Identification of Contiguous-Hour Climatological Wind Modes Utilizing K-Means Clustering Analysis Combined with the V-Fold Cross-Validation Algorithm

Charles J. Fisk
Newbury Park, CA 91320 – cjfisk@att.net

Abstract
The following describes and demonstrates a methodology for identifying and characterizing climatological adjacent-hour resultant wind patterns or "modes", a departure from more conventional explorations of this kind which generally focus on single-hour analysis units (i.e., stand-alone individual hours or solo consolidations of individual hours). Resolving such patterns can be a statistical clustering exercise, utilizing decomposed east-west ("u") and north-south ("v") wind components, and for this purpose, K-means Clustering Analysis integrated with an optimization tool, the V-Fold Cross-Validation Algorithm, is applied on daily series of hourly wind observations for five first-order Canadian weather stations with lengthy periods of records: Victoria, Vancouver, Calgary, Toronto, Montreal, and Victoria/Vancouver (two-station treatment). The resulting analysis units (hourly u and v centroid statistics) are then converted, utilizing the arctangent function, into discrete cluster-by-cluster *progressions* of adjacent hour resultant wind statistics (24 by cluster or mode, for a single station), including resultant wind directions/speeds, mean scalar speeds, and constancies (ratio of resultant wind speed to mean scalar wind speed times 100). The statistics are then depicted graphically using vector icons (representing the resultant winds), colored circular icons (depicting constancy magnitudes), and on the same layout, histograms, depicting the relative frequencies of the idealized patterns, by calendar month.

Key Words: Resultant Winds, Diurnal Wind Variability, K-Means Clustering, Canadian Weather Stations, Statistical Graphics, Data Mining, Climograms

1. Introduction

Climatological wind variability is an element frequently considered in the planning of activities in which direction and changeability can have potentially adverse impacts. As a decision aid, wind rose diagrams, portraying the relative observed frequencies and mean speeds of various directions around the compass, can provide useful insights concerning individual hours of interest. Resultant wind calculations can also be meaningful in producing distilled summary statistics derived from collections of individual observations; these can also be expressed graphically in the form of vector symbolizations.

Just as there are preferred distributions of individual hourly directions as seen in wind-rose diagrams, there should also be multiple sets of contiguous hour-to-hour idealized resultant wind modes extending over prolonged diurnal periods of interest (for example, midnight-to-midnight).

Resolving adjacent hours' climatological wind patterns can be a clustering problem, utilizing decomposed "u" (east-west) and "v" (north-south) statistics, and in this study, K-Means Clustering Analysis combined with a special add-on capability, the V-Fold Cross Validation Algorithm, will be applied. The V-Fold Algorithm is an iterative, automated training sample type procedure that optimizes the number of clusters created, depending on the choice of statistical distance metric (Euclidean, Squared Euclidean, etc.,), percent improvement cutoff threshold (e.g. 5 percent), and other settings. The user also has the option of fixing the number of clusters generated. The methodology has already been applied with good results on multiple individual stations' data in the Southern California area [Fisk, 2012], both for operational and purely informative purposes; also on those of several NOAA Buoys in the Southern California Bight area, local military bases, National Weather Service stations, and automatic weather stations. The multiple station approach has also been explored with two stations in the California High Desert, Victorville and Daggett [Fisk, 2016], and also on a coastal and mountain station pair (Point Mugu Naval Air Base and Laguna Peak tracking station, respectively), separated by three miles and 1500 feet in altitude [Fisk, 2017]. The approach has also found use in rawinsonde (balloon soundings') applications, the modes resolved across altitude or pressure levels rather than hour of the day.

Expanding the geographical scope beyond California and the U.S., the present study investigates idealized mean vector wind patterns for five major Canadian weather stations – Victoria, Vancouver, Calgary, Toronto, and Montreal. The five stations will have their mean vector wind patterns analyzed individually, and a two-station treatment will also be demonstrated on Victoria and Vancouver. These two are a relatively close 76 miles apart, each with sea-breeze/land-breeze regimes but of contrasting character.

2. Methodology

The well-known and frequently utilized K-means methodology was first introduced by Hartigan [1975], the basic technique consisting of assigning observations to a designated number of K clusters such that the multivariate means across the clusters are as different as possible.

One drawback associated with traditional K-Means is that the researcher has to guess how many "true" clusters there are in advance, the ultimate choice of how many there might likely be requiring trial-and-error iterations combined with subjective judgement. One methodological advance, however, adapted in recent years that addresses this kind of uncertainty in the V-Fold Cross Validation Algorithm [Burman, 1989], a training sample type procedure which incorporated into K-Means allows for a more objective determination of an "optimum" number of clusters. Nature being as complex as it is, however, knowledge of the "real" number of clusters in a given application, of course, is an unattainable result; nonetheless, the V-Fold algorithm enhances the methodological objectivity of a clustering technique like K-Means.

The STATISTICA Data Miner Clustering module, which incorporates the V-Fold method, was utilized to create the idealized wind pattern modes. As stated above, the raw input were individual hours' u and v wind components, by day. The V-Fold algorithm rapidly produces in ascending order, 2 to K cluster sets, the iterations ceasing at some "optimal" number K, depending on a choice of statistical distance metric, Squared Euclidean for this study, to go with other settings. In more detail, the V-Fold 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 center centroids [Nisbet, Elder, and Miner, 2009]. The software, however, also gives the option of fixing the number of clusters, and one of the applications in this study goes that route. Preliminary to the clustering, the software automatically normalizes the data internally, by hour, to reduce them to a common scale and lessen the effects of outliers. A single station resultant wind analysis would be one in 48-dimensional space (24 pairs of u and v-components), a two-station analysis one in 96-D space. In a single-station analysis, each of the individual cluster "clouds" would have midpoints or centroids in 48-D space, composed of mean u and v values for each hour of the day. These would be then un-normalized and reconstructed, using the arctangent function, into 24 sets of hourly mean vector wind statistics. Each cluster's 24-member array would be a self-contained "encapsulated" statistical entity, depicting an idealized hour-to-hour *progression* of mean vector wind character over the course of a day. As the clustering process, by its nature, assigns cases of a very similar character to the same cluster, frequently the mean vector hourly wind directions' output could be interpreted with only slight loss of precision, as *average wind directions.* Also, the accompanying mean vector *wind speeds* often would be only slightly less than their counterpart mean scalar wind speed statistics.

The results, cluster by cluster, by station, will be depicted graphically on single-chart layouts, the hourly mean vectors as arrow icons pointing in the direction of flow and proportional in length to their magnitude. The highest magnitude vectors are annotated, and "constancy" statistics (mean vector wind speed/mean scalar wind speeds*100), a measure of collective individual-hour wind observation homogeneity in the given cluster, are represented as colored circular icons ordered from dark blue (highest constancies) to dark red (lowest constancies). In other less ideal lower constancy cases, a diurnal pattern might be portrayed that has inherently lighter and more variable winds, for example, weak diurnal circulation regimes, or naturally occurring diurnal changeover periods (e.g., land-breeze to sea-breeze and vice-versa). Finally, on the layout page below the vector symbolizations, the cluster relative percent frequencies are matched up, month-to-month, as a histogram chart.

While more detailed alternatives are certainly possible, the clustering analysis here is applied to all the calendar months' data as a unit. This "one-size-fits-all" approach is motivated by parsimony (much fewer charts), and as experience has established from applications on other stations, an appealing visual property frequently encountered is that the percent frequency histogram bars commonly exhibit smooth monotonic progressions month-to-month. This been said, an opportunity, nonetheless to visually and subjectively assess the one-size-fits-all approach versus intra-mode month-to-month diurnal mean vector orientations is provided in a later segment (see section 4.7) in the form of hour-by-month resultant wind "climograms" for a few selected modes which exhibit particularly contrasting diurnal variability - namely those offshore/onshore flow cases.

Intra-mode contrasts notwithstanding, the single chart layouts encompassing all months' data as a group permit quick-study generalizations, which can lead to more refined individual months' analyses if deemed warranted.

3. The Data

The data consist of hourly wind observations downloaded and processed from the Integrated Surface Hourly Data online site, maintained by the NOAA Federal Climatic Center in Asheville, North Carolina.

For example, Victoria, British Columbia's available period of record extended from January 1973 into December 2017. The software clustering methodology requires that all 24 hours be represented in the data, those days with even one missing hour excluded from consideration, so to enlarge the sample size, a pre-processing endeavor was undertaken for individual days with single-hour only gaps, namely linear interpolations of their u's and v's relative to those values from the immediately preceding and succeeding hours. In Victoria's case, this enlarged the number of retained observations from 8711 cases (days) to 13919, an increase of nearly 60 percent. Comparison of the clustering results derived from the n=8711 non-interpolated set with those of the latter n=13919 set resulted in the same number of clusters being generated, and they were essentially equivalent in terms of hourly mean u/v properties as well as monthly percent frequencies, so the interpolation scheme was applied on Vancouver, Calgary, Toronto, and Montreal as well. For these four, the same before-interpolations vs. after-interpolations were created – with effectively the same hourly mean u and v properties and percent frequencies. In hindsight, this seemed quite logical and reasonable as the data sets prior to the addition of cases (days) with interpolated hourly u's and v's were very large to begin with. This sample size enhancement strategy could provide useful (or even necessary) for multiple station studies, and for tasks that incorporated additional non-wind variables, like relative humidity.

Vancouver's record extended from January 1978 into December 2017, some 12699 cases retained. Calgary's data dated from January 1957 with 17121 cases considered. Toronto (data back to 1950) had 19056 cases, and Montreal (observations back to 1955 but none available for 1959-1976 and 2000-05) had 12911.

4. Results

4.1. Victoria: Lat: 48.43 N, Long: 123.37 W, Elevation: 38 ft

Five Diurnal Resultant Wind Patterns were generated for Victoria, the two most prominent comprising nearly 70% of the total cases. These seemed to be contrasting warm season vs. cold season "flavors" of the land-breeze/sea-breeze regime. The graphical layouts for these patterns appear below, one after the other, in descending order of importance, in Figures 1 thru 5, respectively.

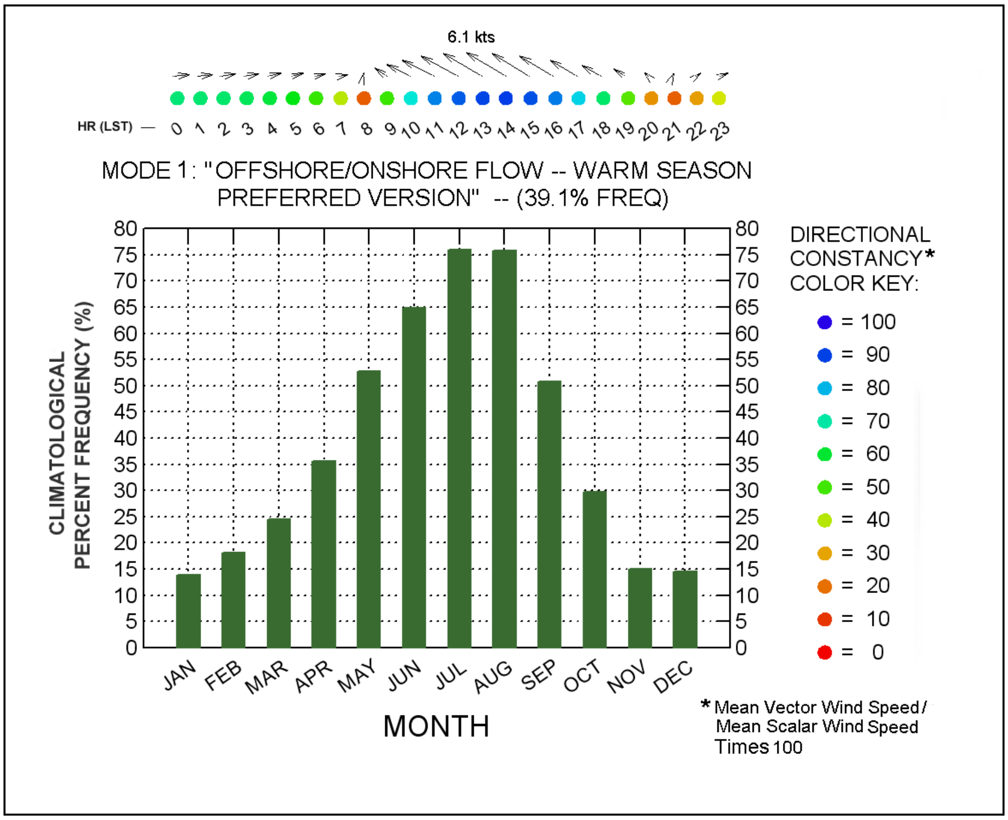

Figure 1: Victoria #1 Ranking Mode – "Warm Season" Offshore/Onshore Flow.

Mode 1, the warm season preferred offshore/onshore version (overall frequency 39.1%) depicts light westerly offshore resultants from midnight to 0700 local time, a one-hour transition period next ensuing at 0800 LST (note the orange-colored low constancy icon), this to be followed by a progressively strengthening southeasterly sea-breeze thru mid-afternoon, the sea-breeze reaching a maximum mean magnitude (6.1 knots) at 1400 LST, and continuing thereafter at decreasing magnitudes through early evening. The 1400 LST constancy statistic is depicted at a high 90 (blue-colored icon) level, indicative of an actual mean scalar wind speed of 6.8 knots. The month-to-month individual percent frequencies display a monotonic rise and fall, attaining peak values (just over 75 percent) for July and August, contrasting with the November-January minima (around 15 percent).

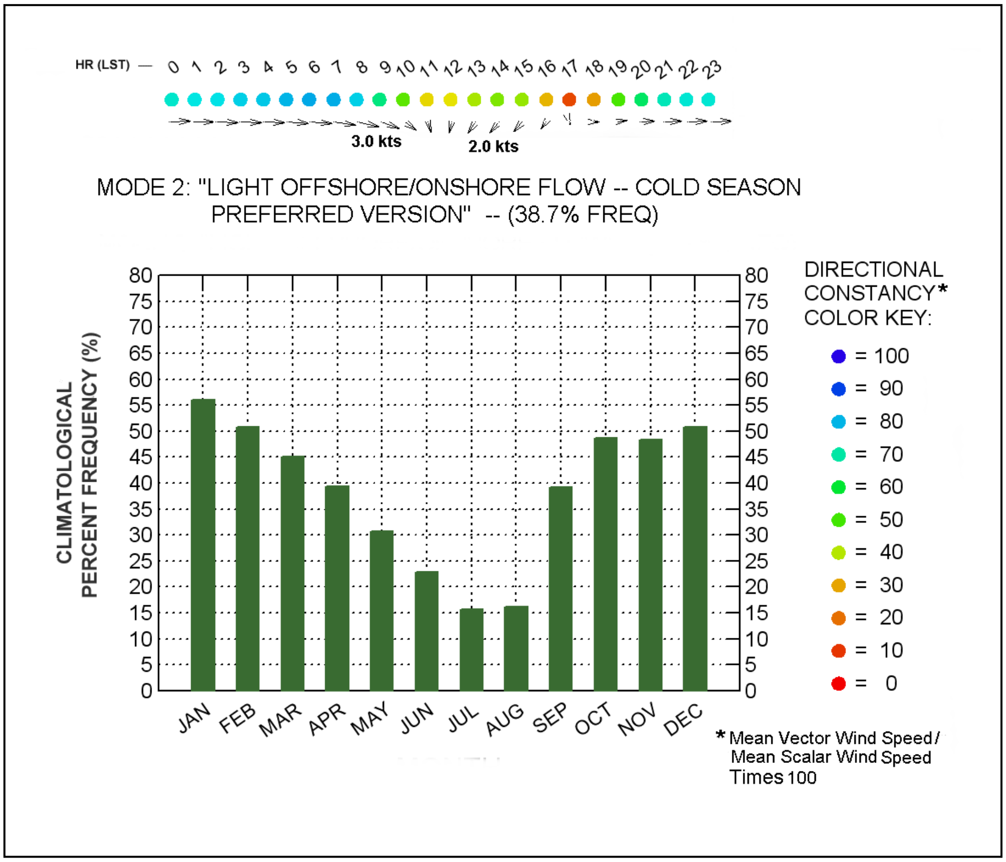

Figure 2: Victoria #2 Ranking Mode – "Cold Season" Offshore/Onshore Flow.

Mode 2, the cold season preferred offshore/onshore version (overall frequency 38.7%), consists of light westerlies (maximum around 3.0 knots) for most of the day, except for light northerlies or northeasterlies from 1100 LST through 1600 LST. A transition back to light westerlies commences at 1700 LST, indicated by the orange/red-colored (low) constancy icon. Maximum climatological frequencies are shown for October-February (48 % to 56 %), the minima for July and August (15 % and 16 %).

