August 30, 1999

**MEMORANDUM**

TO: Stephen Kroner
OSWER/OSW/HWID

From: Sharon LeDuc
ORD/NERL/AMD/MSAB

Subject: Site Selection for HWIR Sampling

The overall goal of HWIR is to implement a methodology suitable for determining constituent-specific waste stream concentrations that represent a threshold below which Subtitle C disposal will not be required. The waste stream may then "exit" the hazardous waste management system. Since HWIR99 is to be a risk-based rule, the intent is to set exit levels such that no significant risk to human or ecological health shall occur as the result of implementing the new exit levels. In characterizing risks, HWIR99 will employ mathematical models to simulate the multimedia release of contaminants from land-based waste management units, their multimedia transport, and the subsequent exposure and risk to human and ecological receptors.

The site selection HWIR99 methodology described here addresses the six general objectives cited earlier. The assumptions that are appropriate and necessary for this section are: 6) calculation of measures of protection is performed at the site level and aggregated over sites to estimate National statistics, 7) waste management units are located as indicated in OSW's industrial Subtitle D database. These assumptions are necessary to assure that:

*The assessment will be national in scale and site-based, that is, risks will be assessed at individual sites across the U.S. where HWIR WMU's may be located. The resulting national distribution of site-based risks will form the basis for establishing exit criteria. For each site, statistically sampled from a national database of WMUs, the simultaneous release of chemicals from the WMU to each environmental medium, the fate and transport of the chemical through a multimedia environment, and the receptor-specific exposures and risks will be simulated. The exit levels will be applicable to all wastestreams and all locations, i.e., nationally.*

While the exit criteria to be developed are intended for National application, the modeling strategy is a "tiered site-based" approach. This reflects an agency decision to base exit levels on an assessment of potential risk occurring at Subtitle D facilities where WMU's may be located. In this approach, the individual site-based assessments are embedded in a two-stage Monte Carlo simulation procedure designed to produce sufficient site-based risk assessments to result in a national-scale statistical distribution of risk. The Monte Carlo analysis also provides explicit quantification of variability and uncertainty associated with the risk estimates and is flexible enough to accommodate alternative policy formulations.

**Subtitle D Survey**

The national database we used to sample the population at risk is the Subtitle D Survey (Schroeder

*et al.*, 1987) which was the basis for previous research by Clickner and Craig (1987). In that study a stratified random sample of industrial establishments was taken from 17 target industry sectors identified from a 1985 Duns Marketing, Inc. list. One objective of the survey was to develop a list of industrial establishments generating industrial Subtitle D solid waste and managing it onsite using specified types of WMUs. Along with the development of the list, an associated objective was to estimate the number of such facilities, nationally and industry-specific.

Since HWIR is an assessment at the national level, it should represent the national risk from the specified WMUs at all such industrial facilities. The sample which was used to estimate the number of facilities, i.e. the Subtitle D Survey, will be basis for estimating the national risk. The Subtile D Survey was based on the stratification by establishment size and by 17 industries. A sample of 18,051 was taken from the entire 1985 Duns Marketing, Inc. list frame of nearly 150,000 establishments. The initial survey on these 18,051 was limited to a telephone screening survey and information collected was not confirmed in the field. The second phase of the survey considered establishment size in the selection of sample for the follow-up survey. Based on EPA studies (EPA, 1983 and S.A.I.C., 1985) larger facilities were believed to be more likely to handle waste onsite. Thus, facility size, as indicated by the number of employees, was used to classify each facility as large, medium, or small. Other better measures of size related to to amount of industrial waste generated (e.g. production volumn) were generally not available at the facility level. Sampling sizes and rates varied with industy. In their paper, Clinker and Craig showed that the stratified sample provided a more efficient estimate of eligible facilities than a simple random sample. The result of this second phase of that survey was a database, the Subtitle D Survey, with 2,718 facilities that had one or more of the four types of waste management units: landfills, surface impoundments, waste piles, and/or land application. The database actually had 2,850 facilities. The additional facilities were not in the statistical sample. Documentation indicated that the data had been provided without being requested as part of the statistical sample. Data for these facilities were archived in the database in spite of the fact that they were not in the statistical sample.

