United States
Environmental Protection
Agency

Research and Development

Atmospheric Research and
Exposure Assessment Laboratory
Research Triangle Park NC 27711

EPA/600/S3-89/031 Sept. 1989

EPA

Project Summary

Materials Aerometric Database for Use in Developing Materials Damage Functions

Ruen-Tai Tang, P. Michael Barlow, and Paul Waldruff

Meteorological and air quality data acquired at field exposure sites have been accumulated into the Materials Aerometric Database (MAD). Task Group VII of the National Acid Precipitation Assessment Program (NAPAP) will use the MAD to develop damage functions for materials exposed at the sites; these functions then will be used in preparing NAPAP integrated assessment reports to Congress. The MAD data cover as many as six and a half years at five materials exposure sites in the eastern United States. Conservative techniques based on secondary-site data, regression predictions, and other information have been applied to the MAD to enhance the quality and usability of the database. The enhanced version of the MAD, as well as the original MAD, have been given to Task Group VII.

This Project Summary was developed by EPA's Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

The EPA's Atmospheric Research and Exposure Assessment Laboratory (AREAL) has undertaken the task of maintaining the Materials Aerometric Database (MAD), which consists of air quality and meteorological data from five test sites to be used in developing damage functions for the National Acid Precipitation Assessment Program (NAPAP) Materials Assessment Program. The research objectives for this project are as follows:

(1) To accumulate and organize an aerometric database (MAD) containing air quality data and meteorological measurements made at five *primary field sites*.
- Develop a uniform format for the MAD data.
- Provide validated tapes of the MAD data to the principal investigators within the Materials and Cultural Effects Task Group (Task Group VII).
- Acquire quality assurance/quality control (QA/QC) data from the site operators.
- Monitor the independent systems and performance audits of the sites, conducted by Research Triangle Institute.

(2) To enhance the database and allow its use in continuous-damage models by making reasonable predictions for missing data points (infilling).
- Acquire secondary-source data for infilling missing primary-source data.
- Provide data tapes of the enhanced air quality and meteorological data to the principal investigators within Task Group VII.

Technical Approach

Five materials exposure sites were chosen for continuously recording air quality, meteorology, particle loadings and chemistry, and rain chemistry measurements (Figure 1):