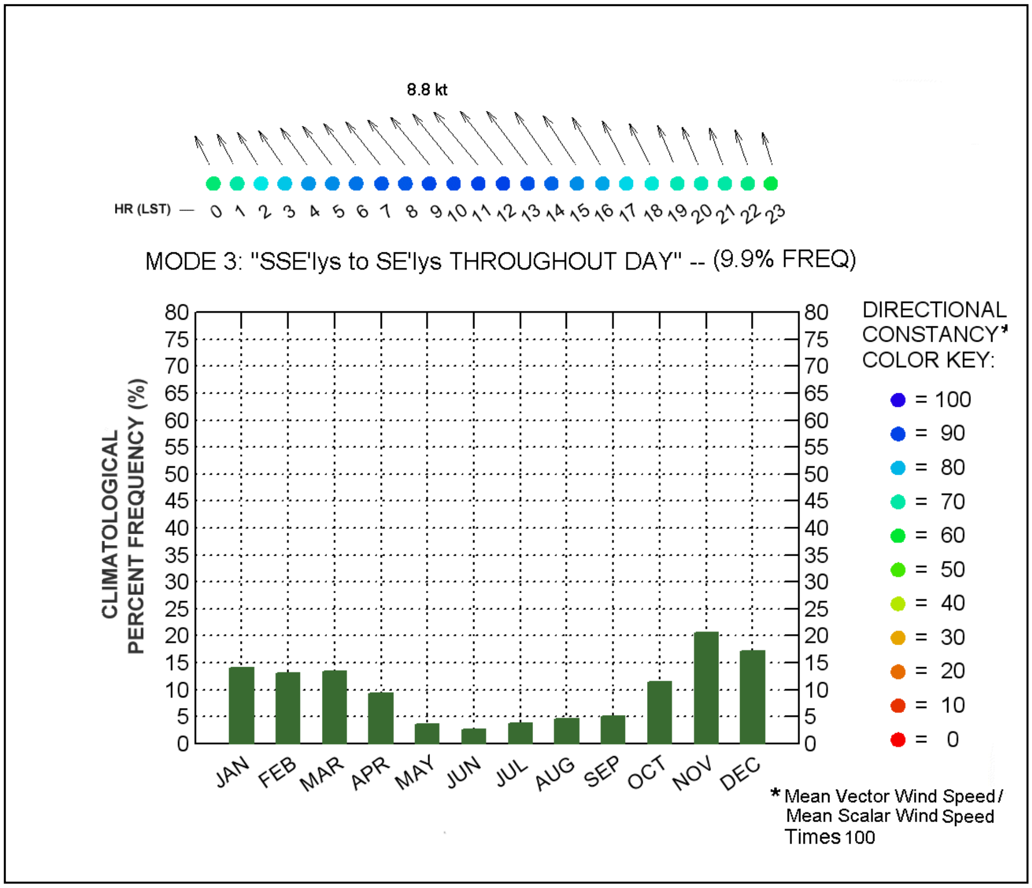

Figure 3: Victoria #3 Ranking Mode – Moderate SSE'lys to Southeasterlies.

The remaining three modes are quite minor in comparative frequency. Mode 3 (9.9% relative frequency), depicting south-southeasterlies to southeasterlies midnight to midnight, likely relates to the effects of larger scale synoptic scale weather features influencing Victoria, the wind flow probably further steered by Puget Sound/Salish Sea area topography. The stronger winds associated with these (8.8 knot maximum mean vector speed at 1100 LST – associated with a 9.8 knot mean scalar speed), obscures the signature of the local-land breeze/sea-breeze, the late-morning maximum speed and constancy indicators running contrary to that expected for a sea-breeze regime. Maximum climatological frequencies are shown for November (20 %) and December (17 %).

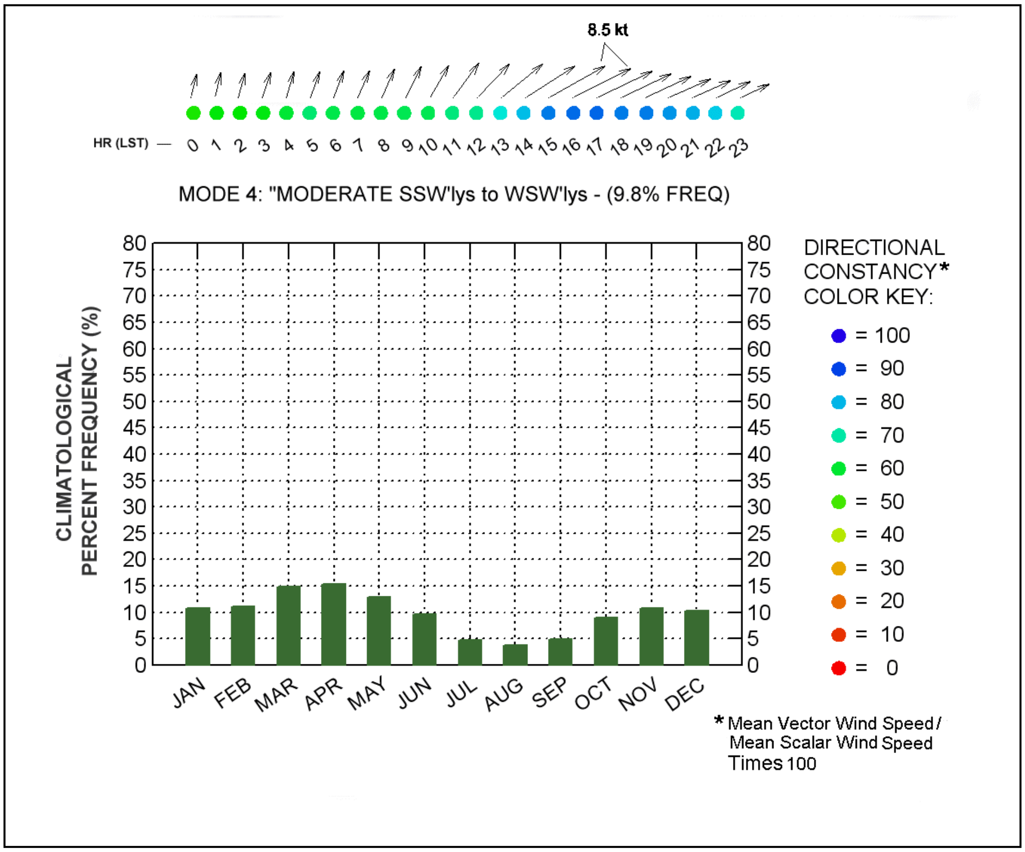

Figure 4: Victoria #4 Ranking Mode – SSW'lys to WSW'lys.

Mode 4 (9.8% overall incidence), representing south-southwesterlies to southwesterlies, also seems to reflect predominant synoptic scale influences, although there is a hint of a sea-breeze effect, the maximum mean vector speed noted over the mid-afternoon hours (8.5 knot mean vector speeds for 1500 LST and 1600 LST, associated with mean scalar speeds of ~10 knots each). Constancy values also reach a maximum ("blue"-icon) level in the late afternoon/early evening (~1500 LST to 1900 LST). Mode 4 appears to be a slightly favored Spring phenomenon with maximum relative frequencies shown for March-May (13 %-15%).

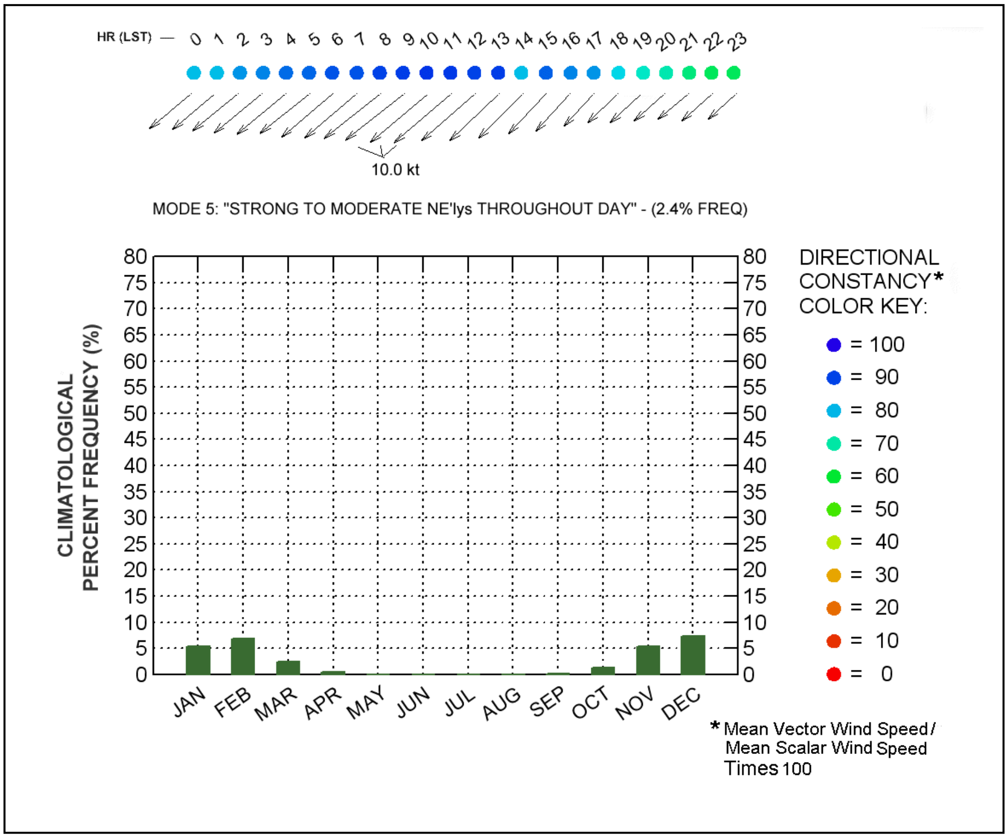

Figure 5: Victoria #5 Ranking Mode – Northeasterlies throughout.

Mode 5 (2.4% overall incidence), quite uncommon and confined almost exclusively to the coldest months of the year, portrays moderate to strong northeasterlies throughout the day, the highest magnitudes evident for the pre-noon hours. This probably reflects influences of strong cyclones to the west of Vancouver Island during the late Fall to early Spring periods. Maximum magnitude mean vectors are 10.0 knots, observed for 0900, 1000, and 1100 LST, each (associated mean scalar speeds: ~ 11 knots), the maximum monthly relative frequencies observed for November-February (5 %-7 %).

4.2. Vancouver: Lat: 49.20 N, Long: 123.19 W, Elevation: 10 ft

Five diurnal mean vector wind modes were generated for Vancouver as well, the two most prominent comprising 61% of the total cases. As with Victoria, the layouts for all the patterns appear below, one after the other, in descending order of importance (Figures 6 thru 10, respectively).

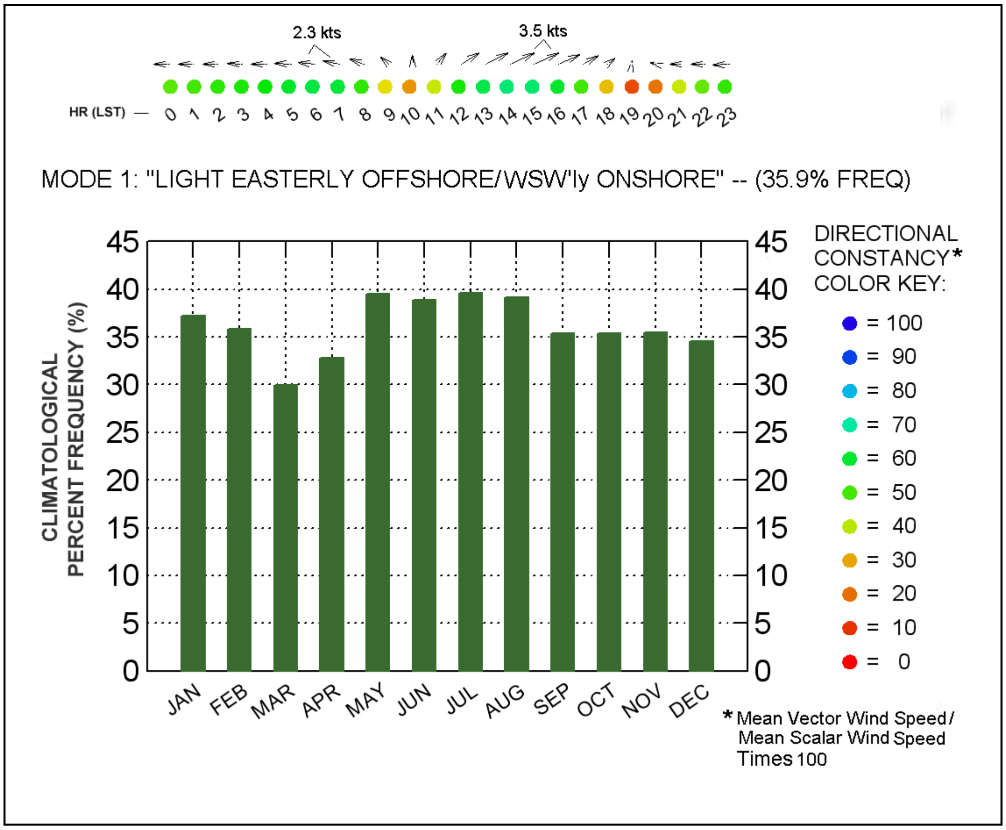

Figure 6: Vancouver #1 Ranking Mode – Land-Breeze/Sea-Breeze.

Mode 1 (overall frequency: 35.9 %) depicts Vancouver's version of the land-breeze/sea-breeze regime, in this case light easterly offshore flow alternating with onshore west-south-westerly onshore flow. Compared to Victoria, the seasonal distribution of relative frequencies is quite uniform with the dichotomous character absent, as indicated by Figures 1 and 2. Highest magnitudes are noted for the contiguous months May-August (frequencies all just less than 40%), the minima only slightly lower at 30% and 33%, for March and April, respectively. The light easterlies transition to light southwesterlies over 0900 LST to 1000 LST, the latter holding sway until about 1800 LST when the orientation switches back to light easterly again.

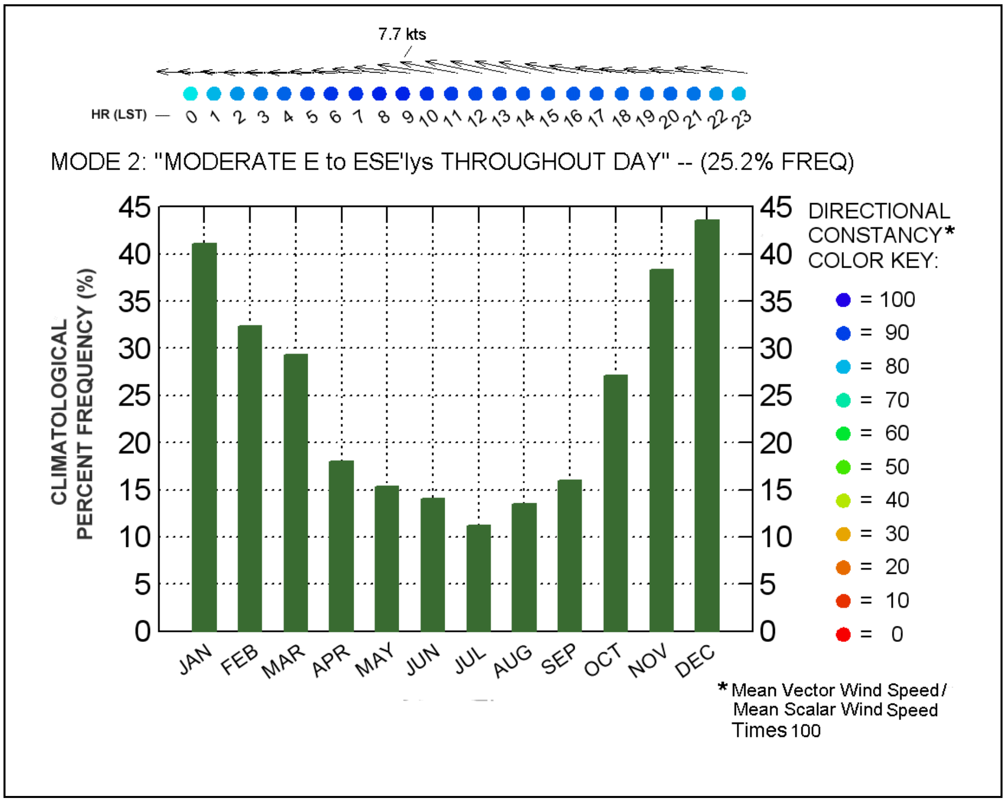

Figure 7: Vancouver #2 Ranking Mode – Moderate East to ESE'lys.

Mode 2 (overall frequency 25.2%), displays moderate east to east-southeasterlies midnight-to-midnight. A clear monotonic month-to-month progression of relative frequencies is indicated, the maxima observed in winter (44% December and 41% for January), the minimum for July (11%). There is no clear diurnal signal, so this is probably more of a synoptic scale influenced pattern, this notion further reinforced by the colder season frequencies' predominance and the relatively uniform constancy magnitudes at relatively high ("blue") magnitudes midnight to midnight. The 7.7 knot maximum vector magnitude (8.5 knot mean scalar counterpart statistic) is noted for 1100 LST.

