TRA
ENVIRONMENTAL
SCIENCES, Inc.

545 Middlefield Road, Suite 200
Menlo Park, CA 94025-3472
Tel: (650) 327-0429
Fax: (650) 327-4024
www.TRAenviro.com

To: Daphne Greene, Off-Highway Motor Vehicle Recreation Division

From: Tom Reid

Date: May 18, 2010

Re: Published Phase 2 Report data does not support claims of association between Oceano Dunes State Vehicular Recreation Area visitor numbers and PM10 downwind.

**Introduction**

The Phase 2 Report (Report) presents the results of a fine particulate (PM10) monitoring program in the Pismo Beach and Nipomo Mesa vicinity, ancillary studies of particulate and substrate composition, and dune sand transport characteristics. The program was intended to determine the extent to which the Oceano Dunes State Vehicular Recreation Area (ODSVRA) contributes to high ambient PM10 concentrations downwind on the Nipomo Mesa. The study collected much useful data, some of which was published by the APCD on the internet, and the Report offers a comprehensive overview of the work done. Subsequent interpretations of the Report, however, have overstated or incorrectly interpreted the Report's findings.

APCD staff have made statements that the Report concluded that ODSVRA activity is responsible for 25% or 30% of PM10 observed on the Mesa. That conclusion is incorrect and is not in any way supported by the Phase 2 study data. This memo explains why.

**Report Conclusions**

The Report draws several conclusions claiming an ODSVRA effect on downwind PM10 summarized as follows:

1) **Direct association with ODSVRA activity.** The Report compares downwind PM10 levels with ODSVRA visitor numbers, seeking to observe an association of high ODSVRA use and high PM10. In the Report, direct comparison of daily activity level and 24-hour PM10 reveals no relationship, see Figure 3.64 – PM10 Concentration as Compared to Number of Vehicles in the SVRA. When the Report claims to compare PM10 levels from a subset of 50 high use days with PM10 levels from a subset of 50 low use days it finds a significant difference in mean PM10 concentrations for Mesa 2, but not elsewhere. That result is reported in Table 3.2, Average PM10 Concentration for the Highest and Lowest SVRA Activity.

2) **Indirect effect of ODSVRA use.** The Report concludes that more PM10 is released from the ODSVRA than from similar dunes with no ongoing vehicle use and that is the cause of greater PM10 concentrations downwind.

The report itself does not draw a conclusion on the quantitative role of ODSVRA activity on PM10 levels -- what percentage of PM10 is attributable to the ODSVRA. In public presentations and in subsequent writings, however, SLO APCD staff have made statements that high ODSVRA activity is responsible for variously a 25% or 30% increase in downwind PM10. Larry R. Allen, Air Pollution Control Officer, Letter of April 19, 2010 to Andrew Zilke, District Superintendent, Oceano Dunes District, California Department of Parks and Recreation (CDPR) "Response to Comments on the South County Phase 2 Particulate Study" states:

> "The ODSVRA attendance analysis we performed compared PM10 concentrations measured at the Mesa 2 site on the 50 highest visitation days to those at Mesa 2 on the 50 lowest visitation days. That analysis shows a statistically significant average increase of over 30% in PM10 levels at Mesa 2 on the 50 highest visitation days compared to the 50 lowest visitation days (PM study, Table 3.2). The CDF site showed an average 17% increase in PM10 levels on the 50 highest visitation days, but was not statistically significant. In contrast, PM10 levels downwind from the Oso site showed no difference between high and low visitation days. "

Although the statement merely associates high visitation with high PM10, it has been incorrectly interpreted to mean that high visitation is responsible for high PM10.

