors exhibit the characteristic northeasterly orientation along with comparatively moderate to high magnitudes relative to the other modes. The mean hourly temperature curve (green trace in Figure 9) shows the "flattest" diurnal amplitude of any of the clusters, consistent, of course, with the damp, cloudy and occasionally rainy conditions associated with this pattern. Mean relative humidities (green trace in Figure 10) are easily the highest of the clusters hour-by-hour, for obvious reasons. Mean scalar wind speeds (Figure 11), curiously, reach their highest levels at mid-morning, with constancy levels (Figure 12) also at their highest levels, just short of Post-Frontal's maxima.

Ranking fifth is Mode 2 or "Cold Front" (incidence 16.5% - down from January's 21.9%). Replicating January's "Cold Front" configurations, the vectors display a counter-clockwise turning over the course of the day from southwesterly to northwesterly, constancy magnitudes reaching their highest levels around Noon LST when the orientations are westerly. Hourly mean temperatures (red trace in Figure 9) match closely those of "Pre-Trough" for much of the day, exhibiting a sharper decline, however, for the evening hours, consistent with a frontal passage. Mean relative humidities (red trace in Figure 10) also match those of "Pre-Trough" closely through the early afternoon (~1400 LST), but depart significantly thereafter as the latter's rise sharply. Also consistent with a "mean" frontal passage, mean scalar wind speeds (red trace in Figure 11) show a marked increase in the early afternoon, matching or falling just short of those for "Post-Frontal" (blue trace) over the balance of hours. Constancy values (red trace in Figure 12), reach mid-80's levels in the early afternoon, just following the frontal "passage"

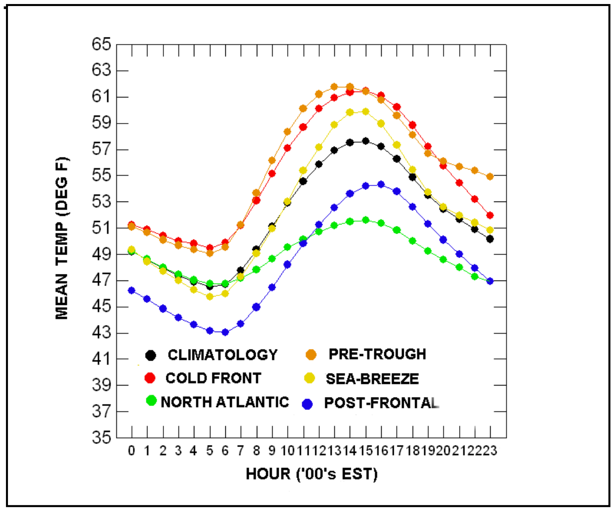

Figure 9 – Mean Hourly Temperatures, by Cluster and for Climatology, for LaGuardia – April

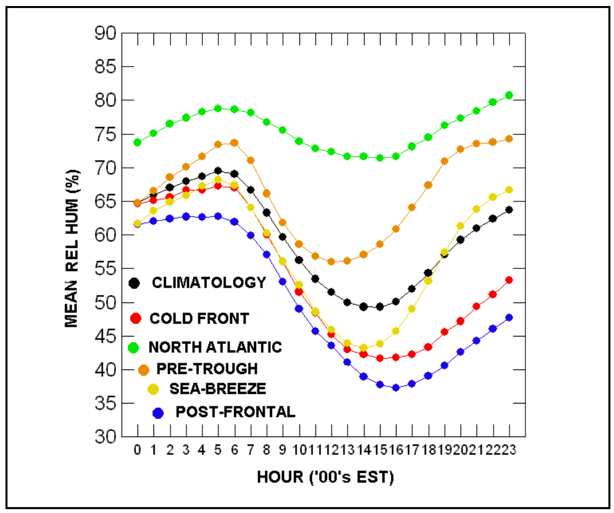

Figure 10 – Mean Hourly Relative Humidities, by Cluster and for Climatology, for LaGuardia - April

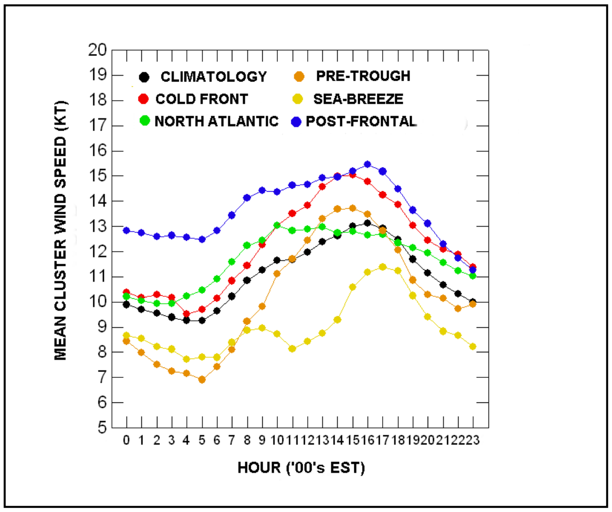

Figure 11 – Mean Hourly Scalar Wind Speeds, by Cluster and for Climatology, for La Guardia - April

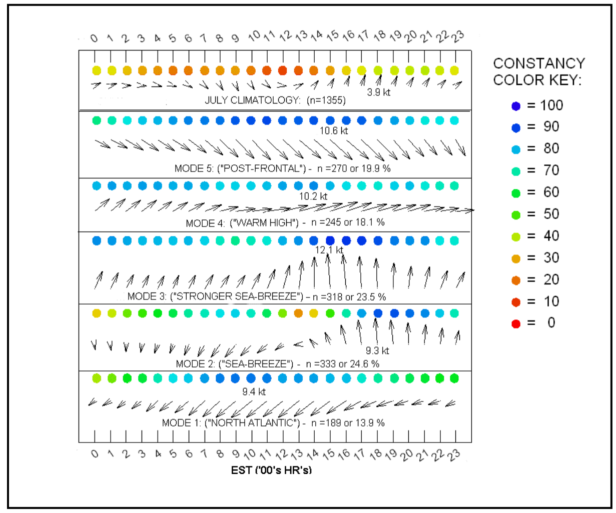

Figure 13 - July Contiguous Hourly Mean Vector Wind Modes (plus Climatology) for LaGuardia

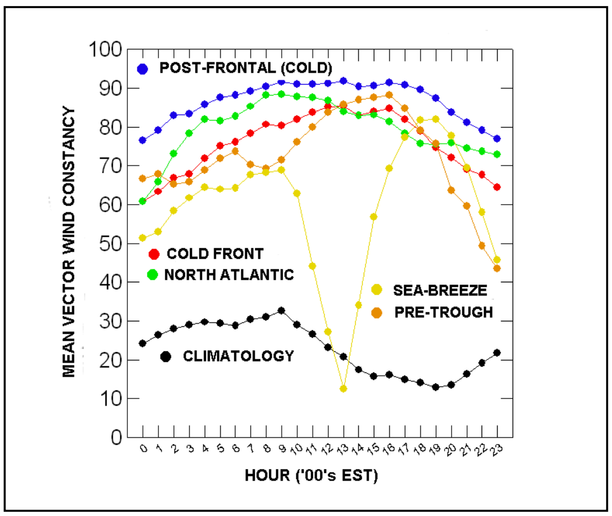

Figure 12 – Mean Vector Wind Constancies, by Cluster and for Climatology, for La Guardia - April

4.4 – *July*

The ISH data base produced 1355 observation-days for July (69.4% of possible) with complete, intact 24-hour observations of winds, temperature, and humidity.