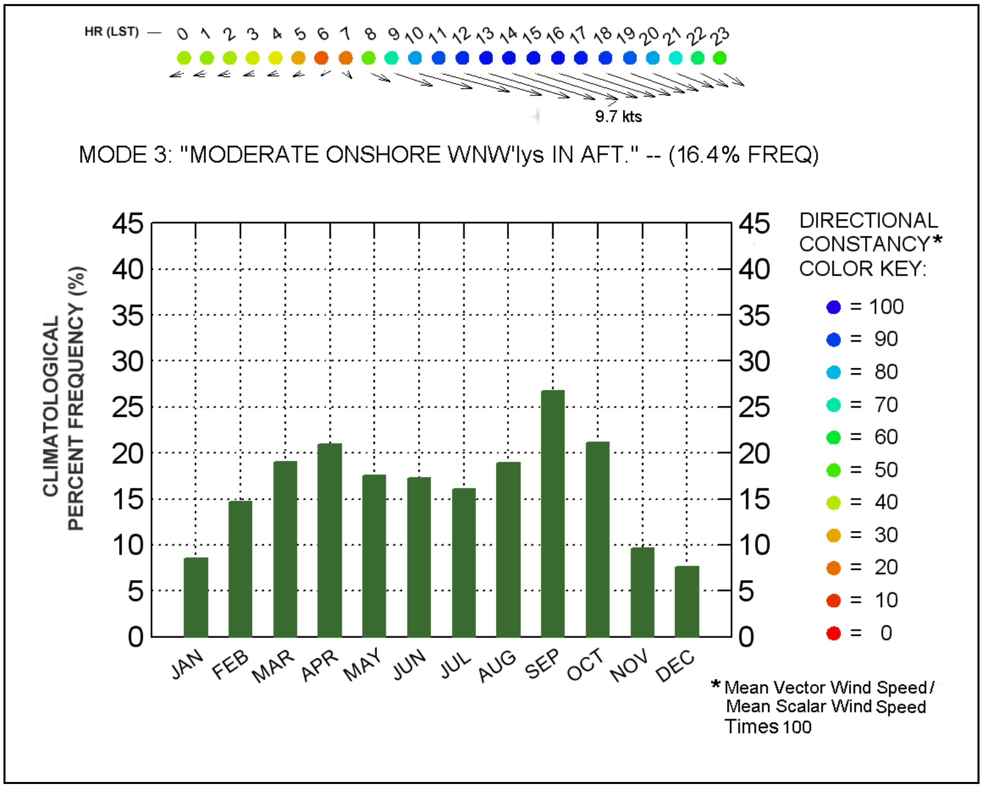

Figure 8: Vancouver #3 Ranking Mode – Moderate WNW'lys in Afternoon.

Mode 3 (overall frequency 16.4 %), depicts a midnight-to-midnight hourly progression with light and variable winds overnight, post-midnight, succeeded by increasing magnitude west-northwesterlies mid-morning thru mid-afternoon (9.7 knot maximum mean vector wind magnitude at 1500 LST, associated with a 10.2 knot mean scalar wind speed statistic), gradually decreasing thereafter. This likely reflects a combination of cold-front passages thru the channel separating mainland Vancouver from Vancouver Island, reinforced by a possible sea-breeze inducing pressure gradient (note the afternoon high "blue" constancies). Monthly frequencies are relatively uniform from March thru August (15 % to 21 %), an uptick noted for September (27 %).

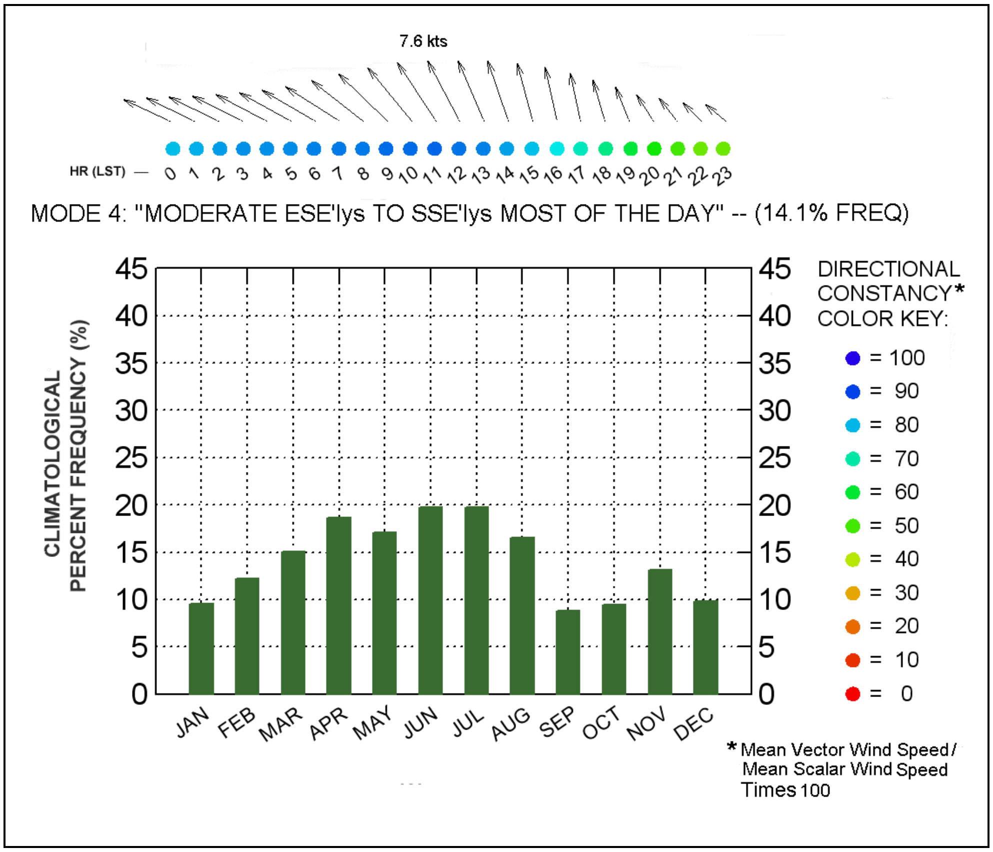

Figure 9: Vancouver #4 Ranking Mode – ESE'lys to SSE'lys Throughout.

Mode 4 (overall frequency 14.1 %), portrays another likely synoptic scale influenced pattern, the east-southeasterly oriented vectors gradually turning clockwise to southeasterly over the day, reaching a maximum magnitude 7.6 knots at 1200 LST, associated with an 8.5 knot mean scalar figure. "Blue" to "light blue" constancy magnitudes are noted prior to the mid-afternoon hours. This mode exhibits a Spring to Summer preference, the maximum relative frequencies observed for the contiguous months April thru August (17% to 20%).

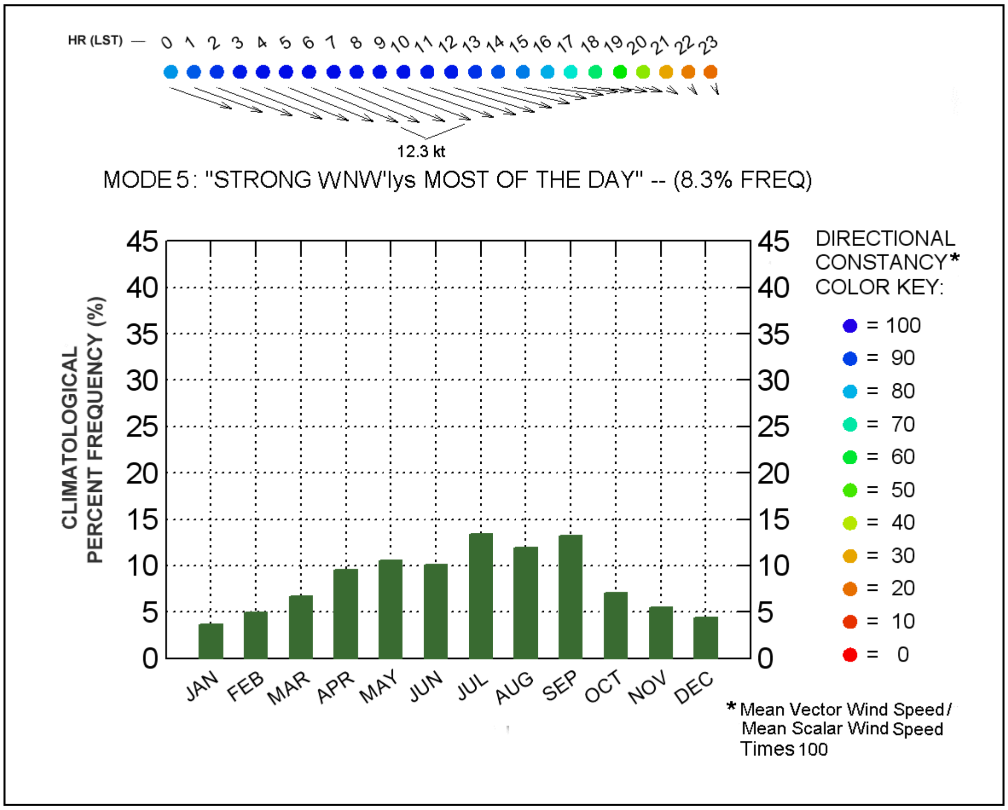

Figure 10: Vancouver #5 Ranking Mode – Strong WNW'lys Most of Day.

Mode 5 (overall frequency: 8.3%) is not unlike that of Mode 3 (see Figure 8), except that the west-northwesterlies are stronger (12.3 knot maximum magnitude vector – corresponding mean scalar figure: 13.0 knots), and in great contrast with Figure 8, concentrated mostly in the forenoon hours. This suggests a greater mix of cold frontal cases, which don't necessarily conform to nominal afternoon maximum sea-breeze character, although the slight July-September relative frequency predominance (12 % to 14 %) also may imply a few occasionally stronger sea-breeze episodes with earlier onset times.

4.3. Calgary: Lat: 51.13 N, Long: 114.10 W, Elevation: 3556 ft

The Calgary data set also produced five clusters. Being an interior continental station located to the lee of the Rocky Mountains, the station's hierarchy of idealized patterns does not include an offshore/onshore mode, the differentiations seemingly related mostly to scenarios relating to the station's proximity to fronts and the resulting contrasts thereof in directions and speeds. Included also is a representation that includes the westerly wind Chinook phenomenon. The layouts again appear in descending order of importance (Figures 11 thru 15, respectively).

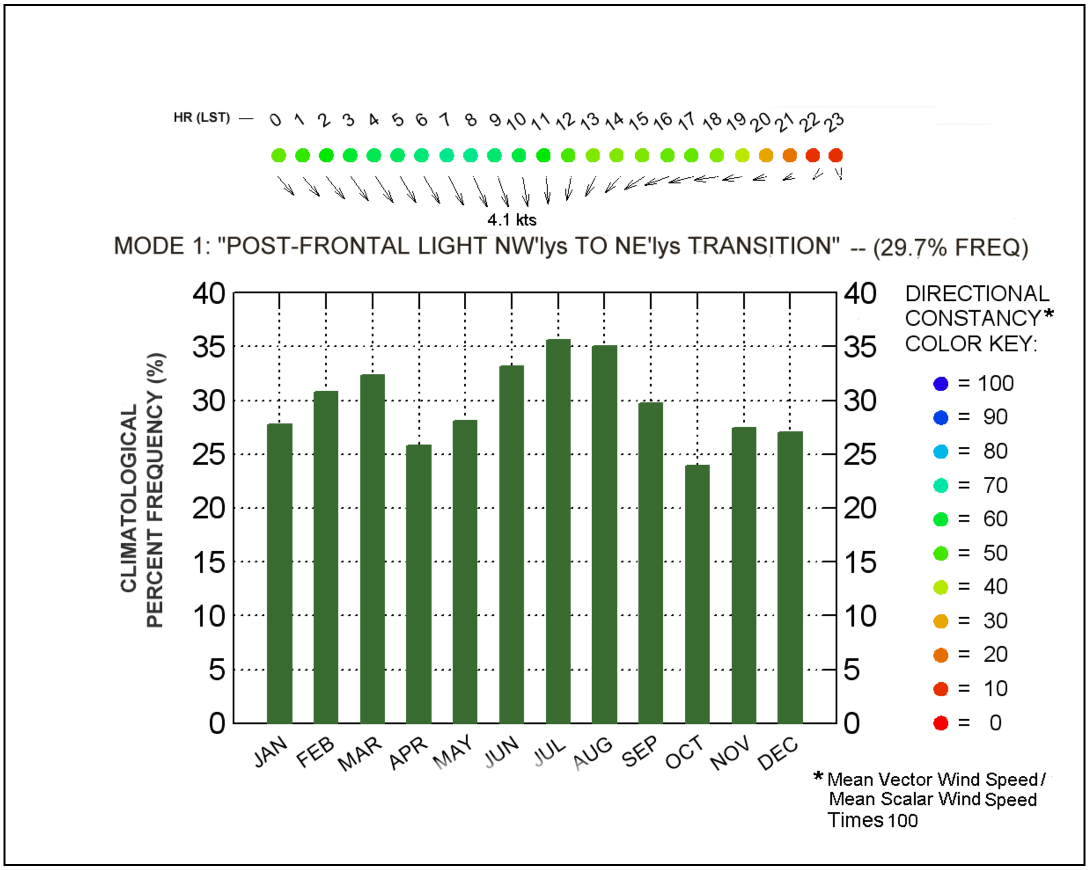

Figure 11: Calgary #1 Ranking Mode – Light Northwesterlies to Northeasterlies.

Mode 1 (overall frequency: 29.7%), with its relatively light winds, is suggestive of a diurnal wind scenario somewhat removed in time from a cold frontal passage, the relatively light northwesterlies to north-northwesterlies giving way to light northeasterlies by afternoon, these becoming less persistent statistically (lower constancies) as midnight approaches. Monthly relative frequencies are comparatively uniform, ranging from 24 % (October) to 36 % (July), but there is an apparent slight summer predominance, the high three noted for June to August (range of these: 33 % to 36 %). Maximum mean vector wind for this pattern is just 4.1 knots, at 0800 LST, corresponding to a 6.1 knot mean scalar magnitude.

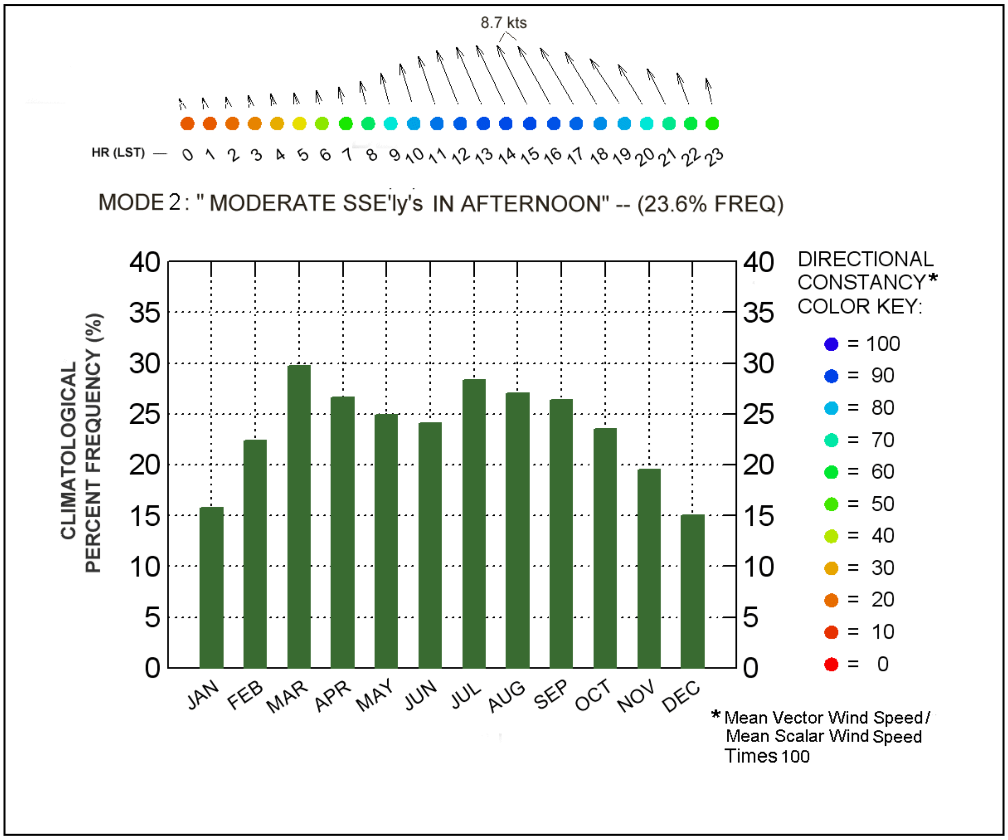

Figure 12: Calgary #2 Ranking Mode – Moderate SSE'lys in Afternoon.

Mode 2 (overall frequency: 23.6%) depicts a regime of light and variable post-midnight nocturnal winds transitioning soon after sunrise to moderate to high constancy south-south-easterlies by afternoon (8.7 knot maximum vector magnitudes at 1500 LST and 1600 LST, each, associated with 9.8 knot mean scalar statistics). Monthly relative frequencies are below the cluster's 23.6% overall average for November thru February, and above average for all the other months, and given the greater insolation received during the latter months, this may be a regime driven mostly by insolation during periods when frontal influences are remote in distance. Still, the monthly percent frequencies are relatively uniform for March to October, the highest figure (30 %), observed for March.