**The problem with Table 3.2**

The inference from Table 3.2 that ODSVRA activity is responsible for 30% of PM10 is incorrect:

1. Table 3.2 as published is incorrect.
   a. Table 3.2 is based on unpublished data. Application of that analysis to the published data produces a different result, showing no significant visitation/PM10 association.
   b. Table 3.2 includes unpublished data from March 2008 when the Oso station was not operating, so a comparison of Mesa 2 and Oso is meaningless.
   c. Table 3.2 is based on an Excel spreadsheet that incorporates substantial formulae errors. The table purports to compare the "high 50 visitor days" with the "low 50 visitor days", but <u>none</u> of the Excel calculations actually use 50 high or 50 low days.
2. Table 3.2 is the wrong way to analyze the data for association.
   a. Lumping and comparing the 50 high and 50 low days ignores over two-thirds of the data and confounds the analysis by disassociating the actual PM10 and visitation numbers within each lumped group.
   b. The correct approach is to use a linear regression where individual days with pairs of visitor and PM10 numbers are examined to see if there is a relationship. There is none.
3. Statistics are being misused
   a. Statistical significance is an important test for variable measured data.
   b. Even if an association was observed (which it was not) it is incorrect to conclude a causal relationship.

These deficiencies are explained below.

**1. Table 3.2 as published is incorrect.**

**a. Table 3.2 is based on unpublished data. Application of that analysis to the published data produces a different result, showing no significant visitor/PM10 association.**

SLO APCD published study data on the internet; http://www.slocleanair.org/air/pdf/PMSecond Data/finalfinalRAWDATA.pdf lists “South County Phase 2 Particulate Study” data in PDF format; the District did not provide computer readable files. When converted to computer readable files (by Ed Waage, checked by TRA staff), the data can be subjected to independent analysis.

The published data cover the 12-month period from April 1, 2008, to March 31, 2009, and are presented as “CDF”, MESA2”, and “OSO”. If the analysis of Table 3.2 is applied to the 12-month period it produces the following result:

**Recreation of Table 3.2 from 12 month published study data:**

| | Highest 50 days for Vehicles | Lowest 50 days for Vehicles | Highest days - Lowest days | Statistical Confidence of Data (1-P) >95% is Significant |
|---|---|---|---|---|
| Average SVRA Vehicles | 3643 | 378 | 3266 | |
| Average Mesa2 PM10 | 29.1 | 24.5 | 4.6 | 85.40% |
| Average CDF PM10 | 28.7 | 27.9 | 0.8 | 70.16% |
| Average Oso PM10 | 28.2 | 28.4 | -0.2 | 51.19% |

The result shows that <u>none</u> of the three stations have sample means that differ significantly between high use days and low use days

**b. Table 3.2 includes unpublished data from March 2008 when the Oso station was not operating, so a comparison of Mesa 2 and Oso is meaningless.**

The version of Table 3.2 published in the report covers 13 months from March 2008 to April 2009. The inclusion of March 2008 data is the reason the table purports to find a significant difference. This result is unreliable.

March 2008 was unusually warm, windy, and dry. Spring break that year was busy at ODSVRA such that 7 out of 50 “high visitor” days come from that one month. The unusual weather could be associated with high levels of PM10 from ODSVRA, agriculture, or natural sources. Unfortunately, the OSO station, considered in the Report as a control for natural dune PM10, was not in operation, so there is no way to know if the PM10 observed in March 2008 at MESA2 had similar high levels at Oso. Indeed, 15 of the “50 high visitor days” have missing data from Oso, so an attempt to compare the two in Table 3.2 is egregious.

Further, any seasonal phenomenon should be analyzed to avoid double counting any season; double counting March is inappropriate. If there is an actual effect as claimed in the Report, it should appear in any set of 12 months of analysis; if it is a real association, it should appear <u>any</u> time it is examined, not just in March 2008.

**c. Table 3.2 is based on an Excel spreadsheet that incorporates substantial formulae errors. The table purports to compare the "high 50 visitor days" with the "low 50 visitor days", but none of the Excel calculations actually use 50 high or 50 low days.**

The Phase 2 Report was published in February 2010. SLO APCD supplied an Excel spreadsheet version of the data to CDPR on May 5, 2010. That version lacks wind speed or other data, but does include the visitor data and 24-hour PM10 for the three stations arranged in columns with a row for each day. Table 3.2 comes from a section of the spreadsheet where formulae in cells calculate the published result. The formulae apply calculations to a span of day rows. Although the table purports to pick the 50 high and 50 low visitor days, the calculations are applied to a different span of days – not 50 high and 50 low. As shown below, the number of days used in calculating the mean, standard deviation and T Test statistics is quite different for the different stations.

