National Environment Protection (Ambient Air Quality) Measure

Technical Paper No. 5

Data Collection and Handling

Prepared by the
Peer Review Committee

May 2001

PREAMBLE

The National Environment Protection Measure (NEPM) for Ambient Air Quality was made in June 1998 with the desired environmental outcome of "ambient air quality that allows for the adequate protection of human health and well-being" across Australia. The NEPM sets national standards against which ambient air quality can be assessed. The NEPM includes a monitoring protocol to determine whether these standards are being met. Each jurisdiction is required to submit to the National Environment Protection Council (NEPC) a monitoring plan consistent with the protocol.

The Peer Review Committee (PRC) was established to assist NEPC in its task of assessing and reporting on the implementation and effectiveness of the NEPM by participating jurisdictions. The PRC includes government experts from all participating jurisdictions, in addition to representatives from industry and community groups. A significant activity of the PRC is the provision of advice to NEPC on the adequacy of jurisdictional monitoring arrangements, to ensure as far as possible that a nationally consistent data set is obtained.

To assure the consistency and transparency of its advisory function, the PRC has developed a set of guidance papers that clarify a number of technical issues in interpretation of the NEPM protocol. These Technical Papers provide the basis for PRC assessment of jurisdictional plans, aimed at assuring the quality and national consistency of NEPM monitoring.

The PRC Technical Papers are advisory for jurisdictions, and they will evolve with time as the science of air quality monitoring and assessment develops and as practical experience with monitoring increases.

M J Manton
Chair
Peer Review Committee

1. PURPOSE

The purpose of this technical paper is to provide guidance and procedures for uniform data recording and handling which ensures nationally uniform ambient air quality data and a nationally consistent implementation of the Ambient Air Quality – National Environment Protection Measure (AAQ NEPM) in Australia.

2. INTRODUCTION

The AAQ NEPM in Clause 17 (Evaluation of performance against standards and goal) specifies that each jurisdiction must evaluate its annual performance as set out in this clause and that for each performance monitoring station in the jurisdiction there must be an evaluation of performance against the standards and goal of the AAQ NEPM. Clause 18 (Reporting) requires that each jurisdiction must submit a report on its compliance with the AAQ NEPM in an approved form to the National Environment Protection Council (NEPC) by the 30 June next following each reporting (calendar) year. The evaluations of performance against the standards and goal are to be included in this annual report. The form of the report and details of the reporting requirements are covered in National Environment Protection (Ambient Air Quality) Measure Technical Paper No. 8, "Annual Reports for AAQ NEPM".

Schedule 1 of the AAQ NEPM specifies the pollutants to be monitored as carbon monoxide (CO), nitrogen dioxide (NO_2), photochemical oxidants (as O_3), sulfur dioxide (SO_2), lead and particles (as PM_{10}). The standards and goal are set out in Schedule 2 and the methods for monitoring are listed in Schedule 3. The averaging periods of the set standards and the methods for monitoring are important factors in developing procedures for data collection and handling.

This technical paper describes the data collection procedures, data handling conventions and computations necessary for determining the extent to which the AAQ NEPM standards are, or are not, met at the performance monitoring stations. Procedures to be used in compliance assessment while making comparisons between reported concentrations and the levels of the AAQ NEPM standards are also outlined. Compliance assessment procedures are specified in greater detail in Technical Paper No. 8.

The AAQ NEPM does not explicitly require meteorological measurements to be undertaken. However, the Peer Review Committee (PRC) recommends that meteorological monitoring be included in the AAQ NEPM monitoring.

Recommendations on meteorological measurements and data handling requirements are covered in National Environment Protection (Ambient Air Quality) Measure Technical Paper No. 6, "Meteorological Measurements for AAQ NEPM".

In view of the importance of uniform data recording and handling for comparability of data, the PRC established a consultancy and a Data Working Group to develop the procedures described in this technical paper. The consultancy which provided the main input to the Data Working Group was carried out by Dr Graeme Lorimer. His report, "Air Quality Data Handling and Reporting," was made to NEPC Service Corporation. This technical paper presents the decisions made by the Data Working Group and the PRC.