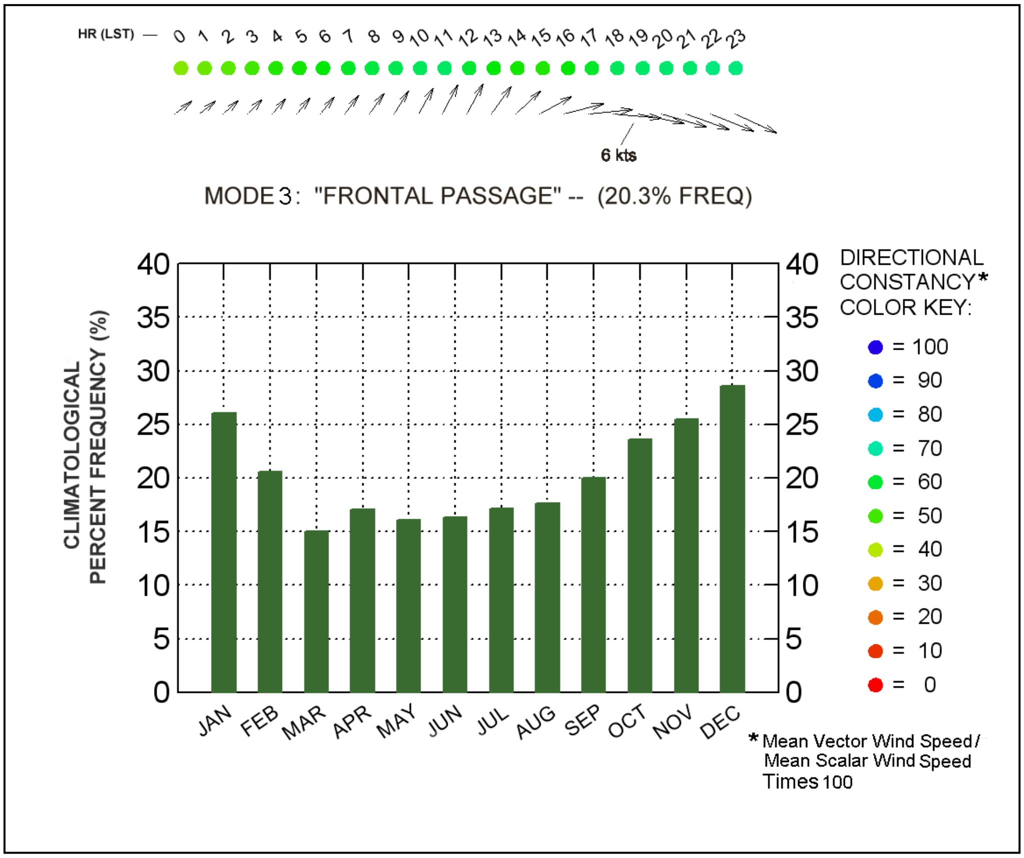

Figure 13: Calgary #3 Ranking Mode – Cold Front Approach and Passage.

Mode 3 (overall frequency: 20.3 %), has a relatively straightforward interpretation, namely the approach of a moderate intensity cold front (southwesterly vectors in the warm sector(?) followed by a shift to westerly and then northwesterly orientations). The vector magnitudes are not especially high, the maximum (6.0 knots) observed at 1800 LST (associated mean scalar speed: 9.5 knots). The month-to-month relative frequencies range from 15 % to near 30 %, and while fronts can occur every month of the year, they are at a maximum in the cold season (November to February). The early evening time of passage is a curious feature, perhaps an artifact of the circular nature of the midnight-to-midnight analytical time frame, or perhaps a reflection of real physical (topographical?) influences in the Calgary vicinity.

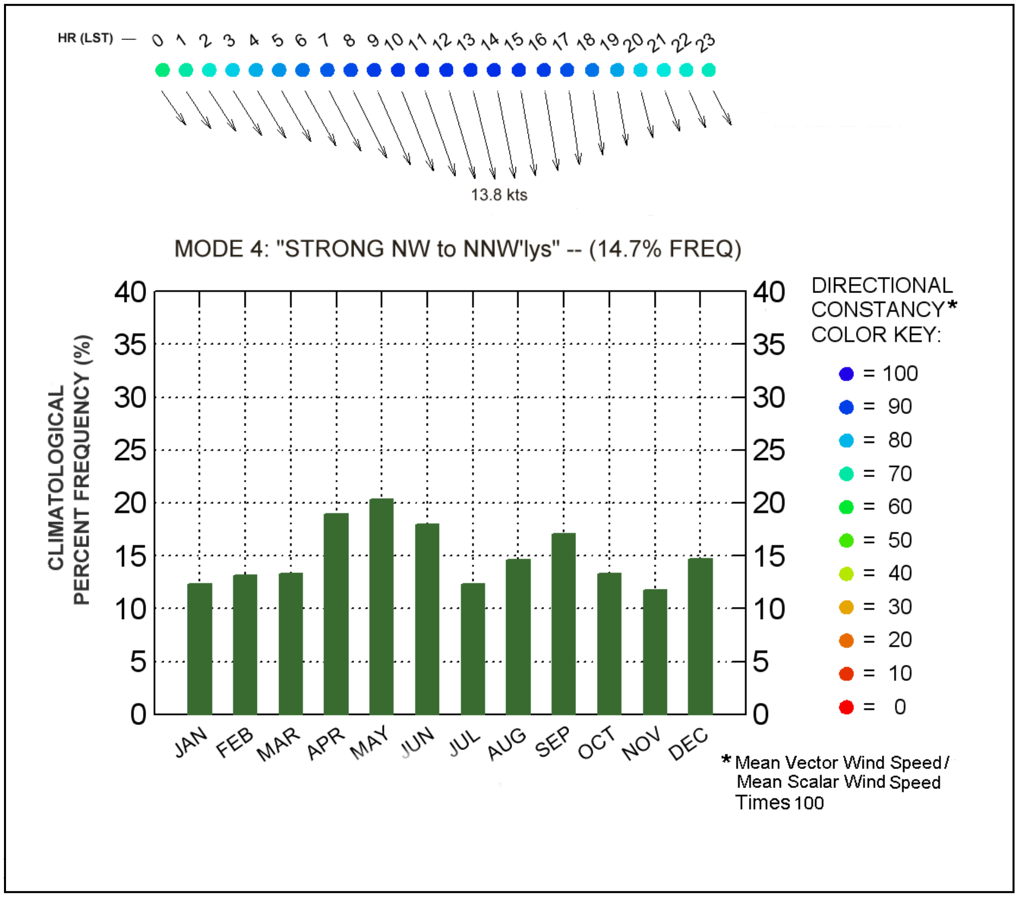

Figure 14: Calgary #4 Ranking Mode – Strong Post-Cold Frontal Passage.

Mode 4 (overall frequency: 14.7 %) relates to the post-frontal character experienced after a strong cold front, with northwesterlies and north-northwesterlies prevalent midnight-to-midnight. There is also a noticeable diurnal component as well. Mean vector magnitudes reach as high as 13.8 knots at 1300 LST (corresponding mean scalar speed 15.0 knots). The high-three individual month frequencies are for the adjacent months April, May and June (17 % to 20 %), not really surprising as the spring months April and May, in particular, are known for their occasionally blustery episodes in the mid-latitudinal areas of North America.

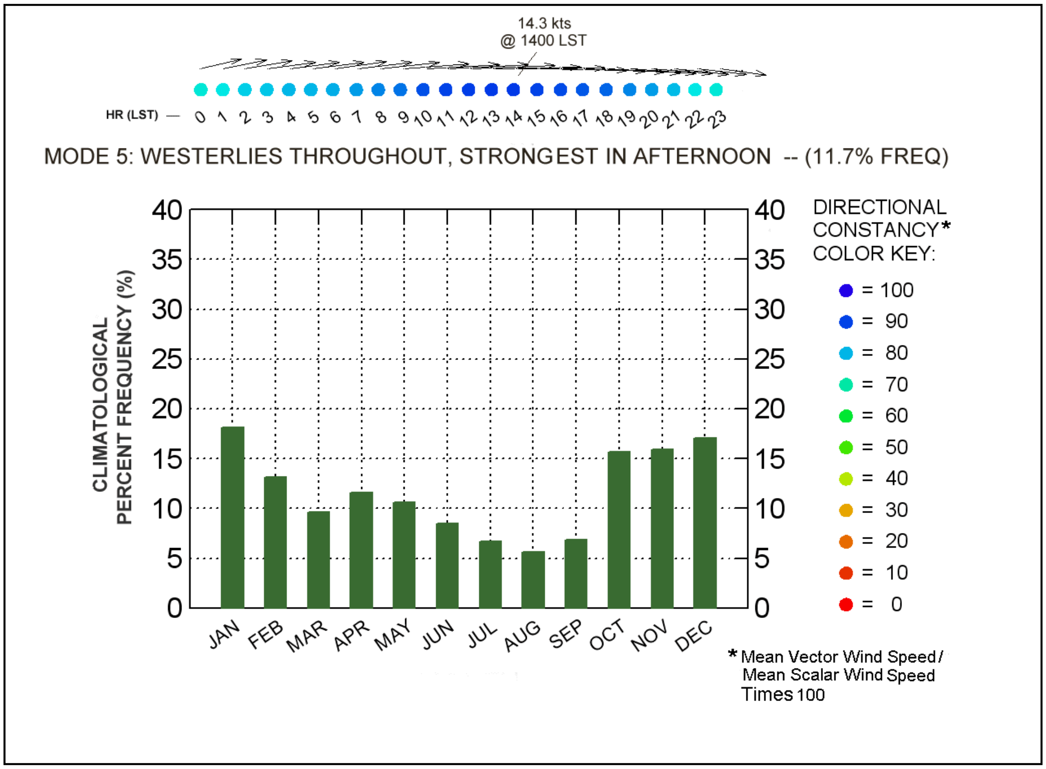

Figure 15: Calgary #5 Ranking Mode – Strong Westerlies.

Mode 5 (overall frequency: 11.7%) is similar in magnitude character to Mode 4 (see Figure 14), except that the orientations are westerly. The maximum magnitude vector (14.3 knots, even higher than Mode 4's), is noted for 1400 LST, associated with a mean scalar speed of 15.7 knots. Maximum individual months' frequencies (between 15 % to 18 %) are noticeably higher for the contiguous months October thru February, This strong westerly mode undoubtedly includes those warm and dry Chinook cases, which are known to be colder-season preferred and accompanied by very strong westerly winds on occasion. Collection of diurnal relative humidity and temperature information for those cases assigned to Calgary cluster#5 would enable identification of those actual Chinook occurrences.

4.4. Toronto: Lat: 43.68 N, Long: 79.63 W, Elevation: 569 ft

Resultant Wind Cluster Analysis of the Toronto data set, that with the most available observations (more than 19000), produced six clusters. Toronto's location bordering on Lake Ontario results in a land-breeze/lake-breeze diurnal mode, not unlike those land-breeze/sea-breeze patterns for Victoria and Vancouver. Toronto's version is also the station's most prominent. The six Toronto layouts appear below in descending order of importance (Figures 16 thru 21, respectively).

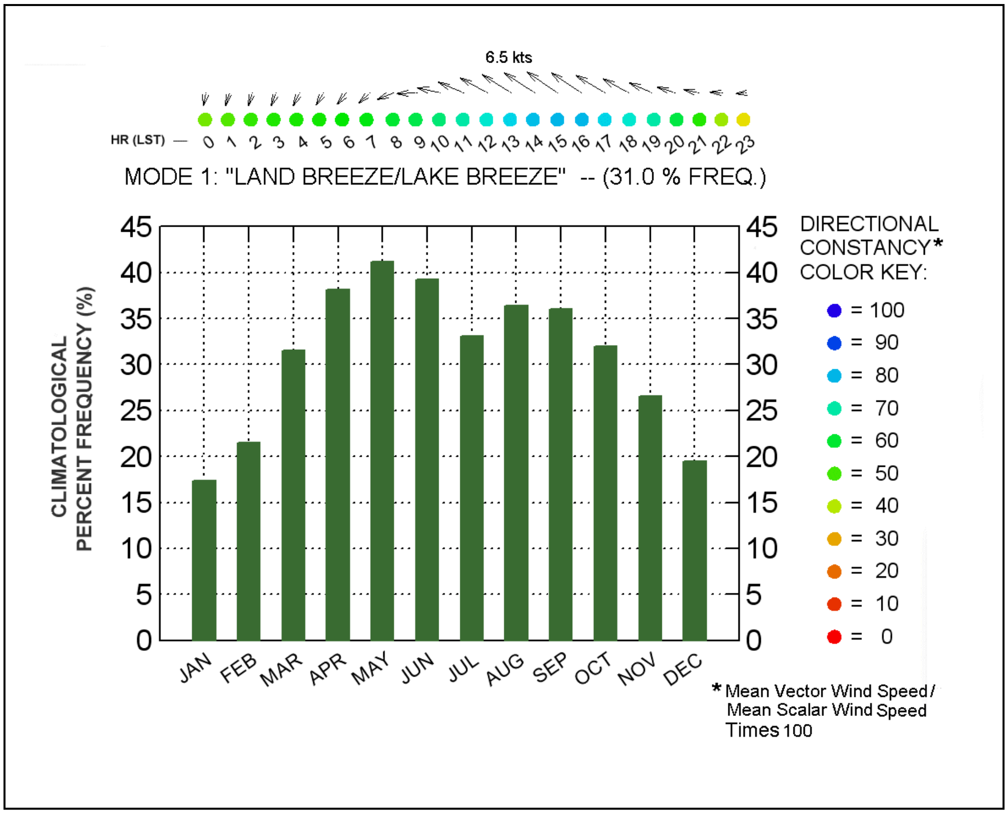

Figure 16: Toronto #1 Ranking Mode – Land-Breeze/Lake-Breeze.

Mode 1, the Land-Breeze/Lake-Breeze pattern (31.0% incidence) is the most frequent by a large margin. It is characterized by light north-northeasterly offshore flow thru about 0700 LST, succeeded by east-southeasterlies to southeasterlies off the Lake. The winds reach a peak resultant magnitude of 6.5 knots at 1500 LST (associated mean scalar speed: 8.2 knots), slowly decreasing thereafter to light, more easterly orientations by late evening. Peak monthly frequencies are noted for April, May, and June (38 % to 41 % range, the lowest for December (20 %) and January (17 %).

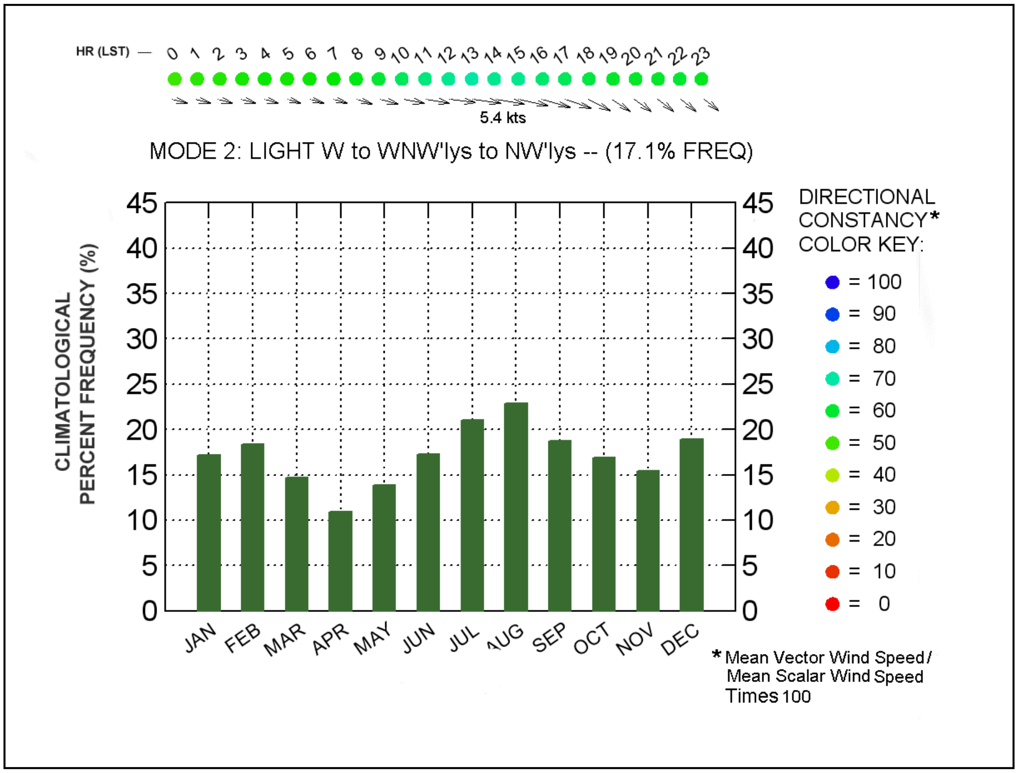

Figure 17: Toronto #2 Ranking Mode – Light West to Northwesterlies.

Mode 2 (17.1 % incidence) exhibits a gradual clockwise turning from light westerly to west-northwesterly to northwesterly over the course of the day, with modest ("green") constancy representations throughout. There is a slight summer predominance in monthly frequencies, July with 21 %, and August with 23 %. Maximum vector magnitude is 5.4 knots at 1400 LST, associated with a mean scalar speed of about 8 knots. This probably represents a relatively quiet period, significantly removed in time from a weak to moderate frontal passage.

