

INTRODUCTION TO THE CART MODEL

CART stands for Classification and Regression Trees. This powerful statistically based model establishes relationships between dependent air quality (i.e. PM10 or PM25) variables and independent meteorological variables such as relative humidity, temperature, stability, precipitation, visibility, etc. The output provides trees that show which meteorological variables contribute to PM concentrations and provides the relevancy of each meteorological variable.

DATABASES USED FOR CART ANALYSIS

The following describe the databases used and a description of the variables:

Air Quality Data

- PM10 – SSI data 1980 to 2000 from AIRS that mostly consisted of every 6th day sampling in the San Joaquin Valley. Three basic sites – Fresno at Fresno Drummond, Bakersfield at a combination of Golden and California Streets, and Stockton at Stockton Hazelton Ave.
- PM25 – FRM data that began collection in 1999 from Fresno – Fresno First, Bakersfield – Bakersfield California Street, and Stockton at Stockton Hazelton.
- The resultant data set consisted, therefore, of 24-hour average PM10 or PM25 concentrations
- Some of the above data were missing and QA was done using SAS to determine any out-of-bound data. Such data were then removed (in only a few instances).

Meteorological Data

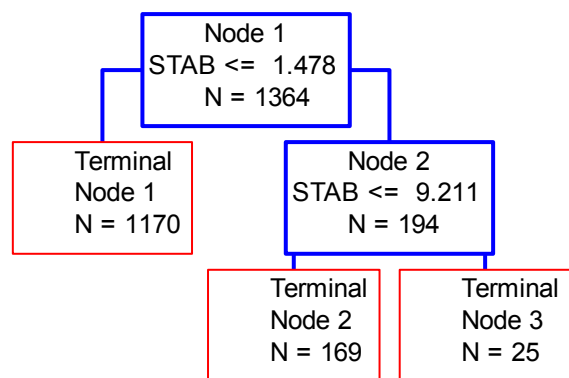
- Meteorological data from the National Weather Service (NWS) were used at the three above locations. All of the meteorological data were located at the local offices in the three cities. These data spanned from 1980 to, in some cases, 1985 to around 2000

- The upper air data was delivered from the San Joaquin Valley District. These data consisted of temperature at 850 mb (or around 5,000 feet in elevation) at Oakland. No other upper air data (such as 500 mb height, etc.) were used. These data extended from 1987 to 2000
- Quality assurance consisting of some range checks in ACCESS to manual viewing of the data. Some data had to be discarded
- The following variables were used in the analysis from the two data sources above:
 - Precipitation – daily amount calculated from summing hourly totals
 - Wetness determined as number of days since rain over quarter of an inch assigned next 7 days as “wet”– in CART, wet days were given a “1” as opposed to “0” for dry
 - Temperature – hourly temperatures were converted to an average for 24 hour period, a maximum, and a minimum
 - Wind speed – calculated average and maximum for daily values from hourly observations
 - Stability – daily stability calculated by the following formula:
 Stability = 850 mb Temperature at Oakland at 12 Zulu time –
 Surface daily minimum temperature (BAK, FRE, or STK)
 - Relative humidity – only average daily RH computed from hourly values at the three stations
 - Visibility – as obtained from NWS automatic airport nephelometers. Hourly values averaged for daily visibility average
- In summary, all meteorological variables were summarized for either daily average, minimum, or maximum

Both data sets were merged by date and location and formatted for CART model runs

CART RUNS FOR SAN JOAQUIN VALLEY

Bakersfield all seasonal periods for PM10 – approximately 1988 to 2000 for PM10 data.

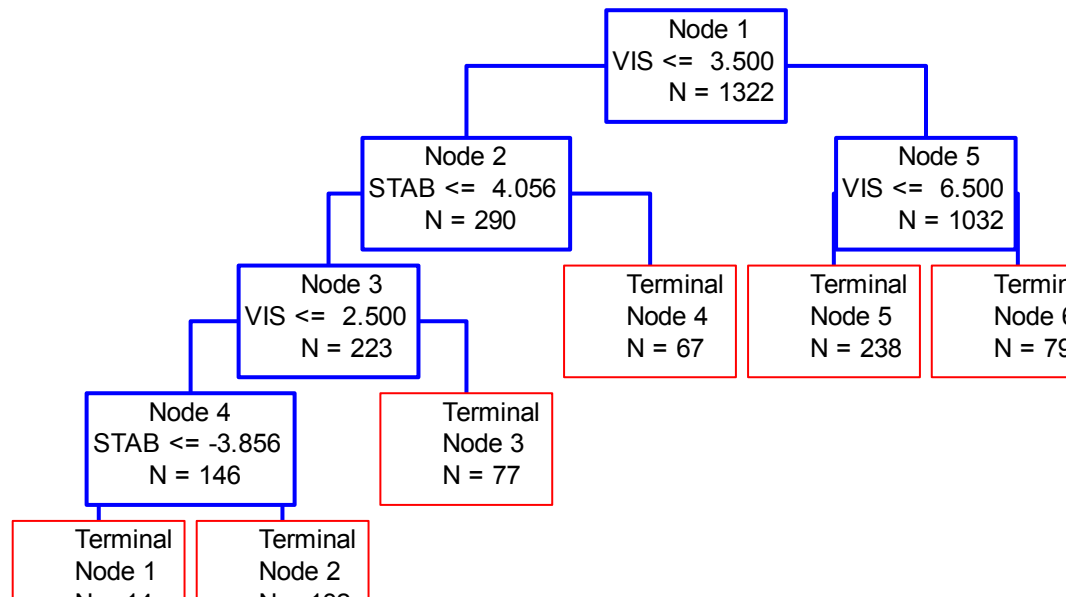


The tree splits according to one variable – stability with three nodes in all. The highest PM10 mean concentration, 95 ug/m³, occurs in terminal node 3, when stability is greater than 9.211, a very high stability index. The lowest mean PM10 concentration, 41 ug/m³, occur in lowest stability, node 1, when stability is less than 1.478. Other variables considered were mintemp, wind speed, and relative humidity. It is apparent overall that high stability, resulting from strong high pressure aloft, is the primary reason for 24-hour high PM10 concentrations at Bakersfield. Stability was 100% in relative importance and mintemp was less than 30%. RH and wind speed had 0% importance. Stability according to CART is by far the predominate variable in determining high PM10 periods.

Summary of Nodes

Node	No. of Occurrences	Bakersfield Average PM10 Concentration (ug/m**3)
1	1170	41.3`
2	169	58.0
3	25	95.1

Bakersfield all seasonal periods for PM25 – approximately 1999 to 2000 for PM25 data.

