

USER GUIDE
for the
WUAQL AethDataMasher
User Guide Version 4.2g / November 20, 2003
Software Version 4.2g / November 20, 2003

DATA POST-PROCESSING SOFTWARE FOR THE MAGEE SCIENTIFIC AETHALOMETER[®]

REGISTRATION, UPDATES AND LIABILITY

This software package is being distributed as freeware. All users are encouraged to register by sending an electronic mail message to -

JRTURNER@SEAS.WUSTL.EDU

Please write "Aeth Masher Registration" in the subject line of the email; the main body should include your contact points (email address, snail mail address). All registered users will be informed of any updates to this software package. In addition, users are encouraged to periodically visit the following web site for information on this software package and other Aethalometer items of interest:

<http://www.seas.wustl.edu/jrtturner/Aethalometer>

Please report any bugs to Jay Turner at the above email address; your suggestions for improvements are also welcome.

The authors assume no liability for any errors in this software package. The disclaimer displayed in the main window of the this software applies to all uses of this package.

INTRODUCTION

The Magee Scientific Aethalometer[™] is available in single-, dual-, and 7-seven wavelength configurations. Details about the aethalometer operating principles and use are available directly from Magee Scientific (<http://www.mageesci.com>).

The WUAQL aethalometer data masher ("AethDataMasher") - developed by the Air Quality Laboratory at Washington University in St. Louis (WUAQL) in collaboration with George Allen (NESCAUM) - can be used to post-process the raw data files obtained directly from the instrument. This program can process data from both 1-channel (AE-16) and 2-channel (AE-16) instruments; it is not configured for processing data from 7-channel instruments.

Raw data from the aethalometer is in 5-minute intervals. This program uses the raw data file(s) as input and generates two processed data files as output: a 5-minute data output file (similar to the raw data file but with additional formatting and data validation); and a 1-hour average output file (which also includes data validation). Both output files are in comma-delimited format which can be easily imported into spreadsheets or other data analysis packages. BC (and UV-C) concentrations are reported as ng/m³ in the raw data files and µg/m³ in the post-processed output files. A log file is also generated which provides important documentation concerning the post-processing.

The software can process multiple input files in a one batch; thus, there is no need to concatenate the daily data files into a single input file prior to processing. There is one constraint on the number of files which can be processed in one batch - the software can handle a maximum time span of 113 days which corresponds to greater than three months (e.g., a calendar quarter of data). If a larger data range is used, the program will terminate without processing the data and an error message will be displayed. The processor cannot handle time stamps prior to January 1, 2000.

The user must specify the instrument type (single channel versus dual channel), deposit spot size (small/circular versus large/oval), and whether an AQS header row is desired for the output file with hourly averages.

- The deposit spot size determines the threshold concentration used in step 6a/6b below. Most of the aethalometers shipped for the NATTS network have large/oval spots and this is set as the default option in the software; older instruments have small/circular spots (as do newer instruments which are special-ordered with his configuration). If you do not know your instrument deposit spot configuration, please contact Magee Scientific and provide your instrument serial number.
- Many users will be providing the hourly-average BC concentrations to AQS. VOCDat - software available from Sonoma Technology, Inc. (<ftp://ftp.sonomatech.com/public/vocdat/>) and distributed as freeware – is a useful tool for migrating data from the hourly-average AethDataMasher output file into AQS format. The user has the option to add a header row to the hourly-average output file which includes the AIRS parameter code for the BC channel and dummy AIRS validation codes for the UV-C channel and the validation code. Note that the AIRS parameter code assigned to the BC channel (84313) is strictly for the AE-21 dual-channel aethalometer with the data reported at STP. Data to be reported at Local T/P must be assigned AIRS parameter code 88313 which this software currently does not support.