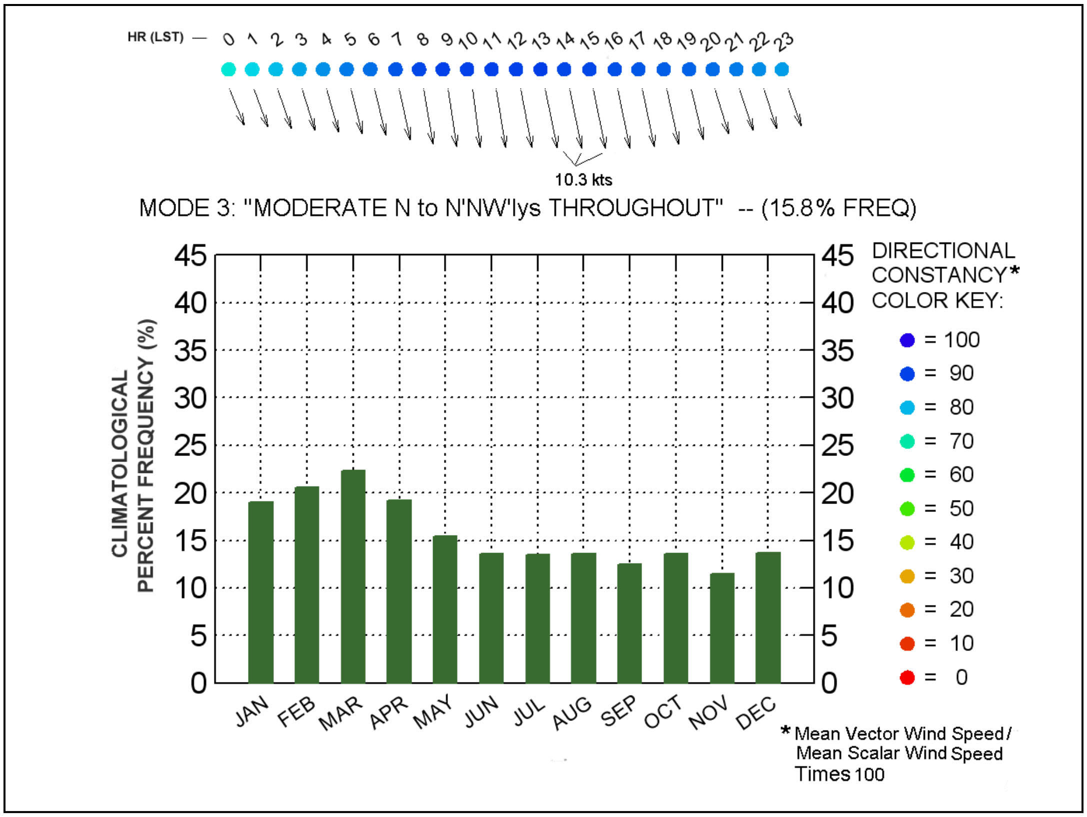

Figure 18: Toronto #3 Ranking Mode – Moderate North to NNW'lys.

Mode 3, exhibiting moderate north to north-northwesterlies midnight to midnight, is a predominant January to April pattern, the highest relative frequencies observed for these four months, the figures ranging between 19 % to 23 %. The 10.3 knot maximum mean vector wind speeds for 1300 LST, 1400 LST, and 1500 LST, each, are associated with 11 knot to 12 knot mean scalar speeds. Compared to Mode 2 (see Figure 17) this may reflect time proximity to stronger cold front passages.

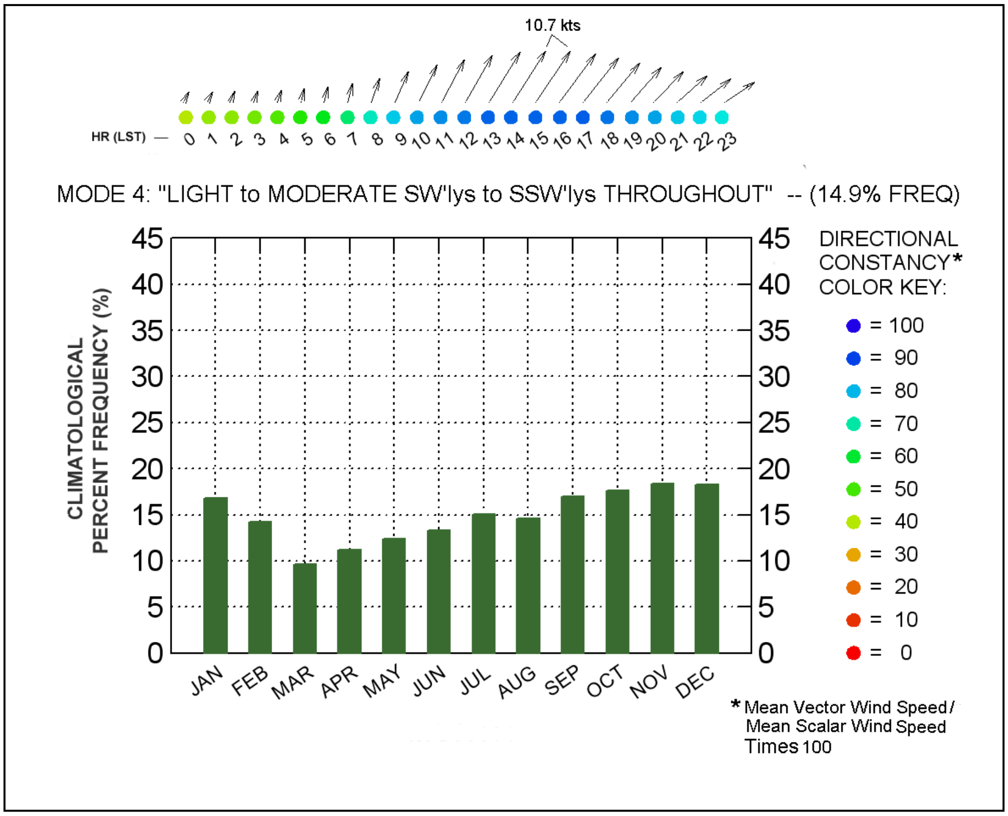

Figure 19: Toronto #4 Ranking Mode – Light to Moderate SW'lys to SSW'lys.

Mode 4 (14.9 % incidence), depicting light to moderate southwesterlies to south-south-westerlies, exhibits a slight preponderance of relative frequencies for the October thru January months (16 % to 18 %). The pattern displays more of a diurnal contrast than the previous two modes (see Figures 17 and 18), mean vector magnitudes relatively light in the early hours at low constancies, but increasing significantly in the morning and afternoon, attaining a maximum 10.7 knot magnitude (related to a 12 knot mean scalar speed). As a speculation, this may be a "fair-weather" pattern with the winds driven by return flow around the back side of a high pressure area retreating to the southeast.

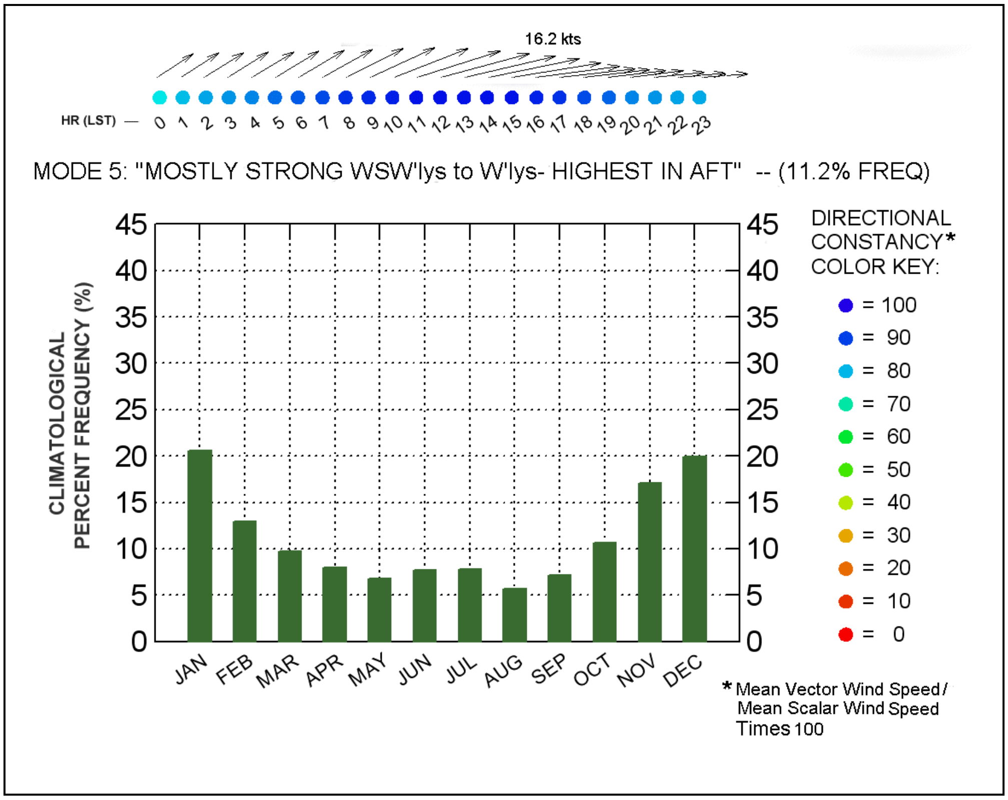

Figure 20: Toronto #5 Ranking Mode – WSW'lys to Westerlies.

Mode 5 (11.2 % incidence) is distinguished by the mostly high magnitude west-southwesterly to westerly mean vectors midnight-to-midnight, in particular the afternoon, exemplified by a 16.2 knot west-southwesterly figure at 1300 LST (17.0 knot corresponding mean scalar figure). In relative frequency terms, this is principally a winter pattern, with the December to February figures dinstictively higher (17% to 20%) than those of most of the other months; summer months' (June to August) figures are only in the 5% to 7% range.

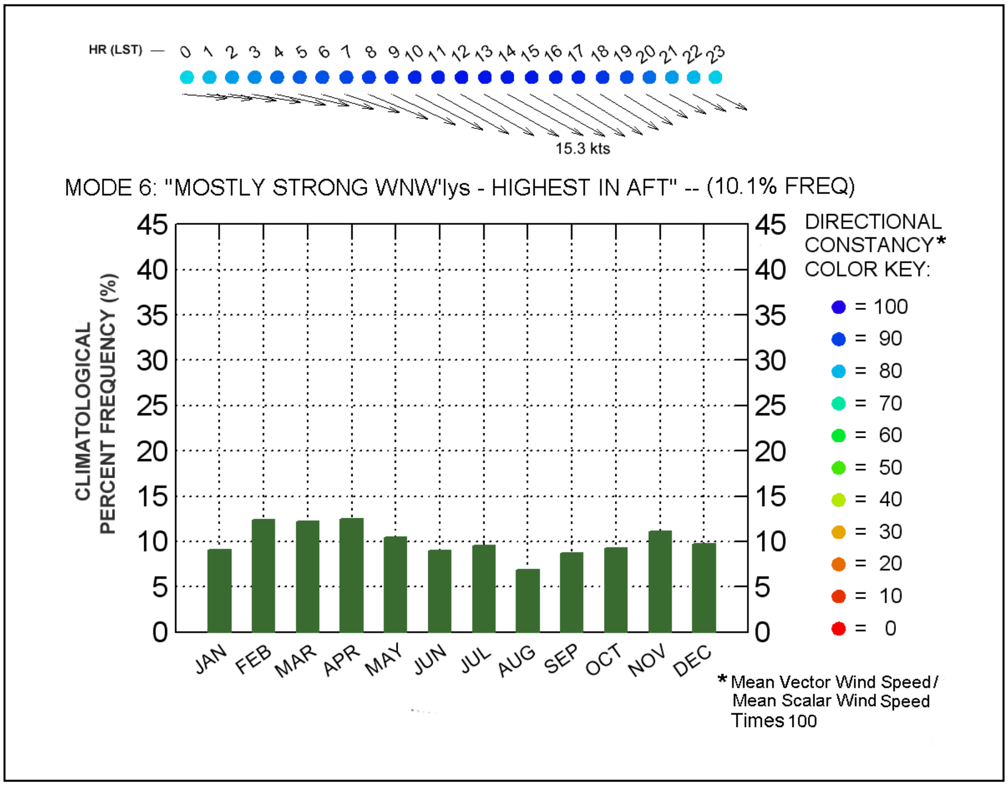

Figure 21: Toronto #6 Ranking Mode – Strong WNW'lys in Afternoon.

Mode 6 (10.1 % frequency) exhibits a clear diurnal contrast, with moderate west-northwesterly vector magnitudes in the forenoon picking up significantly in speed for the afternoon, reaching a peak magnitude 15.3 knots at 1400 LST (scalar mean speed conversion: 16.3 knots). The adjacent months February, March, and April have a very slight preponderance of relative frequencies (12 % for February and March, 13% for April). The relatively high magnitudes and west-northwesterly vector orientations suggest that this is at least in part the signature of a strong post-frontal episode.

4.5. Montreal: Lat: 45.47 N, Long: 73.63 W, Elevation: 98 ft

The Montreal data set also yielded six clusters, ranging in overall frequencies from 26.6 % to 9.9 %. Compared to Toronto, with no lake effect and a greater proximity to the North Atlantic, the idealized pattern hierarchy is noticeably different. As before, the six Montreal layouts appear below in descending order of importance (Figures 22 thru 27, respectively).

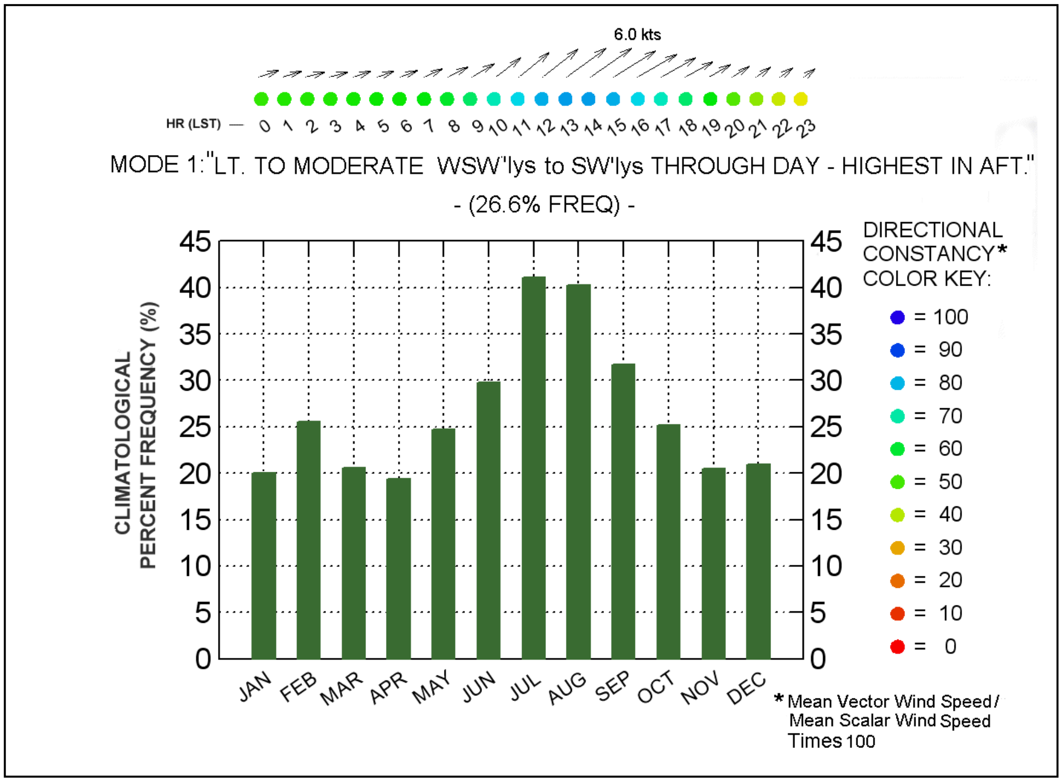

Figure 22: Montreal #1 Ranking Mode – Light WSW'lys to SW'lys.

Mode 1 (overall frequency: 26.6%) is clearly a warm season preferred pattern with the highest individual monthly frequencies visible for the contiguous months June to September – especially July and August, with 41 % and 40 % figures, respectively. Mean vector magnitudes exhibit an afternoon maximum, a 6.0 knot southwesterly noted for 1400 LST (7.4 knot mean scalar speed conversion). This mid-summer peak undoubtedly reflects the flow around the westward extent of the Azores-Bermuda High Pressure area, a semi-permanent feature in summer.

