GUIDELINE DOCUMENT
FOR
BACKGROUND AIR QUALITY
DETERMINATIONS

Department of Environmental Protection
Bureau of Air Quality

5-15-96

TABLE OF CONTENTS

DEFINITIONS

For the purpose of this document, the following terms shall mean:

1. Appropriate Data: With respect to monitoring data, data that reflects the same or comparable spacial and temporal conditions found at the applicant's facility. Where practical, data shall be collected "on site". Data collected "off site" shall meet with the approval of the Division Director prior to being accepted by the Department.

2. Commissioner: Commissioner of the Department of Environmental Protection.

3. Division Director: Director of the Field Services Division of the Bureau of Air Quality.

4. Field Services: The Division of Field Services, Bureau of Air Quality, Department of Environmental Protection.

5. Minimal: The least possible impact or effect from a source based on available data. In no case shall a source contribute 10% or more to a sample being considered. In all cases, the Division Director reserves the right to reject data with measurable source impact.

6. Other suitable analysis: Analysis that shall provide the same or superior qualitative and quantitative results as those provided by existing specified analytical techniques. The Division Director reserves the right to reject any analysis other than that specified by name in this document.

ACKNOWLEDGMENTS

I would like to recognize the efforts of Jeff Emery and Sandy Tate who were responsible for writing and putting together the first guideline document for the Bureau on determining background air quality. Their initial work provided for consistency in how background levels were determined by all AQS staff, and their original document has served us well.

I also want to recognize Merrilee Carlson's contributions to this first revision of the guidelines. Her input from a meteorological, modeling and licensing perspective was very valuable and much appreciated. Special thanks also go to Mary Butler, for without her typing skills in general and mastery of the word processing capabilities of the microcomputer specifically, this revision would still only exist as a jumbled mass of drafts containing innumerable corrections. Thank you for your patience Mary and your collaborative attitude which greatly helped me produce a finished product.

Recognition should also be given to Jeff Emery, Cathy Clayton-Richardson, Denise Cormier, Rob Cote, and Rick Perkins for their contribution to editing the 1993 revision.

Andrew Johnson
March 1987

Donald E. Darling Jr.
November 1993

PREFACE

The need for a revision of the guideline document on background and baseline air quality determinations has grown over the years since it was first formally adopted by the Bureau of Air Quality in July 1983. There appears to be two major reasons. First, as our practical experiences in applying the guidelines grew, so did our awareness of various portions of the document which either needed further clarification and/or expansion, or just simply were not applicable to what was actually being done to accomplish the tasks of determining background and baseline air quality. Second, as the degree of complexity in licensing certain sources increased, so did the level of sophistication required for making the determinations. This resulted in some different approaches being used other than those contained in the existing guidelines. Therefore, the time seemed appropriate for incorporating these experiences, approaches and additional information in revisions of the background and baseline determinations guideline document.

Perhaps the biggest change to be noted is the elimination of all references to determining baseline air quality. Past experience and current practice has shown this determination is best made with air quality modeling by using actual or representative emissions from the larger sources in operation as of the baseline date, and any increases in those emissions in conjunction with the applicant's emissions. In fact, no baseline air quality determinations have been made using air quality data since the guidelines were adopted. Hence, the revised title, "Guideline Document for Background Air Quality Determinations," reflects the focus of these revisions.

This revision is not intended to be all encompassing in terms of dealing with all possible background issues or circumstances. It is intended to provide a starting point from which applicants can develop meaningful background data in support of licensing their new source, or modifying an existing source.

Applicants are encouraged to approach Air Bureau staff with their concerns and questions early in the application process. Background determinations using procedures not defined in this guideline may be acceptable, provided the applicant discusses their approach with Air Bureau staff first.

1. INTRODUCTION

When licensing an air emission source, a finding has to be made showing that the applicant's emissions, in conjunction with existing emissions, will not cause or contribute to a violation of applicable state ambient air quality standards and increment. The purpose of determining background air quality is to provide a qualitative assessment and/or quantification of the "existing" air quality in an area to which the source's impacts can be added. These results will then be compared to the Maine ambient air quality standards (MAAQS). Background air quality is not needed for an increment analysis as increment consumption is determined by modeling. For the purpose of this document, background air quality is defined as existing air quality exclusive of licensed air emissions sources.

The Division of Field Services is responsible for determining background air quality levels for the various pollutants. Background air quality is determined by using ambient air quality monitoring data from the area impacted by the applicant. Background data may be obtained from other sources when necessary. Field Services is also responsible for reviewing background air quality analyses performed by outside parties. When a review is needed, a copy of the air quality analysis must be sent to the Division Director who will then forward it to the appropriate Field Services staff member for evaluation. When a background determination is needed, a written request, containing the information described in Section 2 should be sent to the Field Services Division Director. The appropriate Division staff, after consulting with licensing staff, will then decide whether a qualitative or quantitative background air quality determination is needed (see Section 3). The remaining sections of the guideline document detail the process Field Services uses to determine background air quality.

Background air quality shall not be determined using data older than three years from the date that the Department is notified of an Air Emissions License Application unless no suitable data can be obtained within this time frame.

2. REQUESTS FOR BACKGROUND AIR QUALITY DETERMINATIONS

The written request for a background air quality determination should be sent to the Air Quality Services' Division Director as early in the licensing process as possible. This is important for several reasons. In some instances, representative air quality data may not be available for an area, so preconstruction monitoring (in excess of 1 year) may be required. Air quality problem areas (or potential nonattainment areas) can be identified at the outset. Depending on the type of determination necessary, it can take several months to produce a finished background air quality report. The Field Services Division Director will assign the task of fulfilling the background request to the appropriate regional office supervisor in whose region the applicant is located (or proposing to locate) if the applicant is not responsible for the analysis.

