

---

## ***Atmospheric Model Evaluation Tool (AMET) v1.2 User's Guide***

Prepared under Work Assignment 3-01 of U.S. EPA Contract EP-W-09-023,  
"Operation of the Center for Community Air Quality Modeling and Analysis  
(CMAS)"

---

Prepared for: Bill Benjey  
U.S. EPA, ORD/NERL/AMD/APMB  
E243-04  
USEPA Mailroom  
Research Triangle Park, NC 27711

Prepared by: Wyat Appel<sup>a</sup>, Robert Gilliam<sup>a</sup> and Uma Shankar<sup>b</sup>

<sup>a</sup>U.S. EPA, ORD/NERL/AMD/APMB  
E243-04  
USEPA Mailroom  
Research Triangle Park, NC 27711

<sup>b</sup>Institute for the Environment  
The University of North Carolina at Chapel Hill  
137 E. Franklin St., CB 6116  
Chapel Hill, NC 27599-6116

Date prepared: July 3, 2013

## Contents

<b>Tables .....</b>	<b>iv</b>
<b>1. Overview .....</b>	<b>1</b>
1.1 Overall Objective and Basic Structure.....	1
1.2 Concept of an AMET “Project” .....	3
1.3 Organization of This User’s Guide .....	3
<b>2. Directory Structure.....</b>	<b>4</b>
<b>3. Configuration .....</b>	<b>5</b>
3.1 Perl Configuration File ( <b>amet-config.pl</b> ) .....	5
3.2 R Configuration File ( <b>amet-config.R</b> ) .....	6
<b>4. Datasets .....</b>	<b>7</b>
4.1 Model Data.....	7
4.2 Observational Data.....	8
<b>5. Database Setup .....</b>	<b>16</b>
5.1 MET Setup .....	16
5.2 AQ Setup.....	17
5.3 Basic MySQL Commands .....	18
<b>6. Database Population .....</b>	<b>19</b>
6.1 The wrfExample Project .....	20
6.2 The mm5Example Project.....	21
6.3 The mcipExample Project.....	22
6.4 The aqExample Project .....	23
6.5 Creating a New MET Project.....	24
6.6 Creating a New AQ Project .....	25
<b>7. Analysis .....</b>	<b>26</b>
7.1 wrfExample.....	27
7.2 aqExample.....	29
7.3 Creating a New Analysis Project .....	36
<b>8. CMAS Support for AMET .....</b>	<b>37</b>
<b>Reference .....</b>	<b>37</b>

## **Appendix A: Configuration and Input Files**

A.1	Perl Configuration File .....	A-2
A.2	R Configuration File .....	A-3
A.3	MET Project Setup Input File .....	A-4
A.4	AQ Project Setup Input File.....	A-5
A.5	MET Project Populate Input Files.....	A-5
A.6	AQ Project Script Options .....	A-8
A.7	MET Analysis Input Files .....	A-10
A.8	AQ Network Input File .....	A-17
A.9	AQ Analysis Input Variables .....	A-18
A.10	AQ Variables-Network Mapping for North America .....	A-25
A.11	AQ Variables-Network Mapping for Europe .....	A-26

## **Appendix B: What Is New in AMETv1.2**

## Tables

Table 2-1. Directories under <b>\$AMETBASE</b> .....	4
Table 3-1. Most common variables that need to be changed in <b>amet-config.pl</b> .....	6
Table 3-2. Most common variables that need to be changed in <b>amet-config.R</b> .....	7
Table 7-1. MET analysis scripts. ....	28
Table 7-2. AQ analysis scripts.....	30
Table A-1. <b>amet-config.pl</b> .....	A-2
Table A-2. <b>amet-config.R</b> .....	A-3
Table A-3. MET <b>setup_project.input</b> .....	A-4
Table A-4. AQ <b>setup_project.input</b> .....	A-5
Table A-5. MET <b>populate_project.input</b> .....	A-5
Table A-6. AQ <b>aqProject.csh</b> .....	A-8
Table A-7. MET analysis input variables .....	A-10
Table A-8. AQ network input file settings.....	A-17
Table A-9. AQ analysis input variables .....	A-19
Table A-10. AQ networks, frequency of observations, and chemical species for North America.....	A-25
Table A-11. AQ networks, frequency of observations, and chemical species for Europe .....	A-26

# 1. Overview

## 1.1 Overall Objective and Basic Structure

The Atmospheric Model Evaluation Tool (AMET) (Appel et al., 2011) is a suite of software designed to facilitate the analysis and evaluation of meteorological and air quality models. AMET matches the model output for particular locations to the corresponding observed values from one or more networks of monitors. These pairings of values (model and observation) are then used to statistically and graphically analyze the model's performance.

More specifically, AMET is currently designed to analyze outputs from the PSU/NCAR Mesoscale Model (MM5), the Weather Research and Forecasting (WRF) model, and the Community Multiscale Air Quality (CMAQ) model, as well as meteorological data (surface only) postprocessed by the Meteorology-Chemistry Interface Processor (MCIP). The basic structure of AMET consists of two *fields* and two *processes*.

- The two fields (scientific topics) are **MET** and **AQ**, corresponding to meteorology and air quality data.
- The two processes (actions) are **database population** and **analysis**. Database population refers to the underlying structure of AMET; after the observations and model data are paired in space and time, the pairs are inserted into a database (MySQL). Analysis refers to the statistical evaluation of these pairings and their subsequent plotting.

Practically, a user may be interested in using only one of the fields (either MET or AQ), or may be interested in using both fields. That decision is based on the scope of the particular study.

The three main software components of AMET are **MySQL** (an open-source database software system), **R** (a free software environment for statistical computing and graphics), and **Perl** (an open-source, cross-platform programming language). Figure 1 provides a schematic of the data flow among the various AMET components to help the user navigate the subsequent sections of this User Guide.

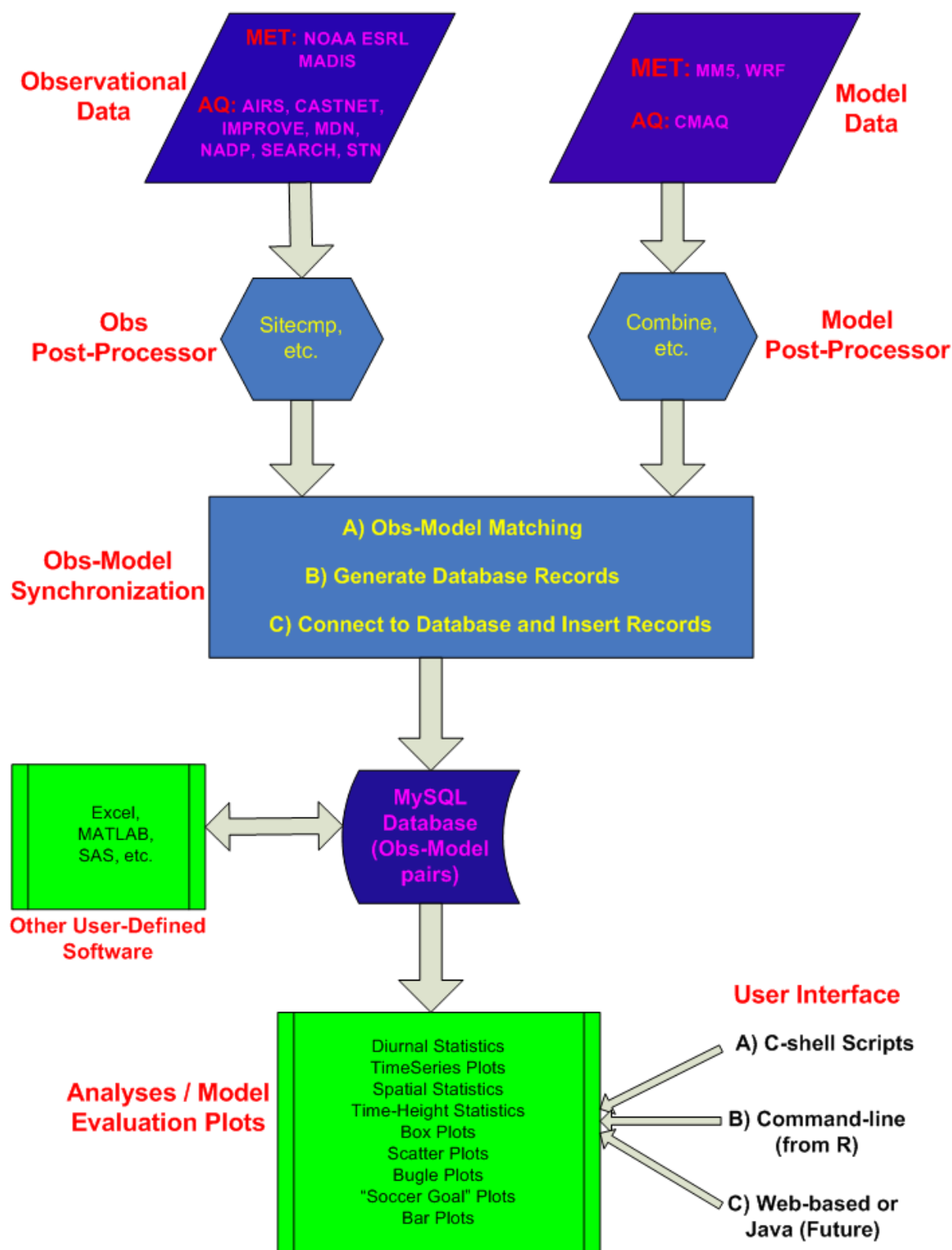


Figure 1. Overview of AMET data flow

## 1.2 Concept of an AMET “Project”

A central organizing structure for AMET applications is a *project*. A project groups a particular model simulation (specific model, physics-set, spatial domain, grid scale, etc.) with all of the AMET database tables that correspond to that simulation, the scripts necessary to populate that database, and the scripts required to analyze that project. For example, you might have one project for a 2002 36-km continental U.S. simulation, and another project for a 2002 12-km Eastern U.S. simulation. A project can be set up for either MET or AQ, but not for a combination. It is essential that you uniquely name each project. It is recommended that you follow the directory structure when creating new projects, by copying one of the three example directories (**aqExample**, **mm5Example**, **wrfExample**) provided with the installation and then renaming it to the new project's name.