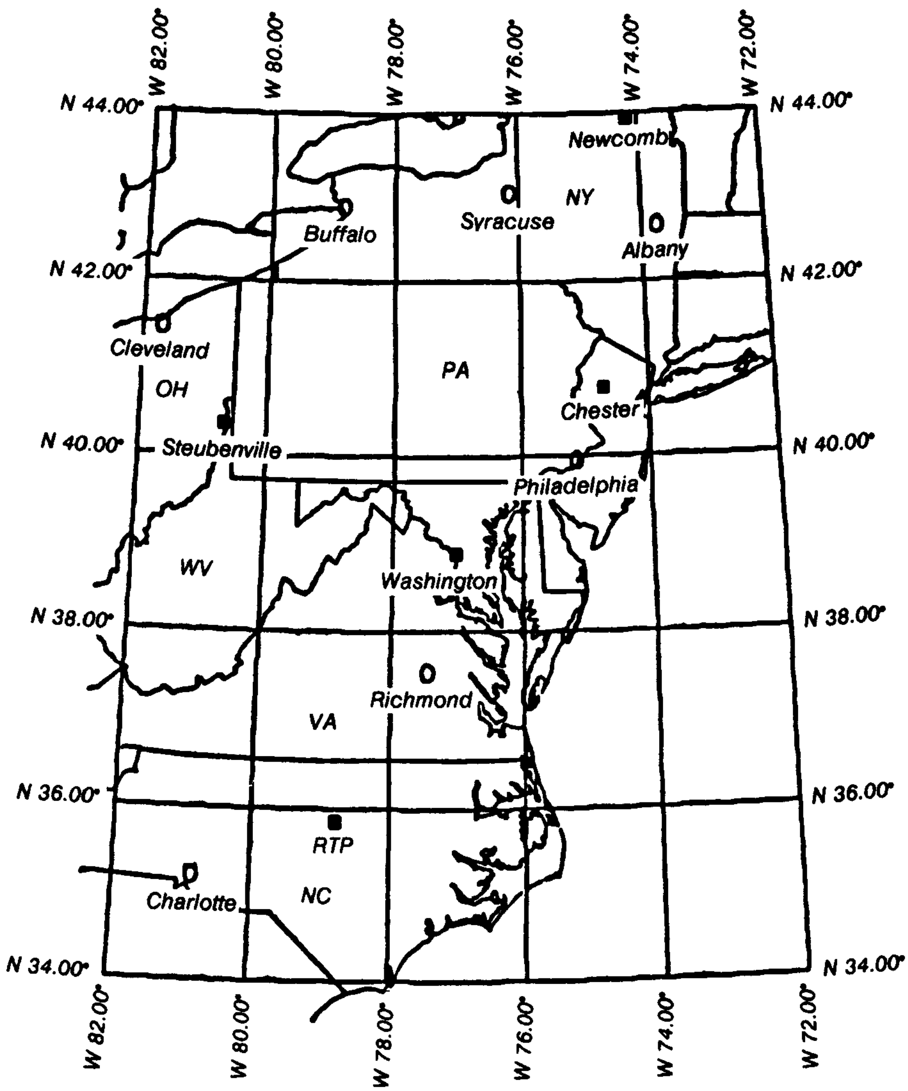

Figure 1. Locations of the five materials exposure test sites (marked with solid squares).

- Adirondack Ecological Center, Newcomb, NY
- Bell Communications Research Center, Chester, NJ
- County Services Building, Steubenville, OH
- West End Library, Washington, DC
- Research Triangle Institute, Research Triangle Park, NC

The following variables were measured in order to quantitatively evaluate the deposition of acids on the materials samples:

Meteorological Variables

- Wind speed average (WSA)
- Wind direction average (WDA)
- Wind direction vector (WDV)
- Temperature (TP)
- Dew point (DP)
- Relative humidity (RH)
- Precipitation (PR)
- Solar radiation (SR)

Air Quality Variables

- Sulfur dioxide (SO_2)
- Ozone (O_3)
- Oxides of nitrogen (NO_x)
- Nitric oxide (NO)
- Nitrogen dioxide (NO_2)

The data were supplied to the EPA in a variety of site-dependent formats. The format used by the Research Triangle Park site was chosen as the standard data storage format (Figure 2), and

```
                   1         2         3         4         5         6         7         8
Col. no :  12345678901234567890123456789012345678901234567890123456789012345678901234567890
          286 043 03   26   15F   3   18   13  320   33  -12   -91   550    0    0  319 9999
          286 043 04   24   13    3   16   15  318   33  -17   -95   553    0    0  318 9999
Contents:  YR DAY HR  SO2  NO2   NO  NOx   O3  WDA  WSA  TP    DP    RH    PR   SR  WDV  WSV
```

Additional information:

Column 1 contains the site ID (1 = DC, 2 = NC, 3 = NJ, 4 = NY, 5 = OH)
Columns 5-7 contain the Julian date.
Column 21 contains an example information flag (discussed below); all are non-numerical characters, except the minus sign.
WSV stands for wind speed vector, a variable reported by only the New York site.

Figure 2. Format for storage of all MAD data (based on the format for the RTP, NC, site).

software was developed to convert all other sites' data to this format.

During the project, some of the data have been lost due to equipment shortages and failures, power outages, etc. A number of secondary sources were found to replace the missing data. These sites were located near the primary sites with similar micrometeorology and are listed in Table 1 with the types of variables recorded at each site.

Each site performed ongoing QA on its own systems and data, based on the format outlined in 40 CFR 58, Appendix A. Also, an independent audit team from the Research Triangle Institute (RTI) has conducted annual or semiannual performance and system audits at the sites since 1985.

As the raw data were received, we performed preliminary statistical analysis and quality assurance checks, identified data problems, and contacted the site operators about them. Wherever possible, these problems were corrected and problem data were replaced by the site.

Using secondary-source data (if available) to infill primary-source data is the most reliable method of infilling, as long as the correlation between them is good and any bias can be removed. We used secondary-source substitution whenever possible, before using alternative forms of infilling. For missing meteorological data, we used only secondary-source substitution, because infilling this type of data using predictions from any form of calculations could corrupt the database. For missing air quality data, however, we employed three predictive infilling methods when no acceptable secondary-source data were available: linear interpolation using good data on either side of the gap; regression using a long-term least-squares prediction; and daytime or nighttime averages each compiled from a month's worth of data.