The written request should contain the following information:

1. Name of applicant/source.
2. Address of applicant/source.
3. Type of license applied for (renewal, modification, new source).
4. Does this application reflect either an emissions increase, decrease or no change?
5. Size of source in terms of BTU/hour capacity.
6. List of pollutants, and their averaging times (if necessary), requiring background determinations.
7. Description of maximum impact areas and what the impacts are, if known.
8. Map showing applicant's location and impact areas.
9. Date by which the background determination is needed.

3. GUIDELINES FOR SELECTING THE TYPE OF BACKGROUND AIR QUALITY DETERMINATION REQUIRED

A finding has to be made when licensing a source that ambient air quality standards will not be exceeded as a result of emissions from the applicant in concert with other sources and existing background. In the past, most background determination requests resulted only from the licensing of new sources or major modifications of existing sources. Now, a determination of existing air quality will be required for all licensed sources. To provide background determinations for all of these sources at the level and scope they have been provided in the past, would not be feasible given existing resources. Instead, two types of background determinations may be provided: qualitative background and quantitative background. These determinations need to be made prior to licensing a new source, or licensing a source undergoing a major modification, or relicensing an existing source. This determination will be valid for three (3) years from the date the determination is issued. Background data older than three (3) years will not be considered representative.

A. QUALITATIVE BACKGROUND AIR QUALITY DETERMINATIONS

This type of determination is perhaps best described as a general characterization of background air quality presented in a narrative style. It may or may not be based on representative air quality data from the area under consideration. Some licensing cases which would only require a qualitative determination are:

1. When a modeling analysis is not required by DEP's Emission License Regulation (Chapter 108);

2. When a previous determination has been made for the applicant that ambient air quality standards will not be exceeded, and the applicant is applying for a renewal;

3. When the applicant's modeled impacts are below de minimus levels (Chapter 100 (88)) and there is nothing else limiting air quality in the applicant's significant impact area.

In summary, a source that can demonstrate that their impacts are not significant everywhere, at all times, need not submit a quantitative background air quality determination.

B. QUANTITATIVE BACKGROUND AIR QUALITY DETERMINATION

Quantitative Background can be described as a detailed analysis of air quality data which ultimately provides a background pollutant concentration value representative of existing air quality in a municipality, area, or region to which a source's modeled impacts can be added. Licensing cases which would require a quantitative background determination are:

1. When an application is for a new source;
2. When an existing source applies for a modification;
3. When an applicant proposes to locate in or has impact on a nonattainment area or an area where air quality increment is limited.

This type of determination requires using increasingly sophisticated methods to further refine the modeling and/or background analysis if initial results show potential air quality problems. (see section 8)

Preconstruction monitoring may be required if the expected emissions from the applicant exceeds the levels stated in Table 1. Expected pollutant emission increases below those listed in Table 1 generated by a new source, or an existing source proposing a modification to their process will require an ambient air quality standards compliance demonstration, but no preconstruction monitoring.

If preconstruction monitoring is required, the applicant shall establish and operate a monitoring network (up to 3 sites) for one year prior to their application. The data gathered from this network will have to meet the Maine Quality Assurance Plan for Ambient Measurement Systems or another quality assurance plan acceptable to the Department. Monitoring periods shorter than one year may be acceptable, but only with prior approval from the Department.

TABLE 1. EPA'S DE MINIMUS IMPACT LEVELS*

POLLUTANT	CONCENTRATION LEVEL		AVERAGING PERIOD
Carbon Monoxide	575 ug/m3	0.503 ppm	8-hr
Nitrogen Dioxide	14 ug/m3	0.007 ppm	annual
Total Suspended Particulate	10 ug/m3		24-hr
Sulfur Dioxide	13 ug/m3	0.005 ppm	24-hr
Ozone	-		-
Lead	0.1 ug/m3		24-hr
Mercury	0.25 ug/m3		24 hr
Beryllium	0.0005 ug/m3		24-hr
Fluorides	0.25 ug/m3		24-hr
Vinyl Chlorides	15 ug/m3		24-hr
Total Reduced Sulfur	10 ug/m3		1-hr
Hydrogen Sulfide	0.04 ug/m3	0.05 ppb	1-hr
Reduced Sulfur Compounds	10 ug/m3		1-hr
PM10	10 ug/m3		24-hr

** From 40 CFR, Part 51, Section 24(i)(8)i*

4. GUIDELINES FOR QUALITATIVE BACKGROUND AIR QUALITY DETERMINATIONS

This type of determination is intended to be a simple, narrative statement consisting of a few paragraphs which characterize the existing air quality in an area without necessarily providing a quantification of any pollutant concentration values. The narrative should address at least the following points:

A. That the location of the source is in a designated attainment area for the criteria pollutant(s) of interest;

B. The mention of any ambient monitoring which may have been done in the area to confirm that air quality levels were below the MAAQS. Describe the pollutant(s)

monitored, the beginning and ending dates of the monitoring period, and the location of the monitoring site(s);

C. The mention of any pertinent information from the air emission licensing files and/or emissions inventory files that may suggest a substantial change in pollutant emissions since the last air quality monitoring study or air quality determination for the area;

D. The mention of any incidental air emission control activities taking place in the area (e.g. street sweeping) which may affect air quality.

5. GUIDELINES FOR QUANTITATIVE BACKGROUND AIR QUALITY DETERMINATIONS

There are several steps to this determination which are described in detail in the following subsections.

A. DETERMINING AVAILABILITY OF AMBIENT AIR QUALITY DATA

1. Ambient air quality data may be available for the proposed an area for determining background air quality. The Bureau of Air Quality maintains a monitoring sites file which should be reviewed to identify if any monitoring stations have been operated in the area under consideration. If no monitoring sites are found, see section 6.