It is recognised that jurisdictions undertake a range of quality assurance activities for their data operations (instrument calibration, data validation and statistical processing of raw data). The resulting data sets can be very extensive. The PRC recommends that each jurisdiction, at the

time of submitting its annual report to the NEPC, has available an AAQ NEPM data set extracted from its comprehensive database. The AAQ NEPM data set should contain all air quality information (air quality data and descriptive information about monitoring stations) needed for the generation of the statistics required for AAQ NEPM annual reporting. The AAQ NEPM data sets must meet common standards set out in this technical paper.

This technical paper provides information that will enable jurisdictions to provide AAQ NEPM data in standard formats for national air quality assessments. This technical paper makes no assumptions as to whether AAQ NEPM data sets will be used to construct a national database.

3. DATA RECORDING

3.1 Sampling rates and logging intervals

CO, NO_2, O_3 and SO_2

Methods for monitoring CO, NO_2, O_3 and SO_2 are specified in Schedule 3 of the AAQ NEPM. These pollutants are measured continuously. Schedule 2 of the AAQ NEPM includes standards with averaging periods less than a day for all of these pollutants. Sampling rates of instruments for continuous monitoring are of the order of seconds. Typically, averages of 1 to 10 second samples are stored by a data logger at intervals from 1 to 60 minutes. Logging intervals of 2, 5, 10, 30 and 60 (1-hour) minutes are either used or being considered by jurisdictions.

- As a minimum requirement, the AAQ NEPM data set should consist of validated 1-hour (clock hour) averages.
- In addition, a validated 10-minute data set is desirable, for assuring quality and for future NEPM development. Shorter averaging times (1, 2 or 5 minutes) which allow construction of 10-minute averages are acceptable alternatives.

Particles (as PM_{10}) and Lead

Schedule 3 of the AAQ NEPM specifies gravimetric methods of measurement for these pollutants. The averaging period for the PM_{10} Standard is 1 day and the averaging period for the lead Standard is 1 year. Particles are measured as PM_{10} (particles with an equivalent aerodynamic diameter of 10 micrometres or less). Daily samples are required for both PM_{10} and lead; every day for PM_{10} and every sixth day for lead.

- As a minimum requirement, the AAQ NEPM data set should consist of validated 24-hour (calendar day) averages, for every day for PM_{10} and for every sixth day for lead.
- When a Tapered Element Oscillating Microbalance (TEOM) is used for measuring PM_{10}, the AAQ NEPM data set may contain hourly averages in addition to 24-hour averages.

3.2 Measurement scales for concentrations

All concentrations are to be stored so as to allow reporting in the same units as specified in the AAQ NEPM, and to the precision of the instrument. Instrument precisions for methods specified in Schedule 3 of the AAQ NEPM which are in current use are listed below.

Pollutant	Data format	Nominal precision
CO:	X.X ppm	tenth of a ppm
NO_2:	X.XXX ppm	thousandth of a ppm
O_3:	X.XXX ppm	thousandth of a ppm
SO_2:	X.XXX ppm	thousandth of a ppm
PM_{10}:	XX.X µg/m^3	tenth of a µg/m^3
Lead:	X.XX µg/m^3	hundredth of a µg/m^3

Schedule 3 of the AAQ NEPM specifies that PM_{10} measurements should be carried out using high volume samplers (Hi-Vol) with size selective inlets or dichotomous samplers. However, Schedule 2 of the AAQ NEPM also requires daily samples to be collected and analysed. Because of the labour-intensive nature of operation of Hi-Vol samplers and the reporting advantages of obtaining continuous measurements, TEOM samplers have almost universally been adopted by jurisdictions for measuring PM_{10}.

The PRC has promoted processes that are leading to a demonstration that the TEOM can provide equivalent information to the standard method (Hi-Vol) for NEPM purposes. Procedures for the handling of TEOM PM_{10} data are described in National Environment Protection (Ambient Air Quality) Measure Technical Paper No. 10, "Collection and Reporting of TEOM PM_{10} Data."