We performed a preliminary survey of the NC data to evaluate the occurrences of missing data. For a given year and variable, we recorded the gap length for each instance of missing data and then developed histograms of this information. In most cases, the data displayed a peak at one hour and then dropped off quickly after two or three hours, as shown in the example in Table 2.

We decided to take the most conservative approach to infilling the data. For one-hour and two-hour gaps, we used interpolation. After these were infilled, the remaining gaps of three or more hours were infilled using a regression prediction, if the R^2 for the regression equation was greater than 0.50. We then infilled any gaps still remaining with the appropriate daytime or nighttime monthly average. Listed below is a summary of the steps we followed in processing each subset of the database to produce the final enhanced database:

Step 1. For both meteorological and air quality data:

Replace all missing data, data below detectable limits, or data above reasonable limits with acceptable secondary-source data, if the correlation between primary- and secondary-source data is high ($r^2 > 0.95$). Use the following mathematical replacement:

$$X_{prim}(t) = X_{sec}(t) + \text{(the difference in their yearly averages)}$$

Step 2. For air quality data only:

For one-hour or two-hour gaps remaining after Step 1, infill with a smoothed value interpolated from the points before and after the gap. For gaps three or more hours long, use regression to replace the data if R^2 for the regression equation is greater than 0.50; otherwise, replace missing data with daytime or nighttime monthly averages (where daytime includes hourly data from 7 A.M. to 6 P.M. and nighttime includes data from 7 P.M. to 6 A.M.)

Step 3. Apply the following special corrections as needed:

- Set solar radiation to zero at night, if not already zero.
- For air quality data only, smooth in filled data into measured data to avoid abrupt slope discontinuities. Use a five-time-step smoothing scheme for infilling done with multiple regressions, secondary-source data, and monthly averages; do not use for infilling done with one- and two-hour interpolations.
- Maintain the NO_x, NO, and NO_2 balance using:

$$\text{Conc}(NO_x) = \text{Conc}(NO) + \text{Conc}(NO_2)$$

An information flag accompanies every data point in the MAD. The flag is a blank for original, untouched data. If the data were modified or infilled, this flag is set to a code describing the infilling method.

Results

Detailed descriptions of the raw data statistics and the site operations for all sites are presented in an April 1988 EPA internal report, *Monitoring and Operations at Materials Effects Sites* (R.T. Tang, P.M. Barlow, and J.W. Spence). Table 3 presents raw-data summary statistics for one site (RTP, NC).

Table 1. *Secondary MAD Data Sources*

Primary Site	*Corresponding Secondary Site(s)*	*Variables Measured*
Newcomb, NY	*State University of New York site 1 km away from primary site*	*Meteorology*
Chester, NJ	*Bell Core Lab, Chester, NJ*	*Meteorology*
Steubenville, OH	*Harvard School of Public Health Study site in Steubenville, OH*	*Meteorology, air quality*
	NOVAA, Mingo Junction, OH	*Meteorology*
Washington, DC	*Washington National Airport*	*Meteorology*
Research Triangle Park, NC	*Raleigh-Durham Airport*	*Meteorology*
	USEPA, Research Triangle Park, NC	*Meteorology, air quality*

Table 2. *Missing-Data Gap-Length Frequencies for 1984 RTP, NC Air Quality Data*

Gap Length	*Number of Gaps in a Variable's Data*				
(h)	O_3	SO_2	*NO*	NO_x	NO_2
1	31	19	12	16	13
2	10	15	5	4	5
3	2	6	3	3	4
4-6	4	15	4	4	4
7-12	5	6	1	1	0
13-24	4	4	1	1	2
>24	6	8	2	2	2

We processed the raw meteorological and air quality data from each site using the steps discussed above, and then performed statistical analyses on the enhanced MAD; the procedures and results are given in *Enhancement of Materials Aerometric Database* (R. T. Tang, P. M. Barlow, and J. W. Spence), a July 1988 EPA draft internal report. Table 4 presents summary statistics for the infilled data for the RTP, NC, site.