There are several known concerns that accompany use of the Subtitle D Survey for estimating national risk. First, the national database was developed from a 1985 industry list. Changes in the frequency, type or location of such facilities may have changed. One type of WMU, aerated tanks, was not considered in the Subtitle D Survey, but was required to be evaluated in HWIR99. Simulated results for aerated tanks are based on the assumption that any site which has a surface impoundment could, for the purpose of establishing risk, have an aerated tank instead. It should be pointed out that some of the industrial sectors included in the list which resulted from the Subtitle D Survey are unlikely to generate hazardous waste, but they will be analyzed as if they do. Commercial waste management facilities are not part of the sampling frame and the population at risk near these facilities is not evaluated.

As mentioned, the purpose of the sampling resulting in the Subtitle D Survey database, was to establish an estimate of the eligibility rate of facilities. For establishments with small numbers of employees it was assumed that 5% would have on site waste units. Resulting sample weights were slightly above 25 in a few cases. For the establishments classified as mid-size it was assumed that 20% would have on site waste units and with this assumption the maximum weight

was 10. These weights were to used to estimate efficiently the rate of eligibility. The Subtitle D Survey database resulted as the list providing details on the establishments which actually had the onsite waste units.

For HWIR, the goal is to estimate the potential risk to human and ecological receptors living in the vicinity of industrial waste sites that could manage HWIR exempted waste. The weights just defined could be used, if several implicit assumptions are generally true. However, because of the small sample, large weights for some of the sites could greatly influence the results. The weights apply to estimating the eligibility rates of facilities, not characteristics that affect risk, e.g. environmental setting.

**HWIR Sample Selection**

The sample of sites to be used for the HWIR assessment was selected from the Subtitle D Survey. A simple random sample was drawn from 2850 in the Subtitle D Survey, i.e. from the facilities handling onsite waste. In drawing this sample, each facility had an equal chance of being selected. There was, however, the possibility that an industry would not be sampled at all. This is particularly true with a small sample size. Since it was important to include onsite waste management from each industry sector or group, the proportion to be drawn from each industry should be the same as its proportion in the Subtitle D Survey. These proportions, based on Clinker and Craig (1987) are called factors and are shown in Table 1. For example, organic chemicals had 78 samples which had onsite WMUs and the total number across all industries was 2718. The factor for organic chemicals was 78÷2718 or about .03 (3%). The size of the sample for the HWIR assessment was set at 200. The size of the sample was not determined because of statistical criteria, but rather for logistical and resource considerations. The sample size could be increased with additional resources. With the sample size set at 200 the number of samples to be selected from the organic chemicals industry was 200 times the organic chemical factor, 0.03, which is 6. The same calculation was used for each industry group. Because sample size was rounded for each industrial sector, the total sample size was 201 rather than 200.

Table 1. Proportionality Factors for Strata

| INDUSTRY GROUP | FACTOR |
|---|---|
| Organic chemicals | .03 |
| Primary iron and steel | .10 |
| Fertilizer & agricultural chem. | .04 |
| Electric power generation | .07 |
| Plastics and resins mfr. | .03 |
| Inorganic chemicals | .07 |
| Stone, clay, glass & concrete | .19 |
| Pulp and paper industry | .07 |
| Primary nonferrous metals | .04 |
| Food and kindred products | .11 |
| Water treatment | .03 |
| Petroleum refining | .08 |
| Rubber & misc. products | .02 |
| Transportation equipment | .04 |
| Selected chem. & allied prod. | .02 |
| Textile manufacturing | .05 |
| Leather and leather prod. | <.01 |

To achieve the random selection of the sample for the HWIR sites, each facility in the Subtitle D Survey was assigned a random number. The facilities were then ordered based on the magnitude of the random number. The facilities for each industry group were selected starting with the lower random numbers. For example the six facilities with the lowest random numbers were selected for the organic chemicals industry. Each industrial group was handled separately. Information on the WMU was critical to the assessment. Nine of the selected sites had to be replaced because the WMU information was inconsistent or missing and five had to be replaced because of missing locational data. The replacement was done by selecting the facility with the next lowest random number as the replacement.