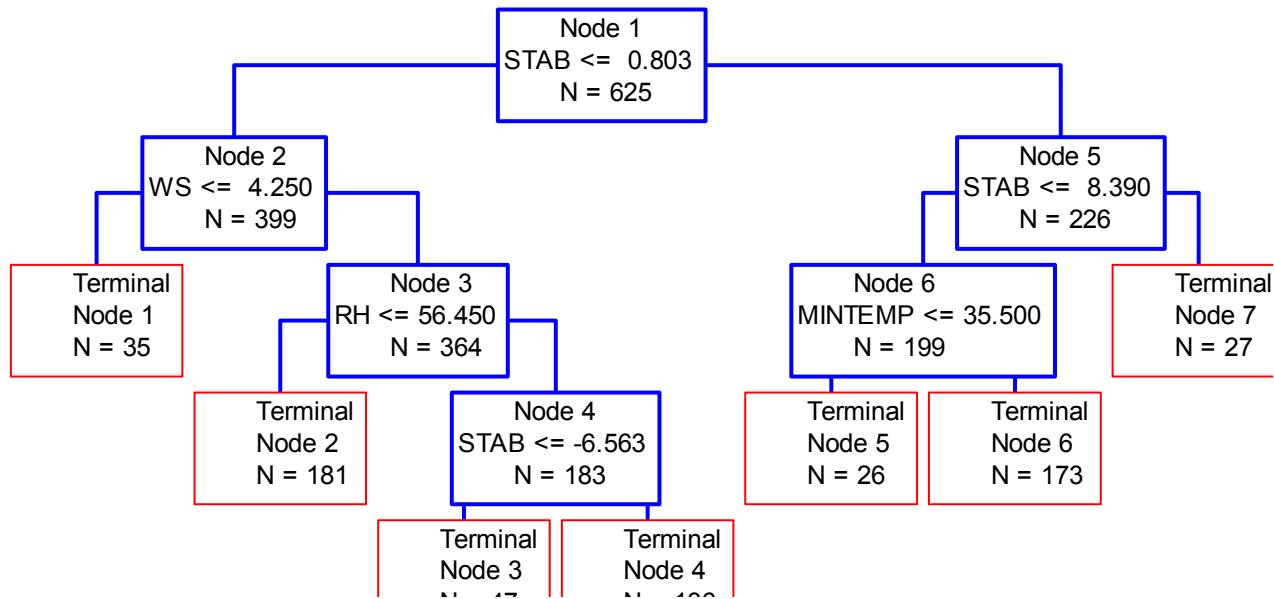


The tree for PM25 is much more complex than the PM10. Rather than stability being the most important variable, visibility is the key splitter with stability being a secondary splitter. Surprisingly, visibility was the top variable of importance (always assumed to be 100% -- a relative scale) while stability came in at near 20%. Another variable, mintemp, was near 50% but did not play any role in defining (or splitting) the data. Terminal node 4 produced the highest PM25 mean concentration, over 71 ug/m³. This node resulted from the lowest visibility (<3.5 miles) and highest stability, >4.1. Relating this to synoptic weather, lowered visibility and moderate stability define high PM25 24-hour concentrations at Bakersfield. The lowest PM25 mean concentration was from Node 6, when visibility was greater than 6.5 miles. The mean PM25 concentration was less than 14 ug/m³ in Node 6. Visibility then stability classify the level of PM25 concentrations at Bakersfield over the long-term of the database.

Summary of Nodes

Node	No. of Occurrences	Bakersfield Average PM25 Concentration (ug/m ³)
1	14	14
2	132	52
3	77	36
4	67	71
5	238	23
6	794	14

Fresno all seasonal periods for PM10 – approximately 1988 to 2000 for PM10 data.



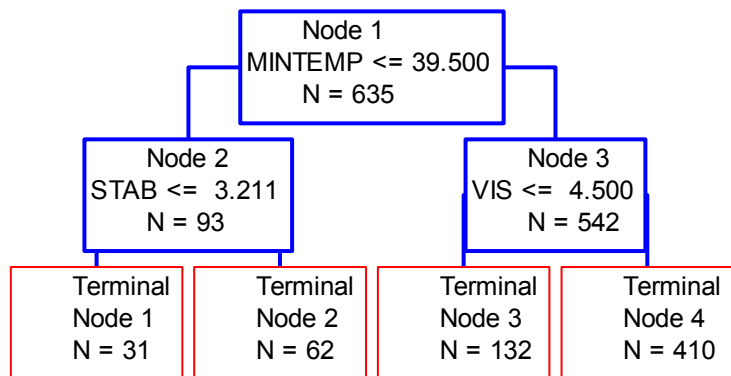
A complex tree is seen at Fresno for PM10 with four predicting meteorological variables – stability, wind speed, relative humidity, and mintemp. The highest terminal nodes are 5 and 7, with node 7 representing very high stability and no other criteria with a mean PM10 concentration of 95 ug/m³. Node 5 represents fairly high stability and mintemp less than 35° F with a mean PM10 concentration of 88 ug/m³. This node might well be influenced by PM25 (because of the very low mintemps) during the winter where PM25 heavily influences PM10 concentrations. The importance of the variables is stability at 100%, wind speed (appears to affect the occurrence of low PM10 conditions) at 82%, mintemp at 66, and % and relative humidity at <25%. The lowest mean PM10 concentration, 14 ug/m³, was from node 3, where stability is less than 6.6, wind speed is greater than 4.25 m/s, and relative humidity is greater than 56.4%. It appears that stability

(characterized by high pressure aloft) is crucial in predicting high PM10 concentrations at Fresno while breezy conditions and moderate stability predict low PM10 concentrations.

Summary of Nodes

Nodes	No. of Occurrences	Fresno Average PM10 Concentration (ug/m**3)
1	35	74
2	181	36
3	47	14
4	136	25
5	26	88
6	173	47
7	27	95

Fresno all seasonal periods for PM25 – approximately 1999 to 2000 for PM25 data.

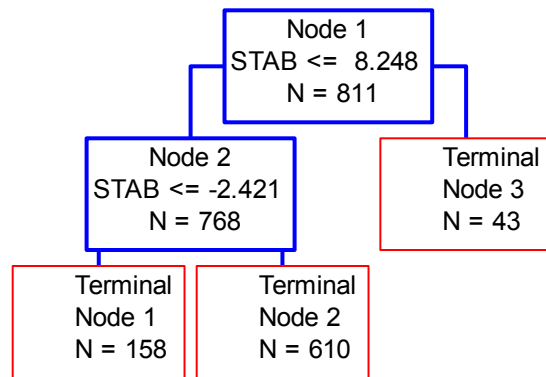


In this case, a straightforward tree was created for Fresno PM25. The splitting variables were mintemp, and then stability and visibility. Node 2, resulting from mintemp less than 39.5° F and stability greater than 3.211, produced the highest mean PM25 concentration of 84 ug/m**3 and the other three nodes are all less than 33 ug/m**3. Thus, high 24-hour mean PM25 is associated with cold overnight lows and moderately high stability. Cold temperatures could be related to high secondary pollutant formation, particularly nitrate. Variable importance (relative) was mintemp at 100%, stability at 74%, and weak third was visibility at 39%

SUMMARY OF NODES

Nodes	No. of Occurrences	Fresno Average PM25 Concentration (ug/m**3)
1	31	33
2	62	85
3	132	33
4	410	11

Stockton all seasonal periods for PM10 – approximately 1988 to 2000 for PM10 data.

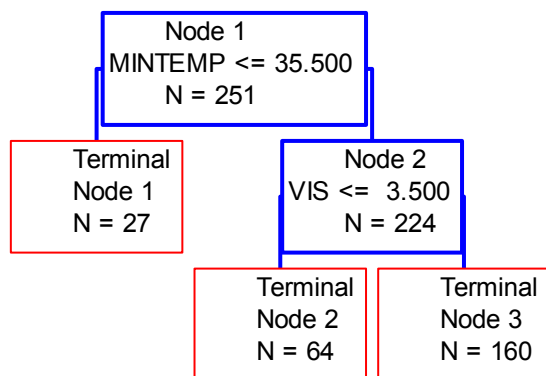


A very simple tree is provided for Stockton’s PM10 concentration – only stability has importance in determining PM10 concentrations. The highest PM concentrations occur in Node 3 when stability is greater than 8.248 with a mean of 63 ug/m³. The other nodes are much less, at 30 ug/m³ or less. Thus, Stockton’s higher PM10 24-hour concentrations are related to the presence of stronger high pressure systems aloft. Relative variable importance is 100% for stability and 11% for visibility (not considered as a splitting variable).

SUMMARY OF NODES

Node	No. of Occurrences	Stockton Average PM10 Concentration (ug/m ³)
1	158	20
2	610	34
3	43	63

Stockton all seasonal periods for PM25 – approximately 1999 to 2000 for PM25 data.