## 1.3 Organization of This User's Guide

The Community Modeling and Analysis System (CMAS) Center has created this user's guide with input from the developers to assist you in applying the AMET system in your work. The CMAS Center obtained the MET and AQ portions of the AMET software separately from EPA, then combined them to create a consistent, integrated AMET v1.1 package that uses the UNIX C-shell interface to perform both MET and AQ model evaluation and analyses. After this integration, we tested the package in multiple environments. Subsequently EPA made several updates to the structure and analyses of AMET to create v1.2. The CMAS Center tested this version in multiple environments, and updated this user's guide to reflect the changes made in creating v1.2. The topics of the remaining sections are listed below:

- Section 2 describes the overall directory structure of the AMET installation.
- Section 3 gives instructions on how to configure the Perl and R configuration files.
- Section 4 contains an overview of the various model outputs and observed data provided with the AMET release.
- Section 5 provides instructions on how to create the AMET MySQL database, with specific instructions for each of the MET and AQ models. Sample MySQL commands are also shown for illustrative SQL queries.
- Section 6 gives instructions on how to populate the AMET MySQL database, with specific instructions for each of the MM5, WRF, MCIP, and CMAQ models, and also on how to create a new MET project and a new AQ project for subsequent analyses.
- Section 7 provides instructions on how to perform model evaluation for the WRF model and CMAQ model, and includes an overview of the functionality of all the MET and AQ evaluation scripts provided.

**IMPORTANT NOTE:** *The set of analyses/evaluation scripts provided in this release are strictly for illustration purposes on the functionality/design of AMET, and are not to be construed as a recommended suite of analyses scripts for model evaluation. We encourage the user community to use the scripts we have provided as examples and as a*

*basis for developing other analyses scripts and contributing them to the modeling community to increase AMET functionality.*

- Section 8 discusses how to obtain support for AMET from the CMAS Center (<http://www.cmascenter.org>).
- Appendix A gives information on the configuration and various input files used in AMET. For each input file a table is provided that includes brief descriptions of all user-defined variables that can be set by the user for a given evaluation.
- Appendix B is a handy reference guide to new or updated features in AMET v1.2 compared with the previous version. The user is referred to this appendix where appropriate within this user's guide.

Before using AMET and this user's guide, you must first install the AMET package on your system. For information on the installation process, please see the separate *Atmospheric Model Evaluation Tool (AMET) Installation Guide* that can be downloaded from the CMAS web site ([https://www.cmascenter.org/help/model\\_docs/amet/1.1/AMET\\_Installation\\_Documentation\\_v1.1.pdf](https://www.cmascenter.org/help/model_docs/amet/1.1/AMET_Installation_Documentation_v1.1.pdf)).

## 2. Directory Structure

In our discussions, we refer to the top of the AMET directory structure as “**\$AMETBASE**”. This environment variable is actually set in many of the scripts discussed below. For example, if you had untarred AMET's tarball in your home directory, then **AMETBASE** would be **~/AMET**.

Under **\$AMETBASE** are the directories shown in Table 2-1.

**Table 2-1. Directories under \$AMETBASE.**

Directory	Description
<b>bin</b>	External executables used by helper scripts.
<b>configure</b>	Configuration files for Perl and R.
<b>model_data</b>	Model output data (contains field-specific [i.e., <b>MET</b> and <b>AQ</b> ] subdirectories).
<b>obs</b>	Observational data (e.g., MADIS, discussed in Section 4.2) (contains field-specific [i.e., <b>MET</b> and <b>AQ</b> ] subdirectories).
<b>output</b>	Output of database population and analysis (contains project-specific subdirectories).



Directory	Description
<b>perl</b>	Perl scripts used primarily for database population.
<b>R</b>	R scripts used for statistical analysis.
<b>scripts_analysis</b>	Project-specific wrapper scripts and inputs for analysis (contains project-specific subdirectories).
<b>scripts_db</b>	Project-specific wrapper scripts and inputs for database population (contains project-specific subdirectories).
<b>src</b>	Source code for third-party software.

*Note:* For large model outputs and for MADIS observations that cover a long period of time, it may be prudent to link these data within the appropriate AMET directories rather than moving or copying them.

### 3. Configuration

After untarring the AMET code and data and installing/building the required three tiers of software components (as discussed in the AMET installation guide referenced above), the next stage is to configure the AMET system. In the **\$AMETBASE/configure** directory, you will find five files along with a README file:

- A Perl configuration file (**amet-config.pl**).
- An R configuration file (**amet-config.R**).
- Three configuration files from the Meteorological Assimilation Data Ingest System (MADIS) (**MET\_matching.conf**, **MET\_matching.MCIP.conf**, and **MET\_matching\_nonPX.conf**). These configuration files describe how MADIS variables “match” (map or translate to) various MET model variables. Most users will not need to change these variable mapping files, so they are not discussed further in this guide. For the AQ side, however, the matching of variables between model output and observations is project specific. See Section 6.6, “Creating a New AQ Project,” for more details on changing the default mapping of AQ observations to model variables.

#### 3.1 Perl Configuration File (amet-config.pl)

The Perl configuration file is used by the underlying Perl programs to populate the database with the data pairs created by AMET’s matching of model output for particular locations to corresponding observed values (these data pairs are referred to in this guide as “model-obs pairs”). Most users will need to modify only a few lines of this configuration file. The most common

variables to change are shown in Table 3-1; a full description can be found in Appendix A, Section A.1.

**Table 3-1. Most common variables that need to be changed in `amet-config.pl`.**

Variable	Description
<code>\$mysql_server</code>	The MySQL server location. Examples are <b>localhost</b> if MySQL is installed on the same machine on which you have installed AMET, or <b>rama.cempd.unc.edu</b> if you have installed the MySQL server on a remote host called <b>rama</b> .
<code>\$root_pass</code>	Password for <b>ametsecure</b> , or your <code>\$root_login</code> (if you changed it from " <b>ametsecure</b> "). This MySQL user will be created later when you are working through Section 5. To provide additional security, AMET is shipped with permissions that allow this file to be read only by the user.
<code>\$perl_lib</code>	Location of additional Perl libraries. In our testing, we added additional libraries through the Comprehensive Perl Archive Network (CPAN) to <b>/usr/local/lib/perl</b> . Below that is the version number directory, and below that are the additional modules.
<code>\$R_lib</code>	All of the R library paths that should be searched by the R scripts. In our installation, they are <b>/usr/local/pkg/R/packages:\$R_dir/lib/R/site-library</b> , since we installed additional R libraries under <b>/usr/local/pkg/R/packages</b> .
<code>\$fslftp_madis</code> <sup>1</sup>	MADIS ftp site (MADIS is discussed in Section 4.2).
<code>\$login_madis</code> <sup>1</sup>	MADIS-provided login.
<code>\$pass_madis</code> <sup>1</sup>	MADIS-provided password.
<code>\$login_ncep</code> <sup>1</sup>	NCEP login; most likely " <b>anonymous</b> ".
<code>\$pass_ncep</code> <sup>1</sup>	Anonymous password for the National Centers for Environmental Prediction (NCEP); this is usually your e-mail address. Note the necessary " <b>@</b> ".

<sup>1</sup> Needed only if using the MET side of AMET.

## 3.2 R Configuration File (`amet-config.R`)

The R configuration file is used by the underlying R programs to perform statistical analysis on

your model-obs pairs. Most users will need to modify only a few lines of this configuration file. The most common variables to change are shown in Table 3-2; a full description can be found in Appendix A, Section A.2.

**Table 3-2. Most common variables that need to be changed in `amet-config.R`.**

Variable	Description
<b>server</b>	The MySQL server location. Examples are <b>localhost</b> if MySQL is installed on the same machine on which you have installed AMET, or <b>rama.cempd.unc.edu</b> if you have installed the MySQL server on a remote host called <b>rama</b> .
<b>passwd</b>	Password for <b>ametsecure</b> , or your <b>login</b> (if you changed it from " <b>ametsecure</b> "). This MySQL user will be created later when you are working through Section 5. To provide additional security, AMET is shipped with permissions that allow this file to be read only by the user. <i>Note:</i> The R and Perl configuration files should have matching passwords and logins.
<b>newLibPath</b>	Should include all R library paths where R packages were installed. In our installation, they are: <b>c("/usr/local/pkg/Rpackages", oldLibPath)</b> , since we installed additional R libraries under <b>/usr/local/pkg/Rpackages</b> . <i>Note:</i> The <b>oldLibPath</b> needs to be included in this variable.