2. The relevance of ambient air data is influenced by how current it is. Generally, data "older" than 3 years is not considered acceptable due to residential, commercial, or industrial growth in the area under consideration and upwind of this area. Some evaluation of growth, and changes in regional air quality may be required to assess the relevance of "older" data.

B. DETERMINING AVAILABILITY AND REPRESENTATIVENESS OF METEOROLOGICAL DATA

Representative meteorological data may be available for use in background determinations. A Bureau staff meteorologist should be contacted to review candidate meteorological monitoring sites and their meteorological data to determine representativeness. The meteorologist will then respond with a determination of which site(s) and corresponding data have been accepted.

Some of the determinations are:

1. That the meteorological data has not been influenced by any unusual topographical features.

2. Whether the data is continuous or instantaneous observations.

3. That the data was collected at the 10 meter level or higher.

C. EVALUATING AIR QUALITY DATA REPRESENTATIVENESS OF BACKGROUND AIR QUALITY

Once a monitoring station has been identified, an evaluation of its data is necessary to determine if it is representative of the applicant's impact area and of background air quality. The evaluation should consist of the following determinations:

1. Determine that the site(s) is(are) representative of the area's land use (such as rural, urban, industrial, commercial, etc.). This information can often be found on the site's identification form.

2. Determine that the site(s) is(are) located such that the applicant's emissions have had minimal influence on the data gathered from it. Minimal influence is somewhat subjective as the nature of the applicants emissions (e.g. process, combustion, or fugitive) and whether it is a continuous process or a batch process may affect the degree of influence on existing air quality data. Obviously, no impact or influence could have occurred if the application is for a new source.

   For the short-term background determinations (3-hour, 8-hour, 24-hour): the site could be source-oriented but it would be better if it was not. Representative meteorological data could be used to select days on which the applicant had no or minimal impact. The applicant is considered to have no impact when wind directions are outside a 90° sector of the azimuth (45° on either side) from the site to the applicant's location.

   Narrower sectors (i.e. <90° compass) may be used provided adequate horizontal stability data (i.e. standard deviation of horizontal wind direction) is available or if by excluding a full 90° sector impacts from other significant sources of pollutants would be eliminated. Horizontal stability data may be used to determine persistent winds carrying emissions that would otherwise pass by the sampling site without impacting on the site.. Broader sectors (i.e. 90°< compass may be required if there are several emissions points at the source. Such a condition may have resulted in the substantial mixing and dispersal of emissions over a greater area, influencing a greater percentage of the data. This condition would be "caught" in more refined reviews of the data.

   For long term standard background determinations (annual): data that was collected at a source-oriented site is more difficult to deal with because the applicant's or source's emissions could have had a large impact on the data. There is an approach for using source-oriented data to determine the annual background concentration. The annual average at a source oriented site can be

used as long as the actual impact from the source for the year can be demonstrated to be minimal. Alternative methods may be used to those described above. Any alternatives should be discussed with Bureau staff prior to their use.

3. Determine that the site(s) scale of representativeness appropriately represents the area being reviewed for background air quality and that it meets EPA siting criteria for that scale and type of pollutant monitoring. Also, note if the site is designated as a NAM, SLAM, SPM or PMS.

4. Determine that the available data had been collected during the three year period preceding the permit application date.

5. Determine if gaseous pollutant data was collected by a continuous monitoring method.

6. If less than one years worth of particulate data is available, assure that the sampling frequency was every day, and that data was collected during a period when maximum background concentrations would most likely have occurred.

7. For applicant's which fall under PSD review, determine that the percent data recovery for the monitoring period is 80% or greater.

8. For particulate data, determine that the site was not unduly impacted by fugitive emissions from a nearby roadway (except data from microscale sites). If impacts are being recorded from several roads in the area, then that would be considered background. If fugitive impacts are regional in nature such as from agricultural activities, etc., then they are considered part of background air quality.

9. Determine that the data meets the measurement completeness criteria for the various pollutant averaging times as outlined in Table 2 (page 8). A review of the monitoring methodology, quality assurance measures, and data reporting procedures will help establish that data also meets the Air Bureau's Quality Assurance Plan criteria.

10. Determine that the data was collected using an EPA approved reference or equivalent method for each specific pollutant monitored.

D. EVALUATING THE QUALITY OF THE REPRESENTATIVE DATA

Once the monitoring data is determined to be representative, then the quality of that monitoring data should be evaluated prior to the actual analysis. Data which is of low or questionable quality may be submitted to the Quality Assurance Committee for review for its use in background analysis. If the quality of the data is determined to be unacceptable, then the applicant may be required to monitor to provide the necessary

air quality data. If the applicant is not required to monitor, data from a site in another representative area or from a regional site may be used to determine background air quality. (see Section 6)

Determine that data collected since January 1, 1981, meets the quality assurance procedures as described in The Quality Assurance Plan, State of Maine Department of Environmental Protection or another plan approved by the Department. All known sampling/monitoring and data handling procedures that may have been used should be identified and listed. This list may be presented to the Quality Assurance Committee for an assessment of the data's quality. In many cases, a detailed list may not be necessary because of the Field Services Division's familiarity with the monitoring program from which the data was generated.