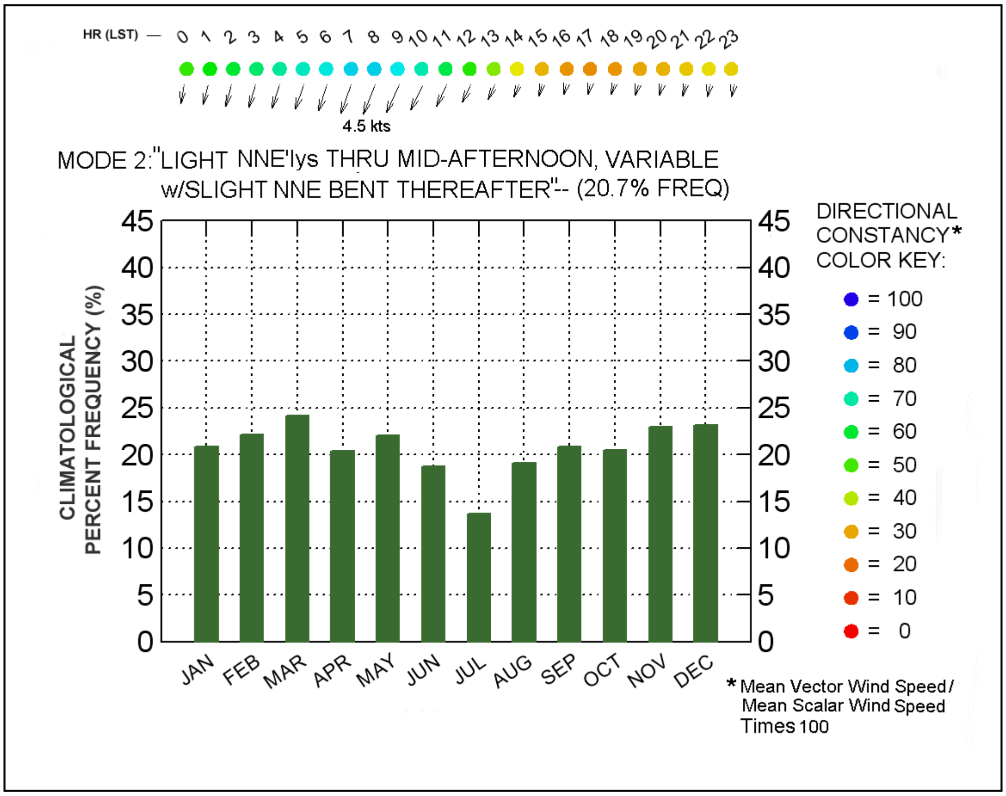

Figure 23: Montreal #2 Ranking Mode – Light NNE'lys in Forenoon.

Mode 2 (overall frequency: 20.7 %) displays light north-northeasterly-oriented vectors at moderate (green to light blue icons') constancy levels in the forenoon hours, the northeasterly bent retained into the afternoon and evening, but associated with weaker constancy (orange to yellow icons') levels. With the exception of July's dip to 14 %, the individual monthly frequencies are quite uniform, ranging from 18 % (June) to 24 % (March). Highest mean *scalar* speed for this mode is only 5.8 knots, calculated for several different hours.

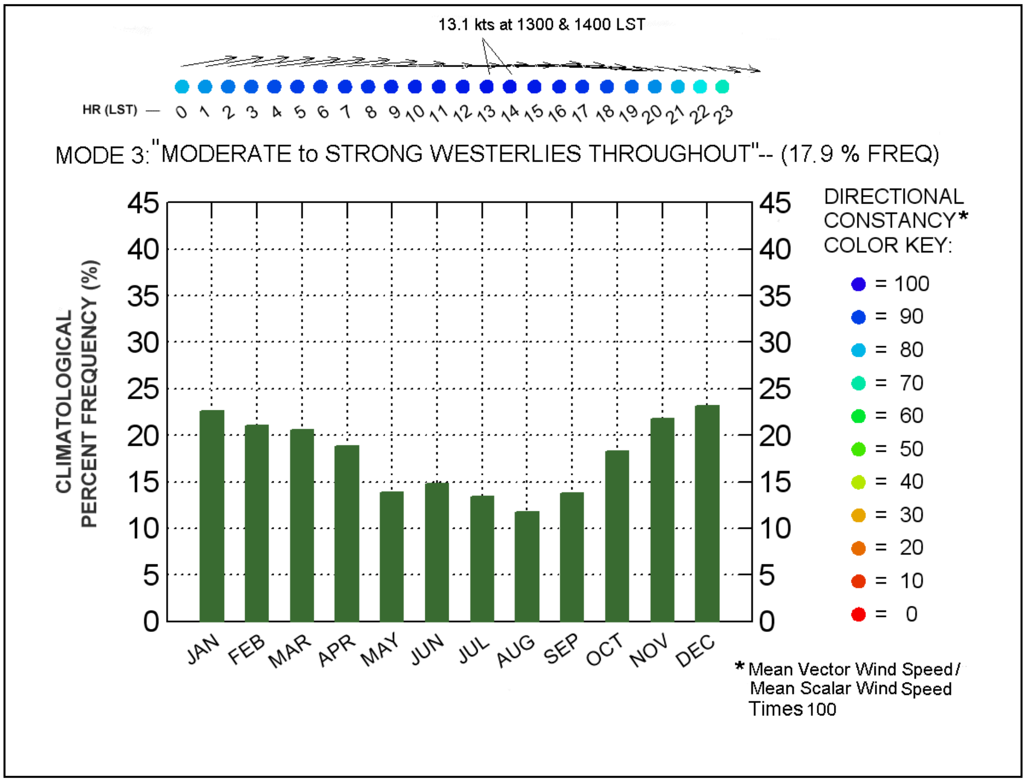

Figure 24: Montreal #3 Ranking Mode – Moderate to Strong Westerlies.

Mode 3 (overall frequency: 17.9 %) portrays a pattern of moderate to strong westerlies persisting throughout the day, the constancy levels at light to dark blue icon levels throughout as well. Cold-season preferred, the maximum individual monthly frequencies encompass the contiguous months November to March (20 % to 25 % range), the minimum figure (12 %) seen for August. Highest mean vector magnitudes (13.1 knots) are realized at 1300 LST and 1400 LST, associated with ~14 knot mean scalar speeds.

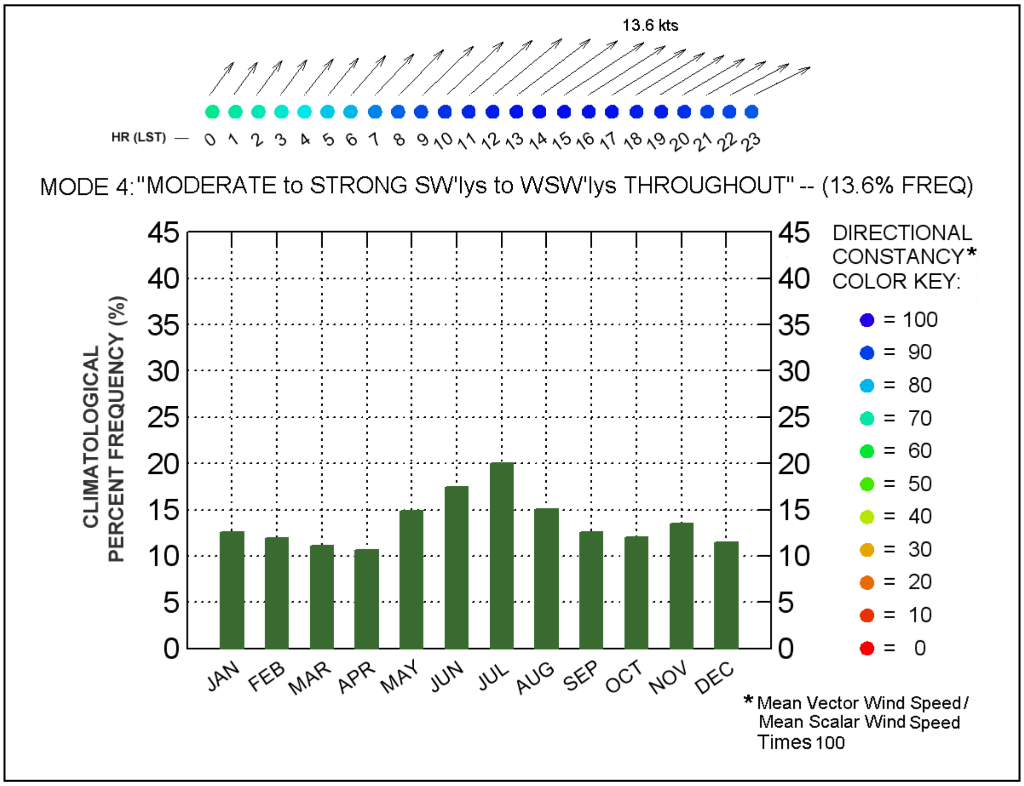

Figure 25: Montreal #4 Ranking Mode – Strong SW'lys and WSW'ly's.

Mode 4 (overall frequency: 13.6 %) is essentially a higher magnitude version of Mode 1 (see Figure 22), with stronger southwesterlies and west-southwesterlies midnight-to-midnight, roughly double the magnitudes of Mode 1. Highest individual monthly frequency (20 %) is noted for July, the lowest (11 %) for April. Maximum magnitude mean vector (13.6 knots) is for 1500 LST, associated with a mean scalar figured of 14.3 knots. Based on the results here along with Mode 1, a generalization that could be made is that the month of July in Montreal has predominant light to strong west-southwesterly to southwesterly flow about 60 % of the time, midnight-to-midnight.

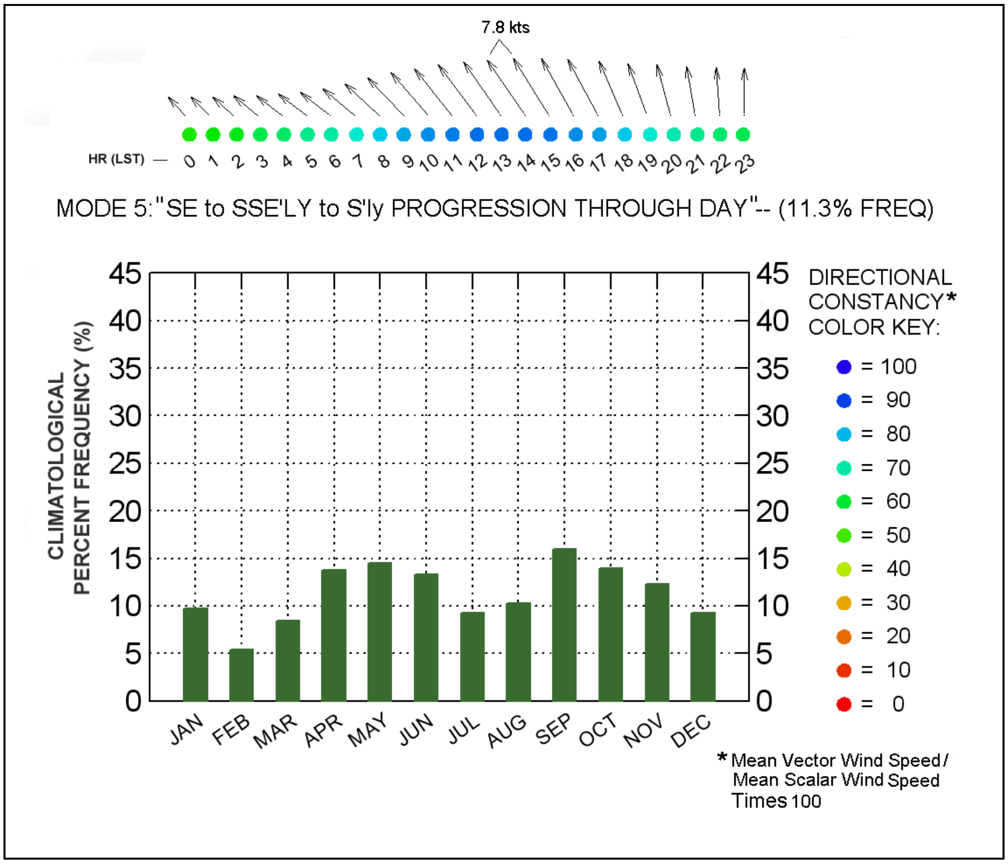

Figure 26: Montreal #5 Ranking Mode – SE'ly to SSE'ly to S'ly Progression.

Mode 5 (overall frequency: 11.3 %) traces a clockwise progression from southeasterlies to southerlies over the course of the day, at light to moderate magnitudes. There is a noticeable diurnal component, the afternoon magnitudes noticeably higher than those in the morning, and at blue-icon high constancy levels. Maximum mean vector magnitude (7.8 knots) is attained for 1400 LST and 1500 LST, each (corresponding to mean scalar figures at ~9 knots). The distribution of monthly relative frequencies shows a slight preponderance for April-June and September-November (all in the 13 % to 16 % range).

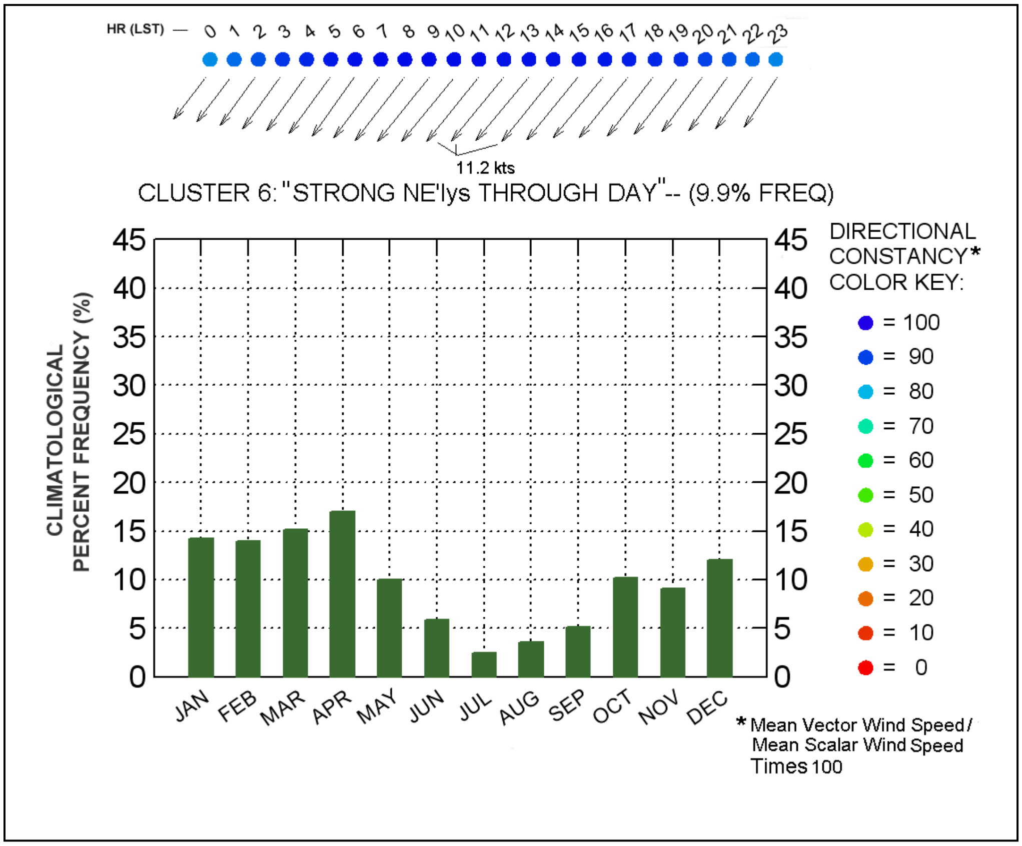

Figure 27: Montreal #6 Ranking Mode – Strong NE'lys throughout day.

Mode 6 (overall frequency: 9.9 %) is to some extent a high magnitude version of Mode 2 (see Figure 23). The mean vectors are northeasterly throughout the day, but with little contrast in the relatively strong magnitudes displayed through the afternoon, morning, and evening. Highest vector magnitudes, 11.2 knots, corresponding to mean scalar figures of about 12 knots, are observed for several adjacent hours confined to the late morning and early afternoon (1100 LST, 1200 LST, and 1300 LST). This pattern's relative incidence is concentrated over the four months January to April, with frequencies ranging from 14 % to 17 %). July and August incidences are both less than 5 %.

4.6. Victoria and Vancouver Two-Station Analysis

Next, a demonstration is provided of a two-station simultaneous analysis, that for Victoria and Vancouver. The initial treatment produced only three clusters, so to refine the results and arbitrarily match the number of clusters generated for the two stations individually, the software was forced to create five. Some 69 % of the cases were assigned to the two most prominent.