Discussion

The MAD is now available for use in predicting damage functions. The raw and enhanced MAD data tapes contain the data for all sites over part or all of the 1982-1988 period. The missing air quality data have been infilled using the algorithm discussed above. However, it was not possible to find secondary sources for all of the sites, so there are gaps in the meteorological data. The use of modeled meteorological data in the development of the damage functions could seriously bias the predicted data values and the statistics that describe them.

Also, a bias has already been found the data from two sites. For data valu below the minimum detectable lim (MDL), the DC site has been reportin the MDL and the NJ site has reporte one-half the MDL. This is acceptable f some EPA uses, but we are current trying to acquire the unmodified data.

Conclusions

We have developed an enhance database that will provide materia assessment investigators with a com prehensive data set of meteorologic and air quality data collected durin materials test exposures. Two tapes, or containing the raw data rewritten in uniform format and the second containin the enhanced database, have been pro vided to the principal investigators withi Task Group VII. The MAD provide essential data for developing damag functions to be used in estimating curre materials damage due to acid rai predicting future damage, and aiding the development of control scenario NAPAP will use this information to de velop reports to Congress.

Table 3. *Summary Statistics for the Raw MAD for the RTP, NC, Site*

		Variable										
Year	*Statistic*	O_3	SO_2	*NO*	NO_x	NO_2	*WSA*	*TEMP*	*DEW PT*	*RH*	*PR*	*SR*
1982*	Mean (ppm)	0.023	0.001	0.009	0.020	0.011	1.25	16.74	11.04	71.22	0.12	13.12
	Std. Dev.	0.019	0.003	0.022	0.024	0.007	1.11	7.96	8.50	17.19	0.97	19.41
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	0.00	−13.50	24.00	0.0	0.0
	Max.	0.091	0.021	0.225	0.250	0.046	6.80	31.80	2.60	99.40	33.0	76.20
	% Missing	3.3	3.2	1.1	2.2	2.3	31.0	0.4	6.5	6.8	0.0	4.4
1983	Mean (ppm)	0.029	0.003	0.008	0.021	0.013	1.63	14.64	6.86	65.07	0.13	14.03
	Std. Dev.	0.023	0.004	0.019	0.023	0.009	1.30	10.07	9.43	20.22	0.85	21.03
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	−16.00	−26.10	16.20	0.0	0.0
	Max.	0.132	0.048	0.295	0.312	0.073	10.00	38.70	23.00	98.60	22.0	78.0
	% Missing	6.5	4.6	3.3	3.3	3.5	0.1	0.9	16.4	17.3	0.0	2.4
1984	Mean (ppm)	0.025	0.004	0.010	0.022	0.012	1.65	15.16	7.79	65.96	0.14	11.38
	Std. Dev.	0.021	0.004	0.024	0.029	0.010	1.20	9.31	9.45	18.85	1.01	17.57
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	−12.70	−17.50	13.40	0.0	0.0
	Max.	0.118	0.040	0.351	0.372	0.065	8.10	35.00	23.10	100.00	32.1	76.0
	% Missing	1.2	42.5	2.8	3.8	3.5	0.1	0.1	10.5	10.5	2.5	0.1
1985	Mean (ppm)	0.025	0.002	0.008	0.022	0.013	1.48	15.29	8.02	61.00	0.12	13.67
	Std. Dev.	0.020	0.002	0.018	0.026	0.010	1.20	9.64	10.58	19.46	0.98	20.27
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	−21.10	−33.50	11.90	0.0	0.0
	Max.	0.119	0.026	0.268	0.307	0.108	10.00	34.80	22.70	98.40	36.1	78.0
	% Missing	3.0	71.2	24.5	16.4	31.2	0.0	0.8	35.3	35.3	1.1	1.6
1986	Mean (ppm)	0.025	0.002	0.010	0.023	0.015	1.44	15.53	5.18	57.05	0.11	13.34
	Std. Dev.	0.022	0.005	0.022	0.025	0.009	1.11	9.77	10.70	20.46	1.15	20.12
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	−12.70	−26.30	10.60	0.0	0.0
	Max.	0.123	0.080	0.410	0.436	0.169	8.10	37.70	22.80	95.30	41.3	76.0
	% Missing	10.3	42.5	42.4	16.4	49.1	0.3	0.3	46.3	46.3	0.0	0.3
1987	Mean (ppm)	0.027	0.003	0.011	0.027	0.014	1.40	14.91	7.98	66.01	0.12	13.27
	Std. Dev.	0.022	0.004	0.028	0.034	0.009	1.23	9.88	9.88	20.06	1.11	19.88
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	−8.90	−16.20	10.70	0.0	0.0
	Max.	0.112	0.054	0.375	0.391	0.059	8.20	37.70	24.40	100.00	64.2	80.0
	% Missing	10.2	13.3	41.6	31.5	41.3	0.8	0.1	1.5	1.2	0.1	1.7