## 4. Datasets

The AMET release includes both model and observational datasets provided as examples. You should have downloaded these into the proper directories during the installation process.

### 4.1 Model Data

For the model data, we have included both meteorological and air quality data. We have organized the data into four example projects: "**mm5Example**", "**wrfExample**", "**mcipExample**", and "**aqExample**". All of the spatial domains cover the continental U.S. and have a 12-km grid resolution.

**On the MET side:** There is a 5.5-d MM5 simulation (July 04 2002 12:00 UTC to July 10 2002 00:00 UTC), a 5-d WRF simulation (July 05 2002 00:00 UTC to July 09 2002 23:00 UTC), and a 5-d MCIP post-processing of the MM5 output above (from July 05 2002 00:00 UTC to July 09 2002 23:00 UTC).

The MM5 data file is

```
$AMETBASE/model_data/MET/mm5Example/MMOUT_DOMAIN1_02Jul04
```

Note that we have bolded “**mm5Example**” in the directory name above to highlight the fact that we are using the project name to organize the model output files into directories.

The WRF data consist of five WRF output files in netCDF format for the same domain as above:

```
$AMETBASE/model_data/MET/wrfExample/
wrfout_d01_2002-07-05_00:00:00
wrfout_d01_2002-07-06_00:00:00
wrfout_d01_2002-07-07_00:00:00
wrfout_d01_2002-07-08_00:00:00
wrfout_d01_2002-07-09_00:00:00
```

The MCIP data consist of five METCRO2D files and one GRIDCRO2D file in netCDF format:

```
$AMETBASE/model_data/MET/mcipExample/
GRIDCRO2D_2002187
METCRO2D_2002186
METCRO2D_2002187
METCRO2D_2002188
METCRO2D_2002189
METCRO2D_2002190
```

*On the AQ side:* We have included two CMAQ output tarfiles. One is for the period July 01 2006 0:00 UTC to July 11 2006 00:00 UTC:

```
$AMETBASE/model_data/AQ/aqExample/
aqExample.2006.tar.gz:
  test.12km.AERO6.aconc
  test.12km.AERO6.dep
```

These two files correspond to the concentration and wet deposition output files from CMAQ, after they have been postprocessed with the **combine** utility. To demonstrate the analysis of statistical correlations of meteorological and air quality model data we have also provided an **aconc** and **WETDEP** output file from a 2002 CMAQ simulation with the temporal range from July 1 2002 00:00 UTC to July 11 2002 00:00 UTC and a spatial domain covering the continental U.S at 36-km resolution; these data were included in the previous AMET release.

```
aqExample.2002.tar.gz:
  test.36km.conc
  test.36km.dep
```

## 4.2 Observational Data

As with the model data, the observations directory structure is divided between MET and AQ

fields. On the MET side, all of the observations come from MADIS, provided by the National Oceanic and Atmospheric Administration (NOAA). If you plan to use AMET for MET analysis, you must first contact MADIS to obtain a MADIS account, login, and password (see <http://madis.noaa.gov> for details). You will use this login information in setting up the Perl configuration file (discussed in Section 3.1).

In the AMET structure, all of the MADIS data are stored under **\$AMETBASE/obs/MET**. The file **stations.csv** is a metadata file that describes the station ID and location of each of the MADIS monitoring sites; this file is in comma-separated-value (.csv) format. The AMET distribution from the CMAS Center does not include any MET observations from MADIS. When you run the database population scripts (Section 6), they will automatically call the MADIS system to download the available monitoring sites' observations for the specified time period via ftp. For example, the radiosonde data will be downloaded to

**\$AMETBASE/obs/MET/point/raob/netcdf**

Each of the observation files is a netCDF file representing one hour's worth of data from all available monitoring sites. The netCDF file can be stored in a gzip compressed format to save space, and unzipped automatically as needed. Note that the default settings for the MADIS obs extraction are set in template files in **\$AMETBASE/bin/madis\_input**. You can modify these template files to extract mesonet data if desired. Also, you can adjust the quality control level from its current highest level. See the MADIS documentation available from their web site for more information on how to customize these template files.

On the AQ side, several new network datasets have been added for North America and Europe in AMETv1.2 (also see Appendix B). We have included observational data for the following networks for North America:

- Clean Air Status and Trends Network (CASTNET)
- Interagency Monitoring of Protected Visual Environments (IMPROVE) network
- National Atmospheric Deposition Program (NADP) network
- Southeastern Aerosol Research and Characterization Study (SEARCH) network
- Chemical Speciation Network (CSN) (previously Speciation Trends Network [STN])
- Air Quality System (AQS) network
- Mercury Deposition Network (MDN)
- National Air Pollution Surveillance Program (NAPS) speciated measurements from Canada

We have also included example observational data from the following European networks and databases:

- AirBase (Europe)
- Automatic Urban and Rural Network (AURN [UK])
- European Monitoring and Evaluation Programme (EMEP) network
- Acid Gas and Aerosol Network (AGANET [UK])

- Acid Deposition Monitoring Network (ADMN [UK])
- National Ammonia Monitoring Network (NAMN [UK]).

The observational datasets have been preprocessed and reformatted (in some instances from their original sources) for access by AMET. The temporal range is network-dependent, and ranges from 2001 to 2011. The monitoring station locations are provided by a series of .csv files under the subdirectory **\$AMETBASE/obs/AQ/site\_lists**. A brief synopsis of each network, along with the steps taken to create these data for AMET, is given below. Note that in the species lists below, each line is of the format “observed species name; model species name (units)”. See Appendix A, Tables A-10 and A-11 respectively, for listings of the available networks in AMET for North America and Europe, along with the frequency of observations, and the species reported in each.

#### 4.2.1. Clean Air Status and Trends Network (CASTNET) Weekly

Source: CASTNET data are obtained through the CASTNET web site:

<http://www.epa.gov/castnet/>. Weekly CASTNET data can be obtained by downloading selecting the link “Prepackaged data” and then the “drychem” file on the CASTNET web site. This will download all years of data available data. No postprocessing of the downloaded data is necessary in order for them to be compatible with the Site Compare (**sitcmp**) software packaged with the AMET system.

Species used with AMET:

tso4; ASO4T ( $\mu\text{g}/\text{m}^3$ )  
tno3; ANO3T ( $\mu\text{g}/\text{m}^3$ )  
tnh4; ANH4T ( $\mu\text{g}/\text{m}^3$ )  
tno3+nhno3; ANO3T+HNO3\_UGM3 (TNO3;  $\mu\text{g}/\text{m}^3$ )  
nhno3; HNO3\_UGM3 ( $\mu\text{g}/\text{m}^3$ )  
wso2; SO2\_UGM3 ( $\mu\text{g}/\text{m}^3$ )

#### 4.2.2. Clean Air Status and Trends Network (CASTNET) Hourly

Source: CASTNET data are obtained through the CASTNET web site:

<http://www.epa.gov/castnet/>. Hourly CASTNET ozone data can be downloaded by selecting “Prepackaged data” and then the files labeled “metdata\_YYYY” on the CASTNET web site, as this file contains several meteorological variables in addition to ozone. No postprocessing of the downloaded data is necessary in order for them to be compatible with AMET’s **sitcmp** program.

Species used with AMET:

Ozone; O3 (ppb)

Additional meteorological variables that could be used with AMET:

Temperature; SFC\_TMP ( $^{\circ}\text{C}$ )

relative\_humidity; RH (%)  
solar\_radiation; SOL\_RAD (watts/m<sup>2</sup>)  
precipitation; precip (mm/hr)  
windspeed; WSPD10 (m/s)  
wind\_direction; WDIR10 (deg)

#### 4.2.3. Interagency Monitoring of Protected Visual Environments (IMPROVE)

Source: IMPROVE data are obtained through the IMPROVE web site:

<http://vista.cira.colostate.edu/improve/>. This web site links to the Visibility Information Exchange Web System (VIEWS) web site, which is an interactive system for downloading various air-quality-related data. IMPROVE data obtained through VIEWS do not require any additional processing to work with AMET and the **sitecmp** program.