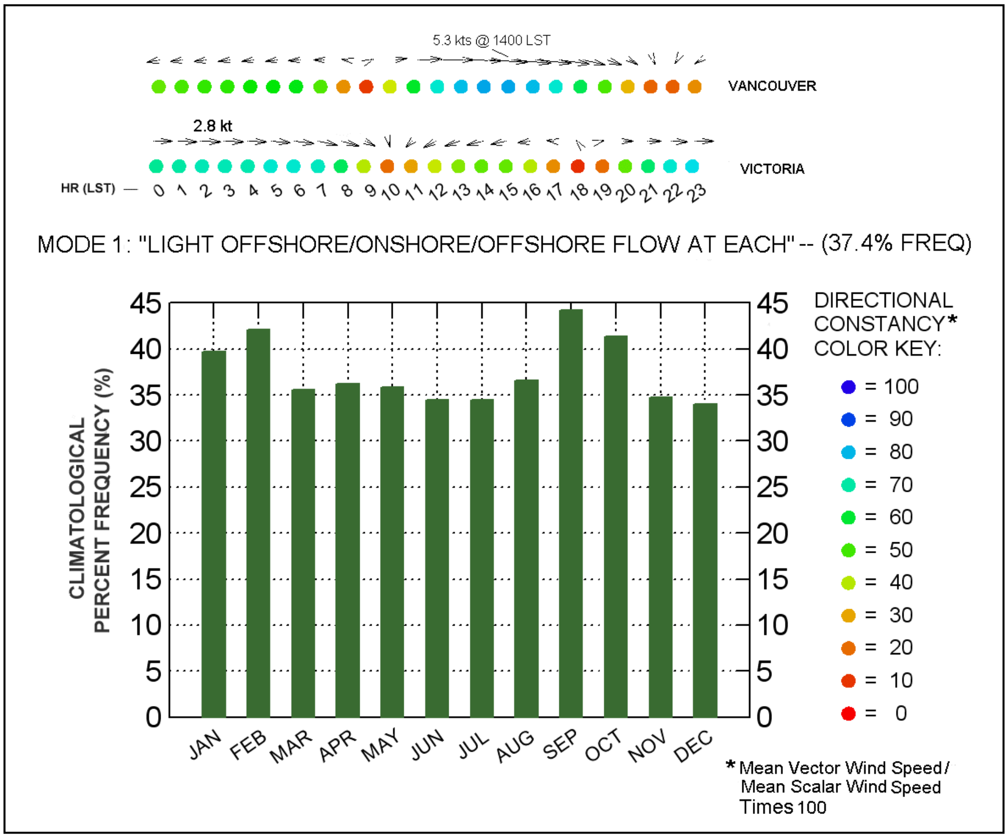

Figure 28: Victoria/Vancouver #1 Ranking Mode –Light Offshore/Onshore/Offshore Flow at Each Station.

Mode 1 (overall frequency: 37.4 %) reflects those cases where land-breeze/sea-breeze influences predominate concurrently at the two stations. Quite interestingly, the orientations and onset/termination times are significantly different, along with the times and durations of the transition intervals. Among other things, the idealized sea-breeze flow at Vancouver is about eight hours long, that at Victoria just four. Monthly percent frequencies range from nearly 45 % (September) to just above or below 35 % (March-July) and November-December).

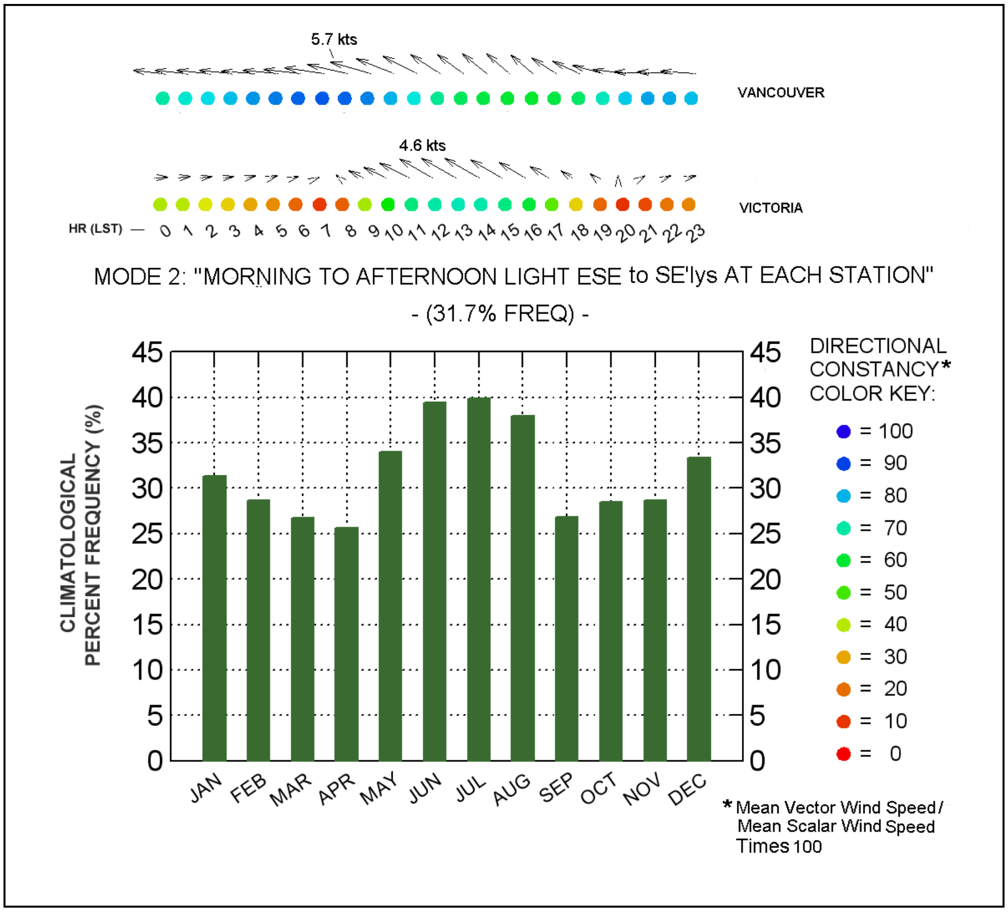

Figure 29: Victoria/Vancouver #2 Ranking Mode –Morning to Afternoon Light ESE'lys to SE'lys at Each Station.

Mode 2 (overall frequency 31.7 %), shows another set of interesting features contrasts, a bit more complex than Mode 1's (see Figure 28). At Victoria, light and variable vector orientations (yellow to orange constancy icons) are evident for the majority of the hours, about half, however, with a net westerly bent. At Vancouver, the orientations are light easterly from midnight to 0600 LST, and 2000 LST thru 2300 LST, but light east-southeasterly to southeasterly and then back again for 0700 LST to 1900 LST. In contrast with Victoria, there are no low constancy transition periods or hours suggestive of light or variable vector character. There is a clear predominance of the highest relative frequencies for the summer months, June to August (incidence percentages between 38 % and 40 %).

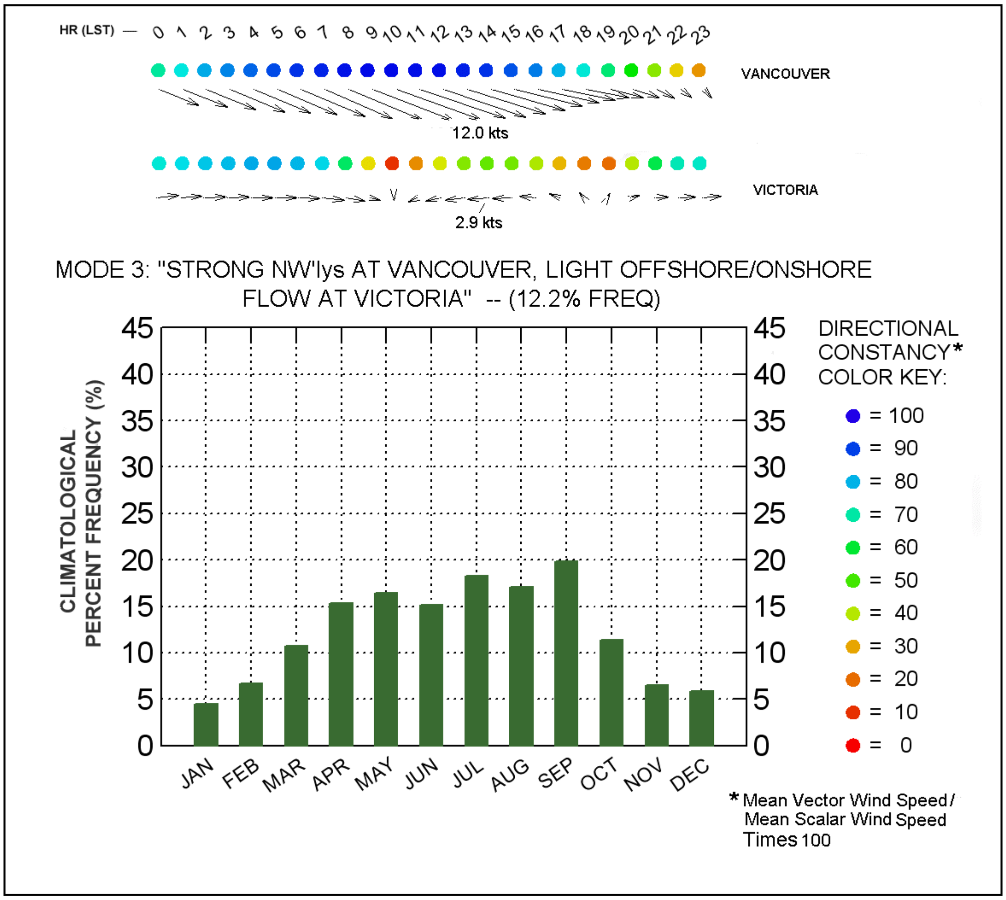

Figure 30: Victoria/Vancouver #3 Ranking Mode – Strong NW'lys at Vancouver, Light Offshore/Onshore Flow at Victoria.

Mode 3 (overall frequency: 12.2 %) portrays another contrast in features – this one very striking. Vancouver shows a pattern of strong northwesterlies for most of the day, including a maximum mean vector magnitude of 12.0 knots at 1000 LST. Victoria, in great contrast, displays a light westerly land-breeze/easterly sea-breeze regime not like that of Mode 1 (see Figure 28), the topography of Vancouver Island obviously having shielded it from the northwesterlies impacting Vancouver. This pattern shows its greatest relative incidence for the contiguous months April thru September (15% to 20% frequencies). September has the 20 % absolute maximum.

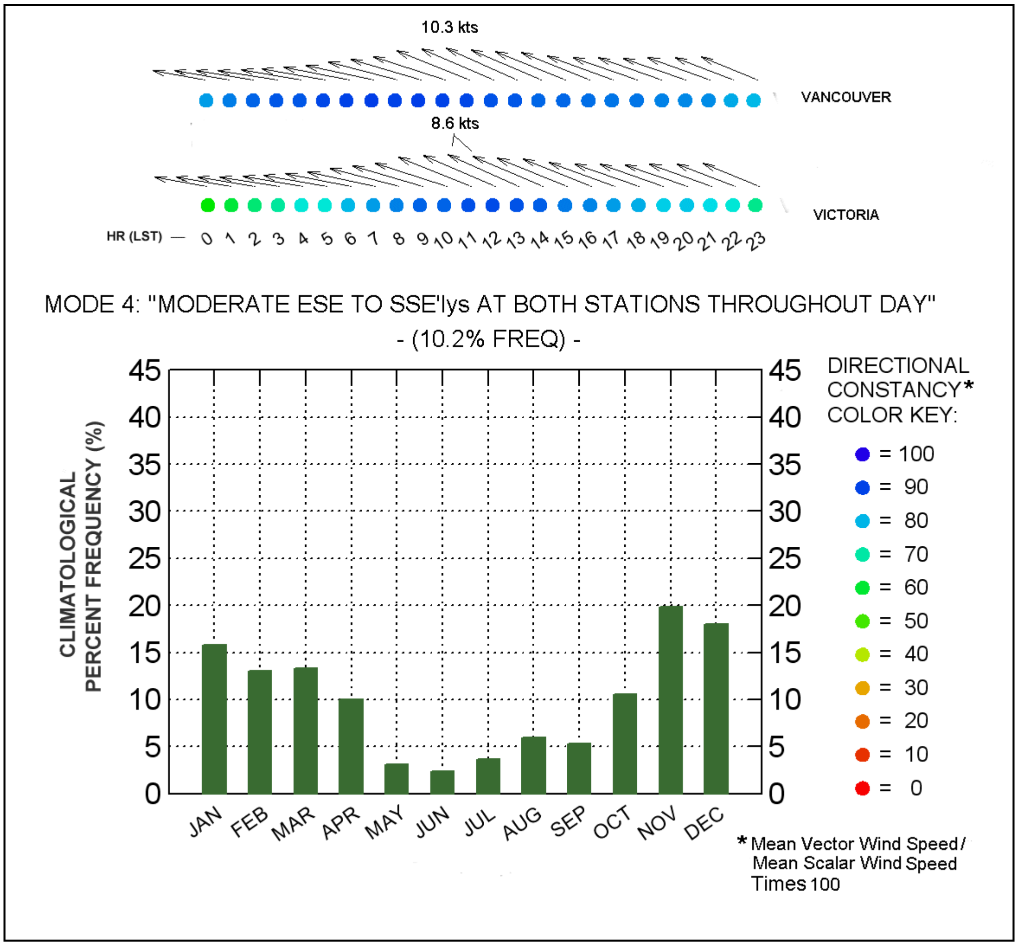

Figure 31: Victoria/Vancouver #4 Ranking Mode – Moderate ESE'ly to SSE'ly Flow at Each Station.

Mode 4 (overall frequency: 10.2 %), in contrast, exhibits very similar mean vector configurations for the two stations, namely moderate east-southeasterly to south-south-easterly orientations at moderate to high constancies midnight-to-midnight, indicative of a more synoptically influenced pattern. This is a preferred colder season regime, the November-January relative frequencies between 15 % to 20 %.

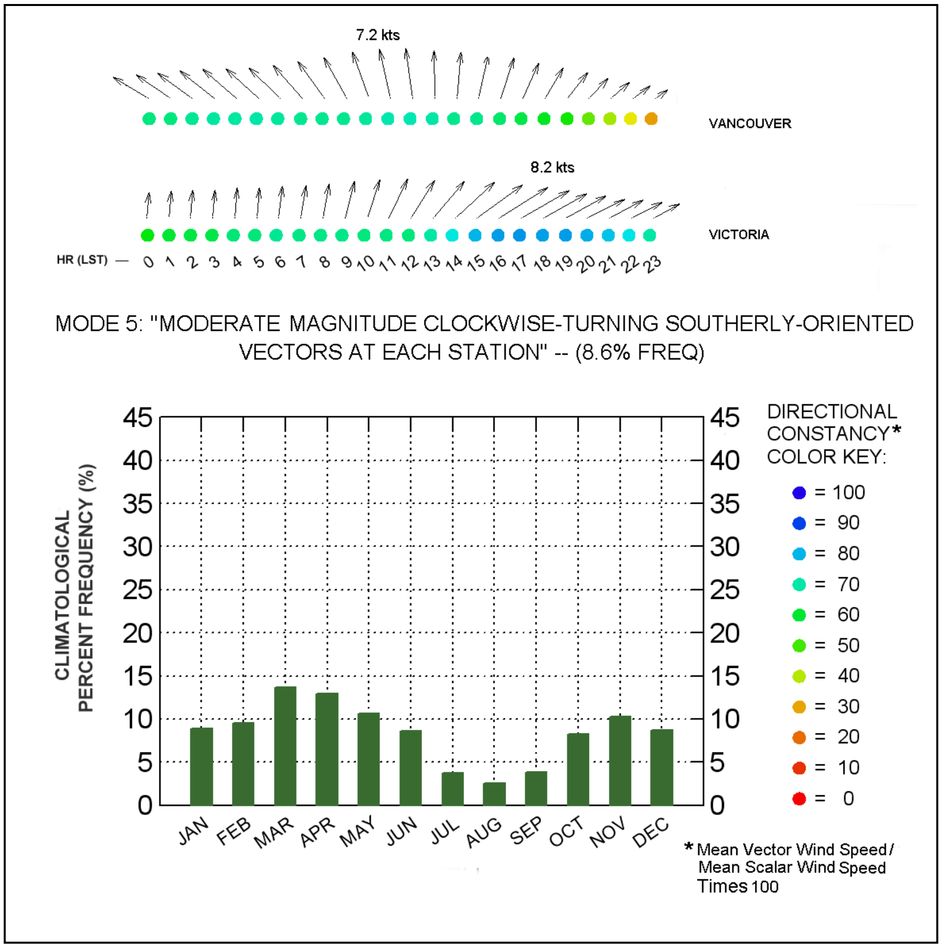

Figure 32: Victoria/Vancouver #5 Ranking Mode – Moderate ESE'ly to SSE'ly Flow at Each Station.

Mode 5 (overall frequency: 8.6 %) displays also a relatively close inter-station similarity in mean vector orientations and magnitudes, with pronounced southerly components for each, the Vancouver ones, however, more southeasterly than Victoria's southerlies for the forenoon hours. Vancouver's afternoon vectors are essentially southwesterly, while Victoria's are more west-southwesterly. This is a Spring preferred pattern, maximum relative frequencies covering March to May (14 %, 13 %, and 11 %, respectively). The Vancouver magnitudes are also noticeably less than those in Mode 4 (see Figure 31).