During the second phase of the Subtitle D Survey, all large establishments were included. Indeed, the survey results reported in Clickner and Craig established that the eligibility rate increased with the size of the facility for 15 of the 17 industrial groups. Since the number of sites for the HWIR

sample is only 201 (much smaller than the 2,718 in the Subtitle D Survey) stratification by establishment size was not possible. However, since more large establishments were in the Subtitle D Survey, it is more likely that the larger establishments are in the HWIR sample. When selecting the HWIR sample, we did not consider the size of the establishment. If probability of 1 had been used as the probability of selecting large establishments (as was done by Clickner and Craig, 1987), no small or medium establishments would have been in the sample, because the sample size was too small.

**HWIR Sample Aggregation and Use of Weights**

The Subtitle D Survey database and the resulting HWIR sample of 201 will be aggregated and interpreted to represent the population at risk. The issue of what weights to use for each of the 201 samples has several important considerations and indeed the sensitivity of the results to the use of different weights is not known. The simplest is the use of equal weights for each of the samples. A second option is the use of the weights available in the Subtitle D Survey. It may be possible to estimate more appropriate weights in the future, but they are not available at this time.

The reasons for not using the Subtitle D Survey weights rests on several considerations. The weights are based on an assumption that there is a direct relationship between the establishment employment size and the site likelihood of having a waste unit. Craig and Clickner (1987) found the assumed relationship was not true in two of the 17 industrial groups. The weights do not consider the population being exposed but implicitly assumes that the same proportions used for the Subtitle D Survey, i.e. 5% for the small establishments and 20% for the mid-sized establishments, are appropriate. For some waste unit types the HWIR samples are much smaller than 201 (land application units exist at only 36 sites) and high weight at even one of these sites could have undue influence on the percentiles. The weight of most of the samples is one with 186 having a weight of ten or less and 10 having a wight of thirty or more. This is a very skewed distribution for the weights. High variability of risk at one of the sites with a high weight, might well lead to a very inappropriate design value.

The equal weight approach ignores the fact that the sites of the establishments which had fewer employees in 1985 had a smaller probability of being in the HWIR sample even though they represent a greater number of establishments. If these sites are in the HWIR sample their influence is less using equal weights than would be indicated based on the Subtitle D Survey weights. If assumptions, made to estimate proportions having similar waste units, are indeed applicabale to proportions of the population at risk, then the equal weights approach ignore information. The validity of these assumptions is not known.

Sample weights for using the Subtitle D Survey weights may be determined. The Subtitle D Survey weights are site specific and each is multiplied by the industry specific adjustment factor from Table 2. The result of this multiplication is applied to the results for each type of waste management unit at that site. The adjustments in Table 2 will maintain the weight of each industry group to be the same as was in the Subtitle D Survey. Column “c) adjustment factor” increases the weight at each facility so that the total weight of that industrial group in the HWIR sample is exactly the same as it was for the Subtitle D Survey. The adjustment factors must be integers so

either the greatest integer no larger than the adjustment factor or the rounded adjustment factor could be used. Whereas the highest Subtitle D Survey weight for a single site in the Subtitle D Survey was just over 25 it is now greater because of the adjustment factor.