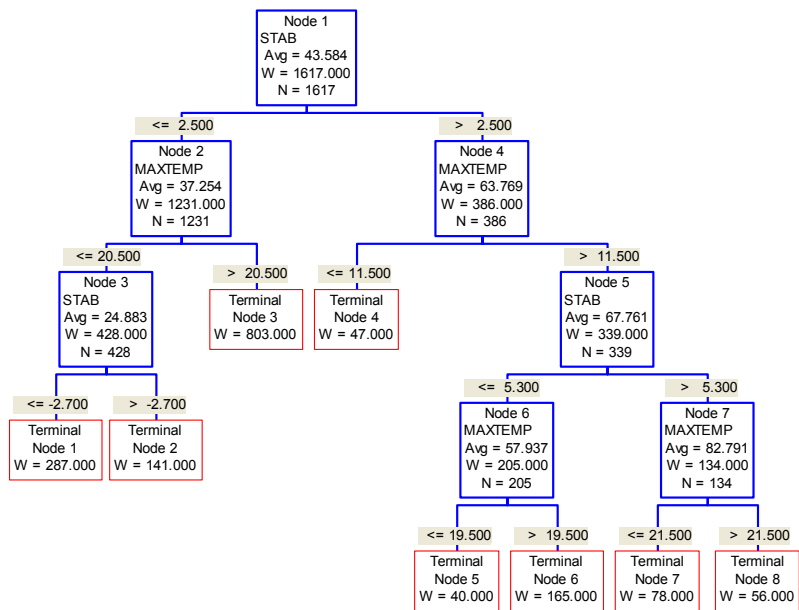


The tree for Stockton PM25 features two splitting variables, mintemp as the primary variable and visibility the secondary variable. There are three terminal nodes all of which are fairly low PM25 concentrations. The highest is node 1 at 45 ug/m**3 and the other two are less than 25 ug/m**3. Node 1 follows the single split where mintemp is lower than 35.5° F. The synoptic pattern for Stockton resembles closely that at Fresno for PM25. The importance of the variables is mintemp at 100% and visibility at 44%. It is interesting that in this rare case that stability does not play a role in determining high or low PM25 concentrations.

SUMMARY OF NODES

Node	No. of Occurrences	Stockton Average PM25 Concentration (ug/m**3)
1	27	49
2	64	24
3	160	10

Corcoran all seasonal periods for PM10 – approximately 1988 to 2000 for PM10 data.

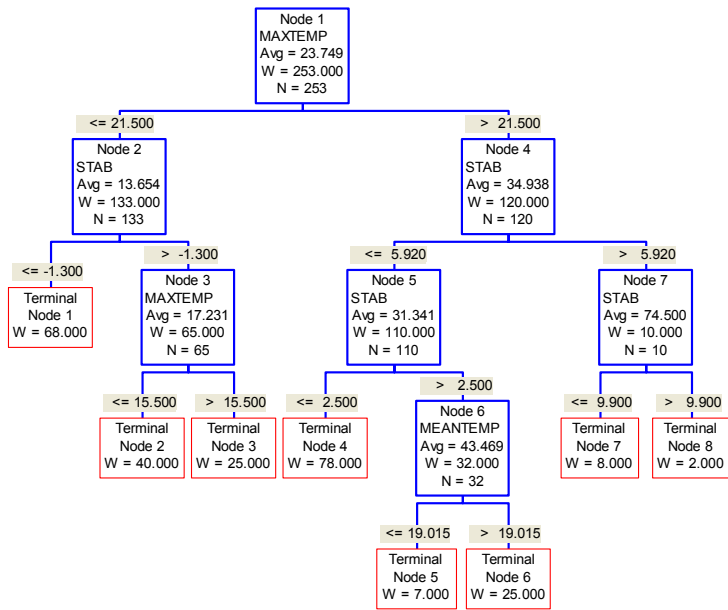


The tree above is quite complex compared to the other urban sites; however only two variables define the tree – stability and maxtemp. The highest PM10 concentration is in node 8 with mean concentration of 96 ug/m**3. Node 8 is when maxtemp is greater than 21 ° C (this station collected temperatures in C rather than F) and stability > 6. Thus, it appears that higher maximum afternoon temperatures and high stability define high mean 24-hour PM10 formation at Corcoran. Node 1 is the lowest mean concentration at 21 ug/m**3 and follows the low maximum temperature and low stability route. Relative variable importance for this tree was 100% for stability and 53% for maxtemp. Several attempts were made to run wind speed in several different averaging techniques but none produced any influence on the tree.

SUMMARY OF NODES

Node	No. of Occurrence	Corcoran Average PM10 Concentration (ug/m**3)
1	287	21
2	141	32
3	803	44
4	47	35
5	40	41
6	165	62
7	78	73
8	56	96

Corcoran all seasonal periods for Coarse PM – PM2.5 to PM10 -- approximately 1988 to 2000 data.



The primary splitting meteorological variable is maxtemp in deg C (it was stability for PM10 total). Stability is the secondary splitter (it was primary for total PM10). The highest PM coarse node was 7 with an average PM coarse of 84 ug/m³. This node occurs when maxtemp is greater than 21.5 deg C and stability is moderate, between 5.9 and 9.9 (interestingly, the other node, 8, was when stability was greater than 9.9 but the average concentration was only 36 ug/m³ – perhaps there was enough instability to create a break through of the mixing layer). This would generally follow a warm, dry, and stable day with light winds, likely a fall seasonal occurrence. The lowest PM coarse concentration occurred with Node 1, where maxtemp is less than 21.5 C and stability is quite low, at less than 1.3. Most likely, it was a cool, unstable day for fall. Relative variable importance was stability at 100% (even though the top splitter was maxtemp, the stability variable had more weight in determining PM coarse concentrations), meantemp at 81 %, and maxtemp at 79%. Again, like for total PM10, several attempts were made to run wind speed as a variable but the CART model found no significance with it.

SUMMARY OF NODES

Node	No. Of Occurrences	Mean PM Coarse (ug/m ³)
1	68	10
2	46	13
3	25	24
4	78	26
5	7	70
6	25	36
7	8	84
8	2	36

SUMMARY OF CONCLUSIONS FOR ALL PERIODS ANALYSIS FOR BAKERSFIELD, FRESNO, STOCKTON, AND CORCORAN

The following table is a summary of the CART results for PM10 and PM25 for the four sites.

Site	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	No Others	93	41	Stab -100
Bakersfield	PM25	Visibility	Stability	71	14	Vis – 100, Stab –20, mintemp – 50 (did not play a role)
Fresno	PM10	Stability	Mintemp, Wind speed, and RH	95	14	Stab –100, wind speed – 82, mintemp – 66, and RH - <25
Fresno	PM25	Mintemp	Stability, Visibility	84	33	Mintemp –100, stab – 74, vis – 39
Stockton	PM10	Stability	No Others	63	<30	Stab – 100, vis – 11

Stockton	PM25	Mintemp	Visibility	45	<25	Mintemp –100, vis – 44
Corcoran	PM10	Stability	Maxtemp	96	21	Stab – 100, maxtemp - 53
Corcoran	Coarse (PM10-PM25)	Maxtemp	Stability	84	10	Stab – 100, Meantemp – 81, Maxtemp - 79

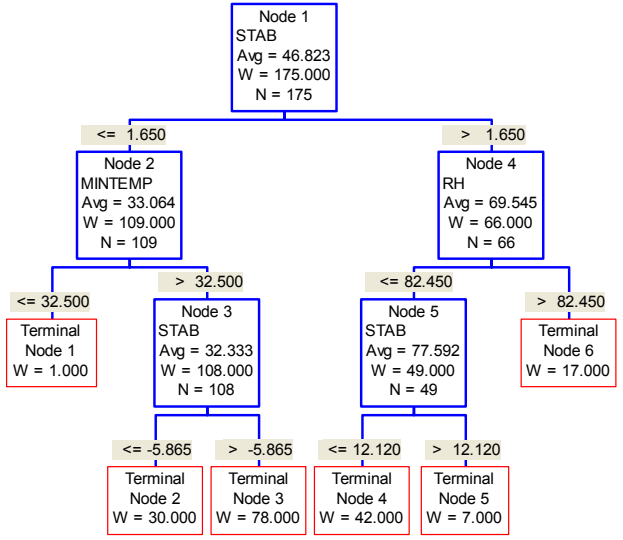
From the above table, it is evident that stability is by far the most important meteorological parameter in defining high PM concentrations, either PM10 or PM25. Other important parameters include mintemp and maxtemp, visibility, and more only rarely RH and wind speed. The maximum node mean PM concentration for PM10 and PM25 ranged from as low as 45 ug/m³ at Stockton for PM25 to a maximum of 96 ug/m³ at Corcoran for PM10. PM concentration means below 20 ug/m³ were reported for the more favorable meteorological regimes (in most cases defined by low stability, higher mintemps, higher wind speed, and normal relative humidity). At Corcoran, very little difference is seen between total PM10 and coarse PM2.5-PM10. Wind speeds make little difference in concentrations at Corcoran.