Species used with AMET:

SO4f\_val; ASO4T ( $\mu\text{g}/\text{m}^3$ )  
NO3f\_val; ANO3T ( $\mu\text{g}/\text{m}^3$ )  
NH4f\_val; ANH4T ( $\mu\text{g}/\text{m}^3$ )  
MF\_val; PM25 ( $\mu\text{g}/\text{m}^3$ )  
OCf\_val; PM\_OC ( $\mu\text{g}/\text{m}^3$ )  
ECf\_val; AECT ( $\mu\text{g}/\text{m}^3$ )  
OCf\_val+ECf\_val; PM\_OC+AECT (TC;  $\mu\text{g}/\text{m}^3$ )

Species used with AMET and CMAQ AERO6:

NAf\_val; Sodium ( $\mu\text{g}/\text{m}^3$ )  
CHLf\_val; Chloride ( $\mu\text{g}/\text{m}^3$ )  
FEf\_val; Iron ( $\mu\text{g}/\text{m}^3$ )  
ALf\_val; Aluminum ( $\mu\text{g}/\text{m}^3$ )  
SIf\_val; Silicon ( $\mu\text{g}/\text{m}^3$ )  
TIf\_val; Titanium ( $\mu\text{g}/\text{m}^3$ )  
CAf\_val; Calcium ( $\mu\text{g}/\text{m}^3$ )  
MGf\_val; Magnesium ( $\mu\text{g}/\text{m}^3$ )  
Kf\_val; Potassium ( $\mu\text{g}/\text{m}^3$ )  
MNf\_val; Manganese ( $\mu\text{g}/\text{m}^3$ )  
 $2.2 * \text{ALf\_val} - 2.49 * \text{SIf\_val} - 1.63 * \text{CAf\_val} - 2.42 * \text{FEf\_val} - 1.94 * \text{TIf\_val}$ ; Soil ( $\mu\text{g}/\text{m}^3$ )

#### 4.2.4. National Atmospheric Deposition Program (NADP)

Source: NADP data are obtained through the NADP/NTN web site:

<http://nadp.sws.uiuc.edu/>. Weekly wet concentration data are downloaded in comma-delimited format directly from the web site. No postprocessing of the downloaded data is necessary in order for them to be used with AMET's **sitcmp** program.

Species used with AMET (from CMAQ deposition file):

Valcode; Valid code (character)  
Invalcode; Invalid code (character)  
NH4; WDEP\_NHX (mg/L or kg/ha)  
NO3; WDEP\_TNO3 (mg/L or kg/ha)  
SO4; WDEP\_ASO4T (mg/L or kg/ha)  
precip; RT (mm)

Species used with AMET and CMAQ AERO6:

Ca; Calcium ion (mg/L or kg/ha)  
Mg; Magnesium ion (mg/L or kg/ha)  
K; Potassium ion (mg/L or kg/ha)

#### **4.2.5. Southeastern Aerosol Research and Characterization (SEARCH) Study**

Source: SEARCH data are obtained through the SEARCH web site:

<http://www.atmospheric-research.com/public/index.html>. The SEARCH data can be downloaded as comma-delimited files for each SEARCH site. In order to be used with **sitcmp** and AMET, the individual site files must first be merged together into a single file. The example SEARCH data file provided with AMET should serve as an example of how the raw SEARCH data need to be combined and formatted in order to work with AMET and the **sitcmp** program.

Species used with AMET:

o3; O3 (ppb)  
co; CO (ppb)  
so2; SO2 (ppb)  
no; NO (ppb)  
no2; NO2 (ppb)  
noy; NOy (ppb)  
hno3; HNO3 (ppb) (available only for some years)  
teom; PM25 ( $\mu\text{g}/\text{m}^3$ )  
ec; EC ( $\mu\text{g}/\text{m}^3$ )  
oc; OC ( $\mu\text{g}/\text{m}^3$ )  
tc; Total Carbon ( $\mu\text{g}/\text{m}^3$ )



no3; ANO3T ( $\mu\text{g}/\text{m}^3$ )

so4; ASO4T ( $\mu\text{g}/\text{m}^3$ )

nh4; ANH4T ( $\mu\text{g}/\text{m}^3$ )

noy; NOY (ppb)

#### 4.2.6. Chemical Speciation Network (CSN) (previously STN)

Source: CSN data are obtained through the EPA's Air Quality System (AQS), located at <http://www.epa.gov/ttn/airs/airsaqs/>, or from the VIEWS web site (<http://views.cira.colostate.edu/web/DataWizard/>), where they are available in the Dataset pull-down menu under the listing "EPA PM2.5 Speciation \_ Daily". The data provided with AMET are a sample of the CSN data that can be obtained through the AQS. No postprocessing of the downloaded CSN data is necessary in order for them to work with AMET and the **sitcmp** program.

##### Species used with AMET:

m\_so4; ASO4T ( $\mu\text{g}/\text{m}^3$ )

m\_no3; ANO3T ( $\mu\text{g}/\text{m}^3$ )

m\_nh4; ANH4T ( $\mu\text{g}/\text{m}^3$ )

FRM PM<sub>2.5</sub> Mass; PM<sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )

oc\_adj; PM\_OC ( $\mu\text{g}/\text{m}^3$ )

ec\_niosh; AECT ( $\mu\text{g}/\text{m}^3$ )

oc\_adj+ec\_niosh; PM\_OC+AECT (TC;  $\mu\text{g}/\text{m}^3$ )

Sodium Ion; Sodium ion ( $\mu\text{g}/\text{m}^3$ )

chlorine; Chlorine ( $\mu\text{g}/\text{m}^3$ )

##### Species used with AMET and CMAQ AERO6:

iron; Iron ( $\mu\text{g}/\text{m}^3$ )

aluminum; Aluminum ( $\mu\text{g}/\text{m}^3$ )

silicon; Silicon ( $\mu\text{g}/\text{m}^3$ )

titanium; Titanium ( $\mu\text{g}/\text{m}^3$ )

calcium; Calcium ( $\mu\text{g}/\text{m}^3$ )

magnesium; Magnesium ( $\mu\text{g}/\text{m}^3$ )

potassium; Potassium ( $\mu\text{g}/\text{m}^3$ )

manganese; Manganese ( $\mu\text{g}/\text{m}^3$ )

2.2\*aluminum – 2.49\*silicon – 1.63\*calcium – 2.42\*iron – 1.94\*titanium; Soil ( $\mu\text{g}/\text{m}^3$ )

[Sodium Ion] + [chlorine]; NaCl ( $\mu\text{g}/\text{m}^3$ )

[FRM PM2.5 Mass] – m\_so4 – m\_no3 – m\_nh4 – oc\_adj – ec\_niosh – [Sodium Ion] – [chlorine] – 2.2\*aluminum – 2.49\*silicon – 1.63\*calcium – 2.42\*iron – 1.94\*titanium;

OTHER ( $\mu\text{g}/\text{m}^3$ )

#### 4.2.7. Air Quality System (AQS)

Source: AQS data are obtained through the EPA's Air Quality System (AQS), located at <http://www.epa.gov/ttn/airs/airsaqs/>. Various species of atmospheric gases are available for download through the AQS; the following is the list of species available in the current AMET package. However, the user can customize AMET to process other gas-phase species besides the ones listed below. .

Species used with AMET:

O<sub>3</sub>; ozone (ppb)  
NO<sub>y</sub>; NO<sub>y</sub> (ppb)  
NO; NO (ppb)  
NO<sub>2</sub>; NO<sub>2</sub> (ppb)  
NO<sub>x</sub>; NO<sub>x</sub> (ppb)  
CO; CO (ppb)  
SO<sub>2</sub>; SO<sub>2</sub> (ppb)  
PM<sub>2.5</sub>; TEOM PM<sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )

#### 4.2.8. Mercury Deposition Network (MDN)

Source: MDN data are obtained through the NADP/MDN network web site: <http://nadp.sws.uiuc.edu/mdn/>. Data are available for download as a comma-delimited file for all sites. No postprocessing of the downloaded data is necessary in order for them to be used with AMET's **sitcomp**.

Species used with AMET (modeled species are from CMAQ deposition file):

HGconc; TWDEP\_HG (ng/L)  
HGdep; TWDEP\_HG ( $\mu\text{g}/\text{m}^2$ )

#### 4.2.9. National Air Pollution Surveillance Program (NAPS)

Source: NAPS data are pollutant measurements available for Canada. Data can be obtained through the NAPS web site: <http://www.ec.gc.ca/rnsps-naps/>. Users who wish to obtain additional NAPS data can explore data resources via the links on this web site. These data are not fully supported by the developers of AMET.

Species used with AMET:

O<sub>3</sub>; ozone (ppb)  
TEOMPM<sub>2.5</sub>; PM<sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )  
TEOMPM<sub>10</sub>; PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )  
CO; CO (ppb)  
NO; NO (ppb)

NO<sub>2</sub>; NO<sub>2</sub> (ppb)

NO<sub>x</sub>; NO + NO<sub>2</sub> (ppb)

SO<sub>2</sub>; SO<sub>2</sub> (ppb)

#### 4.2.10. European Air quality dataBase (AirBase)

Source: Data from the European Environment Agency's Air quality dataBase (AirBase) are provided as an example data set. Users who wish to obtain additional AirBase data can explore data resources available via the AirBase data download web page and links at <http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-7>. These data are not fully supported by the developers of AMET.

Species used with AMET:

PM<sub>10</sub>; PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )

PM<sub>25</sub>; PM<sub>25</sub> ( $\mu\text{g}/\text{m}^3$ )

CO; CO (ppb)

SO<sub>2</sub>; SO<sub>2</sub> (ppb)

O<sub>3</sub>; ozone (ppb)

NO; NO (ppb)

NO<sub>2</sub>; NO<sub>2</sub> (ppb)

NO<sub>x</sub>; NO + NO<sub>2</sub> (ppb)

#### 4.2.11. Automatic Urban and Rural Network (AURN)

Source: AURN data are provided as an example data set. Users who wish to obtain additional AURN data can explore data resources available via the AURN data archive at <http://uk-air.defra.gov.uk/data/>. These data are not fully supported by the developers of AMET.