STEP-BY-STEP DETAILS FOR USING THE SOFTWARE

1. Start the software by double-clicking on the executable file:
"WUAQL_AethDataMasher4.2e_031026.exe"
2. Click the button labeled "Click Here to Continue" to move from the welcome form to the main form. [If you wish to exit the program, do not click on the "X" in the upper-right-hand corner but rather proceed to the main form and exit via the button labeled "Exit the Program". Failure to exit the program properly can cause it to hang up in memory.]
3. Box #1 - Click on the radio buttons to choose the instrument type, deposit spot size, and whether an AQS header is desired in the hourly-average data output file. The options commonly chosen by NATTS users have been set as the defaults.
4. Box #2 – Click on the button labeled "Go to the File Selection Menu" to pull up the form where the input files can be selected. Browse to the drive and subdirectory where the raw data files are housed. The default file list includes only those files with names "BC*.CSV" because the aethalometer daily data files follow this naming format. All data files with extension "CSV" can be displayed by clicking on the button labeled "Change list to display all CSV files (*.CSV)". One-or-more input files can be selected using conventional techniques to highlight the filenames. There is also a button to highlight all files displayed and a button to clear all the files highlighted. Once the desired files have been selected, click on the button labeled "Accept the Selected Files and Return to the Main Menu".
5. Box #3 – The default location for the output files is the subdirectory where the input files are located; click on the button labeled "Select a New Output File Path" to pull up a form which allows you to browse to a different drive/directory for the output files. This is particularly important if you are reading the input files from a floppy disk taken directly from the aethalometer (and on which there might be insufficient room for the output files) or are reading the input files from media where files cannot be directly written by the software (e.g., a CD-ROM).
6. Box #4 – Click on the button labeled "Process the Data" to process the input file(s) selected. You may be prompted with one-or-more message boxes containing warning or error messages; click OK in each case to proceed with the processing. When the processing has been completed, a message box stating "Output Files Created" will appear; click OK. At this point, you can make changes to any of the entries on Boxes #1-#3 and reprocess data, or exit the program by clicking on the button labeled "Exit the Program".
7. Three output files will be created each time input data is successfully processed. The filenames are "PFlog_filename.txt" for the log file, "PF05_filename.csv" for the 5-minute output file, and "PF60_filename.csv" for the hourly-average output file where *filename* is the root filename for the input data file (or the first file listed when multiple files are processed in a single batch). Always

browse the log file for warnings and flags that might influence data validation. Also, it is recommended that both the 5-minute and hourly-average output files be pulled up in a spreadsheet or graphics package and the BC data (and UV-C data, if applicable) plotted as a time series to further check the integrity of the data.

ALGORITHM

This section summarizes the key steps in the post-processing algorithm.

- 1) Read all data records (first pass) from the input file(s) to check for proper formatting.
 - a) For first record in first file, check whether the proper channel option was selected. If it appears the wrong channel option has been selected (single versus dual), a pop-up message alerts the user of this problem and the program will terminate.
 - b) For all records, check the number of commas separating the values and the overall number of columns in the record. For Single channel data should have nine commas and no more than 77 columns; dual channel data should have sixteen commas and no more than 129 columns. If these criteria are not met, a pop-up message alerts the user that the program will terminate and the corrupt records must be cleaned up or removed prior to processing. The log file states which input file(s) contain corrupt records.
- 2) Read all data records (second pass) from the input file(s) into a raw data array.
 - a) Records with incorrectly formatted date stamps are ignored (format should be "dd-mmm-yy" where mmm are characters and dd and yy are numbers); an error message is written to the log file for each improperly formatted date stamp.
 - b) Records with incorrectly formatted time stamps are ignored (format should be "hh:mm" where hh and mm are numbers); an error message is written to the log file for each improperly formatted time stamp.
 - c) If both the date and time formats are incorrect for a given record, the log file message will only identify the date format as incorrect
 - d) While an effort has been made to make the screening of input records robust, users are highly encouraged to use the log file error information - which identifies incorrect date and time stamps - to clean up the raw data files prior to reprocessing; improperly formatted input records can have adverse implications to the data processing.
- 3) Screen each 5-minute record for the BC channel sample beam "lamp on" voltage. If one-or-more values less than 0.35V, send pop-up message to the user and write a warning to the log file. The data is not flagged for this warning, however, which is a potential indicator that the tape is not properly advancing. [Acknowledgement - George Allen]
- 4) Screen each 5-minute record for the BC channel reference beam "lamp on" voltage. If one-or-more values less than 1.5V, send pop-up message to the user and write a warning to the log file. The data is not flagged for this warning, however, which is a potential indicator of a hardware performance problem. [Acknowledgement - George Allen]
- 5) Identify the beginning and ending time stamps in the raw data array. The input file data records do not need to be in sequential order!
- 6) Create an array (hereafter called the "mapped" array) of contiguous 5-minute increments which spans the full range of time stamps in the raw data file (as identified in the previous step).
- 7) Map the raw data records into the mapped array by matching the time stamps
 - a) If the time stamp for any raw data record is not an integer multiple of 5 minutes, then reject that data record; an error message is written to the log file.
 - b) Data records missing in the input file are assigned Null values in the mapped array and are assigned validation code 202
 - c) Data records with "" in the input file (e.g., the reported values during a tape advance) are assigned Null values in the mapped array and are assigned validation code 201
 - d) Replicated records (with respect to the time stamps) in the input file are treated as follows. If a replicated record is encountered:
 - i) The first occurrence data values will be overwritten by the second occurrence data values if the first occurrence contained all "" data values.
 - ii) The first occurrence data values will not be overwritten if the second occurrence contains all

