



Department of Pesticide Regulation



Paul Helliker
Director

MEMORANDUM

Gray Davis
Governor
Winston H. Hickox
Secretary, California
Environmental
Protection Agency

TO: Terri Barry, Senior Environmental Research Scientist
Bruce Johnson, Senior Environmental Research Scientist
Environmental Monitoring Branch **HSM-02030**

VIA: Joseph Frank, Senior Toxicologist
Worker Health and Safety Branch
324-3517

FROM: Sally Powell, Senior Environmental Research Scientist [original signed by S. Powell]
Worker Health and Safety Branch

DATE: September 5, 2002

SUBJECT: COMMENTS ON DRAFT MEMORANDUM TO DPR FROM FUMIGANT ALLIANCE DATED AUGUST 20, 2002, "PROPOSAL FOR DEVELOPING REGIONAL ZONES FOR BUFFER ZONE ANALYSIS"

Chuck sent Joe and me a copy of the August 20 Fumigant Alliance (FA) draft memo to see if we had any comments. Since you two are better qualified to address questions about the regionalization, I'll just relay my concerns to you to use as you see fit. Some of my concerns may simply reflect my ignorance of dispersion modeling.

Meteorological variables for cluster analysis

The input variables for the cluster analysis are not defined precisely in the FA memo. This may be partly because our directions to them were not sufficiently clear. I realize that I'm not wholly clear about it myself. I believe that the unit of analysis for clustering is intended to be the CIMIS station, with all of its data for the year. The distribution of hourly wind speeds, for an example, would be the distribution over the whole year. Am I correct, and if so, is this what we want? The distribution of hourly wind speeds over the year would not reflect any seasonal or daily clumpiness. Number of hours per day of calm hours reflects daily, but not seasonal clumpiness. It could be that for three seasons, a station has conditions that are not problematic for air concentrations, but in the fourth season, it's terrible.

Questions about specific input variables:

- a) Distribution of hourly wind speeds. Does the percent of winds ≤ 7 mph capture everything about wind speed that's important to dispersion? Or should more of the distribution be included, e.g., percent greater than some speed?
- b) Wind direction persistence. I *think* what FA means to do is, for each station, 1) record number hrs each day wind is from each direction, 2) find annual average number hrs per day from each direction, 3) record the maximum of these annual averages. Step 3 is the variable to be used in analysis. Does that variable capture the aspects of wind direction



important for dispersion? If monthly or seasonal effects could be important, the maximum average should be recorded for each month or season. However, I don't think the maximum average is the appropriate measure. The average time in each direction over days reflects the consistency of absolute wind direction over the year, whereas I thought the important thing was consistency within days, regardless of absolute direction. This would suggest the average daily maximum (in any one direction) is more appropriate. Also, would it make sense to subtract from each daily maximum the number of hours that day that the wind was in the opposite direction?

- c) Number of hours per day of calm hours. Should this be represented by distributions, rather than annual averages, which seems to be what FA proposes (it is unclear)? Perhaps percent of days with fewer than k_1 calm hours, and percent of days with more than k_2 calm hours.
- d) Number of hours per day with "F" stability. Same question as for calm hours.
- e) Distribution of hourly air and soil temperatures. I can't judge whether their variables capture the important aspects for dispersion.
- f) Solar radiation. How should this be operationalized?
- g) Number of hours per day with "D" and "F" stability. I don't understand what they are doing with this one.

Other details for the cluster analysis

I'm concerned that relying on only one year of data may give unrepresentative results.

I'm very concerned about the effect of missing data, including those removed because they are judged questionable. If data are missing in different seasons from different stations, as they certainly will be, then spurious differences and similarities between stations are very likely to be introduced. If the amount of missing data is not great, a solution would be to base the analysis on only those days for which every station has data. That would probably reduce the dataset too much, however, so it might be necessary to do some kind of imputation of missing data.

Determination of representative meteorological stations

I'd like to see some detail on the method for selecting representative stations. One method would be to look at the distribution of each input variable within a cluster and choose a station that is near the average on all of them. Another way to select a typical station could be using the

Terri Barry
Bruce Johnson
September 11, 2002
Page 3

station with the smallest distance from the cluster centroid. However, it may be preferable to select a station representing the worst case for expected air concentrations, rather than the average. The best thing, in my view, would be for DPR to make the selections, based on the FA analyses.

cc: Chuck Andrews