Species used with AMET:

0.77649\*NHX; NHX ( $\mu\text{gN}/\text{m}^3$ )

0.22590\*NOY; NOY ( $\mu\text{gN}/\text{m}^3$ )

0.33379\*SOX; SOX ( $\mu\text{gS}/\text{m}^3$ )

PM<sub>10</sub>; PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )

PM<sub>25</sub>; PM<sub>25</sub> ( $\mu\text{g}/\text{m}^3$ )

862.07\*CO; CO (ppb)

0.38162\*SO<sub>2</sub>; SO<sub>2</sub> (ppb)

0.500\*O<sub>3</sub>; ozone (ppb)

0.81486\*NO; NO (ppb)

0.53146\*NO<sub>2</sub>; NO<sub>2</sub> (ppb)

#### 4.2.12. European Monitoring and Evaluation Programme (EMEP) network

Source: EMEP data are provided as an example data set. Users who wish to obtain additional EMEP data can explore data resources available in multiple links to emissions and air quality data on the EMEP web page <http://www.emep.int/>. These data are not fully supported by the developers of AMET.

Species used with AMET:

NHX; NHX ( $\mu\text{gN}/\text{m}^3$ )

NOY; NOY ( $\mu\text{gN}/\text{m}^3$ )

SOX; SOX ( $\mu\text{gS}/\text{m}^3$ )

PM10; PM10 ( $\mu\text{g}/\text{m}^3$ )

PM25; PM25 ( $\mu\text{g}/\text{m}^3$ )

CO; CO (ppb)

SO<sub>2</sub>; SO<sub>2</sub> (ppb)

O<sub>3</sub>; ozone (ppb)

NO; NO (ppb)

NO<sub>2</sub>; NO<sub>2</sub> (ppb)

#### 4.2.13. Acid Gas and Aerosol NETwork (AGANet)

Source: AGANET data are provided as an example data set. Users who wish to obtain additional AGANet data can explore data resources available via links provided on the AGANet web page <http://pollutantdeposition.defra.gov.uk/aganet>. These data are not fully supported by the developers of AMET.

Species used with AMET:

0.22590\*NOY; NOY ( $\mu\text{gN}/\text{m}^3$ )

0.33379\*SOX; SOX ( $\mu\text{gS}/\text{m}^3$ )

NA; Sodium ( $\mu\text{g}/\text{m}^3$ )

CL; Chloride ( $\mu\text{g}/\text{m}^3$ )

0.38803\*HNO<sub>3</sub>; Nitric Acid (ppb)

0.38162\*SO<sub>2</sub>; SO<sub>2</sub> (ppb)

0.92404\*HCL; HCL (ppb)

0.53146\*NO<sub>2</sub>; NO<sub>2</sub> (ppb)

#### 4.2.14. Acid Deposition Monitoring Network (ADMN)

Source: ADMN data are provided as an example data set. Users who wish to obtain additional

ADMN data need to explore data resources on their own. These data are not fully supported by the developers of AMET.

Species used with AMET (model data are from WETDEP1 file):

0.016\*SOX\_conc; SOX (mgS/L or kgS/ha)  
0.014\*NOY\_conc; NOY (mgN/L or kgN/ha)  
0.014\*NHX\_conc; NHX (mgN/L or kgN/ha)  
0.001\*NA\_conc; Sodium (mg/L or kg/ha)  
0.001\*CL\_conc; Chloride (mg/L or kg/ha)  
0.015\*NSS\_SOX\_conc; NSS SOX (mg/L or kg/ha)

#### 4.2.15. National Ammonia Monitoring Network (NAMN)

Source: NAMN data are provided as an example data set. Users who wish to obtain additional NAMN data can explore data resources available via links provided at [http://pollutantdeposition.defra.gov.uk/ammonia\\_network](http://pollutantdeposition.defra.gov.uk/ammonia_network). These data are not fully supported by the developers of AMET.

Species used with AMET:

0.77649\*NHX; NHX (ppb)  
1.53554\*NH3; NH3 (ppb)

## 5. Database Setup

The next step in using AMET is to set up the MySQL database. Please refer to the flow diagram in Figure 1 to understand the overall flow of data among the various modules within AMET. The steps given in Section 5 must be completed before you populate the database with your project-specific data in Section 6. This setup process is required only once for a given AMET installation. There are separate setup procedures for the two fields, MET and AQ. If you are using AMET for only one of those fields, you need to run only the corresponding setup. If you are running AMET for both fields, you will need to run both setups. In the following discussion, we assume the default name of the AMET database, “**amet**”, and the default AMET MySQL user, “**ametsecure**”. If you decide to change either of these, then you will need to update the appropriate variables in the Perl and R configuration files in the directory **\$AMETBASE/configure** (see Section 3). Before you run the setup scripts, you will need to know the “**root**” password for the MySQL administrator. Note that this is not the same as the “**ametsecure**” password that will be created using the scripts discussed below.

### 5.1 MET Setup

Go to the setup directory

```
$ cd $AMETBASE/scripts_db/setupMET
```

Here, you will see a series of C-shell scripts (.csh) and an input file (.input). The input file, **sites\_meta.input**, defines the metadata file for the MADIS stations (station ID, location, etc.).

To initialize the MET side of AMET, you will need to edit and run the **initialize\_met\_db.csh** script. Specifically, you should make sure the value of **AMETBASE** corresponds to your system. If you have already run the setup procedure for AQ, you should set the variable “**new\_db**” to “**N**”; otherwise set it to “**Y**” to create a new **amet** database. Run the initialize script

```
$ ./initialize_met_db.csh
```

This will ask you for MySQL’s “**root**” password. It will then set up the AMET database, “**amet**”, and the AMET user, “**ametsecure**”. It will also create two tables in your **amet** database: **stations** (observation sites metadata) and **project\_log** (contains summary information for each MET project).

The setup directory also contains scripts for removing projects and removing the **amet** database. To delete a specific MET project, use

```
$ ./delete_project.csh
```

This script will ask you for the name of the project to delete. **CAUTION:** *This will delete all of the tables in the database corresponding to that project.*

To delete the **amet** database from MySQL, use

```
$ ./delete_db.csh
```

This script will ask you for the MySQL “**root**” password before proceeding. *Use this script with **EXTREME CAUTION** because this will delete all of the data in the database corresponding to **all** of the projects (**both MET and AQ**).*

## 5.2 AQ Setup

Go to the setup directory

```
$ cd $AMETBASE/scripts_db/setupAQ
```

As with the MET setup, you will see a series of C-shell scripts and an input file. The input file, **sites\_meta.input**, defines the metadata files for the various AQ monitoring networks (station ID, location, etc.).

To initialize the AQ side of AMET, you will need to edit and run the **initialize\_aq\_db.csh** script. Specifically, you should make sure the value of

**AMETBASE** corresponds to your system. If you have already run the setup procedure for MET, you should set the variable “**new\_db**” to “**N**”; otherwise set it to “**Y**” to create a new **amet** database. Run the initialize script

```
$ ./initialize_aq_db.csh
```

This will ask you for MySQL’s “**root**” password. It will then set up the AMET database, “**amet**”, and the AMET user, “**ametsecure**”. It will also create three tables in your **amet** database: **site\_metadata** (observation sites metadata), **project\_units** (species-specific units for each project), and **aq\_project\_log** (summary information for each AQ project).

The setup directory also contains scripts for removing projects and removing the **amet** database. To delete a specific AQ project, use

```
$ ./delete_project.csh
```

This script will ask you for the name of the project to delete. **CAUTION:** *This will delete all of the tables in the database corresponding to that project.*

To delete the **amet** database from MySQL, use

```
$ ./delete_db.csh
```

This script will ask you for the MySQL “**root**” password before proceeding. *Use this script with **EXTREME CAUTION** because this will delete all of the data in the database corresponding to **all** of the projects (**both MET and AQ**).*

### 5.3 Basic MySQL Commands

As you begin to go through the **amet** database setup and the process of populating the database for each project, you may want to query the database directly to see your progress. Here are a few commands to help you interact directly with the MySQL **amet** database. For more specifics, see one of the many MySQL books that are available, or look at the documentation that can be accessed from the web site <http://dev.mysql.com/doc>.

To log onto the MySQL server from the command line, type

```
$ mysql -u ametsecure -D amet -p
```

This will give you a MySQL prompt (“**mysql>**”). Note that all MySQL commands are case insensitive, and they must end with a semicolon (“**;**”).