SEASONAL PM10 AND PM25 ANALYSIS FOR BAKERSFIELD, FRESNO, AND STOCKTON

The following analyses were performed when the data was separated into winter (December, January, and February) versus fall (September, October, and November).

WINTER SEASON ANALYSES

Bakersfield Winter Analysis for PM10 -- approximately 1988 to 2090 data.

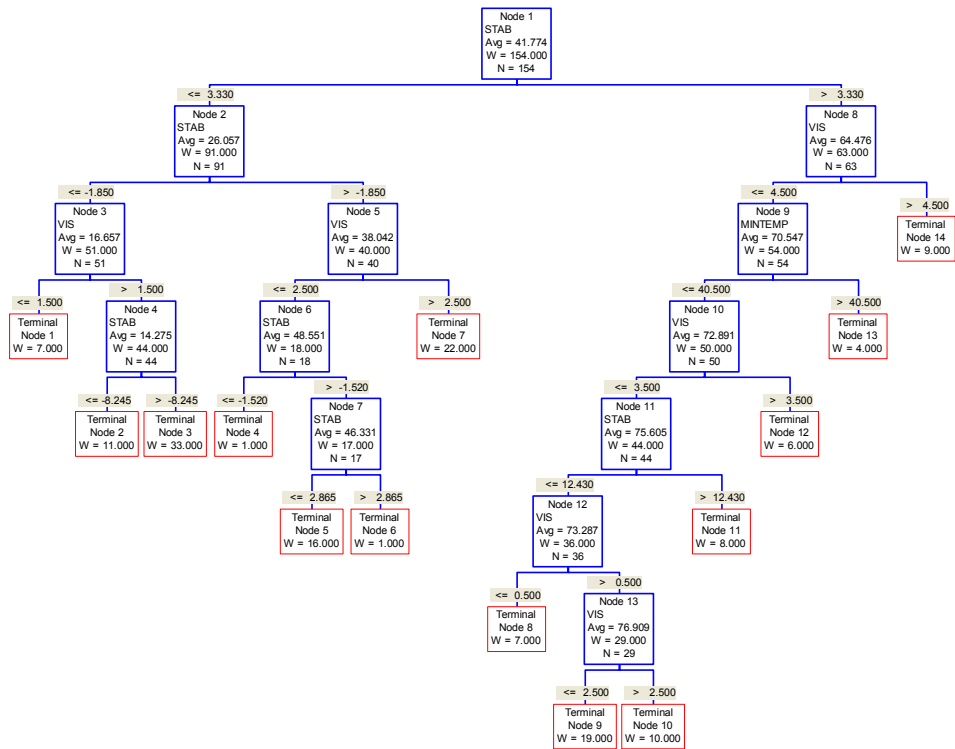


The tree above represents the predicted PM10 mean concentrations for Bakersfield during the winter, defined as December, January, and February. The primary predictive meteorological variable is stability and there are two secondary variables – mintemp and relative humidity. The highest PM10 mean concentrations are from both Node 1 and 5 at 112 ug/m**3. Node 1 occurs when there is moderately low stability and low overnight temperatures. Node 5, on the other hand, occurs under very strong stability and moderately low relative humidity. Thus, high PM10 for winter mean average occurs under cool, stable conditions and low relative humidities. The lowest mean PM10 concentrations occur in Node 2 with a mean of 22 ug/m**3. Node 2 is formed by low stability and elevated minimum temperatures. Variable importance was stability at 100%, mintemp at 56,% and relative humidity at 29%. It appears that either low overnight temperatures or very strong stability define high PM10 concentrations.

SUMMARY OF NODES

Node	No. of Occurrence	Bakersfield Winter Average Concentration (ug/m**3)
1	1	112
2	30	22
3	78	36
4	42	72
5	7	112
6	17	46

Bakersfield Winter Analysis for PM25 – approximately 1999 to 2000 data.

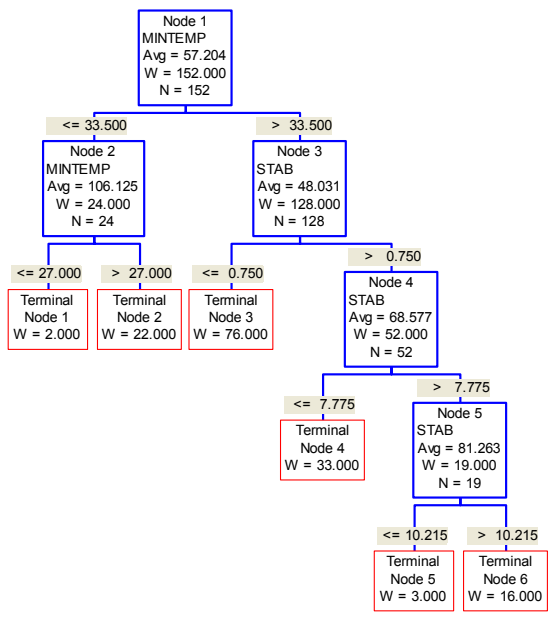


A very complex tree is created for PM25 at Bakersfield for winter. The tree includes primary variables of stability and visibility and a lesser variable of mintemp. Two nodes, 4 and 11, show a mean PM25 concentration of 86 ug/m³. Node 4 occurs under high stability and lower visibility, while node 11 occurs under high stability, lower visibility, and low minimum temperatures. Thus, higher 24-hour mean PM25 winter concentrations are created when stability is high and visibility is less and temperatures are colder. The lowest PM25 mean concentrations occurred under node 2, at 8 ug/m³, when there is low stability and low visibility (a little surprising considering that low visibility also promotes high PM25). Variable importance was visibility at 100%, stability at 97%, and mintemp at 70%. High stability and lower visibility are key to linking high mean PM25 concentrations.

SUMMARY OF NODES

Node	No. of Occurrences	Bakersfield Average PM25 Concentration (ug/m ³)
1	7	32
2	11	8
3	33	16
4	1	86
5	16	45
6	1	73
7	22	29
8	7	58
9	19	82
10	10	68
11	8	86
12	6	53
13	4	41
14	9	28

Fresno Winter Analysis for PM0 -- approximately 1988 to 2000 data.