TABLE 2. COMPLETENESS CRITERIA FOR AIR QUALITY DATA USED IN BACKGROUND AIR QUALITY DETERMINATIONS

CONTINUOUS MEASUREMENT DATA

TIME INTERVAL	MINIMUM NUMBER OF OBSERVATIONS
3-Hour Running Average	3 Consecutive Hourly Observations
8-Hour Running Average	6 Hourly Observations
24-Hour	18 Hourly Observations
Monthly	23 Daily Averages
Quarterly	3 Consecutive Monthly Averages
Yearly	9 Monthly Averages With At Least Two Monthly Averages Per Quarter

NONCONTINUOUS MEASUREMENT DATA

TIME INTERVAL	MINIMUM NUMBER OF OBSERVATIONS
Quarterly	5 Values With At Least 2 Or More Values in Each of 2 Months
Yearly	4 Valid Quarters

E. DETERMINING BACKGROUND CONCENTRATIONS FROM AMBIENT AIR MONITORING DATA

Once the monitoring data has been determined to be both representative and of acceptable quality, the background determinations can be made. The procedures for arriving at the actual background air quality concentrations are as follows:

1. Contact the Division Director and request a report or data file listing the parameters you are interested in. The request should include: the site(s) of interest (including AIRS site code), parameter(s) of interest and the beginning in. You will be provided with a data file on electronic media (or hard copy) with the data you requested. You may have to compile the data manually -- listing pollutant concentrations in descending order from the highest concentration for the averaging time you are reviewing. Note that AIRS data

retrievals are in "block" averages and not sliding (or rolling) averages. We will require data in sliding averages.

2. When analyzing the pollutant concentration values for each separate averaging period, the following meteorological considerations should be employed:

   a. Data collected during "calm" wind conditions should be included for the initial analysis. Calm data may be excluded in subsequent (more detailed) background analysis. The applicant shall provide appropriate documentation to support accepting or rejecting the data. Particulate samples collected over a 24-hour period may prove acceptable provided the "calm" condition existed for less than one hour during the sampling period. Calm conditions existing for greater than one hour during the sampling period may/may not preclude the use of the sample for background if accompanied by petrographic analysis (or other suitable analysis). This analysis should demonstrate no or minimal impact on the sample from the applicant's facility or sources that are to be excluded for the purpose of determining background.

      Gaseous pollutant data may be reviewed more carefully with respect to the effects of calm winds on monitored air quality at a given site. Calm wind conditions may or may not result in an emissions plume extinding out to a given monitoring site. Data collected for determining 1, 3 or 8 hour averages should be scrutinized, considering the distance between the source under consideration and the monitoring site, before rejecting its use in a background determination.

   b. When reviewing gaseous pollutant data collected by a continuous analyzer, hourly averages occurring within the 90° exclusion sector should be deleted. Sliding 24-hour averages may be recalculated and used provided no more than 6 hours of data were deleted during the 24 hour period. Two hours may be deleted from an 8 hour average (e.g. CO data). No hourly averages may be deleted from a 3-hour average.

      Particulate data collected on a day when the winds were within the exclusion sector for brief periods of time may be used for background. Petrographic (or other suitable analysis) may be used to support inclusion or exclusion of samples collected under these conditions. Such analysis should show no or minimal impact reasonably linked to the applicant's source or another source of interest.

3. Beginning with the 1st high data value, either eliminate it based upon the specific criteria for the pollutant and its averaging time, or accept it as a background concentration. If the 1st value is eliminated, continue down the list until a value can not be eliminated. For some determinations, you may not

have to go beyond the 3rd high and in others, you may have to go beyond the 10th high before a value is determined to be a suitable indicator of background air quality. A value is determined to be the background concentration if it can not be eliminated based on the specific criteria for each pollutant and its corresponding averaging time.

Criteria which should be reviewed are the presence of calm wind conditions and possible impacts due to calm conditions, variable wind conditions (resultant direction may be outside of the exclusion sector whereas a portion of the hour would show directions from within the exclusion sector), and filter analysis results which may indicate significant impact from a source due to variable winds or high velocity winds for only an hour or two during the 24-hour period.

SPECIFIC CRITERIA:

1. For 24-hour Particulate Background Determinations.

   a. Impact By The Applicant -- Impact by the applicant does not necessarily eliminate the sample (concentration). The applicant should not have an impact on the sample if the wind direction during the sampling period was greater than 45° on each side (90° total) of the azimuth from the sampler to the applicant. (See subsection E-2 for how to deal with calms and variables) An examination of meteorological data should be undertaken to verify that the source had no or minimal impact on the sample in question. In certain cases where the applicant has had excessive emissions for a considerable period of time it may be impossible to get a sample without a small percentage of the applicant's emissions. In those instances further documentation will be required. The highest concentrations will require petrographic filter analysis (or another suitable analysis) until sufficient documentation exists to show maximum background concentrations excluding identifiable impacts from the applicant. All those filters which are not rejected for background consideration based on the specific criteria used, and are higher than the selected background concentration shall be analyzed petrographically (or by another suitable analytical procedure) to quantify the source impact on the sample.

      If a review of appropriate meteorological data indicates emissions from the applicant had an impact during the sampling period, refined modeling may be undertaken to quantify the applicant's contribution. That contribution can then be subtracted from the sampled value. A consideration in this procedure is that the applicant's impact will be a conservative value. Consequently, by subtracting this value, the remaining concentration will not be conservative. This impact should

be adjusted based on the percentage of time the wind was impacting from the applicant.

In urban areas where multiple sources may exist, quantifying specific source impacts for the purposes of deducting their possible contribution to a given potential background concentration will be difficult. This should not be attempted without discussing the issue with appropriate Technical Services and Field Services staff. Any number arrived at in this manner may be rejected due to it's "non-conservative" nature.

b. Filter Analysis -- Was petrographic filter analysis (or another suitable analytical procedure) performed on any other filters that ran on the same day(s) as those under consideration as a background sample. (NOTE: Past experience has shown that pollen impacts seem to be a regional influence observed across the State.) Results of the analysis may or may not confirm the selected sample concentration as the background number. If the sample was influenced by pollen, insects and/or biologicals, subtract the percent weight due to these and use the adjusted concentration value to confirm the background concentration. The filter may also be rejected.

c. Temporary Emissions -- If there was an unusual occurrence near the site, eliminate the sample (concentration) from consideration. Temporary emissions could be, but not limited to: street sweeping, uncontrolled or non-routine fire, demolition operations, construction, road work/drilling, filling, roofing operations, or impact on the sampler by a combustion source malfunctioning near the site. Impacts from street sweeping may be accounted for (saving the sample).

d. Mineral Impact -- If the major component on an analyzed filter used for background is mineral, special considerations may be used. Those considerations will be applied on a case by case basis.

e. An Outlier Value -- If a value seems "far and above" any other value obtained at the site or in the area, the Grubbs Test for outliers may be used to eliminate or maintain the value. Before using the Grubbs T Test, a frequency plot should be constructed to determine if the data is normal or log-normal. Most air pollution data is log-normal; although, some data is normal, which would eliminate converting the data into logs before applying the test. (See EPA Handbook Vol. 1 Appendix F)

f. Use the highest 24-hour value of all particulate data collected at all sites used in the analysis that were not rejected by 1) through 5) above.