To obtain a list of all the tables in your database, type

```
mysql> show tables;
```

After you have populated all of the example projects (end of Section 6), that command will yield

a table like this:

```
+-----+
| Tables_in_amet |
+-----+
| aqExample       |
| aq_project_log  |
| mm5Example_profiler |
| mm5Example_raob |
| mm5Example_surface |
| mcipExample_profiler |
| mcipExample_raob |
| mcipExample_surface |
| project_log     |
| project_units   |
| site_metadata   |
| stations        |
| wrfExample_profiler |
| wrfExample_raob |
| wrfExample_surface |
+-----+
```

To select every column and row in your **project\_log** table:

```
mysql> select * from project_log;
```

To select the latitude, longitude, and common name columns from the stations metadata table and limit the results to the first 20 rows:

```
mysql> select lat,lon,common_name from stations limit 20;
```

To select all station metadata where the monitor is from the CASTNET network:

```
mysql> select * from site_metadata where network='castnet';
```

To determine which networks are included in the **aqExample** project:

```
mysql> select distinct network from aqExample;
```

## 6. Database Population

The database population phase of AMET must be performed for each new project. As discussed in Section 1.2, the *project* is the organizing structure that we use to group a particular model simulation with the scripts and data used to populate the **amet** tables. If you go to the database populate directory by typing



```
$ cd $AMETBASE/scripts_db
```

you will see four project directories, in addition to the setup directories described earlier. The projects are

- (a) a MET example for the WRF model (**wrfExample**),
- (b) a MET example for the MM5 model (**mm5Example**),
- (c) a MET example for MCIP postprocessed surface data (**mcipExample**), and
- (d) an AQ example for the CMAQ model (**aqExample**).

In the following subsections, we describe how to run each project.

## 6.1 The wrfExample Project

Go to the project directory

```
$ cd $AMETBASE/scripts_db/wrfExample
```

Here, you will see two input files and two C-shell scripts. The **setup\_project.input** file is the input file for initializing the project's database tables (variables in this input file are listed in Appendix A, Table A-3). The only thing you should need to change in this file is the "**\$email**" variable; note that you should use the backslash "escape" character, "\", to prevent Perl from evaluating the "@" in your e-mail address. The **populate\_project.input** file describes specific flags for the model outputs and observations that you want to process. See Appendix A, Table A-5 for information on the specifics relating to each variable. You will likely not need to change anything in this second input file.

The C-shell file **metProject.csh** is a wrapper script for calling the Perl programs that actually populate the AMET database with the project data. You will likely need to verify only that the variable **AMETBASE** has been updated for this project. Run the script by typing

```
$ ./metProject.csh >& log.populate
```

Note that if you are overwriting an existing project (i.e., using the same project name), you cannot redirect the screen output to a file because the script will require input from you.

This C-shell script will create three empty project tables in the AMET database:

**wrfExample\_profiler**, **wrfExample\_raob**, and **wrfExample\_surface**.

These tables correspond to the matches between the model outputs and (1) the wind profiler, aircraft profiles and radar wind (VAD) observations, (2) the radiosonde observations, and (3) the surface observations. After creating these tables, the script then begins the matching process. This consists of auto-ftp-ing data from the MADIS web site for the model's temporal period, unzipping the downloaded data, finding the geographic location of each observation site on the model grid and interpolating to those locations, populating the appropriate table with the model-obs pairs for each variable, and optionally reziping the data for compressed storage. Finally, the script updates the **project\_log** with summary information for the **wrfExample** project.

The second C-shell file, **metFTP.csh**, is a wrapper script for calling Perl programs to download observational data from MADIS for a specific period of time. This allows you to download observational data without having the model output. Make sure that the variable **auto\_ftp** is set to 1 in the **populate\_project.input** file when running this script. Please note that the MADIS data need to be downloaded only once for a given time period and will subsequently be available to all projects.

## 6.2 The mm5Example Project

Go to the project directory:

```
$ cd $AMETBASE/scripts_db/mm5Example
```

Here, you will see two input files and two C-shell scripts. The **setup\_project.input** file is the input file for initializing the project's database tables (variables in this input file are listed in Appendix A, Table A-3). The only thing you should need to change in this file is the "**\$email**" variable; note that you should use the backslash "escape" character, "\", to prevent Perl from evaluating the "@" in your e-mail address. The **populate\_project.input** file describes specific flags for the model outputs and observations that you want to process. See Appendix A, Table A-5 for information on the specifics relating to each variable. You will likely not need to change anything in this second input file. Note that the MM5 model has been run with the Pleim-Xiu (PX) surface model and has the relevant flags for this configuration; see Appendix A for more information.

The C-shell file **metProject.csh** is a wrapper script for calling the Perl programs that actually populate the AMET database with the project data. You will likely need to verify only that the variable **AMETBASE** has been updated for this project. Run the script by typing

```
$ ./metProject.csh >& log.populate
```

Note that if you are overwriting an existing project (i.e., using the same project name), you cannot redirect the screen output to a file because the script will require input from you.

This C-shell script will create three empty project tables in the AMET database:

**mm5Example\_profiler**, **mm5Example\_raob**, and **mm5Example\_surface**. These tables correspond to the matches between the model outputs and (1) the wind profiler observations, (2) the radiosonde observations, and (3) the surface observations. After creating these tables, the script then begins the matching process. This consists of converting the MM5 output file into a netCDF file, auto-ftp-ing data from the MADIS web site for the model's temporal period, unzipping the downloaded data, finding the geographic location of each observation site on the model grid and interpolating to those locations, populating the appropriate table with the model-obs pairs for each variable, and optionally reziping the data for compressed storage. Finally, the script updates the **project\_log** with summary information for the **mm5Example** project.

Note that if your MM5 simulation was configured to use a non-PX surface model, you will need to change the two variables **\$diagnose\_sfc** and **\$match\_config** in the **populate\_project.input** file; see Appendix A, Table A-5 for more information on the specifics relating to each variable.

The second C-shell file, **metFTP.csh**, is a wrapper script for calling Perl programs to download observational data from MADIS for a specific period of time. This allows you to download observational data without having the model output. Make sure that the variable **auto\_ftp** is set to 1 when running this script. Please note that the MADIS data need to be downloaded only once for a given time period and will subsequently be available to all projects.

### 6.3 The mcipExample Project

Go to the project directory

```
$ cd $AMETBASE/scripts_db/mcipExample
```

Here, you will see two input files and two C-shell scripts. The **setup\_project.input** file is the input file for initializing the project's database tables (variables in this input file are listed in Appendix A, Table A-3). The only thing you should need to change in this file is the "**\$email**" variable; note that you should use the backslash "escape" character, "\", to prevent Perl from evaluating the "@" in your e-mail address. The **populate\_project.input** file describes specific flags for the model outputs and observations that you want to process. See Appendix A, Table A-5 for information on the specifics relating to each variable. You will likely not need to change anything in this second input file.

The C-shell file **metProject.csh** is a wrapper script for calling the Perl programs that actually populate the AMET database with the project data. The only variable that you will likely need to change for this project is **AMETBASE**. Also note that the variable **AMET\_MCIP\_GRID** needs to provide the full path to the GRIDCRO2D file used for this project. Run the script by typing

```
$ ./metProject.csh >& log.populate
```

Note that if you are overwriting an existing project (i.e., using the same project name), you cannot redirect the screen output to a file because the script will require input from you.

This C-shell script will create three empty project tables in the AMET database: **mcipExample\_profiler**, **mcipExample\_raob**, and **mcipExample\_surface**. These tables correspond to the matches between the model outputs and (1) the wind profiler observations, (2) the radiosonde observations, and (3) the surface observations. However, AMET currently works only on MCIP surface data, and so the profiler and raob tables will be empty. Further, MCIP METCRO2D files are more limited and do not contain water vapor mixing ratio, and so these are set to missing values in the database. It is recommended that raw WRF netCDF output, rather than MCIP output, be used whenever available. The only reliable evaluation data when using MCIP output are the 2-m temperature and 10-m wind speed. Future releases may

implement the capability to evaluate MCIP output of shortwave radiation and wind direction against measurements.

After creating these tables, the script then begins the matching process. This consists of auto-ftp-ing data from the MADIS web site for the model's temporal period, unzipping the downloaded data, finding the geographic location of each observation site on the model grid and interpolating to those locations, populating the appropriate table with the model-obs pairs for each variable, and optionally reziping the data for compressed storage. Finally, the script updates the **project\_log** with summary information for the **mcipExample** project.

The second C-shell file, **metFTP.csh**, is a wrapper script for calling Perl programs to download observational data from MADIS for a specific period of time. This allows you to download observational data without having the model output. Make sure that the variable **auto\_ftp** is set to 1 when running this script. Please note that the MADIS data need to be downloaded only once for a given time period and will subsequently be available to all projects.

## 6.4 The aqExample Project

Go to the project directory:

```
$ cd $AMETBASE/scripts_db/aqExample
```

Here, you will see one input file, one C-shell script, and the **combine** subdirectory. The **setup\_project.input** file is the input file for initializing the project's database table (variables in this input file are listed in Appendix A, Table A-4). The only thing you should need to change in this file is the "**\$email**" variable; note that you should use the backslash "escape" character, "\", to prevent Perl from evaluating the "@" in your e-mail address.

Note that in this version of AMET the **populate-project.input** file is no longer required. Options that were selected in this file in AMET v1.1 are now incorporated into the **aqProject.csh** script; see Appendix A, Table A-6 for a listing of all the variables that are set in this script. See Appendix B for information on major changes in the AMET v1.2 script relative to AMET v1.1.