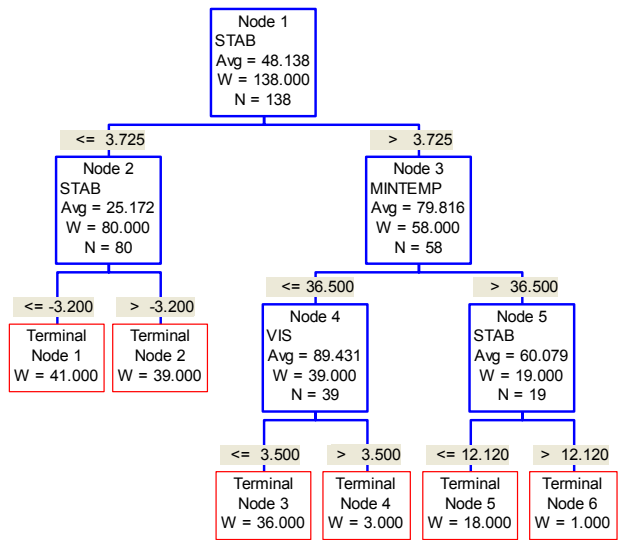


A complex tree results for Fresno PM10 analysis. The primary splitting variable is mintemp and stability is the secondary splitter. The highest mean PM10 concentration is 236 ug/m**3 (much higher than any other tree analyzed) with a secondary maximum at Node 5 of 119 ug/m**3. The tree for Node 1 is simply based on mintemp being less than 27.5° F. Thus, colder overnight low temperatures favor PM10 formation but this might be due to PM25 forming as low temperatures favor higher secondary particle formation. Node 5 is based on higher mintemps but includes moderately high stability. The lowest mean PM10 concentration for winter was from Node 3 at 34 ug/m**3. This node resulted from higher mintemps and low stability values. Variable importance shows mintemp at 100% and stability at 36%. It appears that low nighttime temperatures and high stability are crucial in developing high PM10 concentrations at Fresno.

SUMMARY OF NODES

Node	No. of Occurrences	Fresno Average PM10 Concentration (ug/m**3)
1	2	235
2	22	94
3	76	34
4	33	61
5	3	119
6	16	74

Fresno Winter Analysis for PM25 – approximately 1999 to 2000 data.

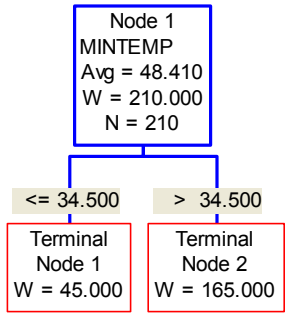


The tree above represents the prediction by CART of the meteorological variables influencing PM25 at Fresno in the winter. There are three variables – the primary is stability followed by mintemp and visibility. The highest mean PM25 s predicted by node 6 at 127 ug/m**3. Node 6 follows very high stability values(>12) and mintemps greater than 36° F. Thus, higher PM25 in winter is favored by very high stability (expected) but higher mintemps. The lowest PM25 mean concentrations are in Node 1 at 16 ug/m**3, occurring during very low stability values (likely stormy conditions). The importance of the variables was stability at 100%, mintemp at 81%, and visibility at 38%. The predicting high PM25 concentration variable of high stability and higher mintemps contradict each other as usually lower than normal mintemps accompany high PM25 events.

SUMMARY OF NODES

Node	No. of Occurrences	Fresno Average PM25 Concentration (ug/m**3)
1	41	16
2	39	35
3	36	94
4	3	35
5	18	56
6	1	127

Stockton Winter Analysis for PM10 – approximately 1988 to 2000 data.

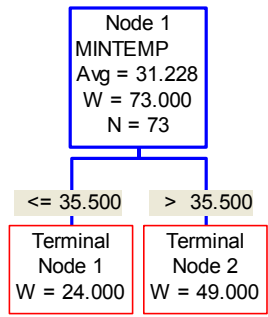


The tree above is very simple – only mintemp is a splitting variable and only two nodes result. The highest mean PM10 concentration is at Node 1 at 75 ug/m**3 and Node 2 is at 41 ug/m**3. The higher node occurs when mintemps are less than 34°.5 F and other node when mintemps are larger than 34.5° F. Variable importance was mintemp at 100% and relative humidity at 31% (which is not a predictor variable).

SUMMARY OF NODES

Node	No. of Occurrences	Stockton Average PM10 Concentration (ug/m**3)
1	45	75
2	165	41

Stockton Winter Analysis for PM25 – approximately 1999 to 2000 data.



This tree is identical to the tree for PM10 at Stockton. Only one variable is used to predict PM25, mintemp. The two nodes segregate mean PM25 levels – Node 1 at 52 ug/m**3 while Node 2 at 21 ug/m**3. Thus, high PM25 at Stockton for the mean 24-hour period may be related to cold overnight conditions, which have been seen at the other locations. The high PM25 node is when mintemp is less than 35.5° F. The variables of importance are mintemp at 100% and visibility at 46% (not used in predicting PM25).

SUMMARY OF NODES

Node	No. of Occurences	Stockton Average PM25 Concentration (ug/m**3)
1	24	52
2	49	21

COMPARISON OF WINTER RESULTS AMONG THE THREE STATIONS – Bakersfield, Fresno, and Stockton

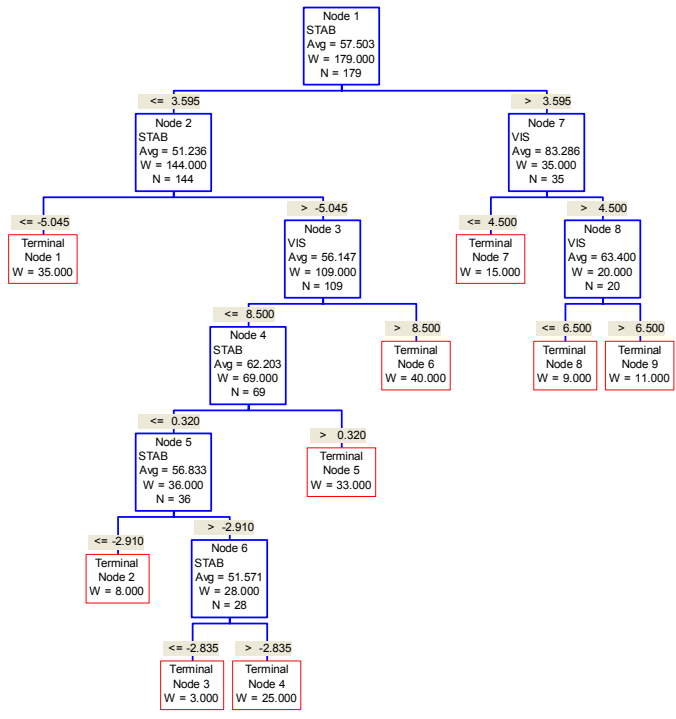
The Table below summarizes the CART analyses for the winter time periods at Bakersfield, Fresno, and Stockton. Provided are the splitting variables, the highest and lowest Nodes and variable importance for each CART run.

Site	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	Mintemp, RH	112	22	Stab –100, mintemp 56, RH – 29
Bakersfield	PM25	Visibility	Mintemp	86	8	Vis – 100, Stab –97, mintemp – 70
Fresno	PM10	Mintemp	Stability	236	34	Mintemp –100, stab - 36
Fresno	PM25	Stability	Mintemp, Visibility	127	16	Stability –100, mintemp – 89 vis – 39
Stockton	PM10	Mintemp	No Others	75	41	Mintemp – 100
Stockton	PM25	Mintemp	No Others	52`	21`	Mintemp –100

For the winter analysis, there is a little more spread of variables that are important; while stability is still quite important, mintemp and visibility take on a greater importance than in the analysis for the entire data set. Another item of interest are the higher concentrations; for example, the Fresno PM10 levels for the highest node are 236 ug/m**3 in the winter dataset versus 95 ug/m**3 in the full dataset. The PM concentrations at Stockton are much lower than the Bakersfield or Fresno urban areas.

FALL SEASON ANALYSES

Bakersfield Fall Analysis for PM10 – approximately 1988 to 2000 data.