Five modes were again created, but with circulation patterns of height-of-summer July being somewhat different than transition-month April, as evidenced by the cluster-by-cluster vector orientations, a few changes were made to titles, "Stronger-Sea-Breeze" and "Warm-High" replacing "Pre-Trough and "Cold-Front". These of course were subjective, imperfect, and probably oversimplified, but expedient to the task of assigning *some* label that possibly captured the overall character of the pattern.

The July designations, in order of importance were "Sea-Breeze" (24.6% incidence), "Stronger Sea-Breeze" 23.5%), "Post-Frontal" (19.9%), "Warm High" (18.1%), and "North Atlantic" (13.9%). Figure 13 depicts the vector winds for July Climatology (top panel) followed below by those for the five modes.

For Climatology, the vectors are at weak magnitudes for all hours of the day, mostly light westerly for the early hours after midnight, light southwesterly for those from mid-afternoon on, and a bit of a mix in-between. Constancy values (black trace in Figure 17), range mostly in low 30's to the 20's, but 15 is noted for Noon LST.

The "Sea-Breeze" pattern (Mode 2), is again the most predominant of the five (incidence: 24.6%, down slightly from April's 25.1%). Very similar to that of January and April, it displays the light north to northeasterly vectors for the pre-noon hours, essentially, followed by a shift to progressively stronger southerlies, reaching maximum magnitudes and constancies late in the afternoon, around 1800 LST. In the Mean Hourly Temperature chart (Figure 14), this event sequence is reflected by a slightly steeper decline in temperature over the evening hours (green trace). Mean hourly relative humidities (green trace in Figure 15) show a correspondingly larger increase over these hours, reflecting a combination of the greater mean cooling and moisture advection from the sea-breeze. Like that for April, "Sea-breeze" mean scalar wind speeds (green trace in Figure 16) are significantly lighter than those for the other clusters, the only July exception being late in the day when the mean speeds slightly exceed those of "North Atlantic". The constancy curve (green trace in Figure 17) shows a dramatic dip to around 20 at 1300 LST, reflecting the sea-breeze changeover.

Second most important pattern, Mode 3 or "Stronger Sea-Breeze" (23.5% incidence) is a new category. Inspection of Figures 13-17 seems to indicate that it is a stronger version of Mode 2, created by virtue of its southwesterly oriented vectors for the mornings (in contrast to the light north and north-easterlies for Mode

2) and the high magnitude south-south-easterlies for the afternoons. This may represent synoptic situations in which local sea-breeze flow is being superimposed on a more regional flow of the same general direction. Also possibly in these instances, the overall flow (local and regional) is oriented more from the ocean rather than over the comparatively rough, by comparison, land surface of Long Island. A look at Mode 3's mean temperature curve (blue trace in Figure 14) shows that while temperatures are warmer than climatology over the pre-noon hours after sunrise, they start to decline sooner – at least a couple of hours so before the other modes' decreases. Mean relative humidities (blue trace in Figure 15) also start their diurnal increase sooner (about Noon), running at levels about 10 % higher than climatology for the rest of the day. Mean hourly wind speeds (blue trace in Figure 16) start increasing rapidly at about 1000 LST, reaching the highest July mean values of any of the clusters over the hours 1400 to 1700 LST, including mean speeds in excess of 13 knots for 1500 LST and 1600 LST. Finally, Constancies (blue trace in Figure 17) show a curious dip at 0900 LST to the high 60's (after readings a few hours previous around 80) , steadily increasing thereafter, however, to low 90's figures by late afternoon, the highest individual readings of the July clusters. Perhaps this dip, albeit much less pronounced than that seen for the other sea-breeze mode, is a subtle earlier in the day changeover signal expressing the onset of "Stronger Sea breeze" flow.

In third place is Mode 5 or "Post-Frontal" (incidence: 19.9% -- down from April's 23.8%). This displays the characteristic northwesterly vectors and high constancies seen in previous related charts connected with conditions following a frontal passage. Mean hourly temperatures (yellow trace in Figure 14) are not much different than climatology except being actually a little warmer for the late afternoons. Mean relative humidities (yellow trace in Figure 15) are the lowest of the clusters for every hour of the day, implying the advection of drier (if not necessarily cooler) air. Mean scalar wind speeds (yellow trace in Figure 16), are at or near cluster maximum values for the forenoon and evening hours, and constancy values (yellow trace in Figure 17) are relatively high for all hours of the day compared to the other clusters.

From the above, "Post-Frontal" conditions, as so defined, appear to have no net cooling impact on temperatures in July. Possibly the associated drier and less hazy/cloudy conditions allow for more insolation and afternoon heating. There also may not be much thermal contrast between air masses either side of fronts at season as well.

In fourth place is Mode 4 or "Warm High" (incidence: 19.1%). This mode is so-named because of the high hourly mean temperatures associated with it (brown trace in Figure 14), mean hourly diurnal temperatures ranging from about 74 F to 88 F, compared to climatology's 70 F to 83 F (black trace). The mean vectors in Figure 13 are mostly southwesterly at comparatively moderate magnitudes, but with nonetheless high constancies (brown trace in Figure 17), those for the hours 0500 to 1300 LST at the highest levels for any cluster – in the 80's (brown trace in Figure 17). Mean relative humidities (brown trace in Figure 15) are noticeably less than climatology for the afternoon and evening hours (attributable at least in part to the higher temperatures), and surpassed in this regard only by the "Post-Frontal" figures. Mean scalar wind speeds (brown trace in Figure 16) typically run a little higher than climatology, in particular, displaying means second in high magnitude only to that of "Stronger-Sea-Breeze" over the mid-afternoon hours. Observations in the "Warm High" category probably reflect circulation patterns that slowly advect air from a more continental origin to the southwest, with some possible down-slope warming effects involved.

Fifth in importance is "North Atlantic" (incidence: 13.9% - down from April's 21.9%). Again, this mode's vectors exhibit the characteristic northeasterly orientations, although at lesser magnitudes than April - still with morning maxima magnitudes, however. To go with damp, cloudy, and occasionally rainy, mean hourly temperature curve (red trace in Figure 14) shows strikingly cooler than average readings for the afternoon hours - no higher than the mid 70's. Mean relative humidities (red trace in Figure 15), to no great surprise, are at maximum cluster levels for each hour. Mean scalar wind speeds (red trace in Figure 16) vye with "Post-Frontal" for highest levels for most of the morning hours, but decrease rapidly after mid-day, and are at cluster-lowest levels for the hours 1700 to 2100 LST, inclusive. A feature also seen for April, "North Atlantic", constancy levels (red trace in Figure 17 are at diurnal maxima levels for the morning hours (in the 80's).