Count of non-zero days included in calculations in Table 3.2 as published

| | Highest 50 days for vehicles | Lowest 50 days for vehicles | Highest days - Lowest days | STDEV High | STDEV Low | TTEST High | TTEST Low |
|---|---|---|---|---|---|---|---|
| Average SVRA Vehicles | 51 | 51 | 0 | | | | |
| Average Mesa2 PM10 | 51 | 51 | 0 | 51 | 51 | 51 | 51 |
| Average CDF PM10 | 48 | 45 | 3 | 48 | 45 | 48 | 48 |
| Average Oso PM10 | 36 | 107 | -71 | 36 | 34 | 36 | 34 |

To illustrate, the PM10 mean at Mesa2 looked at 51 high visitor days; at Oso it looked at only 36 high days. The PM10 mean at Mesa2 looked at 51 low visitor days; at Oso it looked at 107 low days, but the statistical test considered only 34 low days. These errors illustrate the incorrect application of analysis to data.

**2. Table 3.2 is the wrong way to analyze the data for association.**

**a. Lumping and comparing the 50 high and 50 low days ignores over two-thirds of the data and confounds the analysis by disassociating the actual PM10 and visitation numbers within each lumped group.**

Even done correctly, the analysis in Table 3.2 is the wrong way to look for an association between ODSVRA use and PM10 downwind. By lumping the high and low use days, the analysis ignores roughly two-thirds of the data – all the days of the year when visitation ranged from 436 visitors per day to 2,690 visitors per day. By combining the high and low days, the analysis loses sight of what actual PM10 levels were associated with visitor level. That blends data and loses precision.

Table 3.2 tries to draw conclusions from a higher mean PM10 level at Mesa2 on high use days, but the lumping process increases the variance and confounds the comparison, swamping any possible effect, even though the level of visitation differs by almost 10-fold. Consider the days listed below – two examples of "high" and two of "low".

Examples of variability in high/low visitor data

| Date | 24-hr average PM10 | | | Number of Visitors |
|---|---|---|---|---|
| | Mesa2 | CDF | OSO | |
| High Days | | | | |
| 6/7/08 | 114.73 | 34.17 | 90.50 | 3137 |
| 7/5/08 | 18.58 | 22.58 | 27.83 | 5002 |
| Low Days | | | | |
| 4/29/08 | 106.71 | 36.00 | 98.50 | 430 |
| 5/5/08 | 22.32 | 33.88 | 36.67 | 362 |

These days – each less than a month apart – show how much variation there is in the observations. On June 7, 2008, Mesa2 and Oso exceeded the state PM10 standard, while CDF (downwind of ODSVRA) did not; there were 3,137 visitors at ODSVRA. On July 5, 2008, PM10 levels were well below the state standard at all three stations and there were 5,002 visitors at ODSVRA. Yet in the analysis in Table 3.2 all these days are lumped together.

Even if lumping days was a valid approach, it would be useful to test the converse relationship: compare visitation levels during the highest 50 PM10 days with visitation during the lowest PM10 days.

| | Highest 50 PM10 days at Mesa2 | Lowest 50 PM10 days at Mesa2 | Highest days - Lowest days | Statistical Confidence of Data (1-P) |
|---|---|---|---|---|
| Average PM10 at Mesa2 | 74(50) | 9 (50) | 65 | |
| Average Vehicle Visits | 1293.3 (50) | 1384.6 (50) | -91.3 | 66.69%(50,50) |

| | Highest 50 PM10 days at CDF | Lowest 50 PM10 days at CDF | Highest days - Lowest days | Statistical Confidence of Data (1-P) |
|---|---|---|---|---|
| Average PM10 at CDF | 40 (50) | 19 (50) | 21 | |
| Average Vehicle Visits | 1495.0 (50) | 1360.1 (50) | 134.9 | 71.79%(50,50) |

| | Highest 50 days at OSO | Lowest 50 days at OSO | Highest days - Lowest days | Statistical Confidence of Data (1-P) |
|---|---|---|---|---|
| Average PM10 at OSO | 63 (50) | 11 (50) | 53 | |
| Average Vehicle Visits | 1322.5 (50) | 1483.7 (50) | -161.2 | 76.70% (50,50) |

This shows that there were more visitors using the ODSVRA during the lowest PM10 days at Mesa2 than during the highest PM10 days. But the difference is statistically insignificant; it is due to random variation in measurement and is meaningless.