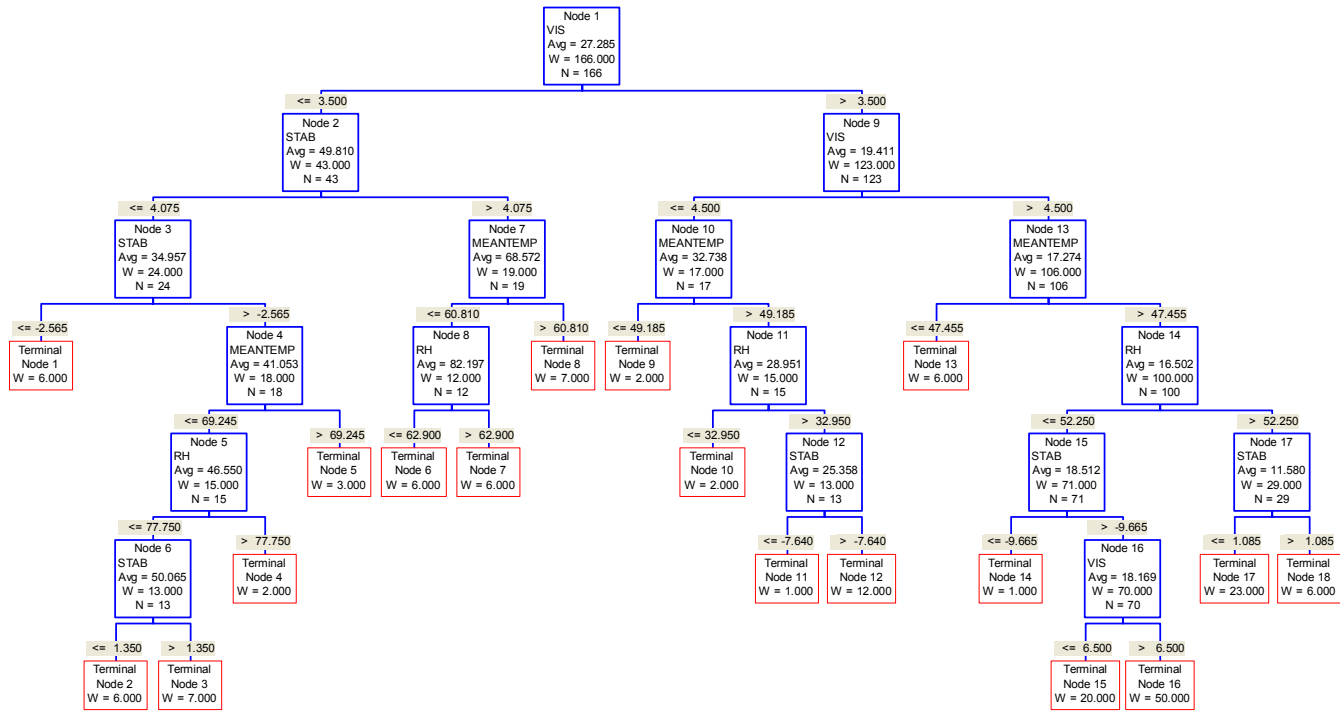


A complex tree results for Bakersfield PM10 during the fall (defined as September, October, and November). Stability is the primary splitting variable and visibility is a secondary variable. The highest PM10 mean concentration is predicted in Node 7 at 110 ug/m³. This node results from stability greater than 4 and visibility less than 1.5 miles. Thus, high stability (high pressure aloft) and lower visibility favor high PM10 concentrations in the mean at Bakersfield. The lowest PM10 mean concentration occurs in Node 1 at 36 ug/m³ where very low stability, <-0.5 is predicted. The importance of the variables is stability – 100 %and visibility – 60%. Again, stability is the key predictor for high PM10 concentrations at Bakersfield and secondly, hazy conditions.

SUMMARY OF NODES

Node	No. of Occurrences	Bakersfield Average PM10 Concentration (ug/m ³)
1	35	36
2	8	75
3	3	24
4	25	55
5	33	68
6	40	46
7	15	110
8	9	78
9	11	52

Bakersfield Fall Analysis for PM25 – approximately 1999 to 2000 data.



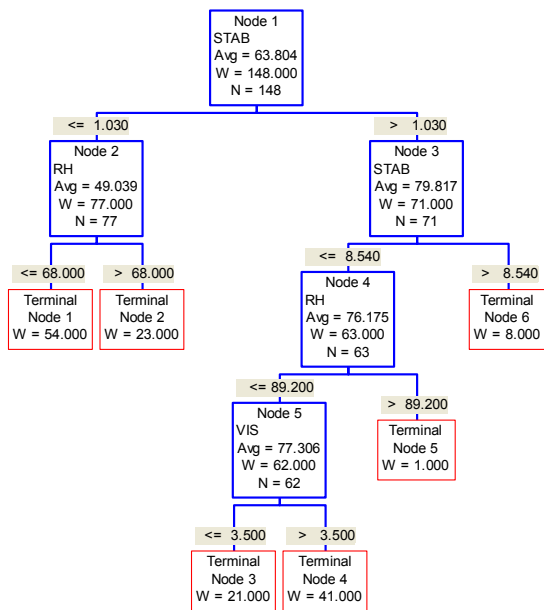
The tree for Bakersfield PM25 fall season is the most complex of the CART model runs. Four variables are involved and include visibility as the primary splitter and meantemp, stability, and relative humidity as secondary splitters. This appears to follow the warmth of the day, the status of the upper atmosphere – is there high pressure with warm air aloft, and how moist is the atmosphere with drier conditions promoting more dust PM while more RH might promote higher secondary aerosol concentrations. The highest mean PM25 concentration was in Node 6 at 94 ug/m³. This node results from visibility less than 3.5 miles, stability greater than 4.1 and relative humidity less than 62%. This would appear to be indicative of high pressure aloft with low wind speeds and hazy conditions. The lowest mean PM25 concentrations occurred with Node 11 at 6 ug/m³. This node occurs when visibility is greater than 3.5 miles, meantemp greater than 49° F, relative humidity greater than 33% , and stability less than 3.6. Variable importance is visibility 100%, meantemp 94%, stability 51%, and relative humidity 38%.

SUMMARY OF NODES

Node	No. of Occurrences	Bakersfield Average PM25 Concentration (ug/m ³)
1	6	17
2	6	57
3	7	44
4	2	24
5	3	14
6	6	94
7	6	70
8	7	45
9	2	61
10	2	52
11	1	6
12	12	27
13	6	30
14	1	43

15	20	22
16	50	17
17	23	10
18	6	18

Fresno Fall Analysis for PM10 – approximately 1988 to 2000 data.

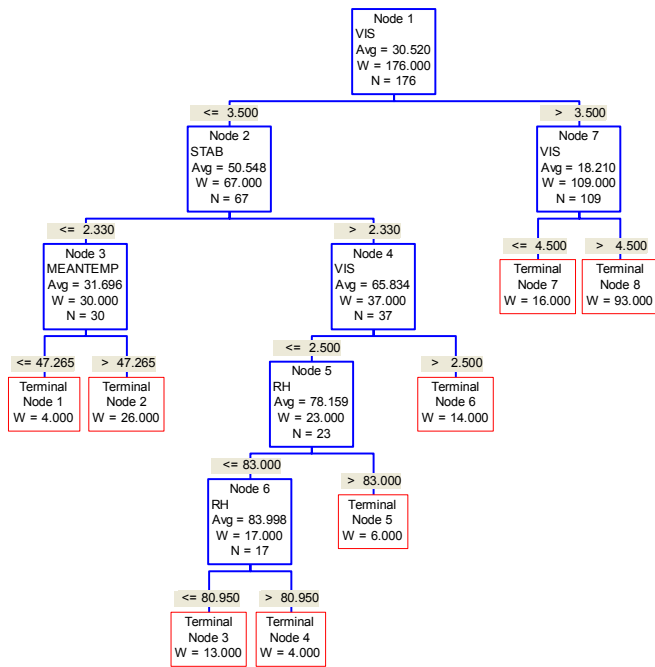


The tree for the Fresno fall PM10 is split by three variables – stability, as the primary variable, and relative humidity and visibility as the secondary variables. The highest mean PM10 concentration was 109 ug/m³ in node 6 and occurred when stability was very high, greater than 8.5 and no other variables were involved. Thus, it appears that strong high pressure aloft defines high PM10 periods at Fresno, with no other contributing meteorological variable. The lowest mean PM10 concentration was 6 ug/m³ in node 2 and occurred when stability was moderate and relative humidity was greater than 89%. The lowest node is a bit puzzling considering that stability is still quite high, between 5 and 8.5.