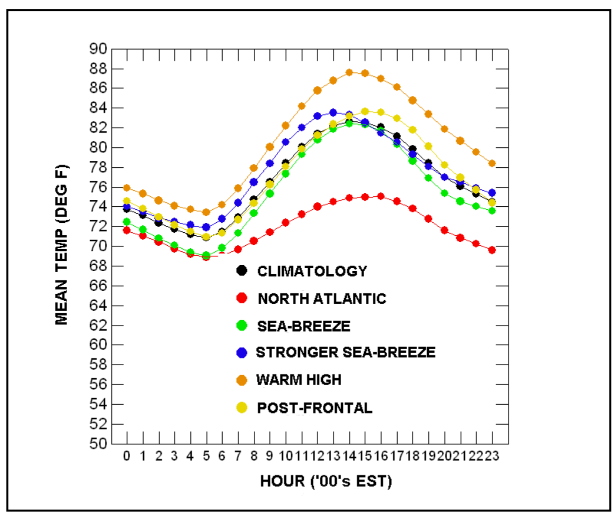

Figure 14 – Mean Hourly Temperatures, by Cluster and for Climatology, for LaGuardia - July

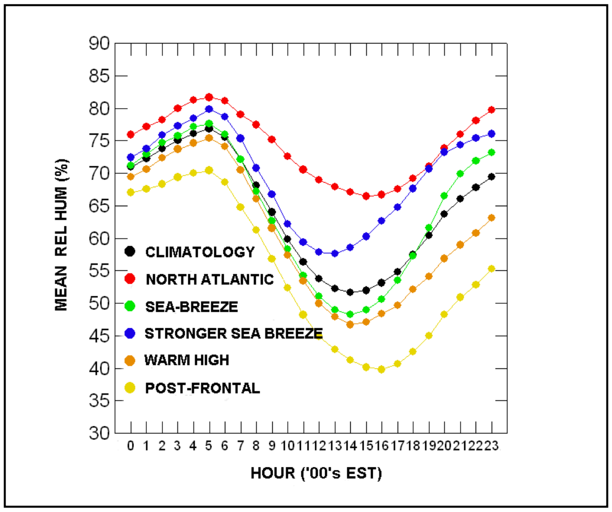

Figure 15 – Mean Hourly Relative Humidities, by Cluster and for Climatology, for LaGuardia - July

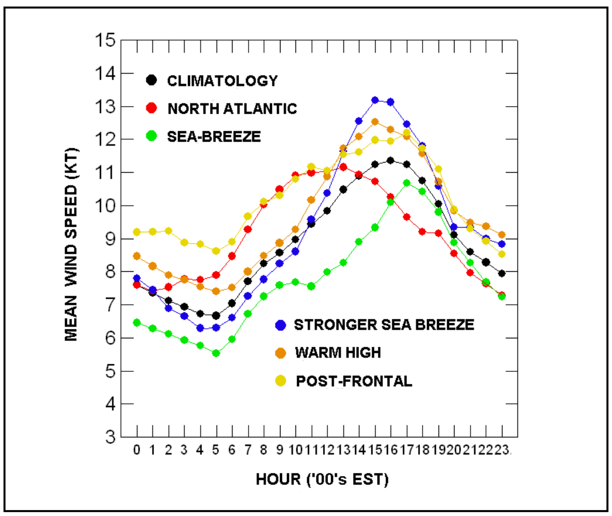

Figure 16 – Mean Hourly Scalar Wind Speeds, by Cluster and for Climatology, for La Guardia – July

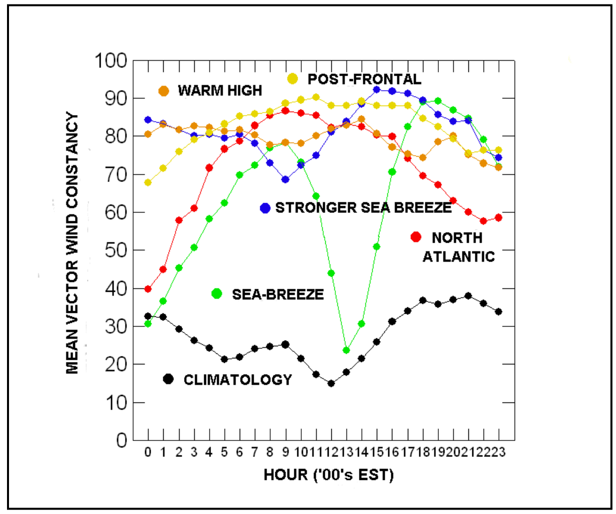

Figure 17 – Mean Vector Wind Constancies, by Cluster and for Climatology, for La Guardia - July

4.5 – *October*

The ISH data base produced 1446 observation-days for October (74.0% of possible) with a full complement of winds, temperature, and humidity.

Five clusters were once again generated by the K-Means/V-Fold algorithm, but October being a seasonal transition month, the panel-by-panel mean vector arrays differed noticeably in some instances from those of July, prompted another change in labeling. In this instance, the “Warm High/Pre-Trough” and “Cold-Front” designations replaced “Stronger Sea-Breeze” and “Warm-High”

Incidence percentages, in rank-order were as follows: “Warm High/Pre-Trough” (21.65% incidence), “Sea-Breeze” (21.58%), “North Atlantic” (19.4%), “Cold Front” (18.74%), and “Post Frontal” (18.67%). The spread in percentages is the least of the four months.

Figure 18 depicts the hourly mean vectors winds for October Climatology (top panel) followed by those for the five modes. The Climatology vectors, once more, are at weak magnitudes throughout, mostly light northwesterly from midnight through mid-afternoon (1500 LST), becoming more westerly for the remaining hours of the day. Constancy values (black trace in Figure 22), range from the low 30’s for the early morning hours to a minimum of 16 at 1700 LST.

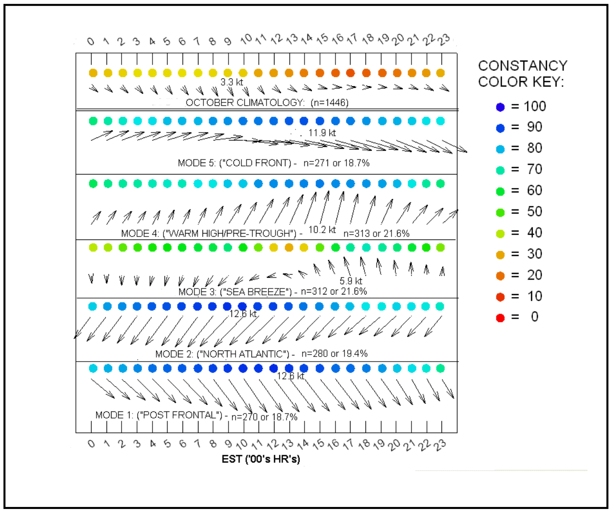

Figure 18 - October Hourly Mean Vector Wind Modes (plus Climatology) for LaGuardia

The newly designated, first in rank "Warm High/Pre-Trough" cluster (Mode 4: 21.65% incidence") is marked by southwesterly oriented vectors throughout the day, the vectors becoming significantly stronger over the afternoon hours with constancies reaching the mid 80's (brown trace in Figure 22). The hourly temperature curve (brown trace in Figure 19) displays the highest mean temperatures for all hours of the day, hence the subjective designation "Warm High/Pre-Trough". The high temperatures could conceivably be associated with Indian Summer ("Warm High") scenarios, or the approach of a long-wave trough with advection of warmer, moister air ("Pre-Trough"). This by no means excludes the possibility of a "Stronger Sea-Breeze" southwesterly sea-breeze signal which might be seen the result of another "sub-cluster clustering" exercise, although there is no indication in the mean of the early decline in temperature (brown trace in Figure 19) that was noted for July (see blue trace in Figure 14). Mean relative humidities (brown trace in Figure 20) do show a increase over the afternoon to early evening hours, not quite as dramatic, however, as the rise shown for the April "Pre-Trough" cluster (brown trace in Figure 10). Mean scalar wind speeds (brown trace in Figure 21) exhibit levels similar to Climatology or a little higher for the afternoon and evening hours, and mean vector constancies (brown trace in Figure 22) are at comparatively low levels in the pre-noon hours, relatively high from the late-afternoon on. No indication of a curve dip is present that would "tipoff" of a climatological changeover in wind direction.