Table 2. Adjustments to Subtitle D Survey Weights Applied to Each Sample

| INDUSTRY GROUP | SUM OF WEIGHTS | | c) ADJUST-MENT FACTOR c = b/a |
|---|---|---|---|
| | a) 201 SAMPLE | b) SUBTITLE D SURVEY | |
| Organic chemicals | 15 | 123 | 8.20 |
| Primary iron and steel | 28 | 595 | 21.25 |
| Fertilizer & agricultural chem. | 38 | 197 | 5.18 |
| Electric power generation | 15 | 399 | 26.60 |
| Plastics and resins mfr. | 6 | 106 | 17.67 |
| Inorganic chemicals | 42 | 399 | 9.50 |
| Stone, clay, glass & concrete | 280 | 4304 | 15.37 |
| Pulp and paper industry | 20 | 476 | 23.80 |
| Primary nonferrous metals | 12 | 416 | 34.67 |
| Food and kindred products | 176 | 2669 | 15.16 |
| Water treatment | 8 | 388 | 48.50 |
| Petroleum refining | 39 | 455 | 11.67 |
| Rubber & misc. products | 3 | 239 | 79.67 |
| Transportation equipment | 45 | 394 | 8.76 |
| Selected chem. & allied prod. | 5 | 96 | 19.20 |
| Textile manufacturing | 15 | 524 | 34.93 |
| Leather and leather prod. | 3 | 64 | 21.33 |

**Representativeness of the HWIR Sample**

Statistics comparing the 201 HWIR sample facilities with the entire set of 2,850 facilities with on

site WMUs indicate that the HWIR sample includes a higher percentage of sites with accurate locations than the full database (Table 3). The sample is similar with respect to the percent of facilities with each type of unit (Table 4), but does not catch the tails of the WMU area distribution for LAUs and waste piles (Table 5). This is not unexpected considering the extremely long tails for these distributions.

**Table 3. Facility Location Method Distribution of 201 Sample Facilities**

| Method | Accuracy (m) | 201 Sample Facilities | | 2,850 Ind. D facilities | |
|---|---|---|---|---|---|
| | | *Number* | *Percent* | *Number* | *Percent* |
| I1, 12 | 2.2 - 2.6 | 5 | 2.5 | 51 | 1.8 |
| | 20.3 - 27.4 | 36 | 17.9 | 463 | 16.2 |
| | 50 | 0 | 0.0 | 12 | 0.4 |
| | >150 | 0 | 0.0 | 3 | 0.1 |
| A1 | 150 | 69 | 34.3 | 818 | 28.7 |
| A4 | 1,000 | 3 | 1.5 | 40 | 1.4 |
| AO | 1,000 | 2 | 1.0 | 18 | 0.6 |
| Z4 | 1,000 | 2 | 1.0 | 22 | 0.8 |
| A6 | 2,000 | 4 | 2.0 | 73 | 2.6 |
| Z2 | 6,000 | 2 | 1.0 | 33 | 1.2 |
| Z1 | 11,000 | 33 | 16.4 | 341 | 12.0 |
| Z1* | 11,000 | 45 | 22.4 | 976 | 34.2 |
| **Total** | | 201 | 100 | 2850 | 100 |

Z1* - zip code centroid assigned by zip code in 1985 Industrial D Screening Survey

**Table 4. WMU Type Distribution of 201 Sample Facilities**

| WMU Type | 201 Sample Facilities | | 2,850 Ind. D Facilities | |
|---|---|---|---|---|
| | *Number* | *Percent* | *Number* | *Percent* |
| landfill | 56 | 20.9 | 790 | 21.3 |
| land application unit | 20 | 7.5 | 310 | 8.4 |
| surface impoundment | 132 | 49.2 | 1,777 | 48 |
| waste pile | 60 | 22.4 | 827 | 22.3 |
| **Total** | 268 | 100 | 3,704 | 100 |

**Table 5. WMU Area (Average) Distribution of 201 Sample Facilities**

| Subtitle D Weight Percentile | Landfill ($m^2$) | | LAU ($m^2$) | | Surf. Imp. ($m^2$) | | Waste Pile ($m^2$) | |
|---|---|---|---|---|---|---|---|---|
| | *Sample* | *All* | *Sample* | *All* | *Sample* | *All* | *Sample* | *All* |
| 25th | 691 | 3,645 | 1,364 | 8,090 | 471 | 405 | 20 | 20 |
| 50th | 11,923 | 14,197 | 10,800 | 59,154 | 2,327 | 2,024 | 243 | 121 |
| 75th | 48,692 | 60,518 | 38,503 | 269,797 | 8,856 | 8,094 | 1,968 | 1,380 |
| 90th | 115,208 | 141,645 | 119,114 | 892,410 | 23,969 | 40,311 | 2,121 | 4,817 |
| 95th | 144,249 | 221,721 | 133,824 | 1,992,022 | 72,134 | 83,836 | 7,493 | 12,100 |
| 99th | 545,579 | 1,034,009 | 984,089 | 12,802,769 | 560,542 | 534,580 | 43,164 | 993,750 |