The C-shell file **aqProject.csh** is a wrapper script for calling the Perl programs that actually populate the AMET database with the project data. The only variable that you will likely need to change for this example project is **AMETBASE**. Run the script by typing

```
$ ./aqProject.csh >& log.populate
```

Note that if you are overwriting an existing project (i.e., using the same project name), you cannot redirect the screen output to a file because the script will require input from you.

This C-shell script will create one empty project table in the AMET database: **aqExample**. After creating this table, the script then begins the matching process. This consists of calling a series of Perl programs and Fortran helper programs. The two Fortran helper programs are **\$AMETBASE/bin/cmp\_airs.exe** and **\$AMETBASE/bin/sitecmp.exe**; the first one

matches the AQS network's data to the nearest grid cell in the CMAQ model, and the second one does the same for the other networks. After each network has been matched to the model, the **aqExample** table is populated with the model-obs pairs. In addition to creating and populating the **aqExample** table, the script updates the **project\_units** table with each network for that project. This table defines the physical units of the species variables for this network (e.g., ppb vs.  $\mu\text{g}/\text{m}^3$ ). Finally, the script updates the **aq\_project\_log** with summary information for the **aqExample** project.

The third element in the project directory, the **combine** subdirectory, is not used for this example; it is discussed later in Section 6.6, "Creating a New AQ Project."

## 6.5 Creating a New MET Project

When you create your own projects, we recommend that you utilize the structure of naming your directories after your projects. If you choose not to do this, you will have to modify the provided run scripts.

To create a new MET project, follow these basic steps:

1. Copy the appropriate example project to a new directory.
2. Rename it after your new project (use the *exact* project name, as many scripts use the project name to navigate directories).
3. Create a new project directory under **\$AMETBASE/model\_data/MET** for the input model data.
4. Change the appropriate variables in the project C-shell script.
5. Run your new populate script.

For example, if we were creating a new WRF project called "**wrfNC2007**", we would use

```
$ cd $AMETBASE/scripts_db
$ cp -r wrfExample wrfNC2007
$ cd wrfNC2007
```

Create a new model data directory and move or link your model data into it, as follows:

```
$ cd $AMETBASE/model_data/MET
$ mkdir wrfNC2007
$ cd wrfNC2007
$ ln -s <model_data> .
```

Here, you would replace "**<model data>**" with the path to your model data file(s). The **metProject.csh** script will perform the model-obs matching on all model outputs in this new directory.

Next, edit the `$AMETBASE/script_db/wrfNC2007/metProject.csh` variables `AMET_PROJECT` ("wrfNC2007") and `AMET_PROJ_DESC` (your description of the project).

Finally, run the populate script:

```
$ cd $AMETBASE/scripts_db/wrfNC2007
$ ./metProject.csh
```

This will create a new MET project in the `amet` database. Specifically, it will create a new row in your `project_log` table and three new tables: `wrfNC2007_profiler`, `wrfNC2007_raob`, and `wrfNC2007_surface`.

## 6.6 Creating a New AQ Project

Before describing the creation of a new AQ project, we need to clarify a potentially confusing issue: the relationship between model species and monitor species. In order for AQ database population to work, there must be a mapping between the model species and the various network species. This mapping is accomplished by postprocessing the CMAQ model data, and through species definitions in the `AQ_species_list.pl` file (located in the `$AMETBASE/perl` directory). The model data used in the `aqExample` section (Section 6.4) were already postprocessed, so we did not need to go through that step when running the example project. In a new project, you will likely need to postprocess your CMAQ data before they are ingested into the `amet` database. This postprocessing is accomplished in the third step of creating a new AQ project (described below), using the `combine` Fortran program.

Also, when you create your own projects, we recommend that you utilize the structure of naming your directories after your projects. If you choose not to do this, you will have to modify the provided run scripts.

To create a new AQ project, follow these basic steps:

1. Copy the appropriate example project to a new directory.
2. Rename it after your new project (use the *exact* project name, as many scripts use the project name to navigate directories).
3. Postprocess the model data using the `combine` Fortran program. (See notes in Appendix B for changes in `combine` relative to AMETv1.1.)
4. Create a new project directory under `$AMETBASE/model_data/AQ` for the input model data.
5. Change the appropriate variables in the project C-shell script.
6. Run the new populate script.

For example, if we were creating a new AQ project called “`aqNC2007`”, we would use

```
$ cd $AMETBASE/scripts_db
$ cp -r aqExample aqNC2007
```

Next you will need to postprocess the raw CMAQ concentration and wet deposition files to map the data to the appropriate species names. To do this, you will use the `combine` Fortran program. Go to the `combine` directory:

```
$ cd aqNC2007/combine
```

Edit the `combine_conc.csh` and `combine_dep.csh` scripts for your model data. For detailed instructions on `combine`, see [http://www.cmascenter.org/help/model\\_docs/cmaq/4.6/EVALUATION\\_TO\\_OLS.txt](http://www.cmascenter.org/help/model_docs/cmaq/4.6/EVALUATION_TO_OLS.txt). Run the two `combine` scripts:

```
$ ./combine_conc.csh
$ ./combine_dep.csh
```

Next, create a new model data directory and move or link your postprocessed model data into it, as follows:

```
$ cd $AMETBASE/model_data/AQ
$ mkdir aqNC2007
$ cd aqNC2007
$ ln -s <model_data> .
```

Here, you would replace “<model data>” with the path to your postprocessed model data file(s).

Next, in the file `$AMETBASE/scripts_db/aqNC2007/aqProject.csh`, edit the variables `AMET_PROJECT` (“aqNC2007”), `AMET_OUT` (if necessary), `START_DATE` (start date of the model data), `END_DATE` (end date of the model data), `CONC_FILE_1` (model concentration file(s)), and `DEP_FILE_1` (model deposition file(s)). Note that the model data files should point to your postprocessed model data. The functions of the other variables in the script are described in Appendix A (Table A-6). Finally, run the `populate` script:

```
$ cd $AMETBASE/scripts_db/aqNC2007
$ ./aqProject.csh
```

This will create a new AQ project in the `amet` database. Specifically, it will create a new row in your `aq_project_log` table, a series of new rows (one for each network) in your `project_units` table, and a new project table: `aqNC2007`.

## 7. Analysis

The analysis phase of AMET consists of performing statistical analyses on the model-obs pairs

and creating plots of the resulting statistics. The basic process is to query the project's database table(s) using a set of SQL criteria; to perform statistical analyses on the returned data; and to create plots, tables, and text file outputs. The current AMET package contains a series of preprogrammed statistical analysis and plotting routines, based on the R language. These scripts are provided strictly as a starting point and as illustrative examples. Because all the model-obs pairs are stored in a MySQL database, an advanced user can decide to access those data in any desired manner, including other software packages. All that is required is a MySQL interface and some exploration of the table structure. We encourage advanced users to extend these R scripts to create more specific or advanced plotting capabilities, to use other languages to expand AMET analysis capabilities, and to contribute these updates to the CMAS community.

As with the database populate phase, the project is the organizing structure that we use to group a particular model run (specific model, physics or chemistry, spatial domain, scale, etc.) with the scripts used to analyze the `amet` tables and with the output from this analysis (plots and data).

A second organizing structure is the grouping of two files for each type of analysis. In the project directories, you will see a C-shell script and an input file with similar names (e.g., `run_timeseries.csh` and `timeseries.input`). These two files set up everything that is necessary to run an underlying R script in `$AMETBASE/R` and then they run that script. The use of the C-shell interface allows users who are not very familiar with R to perform these predefined analyses, shielding them from the actual R code.

The following sections describe the steps to execute the analysis scripts in the WRF example (Section 7.1) and the air quality example (Section 7.2). Section 7.3 provides the steps to create a new analysis project.

## 7.1 wrfExample

Go to the project directory:

```
$ cd $AMETBASE/scripts_analysis/wrfExample
```

Here, you will see a series of C-shell scripts and their accompanying input files. We will go through one of the analysis scripts in detail as an example for running each of the scripts in the project.

The `run_spatial_surface.csh` script creates a series of maps comparing the surface monitors to the model for one or more days. Each plot provides color-coded model performance metrics at each of the monitor locations. In each map, it plots the particular value at the monitor location.

First, edit the `run_spatial_surface.csh` file, by changing the `AMETBASE` variable to correspond with your setup. The corresponding input file, `spatial_surface.input`, will likely not need to be changed.

Run the script:



```
$ ./run_spatial_surface.csh
```

This will print out the location of the plots. In addition to the script outputs, a detailed log file is produced and located in the directory `$AMETBASE/scripts_analysis/wrfExample`. After the script has completed, go to the output directory and view your maps:

```
$ cd $AMETBASE/output/wrfExample/spatial_surface
```

You should see a whole series of plots of the form:

```
wrfExample.<stats>.<variable>.20020705.20020709.pdf
```

A brief summary of each of the C-shell scripts used on the MET side is given in Table 7-1. A partial list of the input variables for the analyses is provided in Appendix A, Table A-7.