Second most important pattern, Mode 3 or "Sea-Breeze" (incidence: 21.58% - down from July's 24.6% figure) had only one fewer case (n=312) than "Warm High/Pre-Trough" (n=313). It shows the recognizable configuration of light north to northeasterly vectors for the morning, shifting to stronger magnitude (sea-breeze) southerlies for the afternoons, although given the fact that this is post-summer October, the changeover comes later in the afternoon, with magnitudes weaker. The mean temperature curve (blue trace in Figure 19) shows the characteristic accelerated decline after about 1400 LST, mean relative humidities (blue trace in Figure 20) also exhibiting the expected pronounced increase over the afternoon hours. Mean hourly "Sea-Breeze" scalar winds (blue trace in Figure 21) are quite light, the contrast vs. Climatology and the other clusters being more pronounced than in April and July. Finally, the constancy curve (blue trace in Figure 22) clearly shows the big dip, although it's less confined to a single hour compared to April and July.

Ranking third is Mode 2 or "North Atlantic", its frequency increasing to 19.4% compared to July's ' 13.9%. It shows the expected Northeasterly vectors, and at higher magnitudes and constancies than July. The mean temperature curve (green trace in Figure 19), exhibits the smallest diurnal range of the clusters, and the mean relative humidities (green trace in Figure 20) easily show highest hour-to-hour figures. Mean scalar wind speeds (green trace in Figure 21) again show their propensity for pre-noon maxima, as do the constancy figures (green trace in Figure 22), a few of the figures exceeding 90.

Fourth in importance (18.7%) is Mode 5 or "Cold Front", reappearing in the October selection after "disappearing" from July's. It displays the recognizable configuration of a slow turning from southwesterly orientations to northwesterly, the northwesterly bent coming sooner in the afternoon than April's, however. Hourly mean temperature-wise (yellow trace in Figure 19), the curve differs only slightly from Climatology's (black curve) a slightly faster decline for the evening hours evident from the chart, however, not unlike that for April. Similar also to April, mean hourly humidities (yellow trace in Figure 20) are relatively low, being exceeded in this regard by only those of the "Post-Frontal" group. Evening mean scalar wind speeds (yellow trace in Figure 21) rival those of "Post-Frontal" for highest levels as do constancy figures (yellow trace in Figure 22).

Fifth in rank is Mode 1 ("Post-Frontal": 18.7% incidence, compared to July's 19.9%). It displays the familiar Northwesterly oriented vectors, although appreciably stronger in magnitude than July's, not unlike that for the other seasonal transition month April. Hourly mean temperatures (red trace in Figure 19 are the lowest of the clusters, as are those for relative humidities (red trace in Figure 20), the only exception to this being a slightly higher figure for "Cold Front" at 2300 LST. Mean scalar wind speeds (red trace in Figure 21) vye with those of "North Atlantic" for maximum values through about Noon, and with those of "Cold Front" for the rest of the day, those for Post-Frontal subsiding noticeably after 2000 LST.

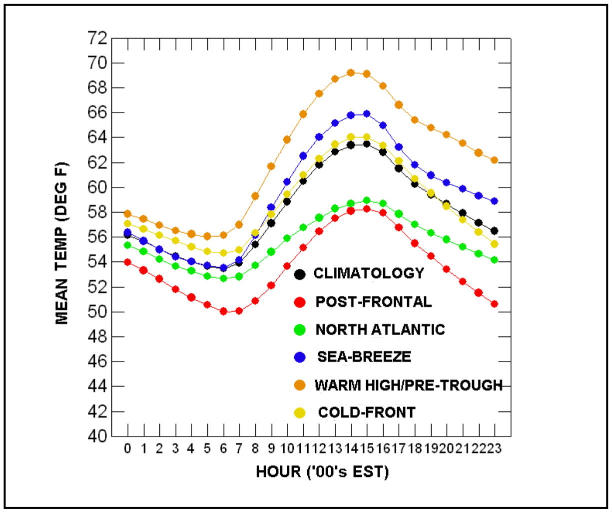

Figure 19 – Mean Hourly Temperatures, by Cluster and for Climatology, for LaGuardia - October

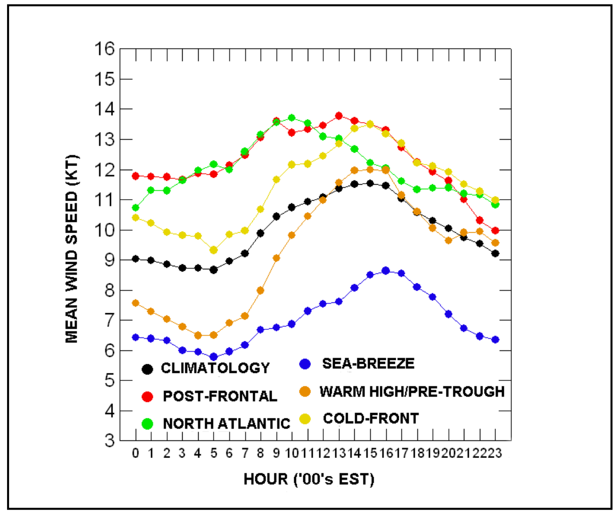

Figure 21 – Mean Hourly Scalar Wind Speeds, by Cluster and for Climatology, for La Guardia – October

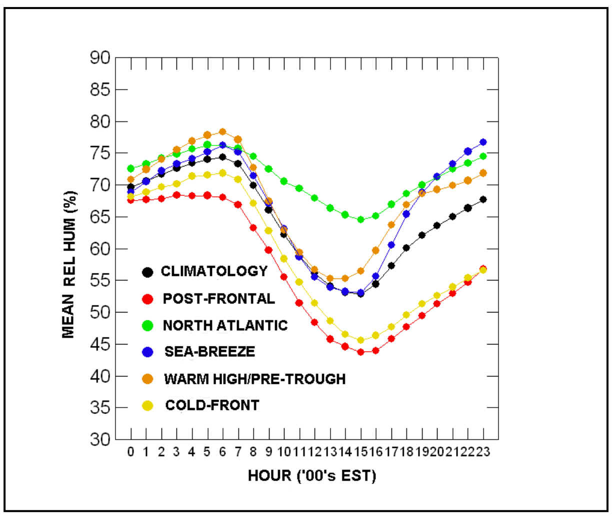

Figure 20 – Mean Hourly Relative Humidities, by Cluster and for Climatology, for LaGuardia - October