**July through December data only.*

Table 4. *Summary Statistics for the Enhanced MAD for the RTP, NC, Site*

Year	Statistic	Variable										
		O_3	SO_2	NO	NO_x	NO_2	WSA	TEMP	DEW PT	RH	PR	SR
1982*	Mean (ppm)	0.023	0.001	0.009	0.020	0.011	1.28	16.77	12.39	77.57	0.12	12.72
	Std. Dev.	0.019	0.003	0.022	0.024	0.007	.98	7.97	8.37	18.45	0.97	19.18
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	0.00	-12.20	24.60	0.00	0.00
	Max.	0.091	0.021	0.225	0.250	0.091	6.80	31.80	23.90	100.00	33.0	76.20
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6
1983	Mean (ppm)	0.029	0.003	0.008	0.021	0.013	1.63	14.50	8.21	69.12	0.13	13.72
	Std. Dev.	0.023	0.004	0.019	0.023	0.008	1.30	10.12	9.84	20.36	0.85	20.89
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	-16.00	-27.20	20.70	0.00	0.00
	Max.	0.143	0.048	0.295	0.312	0.073	10.00	38.70	24.40	100.00	22.0	78.0
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1984	Mean (ppm)	0.025	0.004	0.010	0.022	0.012	1.65	15.17	8.87	68.65	0.14	11.36
	Std. Dev.	0.021	0.003	0.024	0.029	0.010	1.20	9.31	9.87	19 34	1.01	17.56
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	-12.70	-17.20	19.30	0.00	0.00
	Max.	0.118	0.040	0.351	0.372	0.119	8.10	35.00	24.40	100.00	32.10	76.00
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0 0	2.5	0.0
1985	Mean (ppm)	0.025	0.003	0.007	0.022	0.014	1.48	15.18	7.63	63.84	0.12	13.46
	Std. Dev.	0.020	0.002	0.016	0.024	0.014	1.20	9.69	10.54	20.36	0.98	20.18
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	-21.10	-33.30	13.0	0.00	0.00
	Max.	0.119	0.026	0.268	0.307	0.268	10.00	34.80	22.80	100.00	36.10	78.00
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0
1986	Mean (ppm)	0.025	0.002	0.008	0.022	0.015	1.44	15.51	9.31	70.28	0.11	13.31
	Std. Dev.	0.021	0.004	0.017	0.023	0.013	1.11	9.76	10.18	21.89	1.15	20.10
	Min.	0.000	0.000	0.000	0.000	0.000	0.0	-12.70	-22.80	15.70	0.00	0.00
	Max.	0.123	0.080	0.410	0.436	0.278	8.10	37.70	26.10	100.00	41.30	76.00
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1987	Mean (ppm)	0.026	0.003	0.010	0.026	0.016	1.40	14.91	8.63	69.63	0.12	13:27
	Std. Dev.	0.021	0.004	0.022	0.029	0.014	1.23	9.88	10.35	21.80	1.11	19.88
	Min.	0.000	0.000	0.000	0.000	0.000	0.00	-8.90	-20.60	10.40	0.00	0.00
	Max.	0.112	0.054	0.375	0.391	0.276	8.20	37.70	25.00	100.00	64.20	80.00
	% Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.7

*July through December data only.

Ruen-Tai Tang, P. Michael Barlow, and Paul Waldruff are with Computer Sciences Corporation, Research Triangle Park, NC 27709
***F. H. Haynie** is the EPA Project Officer (see below).*
The complete report, entitled "Materials Aerometric Database for Use in Developing Materials Damage Functions," (Order No. PB 89-181 259/AS; Cost: $13.95, subject to change) will be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Atmospheric Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency
ResearchTriangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Official Business
Penalty for Private Use $300

EPA/600/S3-89/031