3.3 Times and dates

All times should be quoted according to local standard time (LST). No daylight saving times (DST) are to be used.

Dates and times should be reported in the following fixed text format:
yyyy/mm/dd hh:mm (or optionally hh:mm:ss)

Schedule 2 of the AAQ NEPM specifies that all concentrations should be referenced by the time at the end of the averaging period. Schedule 2 also specifies that the end time of the averaging period determines the calendar day to which the averaging periods are assigned.

A 24-hour clock time should be used for recording time. Midnight is to be referred to as 24:00 of the day just ended (Hourly averages from 1 to 24 in a day).

3.4 Dealing with data near detection limits

In some situations the concentration of the pollutant being measured may be very close to zero, in which case the measured value (after adjusting for drift of zero and span and any other corrections) may be less than the measurement limit of detection. Various practices have been adopted by jurisdictions in dealing with the measurements near detection limit.

The practice adopted for AAQ NEPM purposes requires that the actual measured values (positive or negative) whether or not below the detection limit are reported. In particular, the common practice of removing small negative measurements is not allowed as it would lead to bias. Large negative values are expected to be invalidated as part of the validation process.

3.5 Invalid and missing data

There are many reasons why the validation process may exclude a datum from further analysis for NEPM purposes; e.g., it may be missing, a calibration point, produced by a faulty instrument and so on. The organisations responsible for processing NEPM monitoring data use various methods for indicating invalid or questionable data, and most of these organisations wish to continue their present practices.

However, it is important for each organisation to provide the AAQ NEPM data in a nationally consistent form for archival and analysis. It is therefore expected that in many cases,

organisations will continue to store their data in the form that best suits their individual needs while they produce the AAQ NEPM data in the recommended format.

In the national standard files for AAQ NEPM monitoring data, each datum that is to be excluded (invalid or missing) should be replaced with "a number less than or equal to –99".

It is noted that the quality assurance procedures for NATA accreditation normally require that any invalidated datum should be traceable to its source, including the reason for invalidating it. This can only be achieved by reference to the original data sets of jurisdictions.

3.6 Rounding and significant digits

In the computations and processing of air quality data whenever rounding occurs, values equal to or greater than 5 should be rounded up.

Air quality data should be stored and reported to the precision of the instrument, as shown in section 3.2. For example, for O_3, NO_2 and SO_2, pollutant concentrations should be stored and reported in ppm to the third decimal place.

When concentrations are compared against the NEPM standards, an exceedence occurs once the concentration reported at instrument precision exceeds the standard. For example, an hourly ozone concentration of 0.101 ppm exceeds the 1-hour standard of 0.10 ppm, therefore 0.101 ppm is an exceedence.

3.7 Metadata

It is important to record the following information for each monitoring site:

- Map coordinates (AMG, latitude/longitude or other), precisely enough for the site to be readily located;
- Site elevation;
- Street address (if applicable);
- Date site established;
- Date site terminated (if applicable);
- List of exceptions to AS 2922-1987 siting guidelines;
- Description of surrounding land use;
- Description of nearby emission sources;
- Names of pollutants monitored;
- Units of concentration measurements;
- Instrument type, make, model, serial number, minimum detection level;
- For open-path instruments, the length and orientation of the path;
- Changes to instruments or conditions;
- Changes to data treatment, e.g., timing and details of adjustments applied to TEOM PM_{10} data;
- Format version (to allow for future upgrading of the format of the data file);
- Other notes.

4. QUALITY CONTROL

4.1 NATA accreditation

Clause 12 of the AAQ NEPM requires monitoring to be accredited by the National Association of Testing Authorities (NATA) or an equivalent system. The criteria for accreditation are

specified in the National Environment Protection (Ambient Air Quality) Measure Technical Paper No.7, "Accreditation of Performance Monitoring". The technical paper concludes that NATA is the only body which currently meets the essential criteria and recommends that NATA be the single accrediting body at this time.