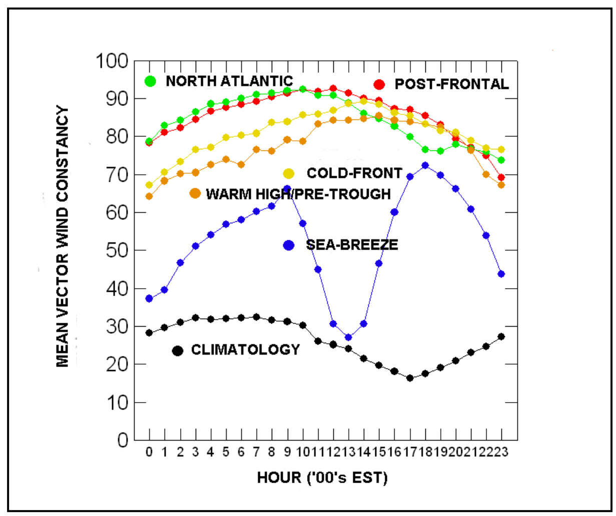

Figure 22 – Mean Vector Wind Constancies, by Cluster and for Climatology, for La Guardia - October

4.5 – *All months combined*

The above analyses seems to have provided useful and physically meaningful information on the different "flavors" of diurnal wind character at LaGuardia for January, April, July, and August, capturing synoptic as well as sea-breeze signatures, with just a few cluster-naming ambiguities.

Based on these results, what of the feasibility of analyzing all the months as a unit? No doubt some of the patterns already seen would reappear again, other new ones might also emerge, with still others disappearing. Probably most importantly, an all-inclusive treatment would allow monthly-to-month relative cluster frequencies to be calculated, creating a more comprehensive picture of diurnal contiguous wind pattern variability over the course of the year.

To this end, the ISH data base was tapped to produce 16857 observation-days (72.9% possible) of intact 24-hour readings of winds, temperature, and humidity.

Figures 23 through 33 depict the mean vector wind results along with the supporting graphical analyses of hourly mean temperatures, humidities, wind speeds, constancies, and additional, new bar charts that depict the individual monthly cluster frequencies.

Six modes were created, and designated as: "Sea-Breeze" (20.8% incidence), "Pre-Trough/Warm High" (19.0%), "Cold Front" (18.0%), "Post Post-Frontal" (14.5%), "Post-Frontal" (13.9%), and "North Atlantic" (13.8%). Compared to individual months' analyses, "Post-Frontal" and "North Atlantic" appeared for all four, "Cold-Front" and "Sea-breeze" for three, and "Pre-Trough/Warm High" for October only. The "Post Post-Frontal" designation is new.

Figure 23 depicts the hourly mean vector winds for Climatology (top panel) followed by the six modes. No great surprise, the Climatology vectors are at weak magnitudes for all hours, mostly light northwesterly from midnight through about Noon LST, essentially westerly for 1500 to 2000 LST, and light northwesterly again for the closing hours. Constancy values (black trace in Figure 27), range from 34 for 0400 LST to 21 for 1600 to 1800 LST, each.

The "Sea-Breeze" pattern (Mode 4), is the most frequent (incidence: 20.8%), having had that distinction also for April, July, and (almost) October. Displayed are the familiar light north to northeasterly mean vectors for the forenoon hours, followed by a shift to progressively stronger southerlies, maximum magnitudes/constancies attained in the late-afternoon around 1800 LST. The hourly mean temperature curve (orange trace in Figure 24) displays the second warmest time-series, attributable to the fact that "Sea-Breeze" has an enhanced frequency during the warmer months (See Figure 28). Its shape is similar to Climatology except for a slightly steeper but expected decline for the late afternoon hours. Mean hourly relative humidities (orange trace in Figure 25) also show the expected sharp rise in the afternoon. Average scalar wind speeds (orange trace in Figure 26), exhibit, as previously seen, the weakest hourly magnitudes (no exceptions),

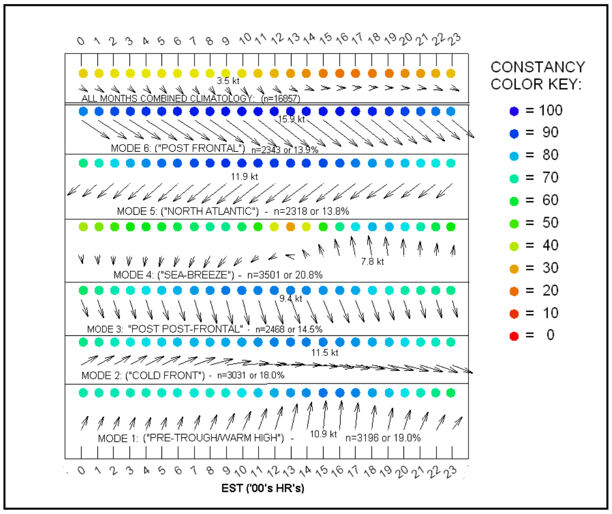

Figure 23 - "All-Months-Combined" Hourly Mean Vector Wind Modes (plus Climatology) for LaGuardia

especially for the afternoon hours. Vector wind constancies (orange trace in Figure 27) as seen before, show a striking decline at 1300 LST, reflecting the statistically favored sea-breeze changeover time. Inspecting Figure 28 at more depth, "Sea-Breeze" frequencies rang from just over 10 percent for December and January to 25-30 percent for May thru August, inclusive.

Second is importance is the "Warm High/Pre-Trough" designation (Mode 1: 19.0% incidence"). Similar to its October namesake, it is marked by southwesterly vectors throughout the day which become significantly higher in magnitude over the afternoon hours. Identical to October, the temperature curve (red trace in Figure 24) has the warmest means for all hours of the day. Mean relative humidities (red trace in Figure 25) show the same accelerated increase during the afternoon hours, and mean scalar wind speeds (red trace in Figure 26) likewise show mostly light mean speeds in the morning, increasing to levels slightly higher than Climatology by mid-afternoon. The constancy curve (red trace in Figure 27) also shows a similar shape as October's. Lastly, Figure 29 shows the monthly incidence frequencies of "Warm High/Pre-Trough". Like "Sea-Breeze", it too is a warm season phenomenon, frequencies at their maxima for June, through August (around 27 percent each), and minima for December through March (between about 10-14 percent). Interestingly, the October and All-Months-Combined clusters by this name have approximately equal incidence (~19 percent).