2. For Annual Particulate Background Determinations

   a. Source Oriented Monitors -- If the site was source oriented, the applicant could have had impact on the site. If the applicant emissions had impact on the data, the following may be done: 1) Recalculate the annual average omitting those days that the applicant had possible impact (impact is based on the 90° exclusion sector). If the number of omitted days is small, data may be used as is and not adjusted; 2) Use data from all sites used in the analysis, when the applicant had no impact (based on exclusion sector) to calculate a sliding 12 month average. The sliding 12 month arithmetic average is used for conservatism and will be considered the annual average.

   b. Non-Source Oriented Monitors -- Use the highest sliding 12 month average of all the sites used in the background data analysis as the annual background concentration. A meteorological wind rose (annual) should be constructed to verify that the applicant had minimal or no impact.

3. For 3-Hour Gaseous Pollutant Background Determinations

   a. Impact From Applicant -- If the applicant had possible impact (based on a exclusion sector) during the 3-hour period, the data should be rejected. If the applicant had no or minimal impact, the value should not be eliminated. Meteorological data used to determine impact should have approval from a Bureau staff meteorologist.

   b. Unusual occurrences affecting the pollutant concentration during the 3-hour period should eliminate any affected that 3-hour value.

   c. Continuous Data -- Only continuous data can be used to obtain the sliding 3-hour background concentration.

   d. An Outlier Value -- If a value seems "far and above" any other value obtained at the site or in the area, the Grubbs Test for outliers may be used to eliminate or maintain the value. Before using the Grubbs T Test, a frequency plot should be constructed to determine if the data is normal or log-normal. Most air pollution data is log-normal; although, some data is normal, which would eliminate converting the data into logs before applying the test. (See EPA Handbook Vol. 1 Appendix F)

   e. Use the highest sliding 3-hour value of all the values collected at all the sites used in the analysis not rejected by (a) through (d) above.

4. For 24-Hour Gaseous Pollutant Background Determinations.

a. Impact From Applicant -- If the applicant had possible impact (based on a 90° sector) during the 3-hour period, the data should be rejected. If the applicant had no or minimal impact, the value need not be eliminated. Meteorological data used to determine impact should have approval from a Bureau staff meteorologist.

b. Unusual occurrences affecting the pollutant concentration during the 3-hour period should eliminate that 3-hour value (i.e. 3-hour average will have to "slide away" from the data affected by the unusual occurrence).

c. Continuous Data -- Only continuous data can be used to obtain 24-hour background concentration.

d. An Outlier Value -- If a value seems "far and above" any other value obtained at the site or in the area, the Grubbs Test for outliers may be used to eliminate or maintain the value. Before using the Grubbs T Test, a frequency plot should be constructed to determine if the data is normal or log-normal. Most air pollution data is log-normal; although, some data is normal, which would eliminate converting the data into logs before applying the test. (See EPA Handbook Vol. 1 Appendix F)

e. Use the highest sliding 24-hour value of all the values collected at all the sites used in the analysis not rejected by (a) through (d).

5. For Annual Gaseous Pollutant Background Determinations

a. Continuous Data -- Only continuous data can be used to obtain annual background concentration.

b. Source Oriented Monitors -- If the site was source oriented, the applicant could have had impact on the site. If the applicant emissions had impact on the data, the following may be done: 1) Recalculate the annual average omitting those days that the applicant had possible impact (impact is based on the 90° exclusion sector). If the number of omitted days is small, data may be used as is and not adjusted; 2) Use data from all sites used in the analysis, when the applicant had no impact (based on 90° sector) to calculate a sliding 12 month average. The sliding 12 month average is used for conservatism and will be considered the annual average.

c. Use the highest annual value of all the values collected at all of the sites used in the analysis as the annual background concentration.

6. GUIDELINES FOR BACKGROUND CONCENTRATIONS WHEN NO MONITORING DATA EXISTS FOR THE APPLICANT'S AREA

When no monitoring data exists, and the applicant is locating in a rural or rural/urban area and there are no other significant sources in the affected area, then the background concentrations may be determined by:

1. Using monitoring data from other areas similar in physical, geographical and population characteristics, and number and type of existing air emission sources, or

2. Using values collected from regional background air quality monitoring sites, or

3. With respect to PM_{10}, TSP annual geometric means may be substituted for PM_{10} annual arithmetic average.

If no data is available that meets one of the above criteria or if the applicant is locating in an urban or industrialized area with other emission sources, then monitoring will be required.

7. GUIDELINES FOR DOCUMENTING AND REPORTING BACKGROUND AIR QUALITY DETERMINATIONS

A written report will be provided by the person responsible for doing the background determinations. Follow the appropriate guidelines for either a qualitative or quantitative determination.

A. QUALITATIVE DETERMINATIONS

A single written report of the characterization of the background air quality, which should include any necessary supporting documentation, shall be provided to the Division Director of the Field Services and Licensing and Enforcement Divisions. This may be in the form of a memorandum.

The Division Directors will review the report and either accept it for release to the project's licensing coordinator or return it to the applicant or Division staff person for revision. If approved, the licensing coordinator may in turn release the background information to the applicant if appropriate. Under no circumstances should any information pertaining to the background determination be released to anyone at any time other than the above Division Directors until approved.