| Equal Weight Percentile | Landfill ($m^2$) | | LAU ($m^2$) | | Surf. Imp. ($m^2$) | | Waste Pile ($m^2$) | |
|---|---|---|---|---|---|---|---|---|
| | *Sample* | *All* | *Sample* | *All* | *Sample* | *All* | *Sample* | *All* |
| 25th | 12,009 | 11,332 | 20,200 | 16,200 | 1,012 | 971 | 53 | 41 |
| 50th | 27,857 | 36,423 | 54,650 | 80,900 | 3,723 | 4,047 | 476 | 445 |
| 75th | 102,819 | 101,175 | 128,750 | 341,750 | 24,666 | 16,188 | 2,040 | 4,050 |
| 90th | 283,290 | 259,413 | 511,600 | 911,400 | 181,866 | 80,940 | 12,510 | 16,200 |
| 95th | 472,487 | 588,838 | 944,950 | 1,984,000 | 337,479 | 202,350 | 28,360 | 47,865 |
| 99th | 934,250 | 1,867,852 | 1,156,990 | 6,479,300 | 1,058,533 | 898,962 | 243,820 | 908,600 |

There are several issues that arise when using the Subtitle D Survey database exclusively. This database may include some units where HWIR exited waste is unlikely to be disposed. These facilities may have large WMUs and thus indicate higher risk than is realistic. One could remove industrial categories that do not appear to be major generators of industrial waste. Also individual facilities could be removed if they did not have a Resource Conservation and Recovery Act (RCRA) identification code, i.e. they had not been identified as a facility that treats, stores or disposes of hazardous waste. Neither of these was done for the HWIR sample. However, once the industrial categories or individual facilities were identified, the weights would not need to be adjusted. The risk would be set to zero.

The other source of error would be the omission of facilities that are not in the 17 industrial sectors, but generate hazardous waste. These facilities would need to be identified and weights would need to be derived. One possible type of facility is that of the commercial non-hazardous waste facility.