- "" data values.
- iii) The first occurrence data values will be overwritten if the first and second occurrences have different BC concentrations. In this case, an error message is written to the log file and the record is assigned validation code 110. No error message will be written to the log file - and the validation code will not be altered if the BC concentrations are identical in both records (assumed to be indicative of truly identical records, not just replicated time stamps).
- 8) Screen 5-minute records for BC values less than $-10 \mu\text{g}/\text{m}^3$; if present, send a pop-up message to the user and write a record to the log file. Negative values of this magnitude most likely result from an instrument problem that should be diagnosed.
- 9) Calculate hourly average BC (and UV-C if 2-channel) concentrations
- If any BC concentration at time stamps of 00 minutes or 55 minutes is less than $-0.1 \mu\text{g}/\text{m}^3$ for small spot aethalometers or less than $-0.3 \mu\text{g}/\text{m}^3$ for large spot aethalometers, void that 5-minute BC concentration and the adjacent 5-minute BC concentration on each side of 00 or 55 minutes, respectively. Write an error message to the log file.
 - If any UV concentration at time stamps of 00 minutes or 55 minutes is less than $-0.1 \mu\text{g}/\text{m}^3$ for small spot aethalometers or less than $-0.3 \mu\text{g}/\text{m}^3$ for large spot aethalometers, void that 5-minute UV concentration and the adjacent 5-minute BC concentration on each side of 00 or 55 minutes, respectively. Write an error message to the log file.
 - All 1-hour intervals for which at least one 5-minute record was voided due to the above **BC criteria** are flagged with validation code 120; all 1-hour intervals for which at least one 5-minute record was voided due to the above **UV criteria** are not flagged!
- 10) Screen the hourly average records for BC values less than $-0.67 \mu\text{g}/\text{m}^3$ for small spot aethalometers or less than $-0.2 \mu\text{g}/\text{m}^3$ for large spot aethalometers; if present, send a pop-up message to the user and write a record to the log file. Negative values of this magnitude most likely result from an instrument problem that should be diagnosed.

VALIDATION CODES

5-Minute Data Output File

- 1st digit
 - 0 = valid
 - 1 = flagged
 - 2 = invalid
- 2nd digit
 - 0 = not flagged
 - 1 = replicated data records with different BC concentrations (retain second-or-greater occurrence)
- 3rd digit
 - 0 = not invalid
 - 1 = value is "" in raw data file
 - 2 = no data record in raw data file

1-Hour Average Data Output File

- 1st digit
 - 0 = average valid
 - 1 = average flagged
 - 2 = average invalid
- 2nd digit
 - 0 = average not flagged, 12 valid/flagged 5-minute records in the 1-hour interval (valid data completeness 100%)
 - 1 = average not flagged, 9-11 valid/flagged 5-minute records in the 1-hour interval (valid data completeness at least 75% but less than 100%)
 - 2 = average flagged, 9-11 valid/flagged 5-minute records in the 1-hour interval but at least one record voided due above criteria concerning BC concentrations $< -0.1 \mu\text{g}/\text{m}^3$.
-