Ranking third is "Cold Front" or Mode 2 (incidence: 18.0%). This replicates the mean vector configurations seen for January, April, and October which exhibit the gradual diurnal turning from southwesterly to northwesterly. As seen likewise for the October "Cold-

Front" cluster, the hourly mean temperature curve (green trace in Figure 24), contrasts only slightly with Climatology (black curve). Also seen previously for the January, April, and October clusters, mean hourly humidities (green trace in Figure 25), especially for the afternoon hours onward, are significantly lower than Climatology, exceeded only in this regard by "Post-Frontal". Replaying a familiar "theme", mean scalar wind speeds (green trace in Figure 26) are also second in high magnitude levels only to those of "Post-Frontal", but constancy levels (green trace in Figure 27), are not at particularly extreme values for any hour of the day. The monthly incidence pattern of "Cold-Front" (Figure 30) is more uniform than "Sea-Breeze" or "Warm High/Pre-Trough"; maximum incidence percentages shown for November-February (23-25 percent), the rest of the months at relatively uniform levels (13 -18 percent).

In fourth place is the new designation "Post Post-Frontal" or mode 3 (incidence: 14.5%). This is so-titled because the "Post-Frontal" pattern, seen previously for all of the four individual months, is well represented as a separate cluster with all the characteristic signatures. "Post Post-Frontal" seems to represent those cases which are some distance removed in time from the more immediate Post-Frontal effects (e.g. high winds, cold temperatures, low humidities, etc.,) but still under the more residual but still identifiable influences of a cold-frontal passage. Figure 23 indicates persistent north-northwesterly vectors with constancy levels from the high 60's to mid 80's (blue trace in Figure 27). The hourly mean temperature curve (blue trace in Figure 24) is quite similar in shape to Climatology (black trace), but at levels several degrees colder throughout. Mean relative humidities are virtually undistinguishable from those of "Cold Front" for all hours. "Post Post-Frontal"'s mean scalar wind speed curve (blue trace in Figure 28) shows atypically slight variation over the day (~1 knot range), values slightly higher than climatology through the late morning hours, but less than Climatology thereafter.. Monthly incidence frequencies (Figure 31) show an even more uniform pattern than that for "Cold Front", with a clear concentration, nonetheless, of maxima for January-March (16-19%) and minima for June-August (11-13%)

Fifth most important is "Post-Frontal" or Mode 6 (incidence: 13.9%). Previously resolved for each of the four individual months, it displays the trademark high-magnitude northwesterlies. Hourly mean temperatures (brown trace in Figure 24) easily rank as the clusters' coldest for all hours of the day, typically about 15 F colder than Climatology, and mean hourly humidities (brown trace in Figure 25) are also, without any individual hourly exceptions, at clusters' lowest levels. Mean scalar winds (brown trace in Figure 26) are also the highest of the clusters, exceeding the 16-knot level from mid-morning to mid-afternoon, and constancies (brown trace in Figure 27) likewise rank at highest levels, with many exceeding 90. Figure 31 shows that "Post-Frontal" is decidedly a colder-season phenomenon, maximum incidence frequencies shown for December-February (22-24 percent) and lowest for June-August, especially August with only 3 percent

Sixth of the clusters is "North Atlantic" or Mode 5 (incidence: 13.8%). Like "Post-Frontal" it was a designated cluster for each of the individual months' analyses, exhibiting the customary northeasterly vectors, with pre-noon maximum magnitudes and constancies. To go with its nature as promoting and accompanying damp, cloudy, and sometimes rainy/snowy conditions, the mean temperature curve (yellow trace in Figure 24), exhibits the smallest diurnal range of the clusters, a property, of course, already seen for the individual calendar months. Mean relative humidities (yellow trace in Figure 25) are highest by a wide margin, especially for the afternoons, and mean scalar wind speeds (yellow trace in Figure 26) again display their tendency for pre-noon maxima. Constancy values (yellow trace Figure 27) with a few exceptions, are second only "Post-Frontal" in high magnitude levels. Lastly, monthly frequencies for "North Atlantic" (Figure 33) show a slight inclination for maximum frequencies during transitional months (16 or 17 percent magnitudes for March, September and October) with relatively distinct minima for June and July (10 and 8 percent values, respectively).

Summing up, the results of the "All-Months-Inclusive" approach are rather impressive, with monotonically increasing/decreasing month-to-month frequencies for four of the six clusters. More specifically, for the "Sea-Breeze" cluster, the 11 percent figure depicted for January seems more realistic than the 4% statistic generated as the result of the January sub-cluster treatment. The "Cold Front" designation, which failed to express itself in the July individual month results, appears as a 17 percent figure for the "All-months-combined" analysis, a much more realistic outcome.

Based on these LaGuardia results, unless there was some compelling reason to do otherwise, it would seem worthwhile to start with an all-months-combined approach first, delving into individual months of interest as warranted.

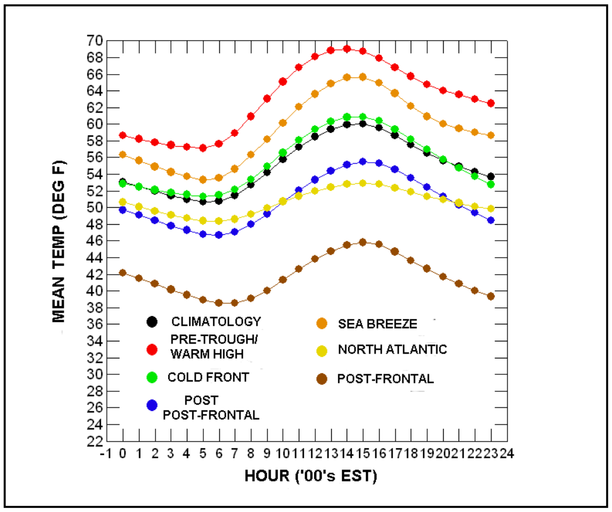

Figure 24 – Mean Hourly Temperatures, by Cluster and Climatology, for LaGuardia – “All-Months-Combined”

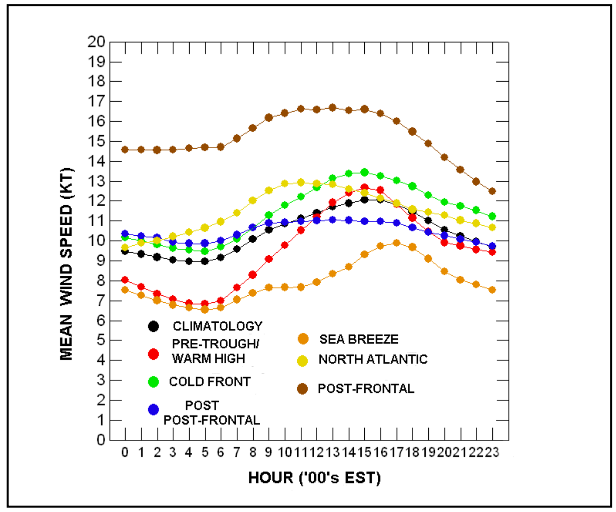

Figure 26 – Mean Hourly Scalar Wind Speeds, by Cluster and Climatology, for LaGuardia – “All-Months-Combined