4.2 Calibration control tolerance

Calibrations should be scheduled and performed to NATA requirements and traceability of measurements to national and appropriate international standards should be ensured.

Calibration control refers to the tracking of variability in an instrument's zero and span readings, as determined from daily or periodic calibrations. A fault is deemed to have occurred when the zero and span readings change from one calibration to the next by an amount that exceeds a specified tolerance. Some of the Australian Standards given in Schedule 3 of the AAQ NEPM specify such tolerances. Tolerances are sometimes also specified in instrument manuals or US guidelines.

The acceptable limits for drift of zero and span readings should be based on Australian Standards wherever possible.

As in the US practice when zero or span drift validation limits are exceeded, ambient measurements should be invalidated back to the most recent point in time where such measurements are known to be valid. Usually this point is the previous calibration (or accuracy audit), unless some other point in time can be identified and related to the probable cause of the excessive drift (such as a power failure or malfunction). Also, data following an analyser malfunction or period of non-operation should be regarded as invalid until the next calibration unless unadjusted zero and span readings at that calibration can support the validity of data.

Provided that the drift of zero and span are within the acceptable range, compensation can be made by linearly interpolating the zero and span values between calibrations and adjusting the measurements accordingly. However, this is not essential if the zero and span readings change by less than the precision of the calibration process.

Jurisdictions appear to have different tolerance limits and different ways of dealing with calibration when outside or within tolerance limits. It is desirable to move towards a consistent approach.

4.3 Data validation

Data validation is a process of checking data by automatic or manual means to verify or to detect inaccuracies. Data are flagged by juridictions to indicate validation status. During the validation process data values may be corrected for known sources of error, e.g., instrument calibration. Data validation procedures vary from jurisdiction to jurisdiction.

The important outcome of the validation process is a set of data, for gaseous pollutants, averaged over intervals of one hour or less (intervals less than an hour are optional), with a uniformly high level of quality assurance. For PM_{10} and lead, the averaging interval is one-day, with one-hour intervals being optional for TEOM PM_{10} data. Only validated data should be used for AAQ NEPM purposes.

The recommended procedure for air quality data validation is listed below. Not all steps apply to all types of data.

- Where necessary, convert instrument electrical output to units of concentration using calibration results;
- Use linear interpolation to correct for observed clock error, zero drift or span drift, when within accepted limits;
- Adjust measurements to STP in dry air (zero Celsius and an absolute pressure of 101.325 kilopascals) where applicable;
- Check for power failures (e.g., by inspecting pen chart traces for readings of zero voltage or current, and the instrument event logger);
- Check for uncharacteristic pen chart traces due to noise, spikes, non-varying data, or influences by other equipment (e.g., air conditioner);
- Check for uncharacteristic shapes of calibration curves;
- Create and examine plots of zero and span outputs from calibrations as a function of time, and compare these against the expected range of variability;
- Invalidate data associated with any calibration run which produces readings of zero or span outside the specified tolerance;
- Check for data outside the plausible range;
- Optionally conduct simple statistical tests on each month's data;
- Check for data that are inconsistent with measurements of other pollutants at the same site, for example NO and O_3 cannot coexist at high concentrations, ozone formation is a function of temperature and solar radiation, and NO_2 and NO levels should never exceed NO_x levels;
- Check measurements for consistency with other monitoring stations in the region;
- Assign validity flags to each datum to indicate its status (e.g., good, questionable, etc.);
- Record all validation steps for quality control, in accordance with the organisation's quality assurance plan.

4.4 Checks and statistical tests

Plausibility checks
Data validation staff must be familiar with the fundamentals of pollutant behaviour and instrument operation and performance, and check the data closely for errant behaviour. Even though site-specific checks can be developed by jurisdictions, some standardisation is desirable.