B. QUANTITATIVE DETERMINATION

After determining that data is representative and of acceptable quality, a tabulation of the sites and data shall be done. Such a table should list the site and type of data

(continuous SO_2, etc.), the agency responsible for data collection, and the time period used in the analysis. This table may be important if an update becomes necessary.

Tabulate data, according to concentration (highest to lowest), concentration, date (starting and ending date/hour, etc.) site and comments as to why a particular value was accepted or rejected as the background concentration candidate.

A single written report describing how the background concentrations were determined and what those concentrations are, along with the supporting documentation such as the above table(s), memos, etc. shall be provided to the Division Directors of Field Services and Licensing and Enforcement. This may be in the form of a memorandum.

For some projects, it will be necessary to produce a more comprehensive report which contains documentation of all phases of the background determination. Such a report should include:

1. a section on the request and information provided for a background determination along with a map of the project's location and impact areas;

2. a statement on the availability of data;

3. a section on the representativeness of the data used;

4. a section on the quality of the data used;

5. a section on the representativeness and quality of the met data used;

6. the report on the background concentrations;

7. appropriate appendices (e.g. graphs, tables, etc.). The Division Directors will review the report and either accept it for release to the project's licensing coordinator or return it to the applicant for revision. If accepted, the licensing coordinator may in turn release the background information to the applicant if appropriate. Under no circumstances should any information pertaining to the background determination be released to anyone at any time other than the above Division Directors until approved.

8. GUIDELINES FOR REFINED MODELING ANALYSES AND QUANTITATIVE BACKGROUND AIR QUALITY DETERMINATIONS

To show MAAQS are met, a source's modeled impacts are added to a background concentration. A conservative method is used for the initial step. If that analysis

shows potential violations of MAAQS, more refined modeling may be used to demonstrate that the applicant's emissions will not result in an exceedance of the MAAQS.

A. METHOD 1

Modeling can be used to determine the approximate distance from the source where its maximum impact would occur. A background concentration may be determined for this area and another area (e.g. where the 2nd highest impact would occur). This analysis may demonstrate compliance with the MAAQS; if not, use method 2.

B. METHOD 2

Further refine the modeling analysis (check with DEP meteorologist before using these methods).

1. For downwashing sources, "direction specific" building dimensions may be used.

2. Compare the data and meteorological conditions for the background concentration with the meteorological conditions leading to the modeled maximum impact from the applicant. If they are different, refine the model to compute the applicant's impact under the same meteorological conditions for the background concentration determination.

C. METHOD 3

Refine the background analysis (check with Field Services Division before using these methods).

1. Determine different background concentrations for different areas, (i.e. rural and urban values).

2. Iterpolate the background concentrations between two monitor locations or different background areas.

D. METHOD 4

"Roll Back" (offset) emissions from the applicant's facility, or from some other source (e.g.: street sweeping for particulate control, additional controls on existing emissions points). Such reductions may offset the added emissions from the applicant and attain compliance with the MAAQS.

Appendix 1

Section No. F
Rev. No. 0
May 1, 1975

APPENDIX F

OUTLIERS

F.1 INTRODUCTION

An unusually large (or small) value or measurement in a set of observations is usually referred to as an outlier in the statistical literature. This terminology also has been adopted in the quality control and engineering literature. Some of the reasons for an outlier in data are:

- faulty instrument or component part
- inaccurate reading of record, dial, etc.
- error in transcribing data
- calculation errors
- actual value due to unique circumstances under which the observation(s) was obtained – an extreme manifestation of the random variability inherent in the data.

It is desired to have some statistical procedure to test the presence of an outlier in a set of measurements. The purpose of such tests would be to:

1. Screen data for outliers and hence to identify the need for closer control of the data generating process.

2. Eliminate outliers prior to analysis of the data. For example, in developing control charts the presence of outliers would lead to limits which are too wide and would make the use of the control charts of minimal, if any, value. In most statistical analysis of data, e.g. regression analysis and analysis of variance, the presence of outliers violates a basic assumption of the analysis. Incorrect conclusions are likely to result if the outliers are not eliminated prior to analysis. Outliers should be reported, and their omission from analysis should be noted.

3. Identify the real outliers due to unusual conditions of measurement, e.g. a TSP concentration which is abnormally large due to local environmental conditions during the time of sample collection. Such observations would not be indicative of the average concentration of TSP, and may be eliminated depending on the use of the data. Ideally, these unusual conditions should be recorded on the field data report. Failure to report complete information and unusual circumstances surrounding the collection and analysis of the sample often can be detected by outlier tests.

It will be assumed in the discussion of the appendix that the measurements are normally distributed and that the sample of size n measurements is being studied for the possibility of one or two outliers. If the measurements are lognormally distributed, such as for concentration of TSP, then the logarithm of the data should be taken prior to application of the tests given herein.

F.2 PROCEDURE(S) FOR IDENTIFYING OUTLIERS

Let the set of n measurements be arranged in ascending order and denoted by

$$X_1, X_2, .., X_n$$

where X_i denotes the ith smallest measurement. Suppose that X_n is suspected of being too large, and that a statistical test is to be applied to the particular measurement to determine whether X_n is consistent with the remaining data in the sense that it is reasonable that it is part of the same population of measurements from which the sample is taken. Consider the following data comparing the results of measurements on NO_3 analysis of 10 high-volume filters by two different laboratories. The differences are given and the largest value 0.2 is questionable and is to be tested as a possible outlier.