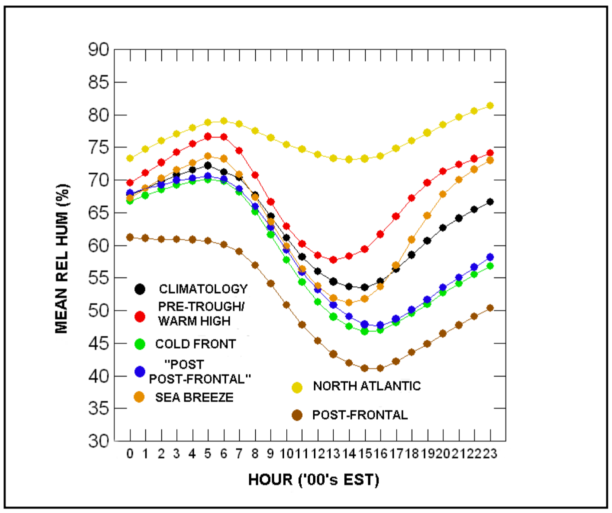

Figure 25 – Mean Hourly Relative Humidities, by Cluster and Climatology, for LaGuardia – “All-Months-Combined”

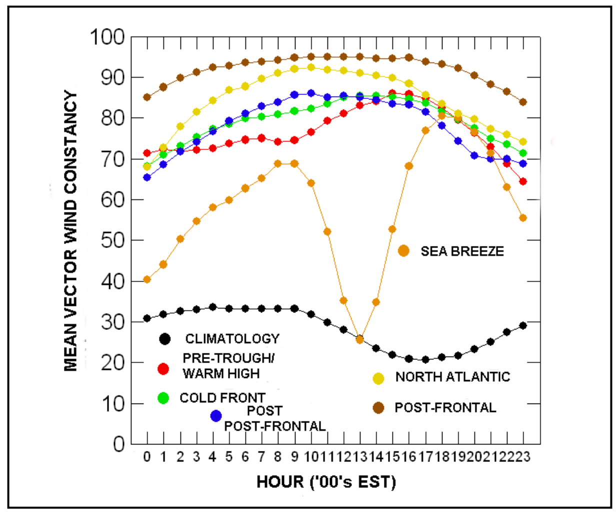

Figure 27 – Mean Vector Wind Constancies, by Cluster and Climatology, for LaGuardia – “All-Months-Combined”

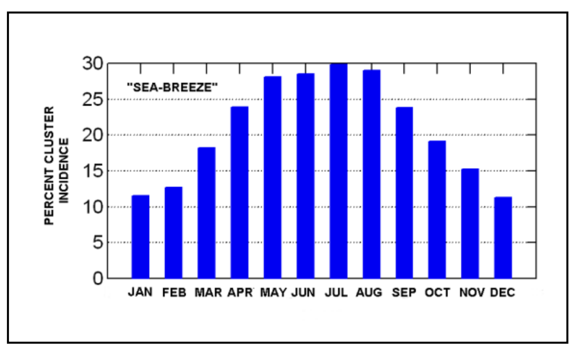

Figure 28 – Incidence Frequencies, by month, of “Sea-Breeze” Cluster – LaGuardia

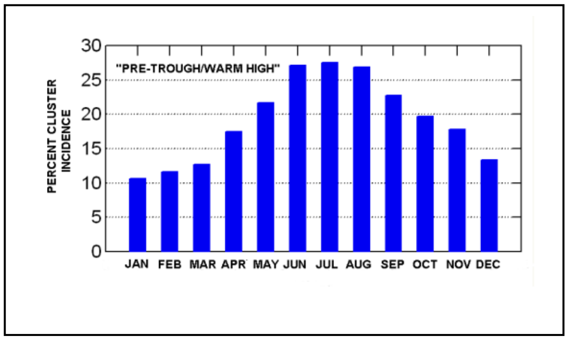

Figure 29 – Incidence Frequencies, by month, of "Warm High-Pre-Trough" Cluster – LaGuardia

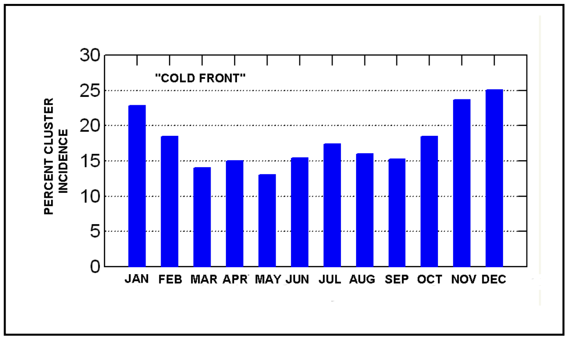

Figure 30 – Incidence Frequencies, by month, of "Cold Front" – LaGuardia

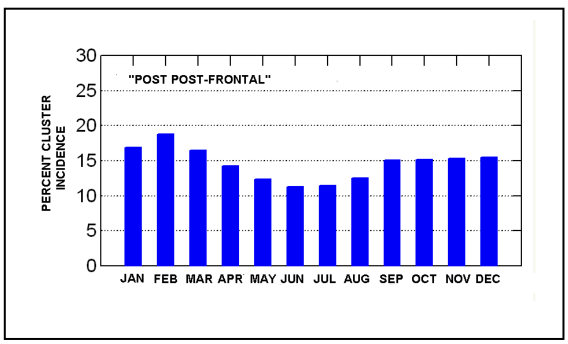

Figure 31 – Incidence Frequencies, by month, of "Post Post-Cold Front" – LaGuardia

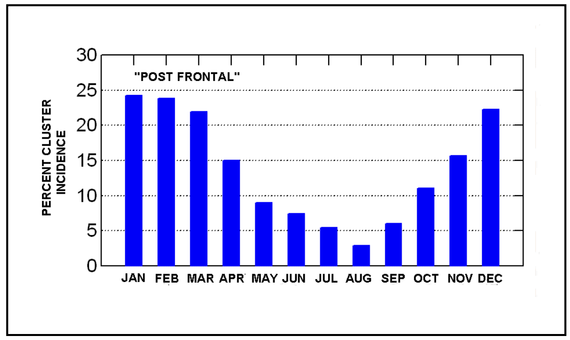

Figure 32 – Incidence Frequencies, by month, of "Post-frontal" Cluster – LaGuardia

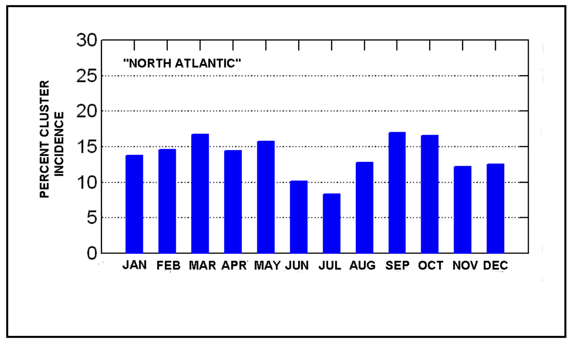

Figure 33 – Incidence Frequencies, by month, of "North Atlantic" – LaGuardia

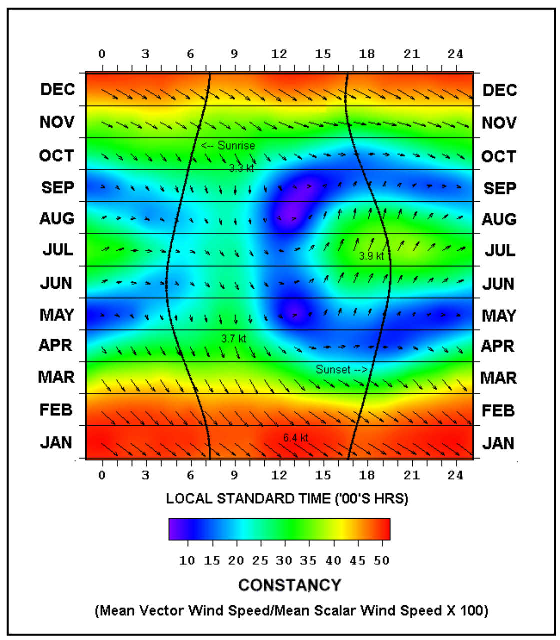

Figure 34 – Mean Vector Winds and Constancies, by Month/Hour, for LaGuardia Airport, New York (1949-2011 Data)