SUMMARY OF NODES

Node	No. of Occurrences	Fresno Average PM10 Concentration (ug/m ³)
1	54	56
2	23	31
3	21	92
4	41	70
5	1	6
6	8	109

Fresno Fall Analysis for PM25 – 1999 and 2000 data.

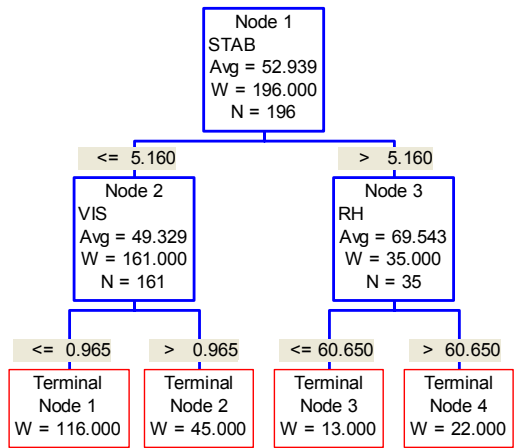


The tree for Fresno PM25 during the fall season is quite complex and involves the primary variable, visibility, and three other secondary variables – meantemp, stability, and relative humidity. It would appear like the other PM25 CART predictions the variables predicting PM25 consist of visibility, temperature, and relative humidity. The highest mean PM25 concentration at Fresno is 113 ug/m³ in node 4 and was predicted when stability was greater than 2.4 (moderate), visibility less than 2.5 miles, and relative humidity between 80 and 83% (or under hazy and humid conditions). The lowest PM25 mean concentration for Fresno is 16 ug/m³ and occurred when visibility was greater than 4.5 miles with no other variables involved. The relative importance of the variables is visibility 100%, meantemp 82%, stability 74%, and relative humidity 64%.

SUMMARY OF NODES

Node	No. of Occurrences	Fresno Average PM25 Concentration (ug/m ³)
1	4	66
2	26	26
3	13	75
4	4	113
5	6	62
6	14	46
7	16	32
8	93	16

Stockton Fall Analysis for PM10 – approximately 1988 to 2000 data.

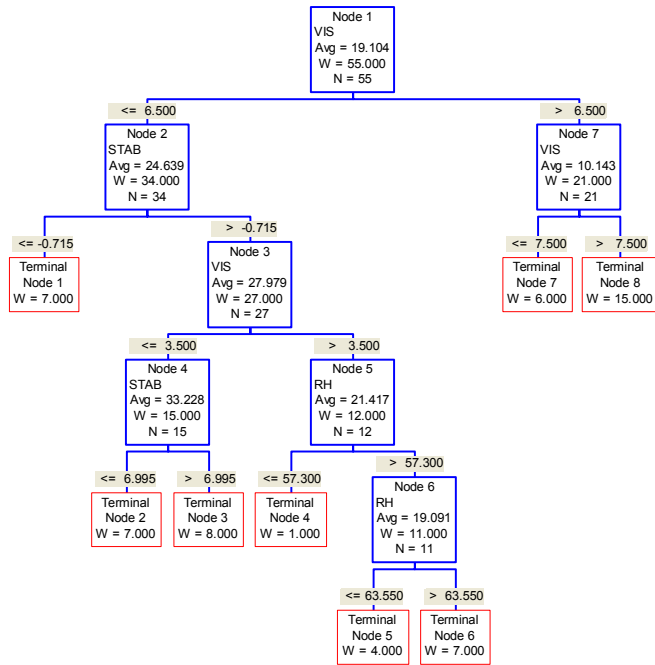


The tree above for Stockton PM10 for the fall season features 3 predictive nodes: stability is primary with two secondary nodes - visibility and relative humidity. The highest mean PM10 concentration for Stockton is 99 ug/m³ at node 3 and this node is where stability is greater than 5 and relative humidity is less than 60% (where a dry and stable weather pattern exists). For low mean PM10 concentrations, node 2 is 32 ug/m³ where low stability (less than 5) and high visibility (>1.0) exist. Variable importance is stability 100%, visibility – 56%, and relative humidity – 51%.

SUMMARY OF NODES

Node	No. of Occurrences	Stockton Average PM10 Concentration (ug/m ³)
1	116	56
2	45	31
3	13	99

Stockton Fall Analysis for PM25 – 1999 to 2000 data.



The tree for the Stockton fall PM25 is fairly complex and includes the primary variable of visibility and two other secondary splitters – stability and relative humidity. The highest mean PM25 concentration is 47 ug/m**3 for node 4. This node results when visibility is less than 6.5 miles, stability is greater than 7.1, and relative humidity is less than 57% (or when the weather is hazy, stable (high pressure aloft), and humid). The lowest mean PM25 concentration for Stockton occurred in node 8 at 8 ug/m**3. This node follows only visibility as a split and occurs when visibility is greater than 7.5 miles. This is probably a result of clear skies and little haze. Variable importance is visibility 100%, stability 64,% and relative humidity 40%.

SUMMARY OF NODES

Node	No. of Occurrences	Stockton Average PM25 Concentration (ug/m**3)
1	7	12
2	7	26
3	8	39
4	1	47
5	4	25
6	7	16
7	6	16
8	15	8

SUMMARY OF FALL SEASONAL CART RUNS FOR BAKERSFIELD, FRESNO, AND STOCKTON

The Table below summarizes the CART analyses for the fall time periods at Bakersfield, Fresno, and Stockton. Provided are the splitting variables, the highest and lowest Nodes and variable importance for each CART run.

Site Name	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	Visibility	110	36	Stab – 100, visibility – 60
Bakersfield	PM25	Visibility	Meantemp, stability, and relative humidity	94	6	Vis – 100, meantemp – 94, stability – 51, and RH – 38
Fresno	PM10	Stability	Relative humidity and visibility	109	6	Stab – 100, RH – 42, and Vis – 38
Fresno	PM25	Visibility	Meantemp, stability, and RH	113	16	Vis – 100, meantemp – 82, stab – 74, RH – 64
Stockton	PM10	Stability	Visibility and RH	99	32	Stab – 100, vis – 56, and RH - 51
Stockton	PM25	Visibility	Stability and RH	47	8	Vis – 100, Stab – 64, and RH - 40

Results of the fall seasonal runs indicate more dependence on visibility than stability as was the case for the winter analysis. Some of the mean PM25 concentrations are quite high, well above the 24-hour federal and state standards. Relative humidity also plays more of a role in defining PM25 concentrations.

HOLIDAY ANALYSIS – CART, SAS INSIGHT, AND STUDENT T TEST

Holiday periods have often been associated with increased fireplace burning, extra traffic, and bad dispersion. The question has been whether holiday periods are necessarily higher than non-holiday periods over the wintertime period.]To help answer that question, three differing techniques were tried – CART, the data analysis tool, SAS Insight, and the Student T test which is a statistical test that measures the difference between the means of two data sets and determine how significant are the differences.

CART

A new variable, Holiday, was used in CART. The period from Monday until Sunday of the week of Thanksgiving, and the two days before and after Christmas, and New Years defined the holiday variable. In CART, this variable was assigned a categorical form where 0=no holiday and 1=the holiday period defined above. This variable was run for all three locations – Bakersfield, Fresno, and Stockton for the period of 1988 to 2000. The output from CART showed no dependence on this variable for PM concentrations. The trees are not shown since the holiday variable was not part of any of the trees predicted.

SAS INSIGHT

Sas Insight is a product that allows data to be visualized in many ways. For this application, the PM10 data were displayed versus all of the various independent meteorological variables. Then the data for holidays were highlighted in red and a cross pattern was used to represent the data during the holiday. The result of this analysis was that the data were not banded or clumped but appeared to be randomly distributed with the rest of the non-holiday period. There did appear to be some bias towards the high PM periods however.