Example F.1

NO_3 Analysis (mg/filter)

Lab 1	**Lab 2**	**Difference**
1.7	2.05	-0.35
2.2	2.4	-0.2
3.9	3.7	0.2
3.3	3.6	-0.3
2.7	3.3	-0.6
3.5	3.8	-0.3
0.9	1.46	-0.56
1.3	1.5	-0.2
6.1	6.4	-0.3
2.9	3.2	-0.3
		-2.91

One test procedure is to compare the value 0.2 with the remaining data through use of a test by Dixon.[1]

$$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_2} = \frac{0.2 - (-0.2)}{0.2 - (-0.56)} = \frac{0.4}{0.76} = 0.53$$

This value is compared to tabulated values given by Dixon[1], and it is indicated that the value r_{11} is significant at the 0.05 level, i.e. it is larger than 0.477. See Table F1 for criteria for testing for outliers. Thus, one infers that the measured difference is subject to question, and that either Lab 1 obtained an unusually large result, or Lab 2 obtained a small value. In any case, some investigation of the measurements should be considered, particularly if there are recurring results of this nature which may indicate poor quality data.

Another commonly used test procedure[2] requires additional computation and is given by

$$T_n = \frac{(X_{n-} \overline{X})}{s}$$

where:

$\overline{X}$ is the sample mean

s is the sample standard deviation, i.e.

$$\overline{s} = \{\Sigma(x-X)^2 / (n-1)\}^{½}$$

For the data set previously given,

$\overline{X}$ = -0.291

s = 0.22

and hence:

$$T_n = \frac{0.2 - (-0.291)}{0.22} = \frac{0.491}{0.22} = 2.23$$

which is significant at the 0.05 level, i.e. larger than 2.176 which is the tabulated value for this level from Table F.2. This test result is in agreement with the previous one, indicating that this measured difference is larger than expected in sampling from a population which is normally distributed. To aid in computing T_n (or T_1 for testing the smallest value), Table F.3 may be used. In many examples

it will be obvious that a particular value is an outlier, whereas in Example F.1 this is not the case. A plot of the data on a straight line, x-axis is often helpful in examining a set of data.

After rejecting one outlier using either T_n or T_1, the analyst may be faced with the problem of considering a second outlier. In this case, the mean and standard deviation may be recomputed and either T_n or T_1 applied to the sample of n-1 measurements. The user should be aware that the test T_n or T_1 is not theoretically based on repeated use.

Grubbs[2] also gives a test procedure for simultaneously testing the two largest or two smallest values. This procedure is not given here. Grubbs also gives tables for the critical values for this procedure.

The use of the procedures given in Table F.1 requires very little computation and would be preferable on a routine basis. Grubs[3] gives a tutorial discussion of outliers and is a very good reference to the subject. Another reference[4] on outliers contains Chauvenet's criterion and the appropriate table of critical values. This test is similar to the T_n or T_1 test given here. Chauvenet's criterion for n larger than five rejects outliers more frequently than using the T_n or T_1test procedure, i.e. it would be less conservative for these larger sample sizes.

The determination of outliers when, for example, several laboratories measure two or more samples in an interlab study, is made by more complex methods.[5,6]

Table F.1 DIXON CRITERIA FOR TESTING OF EXTREME OBSERVATION (SINGLE SAMPLE)*

n	Criterion		10%	5%	1%
3	$r_{10} = \frac{X_2 - X_1}{X_n - X_1}$	if smallest value is suspected	.886	.941	.988
4			.679	,765	.889
5			.557	.642	.780
6	$r_{10} = \frac{X_n - X_{n-1}}{X_n - X_1}$	if largest value is suspected	.482	.560	.698
7			.434	.507	.736
8	$r_{11} = \frac{X_2 - X_1}{X_{n-1} - X_1}$	if smallest value is suspected	.479	.554	.683
9			.441	.512	.635
10			.409	.477	.597
	$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_2}$	if largest value is suspected			
11	$r_{21} = \frac{X_3 - X_1}{X_{n-1} - X_1}$	if smallest value is suspected	.517	.576	.679
12			.490	.546	.642
13			.467	.521	.615
	$r_{21} = \frac{X_n - X_{n-2}}{X_n - X_2}$	if largest value is suspected			
14	$r_{22} = \frac{X_3 - X_1}{X_{n-2} - X_1}$	if smallest value is suspected	.492	.546	.641
15			.472	.525	.616
16			.454	.507	.595
17			.438	.490	.577
18			.424	.475	.561
19	$r_{22} = \frac{X_n - X_{n-3}}{X_n - X_3}$	if largest value is suspected	.412	.462	.547
20			.401	.450	.535
21			.391	.440	.524
22			.382	.430	.514
23			.374	.421	.505
24			.367	.413	.497
25			.360	.406	.489

** Reproduced with permission from W.J. Dixon, "Processing Data for Outliers", Biometrics, March 1953, Vol. 9, No. 1, Appendix, Page 89. (Reference 11)*

$X_1 \le X_2 \le X_3 \le \cdots \le X_{n-2} \le X_{n-1} \le X_n$

Criterion r_{10} applies for $3 \le n \le 7$

Criterion r_{11} applies for $8 \le n \le 10$

Criterion r_{21} applies for $11 \le n \le 13$

Criterion r_{22} applies for $14 \le n \le 25$

Table F.2 Table of Critical Values for T (one-sided test of T1 or Tn) When the Standard Deviation is Calculated From the Same Sample

Number of Observations	Upper .1% Significance Level	Upper .5% Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
3	1.188	1.188	1.188	1.188	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.715	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.274	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.705	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.867	2.699	2.607	2.462	2.331	2.17S
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2 997	2.806	2.705	2.549	2.409	2.247
16	3 052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.62	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.968	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	Z.730	2.549
30	3.507	3.236	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.546	3.270	3.135	2.938	2.773	2.591
33	3.565	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.343	3.204	3.003	2.835	2.650
38	3.646	3.356	3.216	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.866	2.682
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
43	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768

Use $T_1 = \frac{\bar{X} - X_1}{s}$ when testing the smallest value, X_1. Use $Tn = \frac{X_n - \bar{X}}{s}$ when testing the largest value, X_n in a sample.