4.6 – *Identification of the Most Extreme Individual Days' Patterns*

An additional interesting side-application of a cluster analysis of this kind is identification of extreme individual patterns, utilizing statistical distance information. In the LaGuardia application, cluster memberships of (normalized) individual observations were determined by comparing their squared Euclidean distances to each of the 48-D cluster centroids; the cluster associated with the least distance would be that to which the observation was assigned.

Large statistical distances within a cluster would reflect "extreme" sets of 24-hour wind observations that were still considered affiliated statistically with that cluster. Since the La Guardia analyses segregated hour-to-hour observations of the same directional character, extreme distances probably represented cases which exhibited unusually high scalar wind speeds within the acceptable "boundaries" of the idealized pattern.

Extending this to an entire data set, ranking all the distances collectively, irrespective of cluster, could be a means of assessing in a relative way the most extreme of these extremes.

Utilizing the "All-Months-Combined" data set again, and from the STATISTICA output, Figure 35 is a histogram of the distribution of distances for each of the 16857 individual observations.

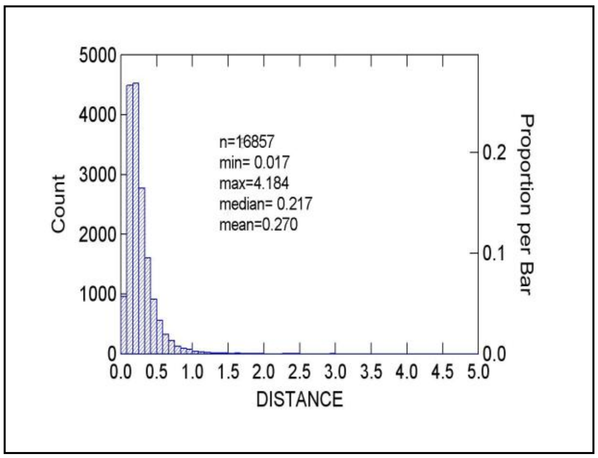

Figure 35 – Histogram of Statistical Distances of Individual Observations to Respective (Six) Centroids, "All-Months-Combined" Clustering Analysis, LaGuardia

In general, the distances (statistical departures of the individual days' normalized u's and v's from parent centroids) are low (mean: 0.270), indication that the K-Means/V-Fold algorithm did a good job resolving clusters and assigning individual observations. Nonetheless, there are still some outliers, the most of extreme of which are of interest here. Ranking highest was the 4.184 distance generated for 25 Nov 1950, a day assigned to the "North Atlantic" cluster.

HRLST	WDIR	WSPD (kts)	TEMP (deg F)	REL HUM (%)
0	ENE	23	54	90
1	ENE	23	54	90
2	E	26	56	83
3	E	30	57	77
4	E	35	57	77
5	E	34	57	77
6	E	36	54	93
7	E	45	57	77
8	ENE	43	55	86
9	ENE	40	55	90
10	E	43	55	93
11	E	45	55	93
12	E	45	57	90
13	SE	43	58	90
14	E	49	58	90
15	ESE	50	59	87
16	E	54	60	87
17	ESE	45	59	96
18	SSW	15	49	100
19	SSW	21	48	93
20	SSW	21	44	82
21	SSW	23	42	79
22	SSW	21	38	73
23	SSW	14	36	67

Figure 36: Selected Hourly Weather Conditions for 25 Nov 1950 at LaGuardia Airport, New York

Figure 36 above lists the hourly observations of Wind Direction, Wind Speed, Temperature and Relative Humidity for that day.

From the columns, 25 Nov 1950 experienced a very strong Nor'easter, winds East to East-North-Easterly through Noon LST at sustained speeds up to 45 knots. They picked up even more through the afternoon hours, reaching 54 knots (Easterly) at 1600 LST, then slackened as directions shifted to South-Southwesterly and temperatures tumbled. The very high speeds for virtually the whole day, almost all from the East quarter, along with the shift to a significantly different direction in the evening (but at still relatively strong speeds) are responsible for the high distance statistic.

Interestingly, three of the five highest statistical distances associated with the "All-Months-Combined" data set were members of the "North Atlantic Cluster", but it should also be reminded that only 72.9% of the possible daily observations at LaGuardia for 1949-2011 were utilized.

It should of course be repeated that this extreme, while interesting in its own right, was based on a distance to its own cluster centroid, not the *overall* data set centroid (presumably a 48-D array with 0.5 normalized statistics, each). Distances calculated to the latter might or might not show similar results, although the "Northeast Atlantic Cluster", with the strong easterly components displayed for many of its cases, is quite anomalous in this regard relative to the overall wind climatology for LaGuardia, which has a more westerly bent (see Figure 34).

5. SUMMARY AND CONCLUSION

Utilizing the clustering tool K-Means, accompanied by the V-fold cross validation algorithm. The existence and identification of contiguous midnight-to-midnight hourly wind patterns were explored for La Guardia Airport, New York using 63 years' data for the months January, April, July, October, and all the calendar months as a unit. Inputs were normalized u/v wind components. Results resolved four clusters (or "modes") for January, five for April, July, and October each; and six for the "All-Months-Combined" data set. Signatures of mean vector wind orientations belonging to synoptic event sequences (e.g., Pre-Trough, Cold Front, Post-Cold Front, etc.,), "Back-Door Cold Fronts", "Nor'easters", as well as the sea-breeze were clearly isolated in most cases, the interpretations supported by line graphs portraying hourly mean temperatures, relative humidities, scalar wind speeds, and vector wind constancies.

For the "All-Months-Combined" application, most of the six modes showed monotonic increases or decreases in frequency by month. Also utilizing this data set and an array ranking its observations' statistical distances from cluster centroids, a particularly "extreme" individual diurnal wind pattern (25 Nov 1950) was identified and discussed.

Low constancies (i.e,, high wind variabilities) are the rule for most stations in the United States, particularly those removed from westerly or southerly facing coastlines, situated in the interior, and near mean storm tracks and frontal boundaries. The LaGuardia results being any indication, the K-Means/V-Fold clustering of u and v components should be able to likewise sort out and delineate clearly the various pattern modes that combine to produce a given station's overall diurnal wind climatology.

6. REFERENCES

Fisk, C.J., 2012: "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 – 10th Conference on Artificial Intelligence Applications to Environmental Science, American Meteorological Society, 22-26 January 2012, New Orleans https://ams.confex.com/ams/92Annual/webprogram/Paper192739.html

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