Checks for outliers and exceptional events
Exceptionally high values would naturally receive individual attention in data validation even if their durations are much less than the relevant averaging times specified in the AAQ NEPM. Exceptional events that exceed the AAQ NEPM Standards would normally be explained in the annual NEPM report. All valid measurements, in particular outliers and exceptional events should stay in the database.

Statistical tests
It is desirable to incorporate some statistical tests into the national software for processing hourly NEPM air quality data. The appropriateness of the statistical tests used in the US could be investigated.

The data that fail the tests should be investigated, but not necessarily invalidated.

4.5 Data availability

Clause 18 (5) of the AAQ NEPM specifies that the annual report for a pollutant must include the percentage of data available in the reporting period.

An average concentration can be valid only if it is based on at least 75% of the expected samples in the averaging period. This rule applies to all averaging periods, from the hourly concentrations that make up basic air quality data to annual averages.

It is essential that data loss is kept to an absolute minimum. For representative monitoring data and for credible compliance assessment it is desirable to have data capture rates higher than 95%. 75% data availability is specified as an absolute minimum requirement for data completeness.

The annual data availability should be based on the number of valid hourly (or daily for PM_{10} and lead) data. For lead, data availability should be based on the number of scheduled sampling days, allowing for monitoring on one in six days. In addition, the annual percentage of valid 4-hourly, 8-hourly and 24-hourly data, for pollutants with standards for these averaging times, may be provided in reports where relevant.

Compliance is assessed using valid averages. There is a USEPA rule which allows for exceedences based on less than 75% data availability. According to this rule, in the event of less than 75% data availability, if replacing the missing values with zeroes produces a concentration above the standard, a valid exceedence of the standard is reported. The PRC has considered this practice but decided against adopting it. Such a rule would lead to inconsistencies in reporting. Furthermore, in view of the requirement for high data availability from performance monitoring stations, such occurrences are expected to be rare.

For demonstrating annual compliance, in addition to the 75% annual data availability requirement, at least 75% data per calendar quarter are required to cover possible seasonal effects in pollutant behaviour.

Years with less than 75% data availability can still demonstrate non-compliance if sufficient exceedences of the standard are reported. For example, it can be stated that "non-compliance is demonstrated" at a performance monitoring station with an annual 60% data availability for CO, if the 8-hour concentrations exceed 9.0 ppm on more than one day in a year.

5. DATA HANDLING CONVENTIONS

5.1 Computing averages

Schedule 2 of the AAQ NEPM specifies that "average" refers to an arithmetic mean and that 4-hour and 8-hour running averages are based on 1-hour averages.

Each hourly average concentration is to be calculated by averaging all available measurements taken within the hour.

Four-hour, 8-hour and 24-hour averages are to be based on hourly averages. Four-hour and 8-hour averages are running averages and are to be referenced by the last (end) hour of the period. Daily (24-hour average) values refer to concentrations calculated from midnight to midnight (local standard time).

Annual averages are to be calculated from hourly averages. To demonstrate compliance, a valid annual average must be based on hourly data that are at least 75 percent complete in each calendar quarter.

5.2 Computing running averages

Running 4-hour and 8-hour averages should be computed from the hourly concentrations for each hour of the year and the results should be referred to the end-hour of the 4-hour or 8-hour periods. There is no requirement for jurisdictions to provide data sets of running averages. Jurisdictions may wish to store the computed running averages as running average data sets for use in compliance assessments and analysis, as computations can be time consuming.

The computed 4-hour average ozone concentrations should be reported to three decimal places and the computed 8-hour CO concentrations should be reported to one decimal place, as specified in section 3.2.

A 4-hour or 8-hour running average is valid if at least 75% of the hourly averages for the 4-hour or 8-hour period are available.

5.3 Handling of TEOM PM_{10} data

Technical Paper No 10 specifies that where the TEOM site is known to experience a significant contribution from volatiles, adjustments should be applied to validated 24-hour averages, or optionally to validated hourly averages. Where adjustments are applied, it should be clearly stated that TEOM data are the adjusted TEOM data. The procedures for the handling of TEOM PM_{10} data are described in Technical Paper No 10.