**b. The correct approach is to use a linear regression where individual days with pairs of visitor and PM10 numbers are examined to see if there is a relationship. There is none.**

The best way to search for an association is to compare data in a linear regression. There, each pair of data – PM10 and visitation – for each day is plotted to see if variation in one produces consistent variation in another. If you had data pairs X and Y and found that Y plotted against X showed that as X was large, Y was also large, you could conclude that X and Y were correlated. This doesn't mean that large X produces large Y, just that they occur together. The approach allows prediction, even if there is no causation, and is commonly used for a data set such as the Phase 2 Report.

The Report does test this correlation. Figure 3.64 shows an X-Y plot of Mesa2 PM10 against visitation. When all the data are taken as pairs, there is no correlation: visitation is flat against PM10. Figure 3.64 uses the incorrect 13-month March 2008 data; when the correct published data are used, the results are essentially the same – no correlation.

Figure 3.64 also reports an important statistic, R-squared. R-squared is a measure of the extent of variation in one variable that is explained by variation in the other. In these cases, using the original analysis and using the corrected analysis, the R-squared value is less than a tenth of a percent, meaning that less than a tenth of a percent of variation in PM10 levels can be explained by variation in visitation levels.

This is an astonishingly miniscule measure of correlation. No responsible conclusion on association of PM10 and visitation can be drawn in the face of this result.

**3. Statistics are being misused**

**a. Statistical significance is an important test for variable measured data.**

When measurements vary over time – PM10 is high on some days and low on others – it is important to test to see whether any observed differences between one group of observations such as high visitation days may be the result of random variation in the measurement or may be attributable to a real difference between sample groups. If the difference is so great as to be unlikely to be mere random variation then it is called statistically significant.

There are several basic tests to be used. The Phase 2 Report applies and reports the results from a T test in Table 3.2. Generally, when a probability result (1-P) is less than 95%, the difference is considered less than significant because it means that there is a less than 95% chance that the observed differences could be the result of random variation in measurements. The 95% level of significance is well accepted in engineering and scientific work. Use of a lower significance test must be justified.

It is incorrect for the APCD staff to cite results shown to be statistically insignificant as if they were proven fact. Trends, hunches, guesses, common wisdom, gambler's luck and the like can rely on statistically unproven results, but a scientific study should be held to accepted standards. The APCD staff should not cite results that are insignificant. None of the associations between PM10 and visitation in the Phase 2 Report and study are significant.

**b. Even if an association was observed (which it was not) it is incorrect to conclude a causal relationship.**

Merely because two phenomena vary in the same way does not necessarily mean that there is a cause-and-effect relationship. As an example, if reading ability in elementary school children were compared to shoe size, it may be concluded that children with larger feet read better than children with smaller feet. While the association may be true, it does not mean that large feet improve reading; they are both correlated to another factor, in this case grade level.

It is extremely misleading for APCD staff to conclude that an association of visitation and PM10 necessarily means that high visitation causes high PM10. Even if found (which the data show were clearly not found), each could be the result of other factors such as good weather, season of the year, etc. that vary in the same way.

However, when two variables lack any correlation, it is a strong conclusion that they are not causally related. If ODSVRA use directly resulted in PM10 emissions and hence high PM10 downwind, you would expect the data to show it. Unless there are significant defects in PM10 measurements, it would be expected to show an increase in proportion to use. It does not.

**Conclusion**

Direct ODSVRA emissions may contribute to downwind PM10, but the Phase 2 Report does not provide an estimate of how much that may be. Clearly, no quantitative conclusion can be drawn from the data published and presented in the Phase 2 Report. Any statement relying on Table 3.2 for support is wrong. As demonstrated above, claiming causation from association is unsupported: Table 3.2 is based on an erroneous analysis of an incorrect 13-month data range, uses a method that is inappropriate to the purpose claimed, and produces no statistically significant result.