•3rd digit

0 = average not invalid

1 = average invalid, 1-8 valid/flagged 5-minute records in the 1-hour interval (valid data completeness > 0% but less than 75%)

2 = average invalid, no valid/flagged 5-minute records in the 1-hour interval (valid data completeness = 0%)

HINTS AND CAUTIONS

- **Required DLLs** Depending on your operating system, upon execution of this program you might encounter an error message stating that a system file cannot be found or accessed (e.g., VB40032.DLL). Future versions of this manual will provide troubleshooting tips; at this time, however, please contact Jay Turner for troubleshooting.
- **File Names** The input files should follow the standard comma-delimited aethalometer format (see below) with filename "*.csv". The output files will be named as follows: "PFlog_filename.txt" for the log file, "PF05_filename.csv" for the 5-minute data, and "PF60_filename.csv" for the hourly-average data where *filename* is the root filename for the input file (or the first file if multiple files are processed in one batch).
- **Raw Data Format** If the raw data files are pulled up in a spreadsheet (e.g., Excel) and this file is saved, it will likely change the format for the date and time stamps such that they cannot be recognized by this program! If you do desire to pull up the raw data files in Excel, do not save over the original file. The proper format for a 2-channel input file record is shown below. If the program crashes while reading in the data, record structure by pulling up the raw data file in a text processor (e.g., NotePad, WordPad) rather than a spreadsheet to verify the formatting.

"dd-mmm-yy","hh:mm", 15 comma-delimited numerical values

e.g.,

"24-jun-01","12:55", 804, 579, 4.2, 0.0220, 1.5134, 0.0220, 2.4555, 1.00,19.343, 0.0220, .3615, 0.0220, .8324, 1.00,43.902

- **Raw Data File Concatenation.** The aethalometer archives raw data to disk as daily files. This version of the data processing software can handle multiple files as simultaneous inputs. The data is automatically sorted by time stamp and ignores replicated records unless the data values are different for records with the same time stamps. In the latter case, the records are handled according to the algorithm described above.
- **Screening Raw Data Against Field Operations Activities.** It is very important to screen the raw data for any periods when maintenance was performed on the instrument while it was running. For example, a leak check - while taking less than one minute - can cause a very large spike in the concentration which carries through to the 5-minute average and even hourly-average.
- **Known Aethalometer Bug - Midnight Tape Exchange.** The aethalometer software incorrectly assigns time stamps to records with tape advances at midnight. Tape advances are typically 15 minutes (3 records) in duration with all data values listed as "". If the tape advance starts at 23:50, then the 00:00 record will not roll over to the next day; if the tape advance starts at 23:55, then the 00:00 and 00:05 records will not roll over to the next day; and if the tape advance starts at 00:00, then the 00:00, 00:05 and 00:10 records will not roll over to the next day. This program will not overwrite existing data with these tape advance null values; the time stamps during the tape advance are assigned validation code 202 (as if no data records existed).
- **Handling Large Negative Concentrations in the 5-Minute Records.** The above description of the algorithm includes the methodology for handling large concentration values (i.e. more negative than -100 ng/m³). Such large negative values are often accompanied by a large positive value for the record either immediately before or immediately after the large negative value. Thus, in most cases these fluctuations balance out to yield a generally valid hourly average. An exception occurs when the top of the hour splits these negative and positive swings, with the negative value being assigned to one hour and the positive value being assigned to a different hour. While there are various ways to

handle this case, when calculating hourly averages this program invalidates the 5-minute records on either side of the large negative value (i.e. it does not screen for the magnitude of such adjacent data records) when the large negative value occurs at 55 minutes or 00 minutes. This program does not alter the 5-minute data output stream for this issue. [Acknowledgement - George Allen]