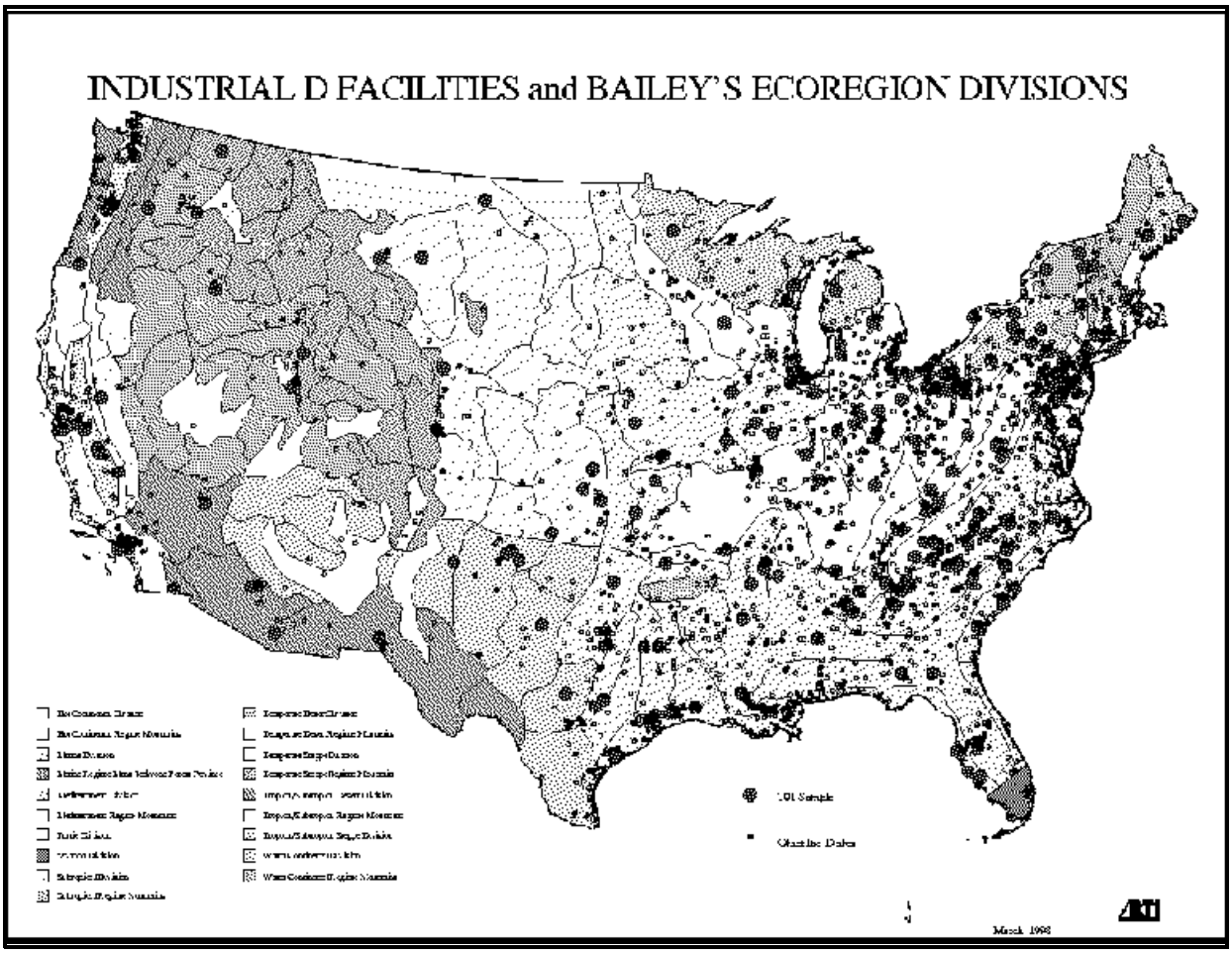


**Figure 1**. Subtitle D Survey Sites (small open circles) and HWIR Sites (large solid circles)

The geographic representativeness is of interest as well. The locations of the 201 sample facilities with respect to the 2,850 facilities with onsite WMUs in the Industrial D screening survey indicate that the sample provides good geographic representation and also good ecosystem diversity. In Figure 1 these locations are overlain on Baileys ecoregion divisions and both the sample sites and the full data set locations shown. In Figure 2 Theissen polygons are shown around the 218 continental US meteorological stations. The spatial coverage of the 201 sample sites, in terms of basic physiographic/climatic regions, as determined by Baileys ecoregions, and the meteorological stations being used appears to be in good agreement with the complete sample.