STUDENT T TEST

A simple statistical test is the student t test where the means of two samples are calculated and the difference is assessed by examining the level of confidence about the difference. For the target sites (Bakersfield, Fresno, and Stockton), the

holiday mean was significantly higher than the non-holiday (Bakersfield was borderline) at the 95% confidence level. For Bakersfield, the holiday mean was about 10 ug/m³ higher, for the other sites, the holiday mean were about 20 ug/m³ higher.

SUMMARY OF CONCLUSIONS

1. For All Seasons Analysis – 1988 to 2000

- PM10 – stability is the primary and overwhelming splitting variable for all four sites – Bakersfield, Fresno, Stockton, and Corcoran. For 2 sites, stability is the only variable (Bakersfield and Stockton). At Fresno, secondary variables include mintemp, wind speed (for low PM10), and relative humidity. At Corcoran, maxtemp is the only secondary splitting variable.
- PM25 – more complex trees with mintemp and visibility for the primary splitting variable at 2 out of 3 sites (Stockton and Fresno) and at the other, visibility. Other secondary variables include further splitting by stability and visibility.
- Stability greater than about 8 (stability is the temperature at 850 mb (roughly 5,000 ft) minus the surface minimum temperature) defines in most cases very high mean PM10 concentrations (over 90 ug/m³ except only 63 ug/m³ at Stockton).
- Low visibility, low mintemps, and high stability seem to predict accurately high mean PM25 concentrations, with mean PM25 concentrations from 45 to 84 ug/m³ (highest at Fresno)

- CART, when it segregates terminal nodes, simulates very high PM concentrations with a very low number of occurrence, which correctly matches reality
2. Winter Season Runs (for December, January, and February) for 1988 to 2000 period.
- Winter season splitting variables for PM10 are not so dependent on stability – other primary splitting variables include mintemp at 2 sites, Fresno and Stockton. Secondary splitting variables include mintemp, stability, and relative humidity.
 - For PM25, the primary splitting variables were visibility, stability, and mintemp (at two sites, Bakersfield and Stockton). Secondary variables include mintemp and visibility.
 - The highest Bakersfield mean PM10 concentration nodes, 112 ug/m³ is defined by stability greater than 12 and relative humidity less than 82% or, in another node, under low to moderate stability (<1.6) and low overnight temperatures (<32° F) (this might be a result of a high PM25 contribution to PM10). At Fresno, the highest mean PM10 concentration in a node was 236 ug/m³, the highest in any of these analyses. This node was predicted by a single variable, when overnight low temperatures were less than 27° F. Again, it can be surmised that high PM25 concentrations are contributing to this high mean PM10 concentration in this node. At Stockton, the highest PM25 mean node concentration was 52 ug/m³ and occurred when minimum temperature of less than 35.5° F was predicted.
 - The model seems to do better in realistically simulating PM10 and PM25 concentrations with the correct meteorological independent variables for winter than for the entire season
3. Fall Season Runs (for September, October, and November) for 1988 to 2000 period.

- a. Because “fall” and transition to “winter” occurs at different times from year to year, definition of fall could be argued. For example, fall might end in October or may not end until late December. For these analyses, fall is between September and November.
- b. Stability was the primary splitting variable for all three urban locations for PM10. Other secondary variables were visibility and relative humidity.
- c. For PM25, the primary splitting variable was always visibility at the three sites. Other secondary variables included meantemp (as opposed to mintemp in the winter analysis), stability, and relative humidity.
- d. The highest fall PM10 mean node concentration was close among the three location at 99 ug/m³ (Stockton), 109 ug/m³ (Fresno), and 110 ug/m³ (Bakersfield). At Bakersfield, stability greater than 4 and visibility less than 1.5 miles resulted in the highest PM10. At Fresno, stability greater than 8.5 was the sole variable determining the highest mean PM10. At Stockton, the highest PM10 occurs when stability is greater than 5 and relative humidity is less than 60%.
- e. The highest fall PM25 mean node concentration was 113 ug/m³ at Fresno, followed by 94 ug/m³ at Bakersfield and 47 ug/m³ at Stockton. At Fresno, the highest PM25 concentration is determined by stability greater than 2.4, visibility less than 2.5 miles, and relative humidity between 80 and 83%. At Bakersfield, the highest fall mean PM25 concentration occurs when visibility is less than 3.5 miles, stability greater than 4.1 and relative humidity less than 62%, conditions similar to Fresno. At Stockton, the highest fall PM25 concentration is predicted by visibility less than 6.5 miles, stability greater than 7.1 and

relative humidity less than 57%. The requirements for a Stockton high PM25 concentration appear to be much stability dependent than the other two locations.

4. Holiday Seasonal Analysis – 1988 to 2000

- **CART**

The output from CART showed no dependence on the holiday period for PM10 or PM25 concentrations.

- **SAS INSIGHT**

The result of this analysis was that the data were not banded or clumped but appeared to be randomly distributed with the rest of the non-holiday period. There did appear to be some bias towards the high PM periods however.

- **STUDENT T TEST**

For the target sites (Bakersfield, Fresno, and Stockton), the holiday mean was significantly higher than the non-holiday (Bakersfield was borderline) at the 95% confidence level. For Bakersfield, the holiday mean was about 10 ug/m³ higher, for the other sites, the holiday mean were about 20 ug/m³ higher.

5. OVERALL CONCLUSIONS

- The categorical and regression tree (CART) model, a statistical based model, was used in the San Joaquin Valley to determine the relationship between PM10 or PM25 and the various independent meteorological variables such as wind speed, relative humidity, stability, visibility, precipitation, wetness, and mintemp, maxtemp, and meantemp. These meteorological variables were derived from either district or ARB data, or, preferably, the National Weather Service, who maintains a better and broader array of measurements than anyone else. For all of the analysis except Corcoran, NWS data were used. At Corcoran, district meteorological data were used. These meteorological variables and PM10 or PM25 (dependent variables) were run in CART to determine only the most statistically significant meteorological variables affecting PM concentrations. Terminal nodes resulted in each model run where a mean PM10 or PM25 concentration was computed and the tree path developed (for example, stability >10 and visibility < 3 miles might define a terminal node where the mean PM10 was 100 ug/m³). The terminal nodes were as many as 18 to as few as 2. Each node described a unique concentration and set of corresponding meteorological variables in the valley for a long-term 24 hour average. Very high PM concentrations were mostly associated with very few occurrences.
- Variables quickly found to be insignificant were precipitation and wetness. These parameters should be significant in describing suppression of geological dust production. The inability of CART to use these variables probably is not a model failure, rather the fact that for these runs, CART analyzed long-term trends (over 2 years for PM25 and over 13 years for PM10) of PM concentrations and meteorological variables. Precipitation impacts are shorter-term (i.e. likely less than a month). When the CART model is used on the 2000-2001 BAM data during the CRPAQS period, perhaps more correlation with precipitation and PM will be established.

- CART turned out to be useful in defining what meteorological parameters are important in determining the range of PM10 and PM25 concentrations in the San Joaquin Valley. CART seemed to handle the three parts of the valley especially well, with the north being represented by Stockton data, the central being represented by Fresno data, and the south being represented by Bakersfield data.
- CART predicted reasonably well regional events where the meteorology was uniform over a large scale (where part of or most of the valley is being affected by the same conditions) so PM25 concentrations seemed very realistic for example.
- Where CART did not do well was localized conditions where, for example, fugitive dust might be a problem. CART did not consider wind speed, for example, to be a significant variable and this short-coming is especially evident in the fall analyses. This was especially evident at Corcoran where both PM10 and PM coarse were analyzed with CART.
- Future planned work includes in order of priority:
 1. Incorporate Beta attenuation monitor (BAM) data for 2000 and 2001 (the CRPAQS period) and let CART predict the best meteorological parameters for short-term episodic conditions
 2. Include more historical pm10 and dichot data back to 1980 to obtain a better historical perspective of the San Joaquin Valley PM10 and PM25 average concentrations