- **Detection of Possible Instrument Problems.** One indicator of a possible problem with the instrument – either hardware, software, or the environment in which it is deployed – is large negative concentrations. Therefore, the algorithm attempts to identify such cases and alert the user via pop-up messages and error records in the log file. The thresholds for bringing attention to large negative values is subjective; this algorithm uses than $-10 \mu\text{g}/\text{m}^3$ for the 5-minute BC data, $-0.67 \mu\text{g}/\text{m}^3$ for small spot aethalometer hourly BC data, and $-0.2 \mu\text{g}/\text{m}^3$ for large spot aethalometer hourly BC data. If a pop up message is displayed during data processing which alerts the user of such negative values, the user is strongly encouraged to investigate possible problems with the instrument performance.

ANTICIPATED FUTURE IMPROVEMENTS

- **Enhanced Input Data Format Checking.** Code will be added to determine if single channel data is being read in under the dual channel option (or vice versa) for the first record in all input files (not just the first input file as currently checked).
- **BC Channel Lamp Voltage Checks.** The program currently issues a warning if one-or-more BC channel sample beam "lamp on" voltage are less than 0.35V (possibly indicating a failure to advance the tape) or one-or-more reference beam "lamp on" voltages are less than 1.5V. The program likely will be modified to flag such 5-minute data as well as the corresponding 1-hour average data.
- **UV Channel Validation Codes.** The hourly-average validation codes are primarily keyed to the BC channel as described above. The program will be modified to provide similar validation codes for the UV channel.
- **Enhanced Data Quality Diagnostics.** The program will be modified to track the frequency distribution of changes in adjacent 5-minute values and alert the user to distributions which suggest possible instrument performance problems. The program will also be modified to monitor for fluctuations in the reference lamp voltage intensity which might be indicative of an instrument performance problem.

SUMMARY OF CHANGES BETWEEN SOFTWARE VERSIONS

Version 4.2.

- Output file extensions changed from ".out" to ".csv".
- Attempts to process data spanning more than 113 days between the two extreme records will terminate the program with an error message.

Version 4.2a.

- Output files can be written to a subdirectory different from where the executable file resides.

Version 4.2b.

- Change instrument selection method from a list to option buttons with the dual channel instrument set as the default.
- Multiple input files can be read and processed in a batch; this avoids the need to concatenate the daily data files prior to processing.

Version 4.2c.

- Move the data file selection to a separate form.

Version 4.2d.

- Cleaned up various user interface elements.

Version 4.2e.

- Added option buttons for deposit spot size and AQS format header.
- Modified algorithm to use $-0.3 \mu\text{g}/\text{m}^3$ as the threshold for data filtering at the hourly-average level of aggregation for large spot instruments (retain $-0.1 \mu\text{g}/\text{m}^3$ as the threshold for small-spot instruments).

Version 4.2f.

- Cleaned up various user interface elements.

Version 4.2g.

- Added error message boxes and log file records for: (a) 5-minute records with $\text{BC} < -10 \mu\text{g}/\text{m}^3$; and (b) hourly-average records with $\text{BC} < -0.2 \mu\text{g}/\text{m}^3$ for large spot instruments $\text{BC} < -0.67 \mu\text{g}/\text{m}^3$ for small spot instruments. These changes do not affect the data processing but rather provide notice to the user of a potential instrument problem that should be diagnosed.
- Added code to check the first record of the first input file to determine whether single channel data is being read as dual channel data (or vice versa). If it appears the improper channel type was selected (single versus dual), program exits with an error message.
- Added code to check prior to processing for proper number of commas and proper maximum column length in each record. If criteria are not met, program exits with an error message.
- Added code to check for reference beam "lamp on" voltages $< 1.5\text{V}$, which may indicate a hardware performance problem.

CONTACT INFORMATION

Jay R. Turner, D.Sc.
Associate Professor
Environmental Engineering Program
Campus Box 1180
Washington University
One Brookings Drive
St. Louis, MO 63130-4899

Tel: 314-935-5480
Fax: 314-935-5464
Email: JRTURNER@SEAS.WUSTL.EDU