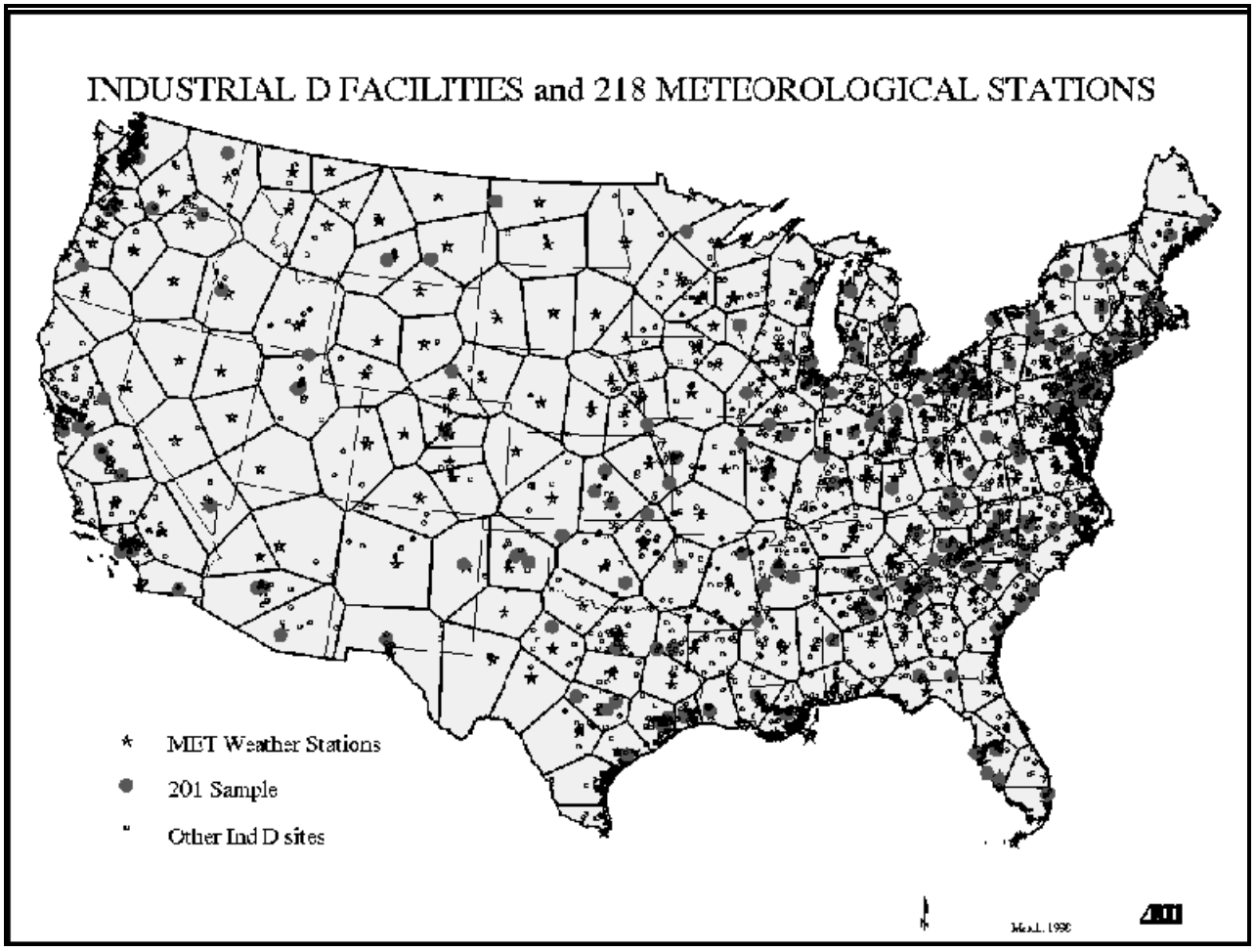


Figure 2. Same as Figure ! along with location of Meteorological Stations (stars)

**Reference**

Clickner, Robert P. and Jim Craig. 1987. "Using Business Establishment Size in an Environmental Survey", Proceedings of the American Statistical Association: Survey Research Section, 5 pp, American Statistical Association, 1429 Duke St., Alexandria, VA 22314-3402, 5pp.

Clickner, Robert, 1988. "Sampling weights for the industrial Subtitle D screening survey", July 28 , 1988, Westat memorandum to Zubair Saleem, U.S. EPA, 2pp.

EPA (United States Environmental Protection Agency), 1983. "Surface Impountment Assessment National Report", EPA 570/9-84-002.

Schroeder, Kirsten, Robert Clickner and Esther Miller, 1987. "Screening Survey of Industrial Subtitle D Establishments, Draft Final Report," submitted to EPA Office of Solid Waste, by Westat, Inc. 1650 Research Blvd. Rockville, MD 20850, under contract # 68-01-7359, December 29, 1987, 154pp.

S.A.I.C. (Science Applications International Corporation), 1985. "Summary of Data on Industrial Non-Hazardous Disposal Practices", prepared for the U.S. EPA under contract # 68-01-7059, December 5, 1985.