6. STANDARD DATA FORMATS

A national standard format for AAQ NEPM data is intended to facilitate:

- Transparency and accessibility of data for people outside the organisation that collects the data;
- The establishment of a future national database and archive of all NEPM data in a form that will be easily read and interpreted; and
- Potential input to a standard computer program to provide identical assessment of AAQ NEPM compliance across all States and Territories of Australia.

It is recommended that AAQ NEPM data are made available and exchanged as separate files for each performance monitoring station. Data files should contain validated data for a calendar year for all pollutants with a common averaging period. The national standard data format is an ASCII tab delimited file.

AAQ NEPM data are to be available as:

- Hourly (1-hour) averages for CO, NO_2, O_3 and SO_2; and
- Daily (24-hour) averages, everyday for PM_{10} and every sixth day for lead.

Additionally, jurisdictions may provide data with shorter averaging periods.

For each calendar year and performance monitoring station there may be a number of data files generated depending on the averaging periods of the pollutants measured. For example, a station may generate two files, one for hourly data and one for daily (24-hour) data. Additionally, if 10-minute averages are available, a third file could optionally be generated. Note that in this context, hourly TEOM PM_{10} data could be included in the data file containing other hourly averages.

Proper interpretation of air quality monitoring data requires information other than just the concentration values. Metadata is the term used for any information that may assist in

interpretation, quality control, quality assurance or processing of concentration data. Such data can be voluminous, and are stored by different monitoring organisations in many different ways. Only basic metadata need to be provided in national data files, as text files, using the list in section 3.7 as a guide.

Metadata for each performance station should be in a separate meta file.

STANDARD FORMAT SPECIFICATIONS

An example of the general structure of the national standard data file is shown in Figure 1.

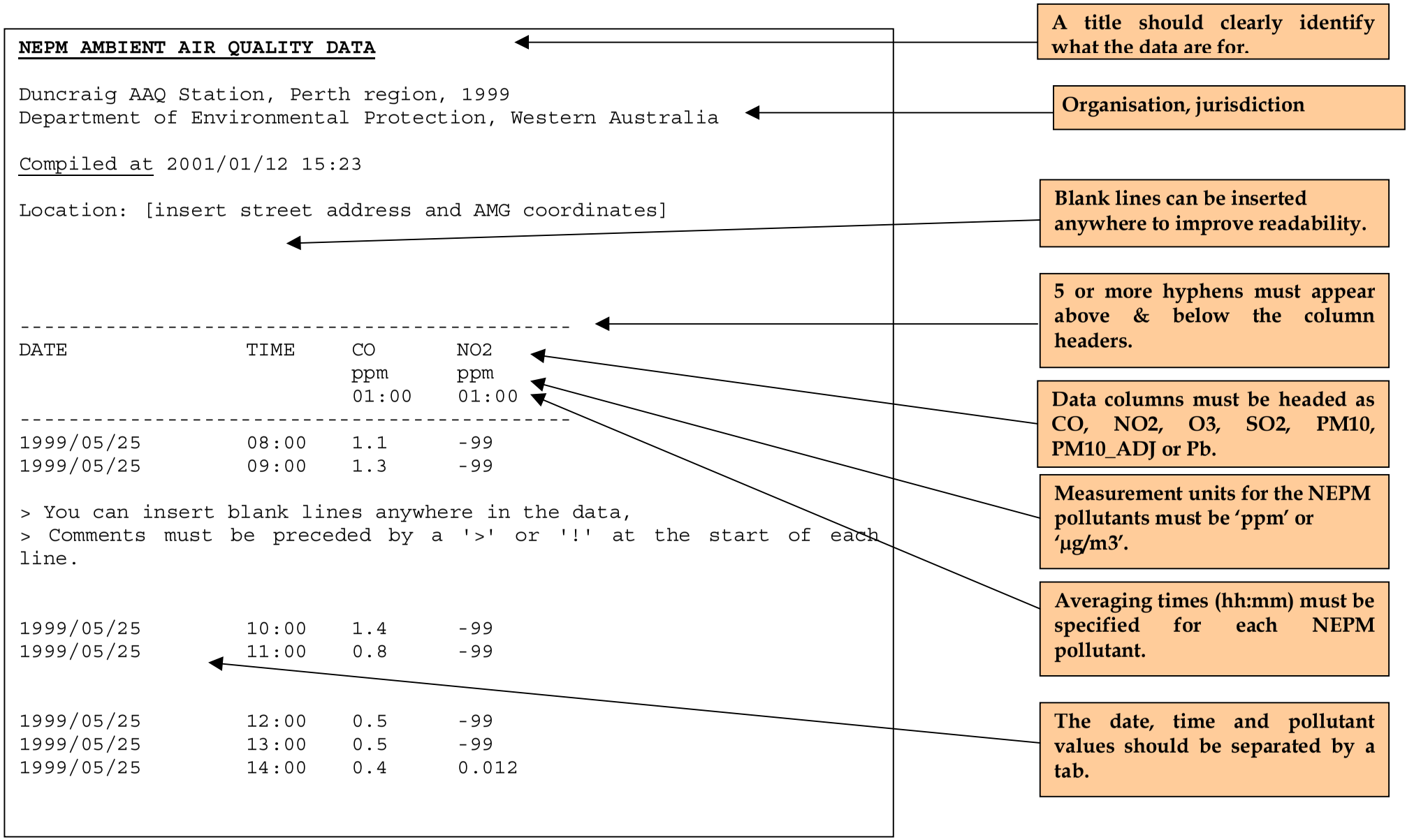
Standard file content:

1st record must commence with 'NEPM AMBIENT AIR QUALITY DATA'
2nd nonblank record contains a title that includes the monitoring station name, region and year.
3rd nonblank record contains the names of the custodian organisation and the State or Territory.
4th nonblank record must commence with 'Compiled at', followed by date and time.
5th nonblank record should give the station location in a form that allows an easy, precise map search.
A heading block for the columns of data should follow, bounded above and below by lines of at least five hyphens. Date, time and pollutant values should be separated by a tab. Leading spaces are allowed. Data columns must be labeled as CO, NO2, O3, SO2, PM10, PM10_ADJ or Pb, where PM10_ADJ refers to adjusted TEOM data. Averaging periods for gases are 1-hour, and for PM10 and lead are 24-hours. Optionally, additional files with data for shorter averaging periods may be provided.
Data records follow, with invalid data indicated by "a number less than or equal to –99". They may be interspersed with blank lines or comment lines; e.g., to indicate where there was a change of instrument or an incident of bushfire smoke. These comment lines do not obviate the need to provide relevant information in the metadata file.

A possible file naming convention that allows the identification of the jurisdiction, performance monitoring station, year and the averaging period is as follows.

- NEPM_AAQdata_station_region_jurisdiction_year_averaging period.txt
- NEPM_Metadata_station_region_jurisdiction_year.txt

Figure 1. Example file following the standard data format. Underlined text must appear verbatim.



7 ARCHIVAL OF DATA

Jurisdictions are responsible for archiving their air quality data to ensure that AAQ NEPM data can always be generated and provided. Where preferred, jurisdictions can also archive NEPM data in the standard format.

8. SOFTWARE DEVELOPMENT

A computer program called Air National Environment Protection Measure Analysis Program (ANEPMAP) in Fortran has been written as part of the data consultancy to implement the recommendations and apply the procedures specified in this technical paper. The program provides a nationally consistent means of processing AAQ NEPM air quality data and determining AAQ NEPM compliance. This computer program is available to all jurisdictions.

To use this program, raw data will first have to be validated as specified in Section 4, and then cast into the standard format specified in Section 6. The program optionally generates files containing concentrations averaged over all the periods relevant to the AAQ NEPM, using the methods approved by the PRC. It also produces summary information of AAQ NEPM compliance status. Full instructions for the program are presented in a separate manual.