Table F2 (continued)

Number of Observations	Upper .1% Significance Level	Upper .5% Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
51	3.798	3.491	3.345	3.136	2.964	2.775
52	3.808	3.500	3.353	3.143	2.971	2.783
53	3.816	3.507	3.361	3.151	2.978	2.790
54	3.825	3,516	3.368	3.158	2.986	2.798
55	3.834	3.524	3.376	3.166	2.992	2.804
56	3.842	3.531	3.383	3,172	3.000	2.811
57	3.851	3.539	3.391	3.180	3.006	2.818
58	3.858	3.546	3.397	3.186	3.013	2.824
59	3.867	3.553	3.405	3.193	3.019	2.831
60	3.874	3.560	3.411	3.199	3.025	2.837
61	3.882	3.566	3.418	3.205	3.032	2.842
62	3.889	3.573	3.424	3.212	3.037	2.849
63	3.896	3.579	3.43	3.218	3.044	2.854
64	3.903	3.586	3.437	3.224	3.049	2.860
65	3.910	3.592	3.442	3.230	3.055	2.866
66	3.917	3.598	3.449	3.235	3.061	2.871
67	3.923	3.605	3.454	3.241	3.066	2.877
68	3.930	3.610	3.460	3.246	3.071	2.883
69	3.936	3.617	3.466	3.252	3.076	2.888
70	3.942	3.622	3.471	3.257	3.082	2.893
71	3.948	3.627	3.476	3.262	3.087	2.897
72	3.954	3.633	3.482	3.267	3.092	2.903
73	3.960	3.638	3.487	3.272	3.098	2.908
74	3.965	3.643	3.492	3.278	3.102	2.912
75	3.971	3.648	3.496	3.282	3.107	2.917
76	3.977	3.654	3.502	3.287	3.111	2.922
77	3.982	3.658	3.507	3.291	3.117	2.927
78	3.987	3.663	3.511	3.297	3.121	2.931
79	3.992	3.669	3.516	3.301	3.125	2.935
80	3.998	3.673	3.521	3.305	3.130	2.940
81	4.002	3.677	3.525	3.309	3.134	2.945
82	4.007	3.682	3.529	3.315	3.139	2.949
83	4.012	3.687	3.534	3.319	3.143	2.953
84	4.017	3.691	3.539	3.323	3.147	2.957
85	4.021	3.695	3.543	3.327	3.151	2.961
86	4.026	3.699	3.547	3.331	3.155	2.966
87	4.031	3.704	3.551	3.335	3.160	2.970
88	4.035	3.708	3.555	3.339	3.163	2.973
89	4.039	3.7i2	3.559	3.343	3.167	2.977
90	4.044	3.716	3.563	3.347	3.171	2.981
91	4.049	3.720	3.567	3.350	3.174	2.984
92	4.053	3.725	3.570	3.355	3.179	2.989
93	4.057	3.728	3.575	3.358	3.182	2.993
94	4.060	3.732	3.579	3,362	3.186	2.996
95	4.064	3.736	3.582	3.365	3.189	3.000
96	4.069	3.739	3.586	3.369	3.193	3.003
97	4.073	3.744	3.589	3.372	3.196	3.006
98	4.076	3.747	3.593	3.377	3.201	3.011
99	4.080	3.750	3.597	3.380	3.204	3.014
100	4.084	3.754	3.600	3.383	3.207	3.017

(Source: Grubbs, F.E., and Beck, G., Extensions of Sample Sizes and Percentage Points for Significant Tests of Outlying Observations, Technometrics, Vol. 14, No. 4, Nov. 1972, pp. 847-854.

Table F.3 CHECK FOR OUTLIERS USING T_1 OR T_n (SINGLE SAMPLE)

Sample No.	Value X	Rank *
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Sample No.	Value X	Rank *
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

TOTAL = ΣX = _____

Calculations:

Mean = $\bar{X} = \frac{\sum X}{n} =$ _____

Variance = $s^2 = \frac{\Sigma(X - \bar{X})}{(n-1)} = \frac{[\Sigma X^2 - \frac{(\Sigma X)^2}{n}]}{(n-1)}$ = _____

ΣX^2 = _____

$(\Sigma X)^2/n$ = _____

Difference: n – 1 = _____

Divide: s^2 = _____

Standard Deviation = s = $\sqrt{s^2}$ = _____

If Largest Value is Suspect: calculate $T_n = \frac{X_n - \bar{X}}{s} =$ _____

If Smallest Value is Suspect: calculate $T_1 = \frac{\bar{X} - X_1}{s} =$ _____

Compare computed value with tabulated value in Table F.2 for acceptable risk. If T_n> tabulated critical value, the outlier is suspect.

** Rank values from smallest to largest value, e.g. 1 to 20. Actually need to rank 3 smallest and 3 largest values.*

F.3 REFERENCES

1. Dixon, W.J., Processing Data for Outliers, Biometrics, Vol. 9, No. 1, March 1953, pp. 74-89.

2. Grubbs, F.E. and Beck, G., Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations, Technometrics, Vol. 14, No. 4, Nov. 1972, pp. 847-854.

3. Grubbs, F.E., Procedures for Detecting Outlying Observations in Samples, Technometrics, Vol. 11, No. 1, Feb. 1969, pp. 1-21.

4. Guide to Statistical Problem Solving, prepared for U.S. EPA Air Pollution Training Institute.

5. Anscombe, F.J. and Tukey, J.W. The Examination and Analysis of Residuals, Technometrics, Vol. 5, No. 2, May 1963.

6. Stefansky, W., Rejecting Outliers in Factoral Design, Technometrics, Vol. 14, 1972, pp. 469-480.

DDguidel-mgb