4.7. A Visual Scrutiny of the "One-Size-Fits-All" Approach – Four Resultant Wind Hour-by-Month Climograms presented for the Two Victoria Land-Breeze/Sea-Breeze Patterns, the Vancouver Land-Breeze/Sea-Breeze Pattern, and the Toronto Land-Breeze/Lake-Breeze Pattern

Addressing the one-size-fits-all hypothetical described in the Methodology (See Section 2), this concerning the performance of the wind clustering analyses on all calendar months as a unit, Figures 33 to 36 are resultant wind hour-by-month Climograms for Victoria's two land-breeze/sea-breeze modes, Vancouver's version of the same, and Toronto's land-breeze/lake-breeze pattern, respectively. The Climogram methodology dates from [Fisk, 2004] and the data come from the stations' four subsets which include the observations assigned to the particular clusters of interest. Like those in Figures 1 to 32, the Climogram hour-by-month vector orientations are depicted by arrows oriented in the direction of flow, their lengths proportional to their magnitudes. For continuity purposes the 0000 LST vectors near the left edge are repeated as 2400 LST vectors near the right edge, and to add a diurnal physical perspective, sunrise/sunset demarcation traces are also overlain. The diurnal orientations of the vectors, by month, in Figures 33 to 36 can be compared with those in the layouts, in this case Figures 1, 2, 6, and 16, respectively. In contrast, with Figures 1 to 32, though, the Climogram constancy colorizations are reversed, ranging from blue (lowest values) to red (highest values). Thus, without going into great detail, it is apparent that there are some intra-mode feature dissimilarities across the hours/months of the four. This is not particularly surprising, as offshore/onshore circulation regimes in these cases are driven by solar insolation influences, which themselves are affected by time of year (length of day and sun angle). The most prominent signatures in this regard are the arc-like constancy feature contrasts between the red (high) and light/dark blue (low) areas, the latter depicting changeover periods in direction; also, on occasion, noticeable colder season vs. warmer season contrasts in vector orientations are discernible. While it is useful to present these differences, they do not appear drastic, and subjectively seem to reinforce the parsimony trade-off motivation described in Section 2. In any case, the Resultant Wind Hour-by-Month Climogram tool demonstrated here could serve as a useful supplement - an alternative to follow-up exhaustive and time-consuming individual month treatments. It should also be said that other one-size-fits-all modes with less inherent diurnal (synoptic-scale oriented) variability would show considerably less month-to-month variability in their own Climograms.

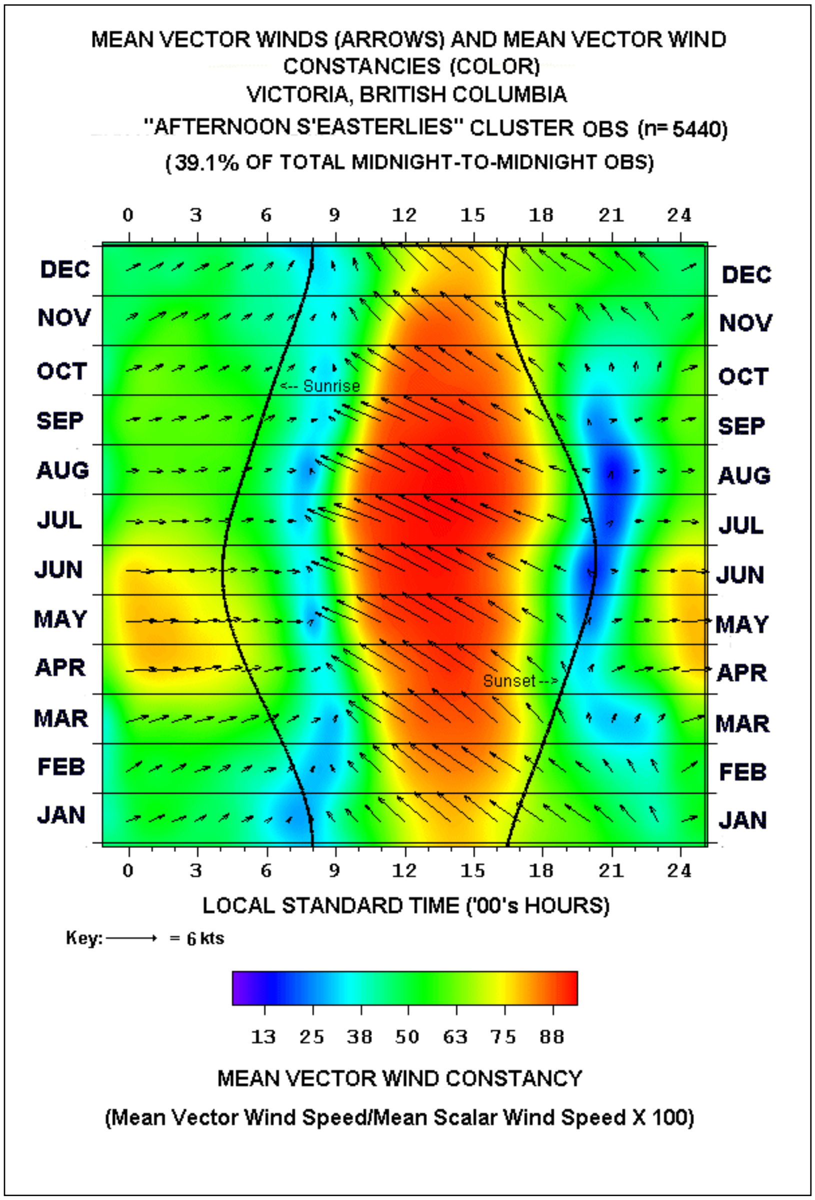

Figure 33: Resultant Wind Climogram for Victoria Mode 1 (See Figure 1).

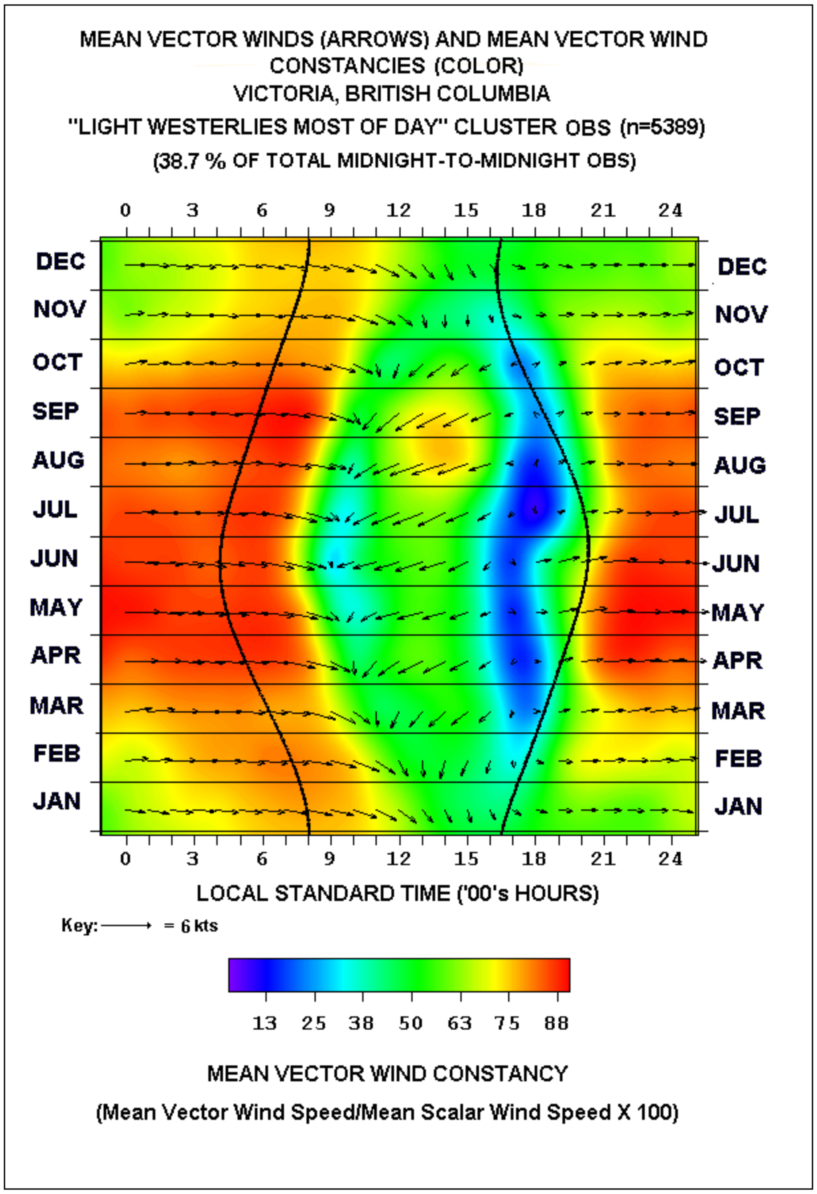

Figure 34: Resultant Wind Climogram for Victoria Mode 2 (See Figure 2).

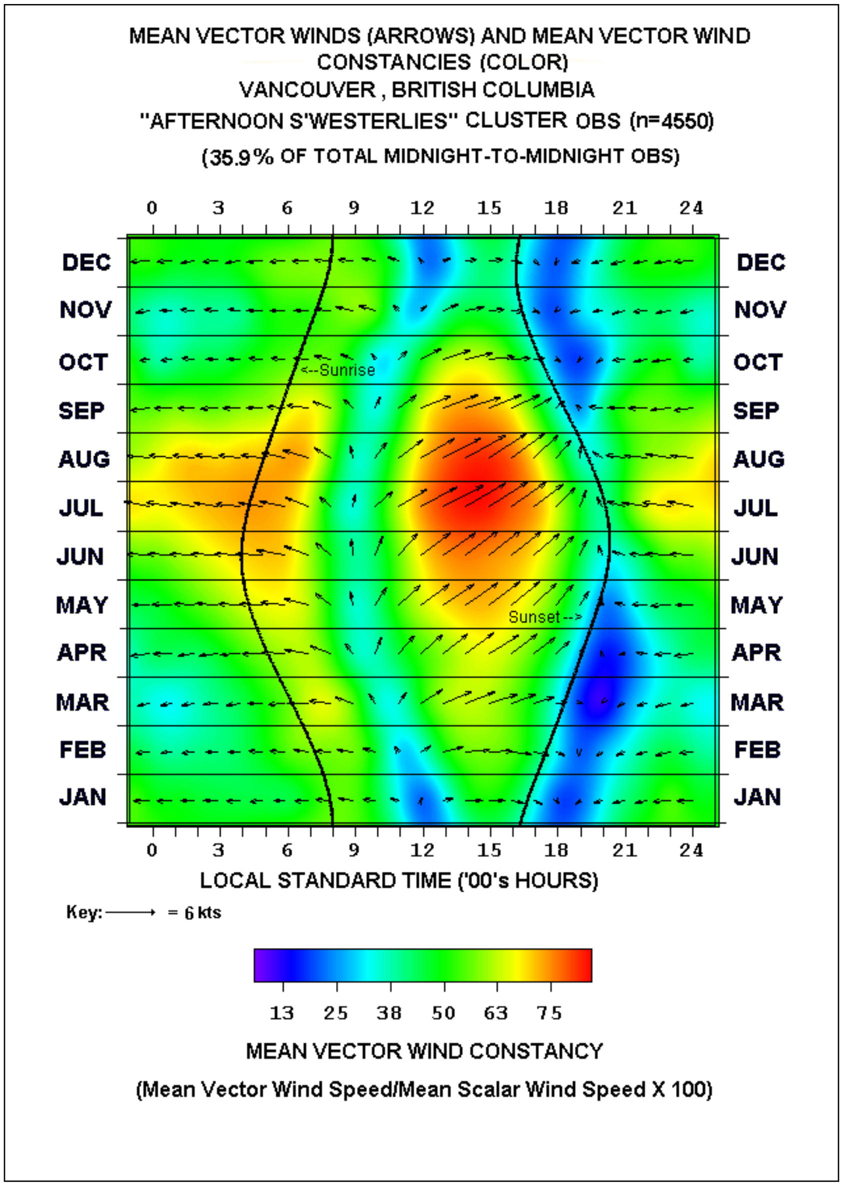

Figure 35: Resultant Wind Climogram for Vancouver Mode 1 (See Figure 6).

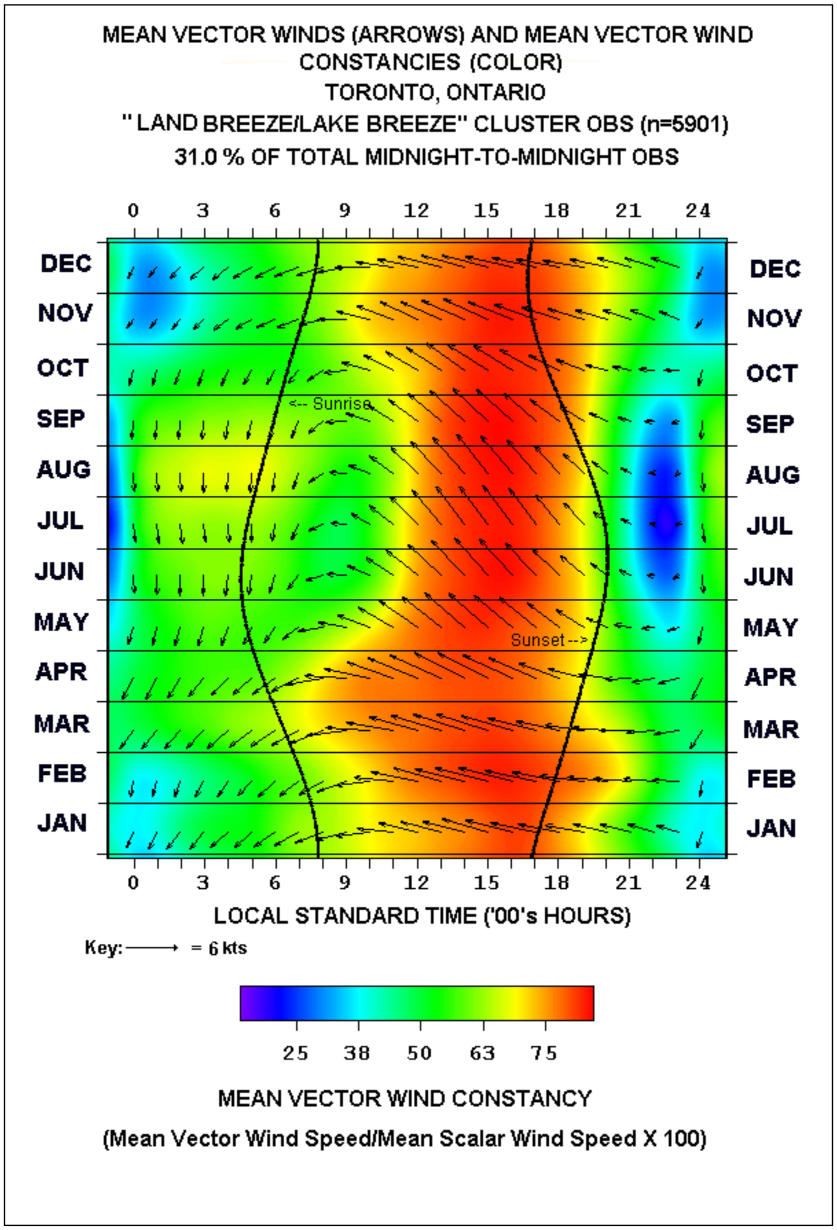

Figure 36: Resultant Wind Climogram for Toronto Mode 1 (See Figure 16).

4.8. Summary and Conclusion

The foregoing demonstrated K-Means clustering accompanied by a special add-on software feature, the V-Fold Cross Validation Algorithm, as a means of identifying optimally numbered climatological resultant wind patterns for five Canadian meteorological stations, the key distinction relative to more traditional single-hour approaches being that the analytical units are self-contained “encapsulated” *progressions* of resultant wind statistics, derived from reconstructed east-west (“u”) and north-south (“v”) centroid statistics. The clustering methodology seemed to capture effectively the idealized diurnal offshore-onshore patterns for Victoria and Vancouver (land-breeze/sea-breeze), and Toronto (land-breeze/lake-breeze). Useful insights were also provided on the modes for concurrent Victoria/Vancouver observations, including the contrasting land-breeze/sea-breeze vector orientations and onset/termination times between the two stations. In other cases where additional type influences predominated, signatures such as higher magnitude vectors during the afternoon (solar insolation effects), frontal influences (pre-frontal, frontal, and post-frontal; for example, Calgary), higher speed versions of the same mode (re: Montreal northeasterlies), Azores-Bermuda High summertime influences (re: Montreal), and Chinook influences (re: Calgary westerlies) were described.

5. Acknowledgements

Burman, P., 1989,. "A Comparative Study of Ordinary Cross-Validation, V-Fold Cross Validation, and the Repeated Learning Testing Methods", *Biometrika, Volume 76, Issue 3.*

Fisk, C.J., 2004. "Two-Way ("Hour-Month") Time-Section Plots as a Tool for Climatological Visualization and Summarization", *American Meteorological Society 14th Conference on Applied Climatology, Seattle.*

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", *American Meteorological Society 10th Conference on Artificial Intelligence Applications to Environmental Science, New Orleans.*

Fisk, C.J., 2016. "Identification and Analysis of Contiguous Hours' Climatological Wind Pattern Modes Utilizing K-Means Clustering Analysis Combined with the V-Fold Cross-Validation Algorithm", *American Meteorological Society 23rd Conference on Probability and Statistics in the Atmospheric Sciences, New Orleans.*

Fisk, C.J., 2017. "Identification and Analysis of Contiguous Hours' Climatological Wind Occurrences for Two Nearby Coastal and Mountain Stations in Southern California", *American Meteorological Society 15th Symposium of the Coastal Environment, Seattle.*

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