**Table 7-1. MET analysis scripts.**

C-shell Script	Input	Description
<code>run_daily_barplot.csh</code>	<code>daily_barplot.input</code>	Creates a series of daily barplots for all sites within your domain.
<code>run_met_aq.csh</code>	<code>met_aq_coupled.input</code>	Compares AQ and MET data. Creates a spatial plot of correlation and a time series plot of bias. <i>Note:</i> You need to setup both MET and AQ models and define the MET variables and the AQ species and network of interest.
<code>run_plot_srad.csh</code>	<code>plot_srad.input</code>	Generates three analyses of model shortwave radiation compared to SurfRAD observations: time series, diurnal and spatial statistic.
<code>run_plot_prof.csh</code>	<code>plot_prof.input</code>	Creates time-height maps of wind direction and speed profiles statistics as well as hourly profiles, boxplots of metrics and a RMSE histogram. Also can create spatial plot of site locations and names to help identify sites in the domain.
<code>run_plot_raob.csh</code>	<code>plot_raob.input</code>	Creates plots of rawinsonde sites. It creates a series of maps of all sites and vertical profile for user-specified site.
<code>run_spatial_surface.csh</code>	<code>spatial_surface.input</code>	Creates a series of maps comparing surface monitor and model values for one or more days.

C-shell Script	Input	Description
<code>run_summary.csh</code>	<code>summary.input</code>	Creates a summary plot of a series of MET variables. This includes a time series plot and an “ametplot,” which includes a scatterplot, time series, boxplot, and for wind direction, a histogram and directional scatterplot. Note that you may want to change the variable <b>AMET_RUNID</b> to change the file naming convention. Also, the <b>AMET_CRITERIA</b> is an additional SQL command that further limits the data used. This can be customized by the user.
<code>run_timeseries.csh</code>	<code>timeseries.input</code>	Creates a time series plot for one or more monitoring stations. Creates a time series plot for temperature, mixing ratio, wind speed and direction. You control which monitors are plotted through the <b>AMET_SITEID</b> variable. You can also control date ranges and provide additional SQL criteria to further limit the data. <i>Note:</i> Multiple sites can either be averaged or used to create separate plots.
<code>run_wind_prof.csh</code>	<code>wind_prof.input</code>	<u>Creates wind arrow profiles for a set of specific monitoring sites.</u>

Under the **\$AMETBASE/scripts\_analysis** directory, you will notice that there is only a **wrfExample** project on the MET side. We did not include the **mm5Example** project nor the **mcipExample** project because the different inputs are basically treated the same as in the **wrfExample** project. You could easily add a new project directory for **mm5Example** or **mcipExample** by following the instructions in Section 7.3.

## 7.2 aqExample

Go to the project directory:

```
$ cd $AMETBASE/scripts_analysis/aqExample
```

Here, you will see a series of C-shell scripts and their accompanying input files. We will go through one of the analysis scripts in detail as an example for running each of the scripts in the project.

The **run\_scatterplot.csh** script creates a scatterplot for one species from one or more monitoring networks. It compares the observed values to the corresponding model values. The AQ side has an added complication: not all monitoring networks monitor all species. Therefore, the user needs to know which network(s) has the species of interest; if that species is not

available in the network(s) specified, the analysis scripts will likely fail. See Section 4.2 and Appendix A, Table A-10 for more details on the various species that are monitored by (or available from) each AQ network.

First, edit the **run\_scatterplot.csh** file, by changing the **AMETBASE** and **AMET\_DATABASE** variables to correspond with your setup. Note that the species selected is SO<sub>4</sub> and you are plotting two networks: IMPROVE and CASTNET. The corresponding input file, **scatterplot.input**, will likely not need to be changed.

Also note that all AQ analysis scripts make use of the **Network.input** input file. This file contains information about each observational network available to the project that is needed by the R scripts. More information about this file can be found in Appendix A, Table A-8.

Run the script:

```
$ ./run_scatterplot.csh
```

This will print out the location of the plots. In addition to the script outputs, a detailed log file is produced and located in the directory **\$AMETBASE/scripts\_analysis/aqExample**. After the script has completed, go to the output directory and view your plots:

```
$ cd $AMETBASE/output/aqExample/scatterplot
```

You should see files of the form:

```
aqExample_SO4_scatterplot.pdf
```

A brief summary of each of the C-shell scripts used on the AQ side is given in Table 7-2. Several modifications have been made to these scripts relative to AMET v1.1. See Appendix A, Table A-9, for a partial list of the input variables and their descriptions. Appendix B details the updates in the AMET v1.2 analyses relative to AMET v1.1.

**Table 7-2. AQ analysis scripts.**

C-shell Script	Input	Description	Scope
<b>run_boxplot.csh</b>	<b>boxplot.input</b>	Creates a box plot of model-obs quartiles.	single network; single species; multiple simulations
<b>run_boxplot_DofW.csh</b>	<b>boxplot_dofw.input</b>	Similar to the <b>run_boxplot.csh</b> , but instead creates a box plot of model-ob quartiles separated by the day of the week.	single network; single species; multiple simulations

C-shell Script	Input	Description	Scope
<code>run_boxplot_hourly.csh</code>	<code>boxplot_hourly.input</code>	Creates side-by-side boxplots to create a diurnal average curve. Hourly data only.	single network; hourly species only; single simulation
<code>run_boxplot_roselle.csh</code>	<code>Boxplot_roselle.input</code>	Designed specifically for analyzing multiple simulations. Creates a single box plot for the observations and each model simulation. Summary statistics are also provided on the plot for each simulation.	single network; single species; multiple simulations
<code>run_bugleplot.csh</code>	<code>bugleplot.input</code>	Model performance criteria are adjusted as a function of the average concentration of the observed value for that species. As the average concentration of the species decreases, the acceptable performance criteria increase. Creates a bias and error plot.	multiple networks; single species; single simulation
<code>run_historgram.csh</code>	<code>histrogram.input</code>	Creates two histograms, one of the model and observed data, and the other of the model-to-ob bias.	single network; single species; multiple networks
<code>run_met_aq.csh</code>	<code>met_aq_coupled.input</code>	Compares AQ and MET data. Creates a spatial plot of correlation and a time series plot of bias. Note that you need to setup both MET and AQ models and define the MET variables and the AQ species and network of interest.	AQS; O <sub>3</sub> ; single MET/AQ simulation pair

C-shell Script	Input	Description	Scope
<code>run_monthly_stat_plot.csh</code>	<code>monthly_stat_plot.input</code>	Essentially creates three time series plots, one with the modeled and observed values, one with the NMB, NME and Correlation, and one with the MdnB, MdnE and RMSE. Designed to work with one month to multiple years of data.	single network; single species; single simulation
<code>run_pave_overlay.csh</code>	<code>pave_overlay.input</code>	Creates a data file that can be used by the program <b>bldoverlay</b> to create an overlay file. This file can be used in PAVE to overlay over CMAQ model output. Hourly data only.	single network; hourly species only; single simulation
<code>run_plot_spatial.csh</code>	<code>plot_spatial.input</code>	Plots the observed value, model value, and difference between the model and obs for each site. Multiple values for a site are averaged to a single value for plotting purposes.	multiple networks; single species; single simulation
<code>run_plot_spatial_diff.csh</code>	<code>plot_spatial_diff.input</code>	Designed specifically to compare two model simulations, so two simulations are required. Provides plots of the maximum, minimum and average difference in error and bias between two simulations. Useful for seeing how the bias and error have changed between two simulations.	multiple networks; single species; multiple (two) simulations

C-shell Script	Input	Description	Scope
<code>run_plot_spatial_mtom.csh</code>	<code>plot_spatial_mtom.input</code>	Designed specifically to compare two model simulations, so two simulations are required. Provides plots of the maximum, minimum and average difference of two simulations at the network observation sites. Unlike the <code>plot_spatial_diff</code> script, no observation data are used.	multiple networks; single species; multiple (two) simulations
<code>run_scatterplot.csh</code>	<code>scatterplot.input</code>	Creates a single model vs. obs scatterplot. This script will plot a single species from up to three networks on a single plot. Summary statistics are also included on the plot.	multiple networks; single species; multiple simulations
<code>run_scatterplot_bins.csh</code>	<code>scatterplot_bins.input</code>	Creates a “scatter plot” of model mean bias and RMSE binned by the observed concentration. Points are plotted and interquartile ranges with the mean and median values marked.	single network; single species; multiple simulations
<code>run_scatterplot_density.csh</code>	<code>scatterplot_density.input</code>	Creates a scatter plot that codes the density of the points by different colors. Useful when plotting a large number of points that have a significant amount of overlap.	single network; single species; single simulation
<code>run_scatterplot_MtoM.csh</code>	<code>scatterplot_MtoM.input</code>	Creates a single model-to-model scatterplot. <i>Note:</i> The model points correspond to network's site locations only.	multiple networks; single species; multiple simulations

C-shell Script	Input	Description	Scope
<code>run_scatterplot_multi.csh</code>	<code>Scatterplot_multi.input</code>	Creates a scatter plot with multiple simulations that uses separate colors for each simulation. Summary statistics are provided for each simulation. Unlike the basic scatter plot, the multi scatter plot can accept more than two simulations, but only one network.	single network; single species; multiple simulations
<code>run_scatterplot_percentiles.csh</code>	<code>Scatterplot_percentiles.input</code>	Creates a scatter plot of the 5%, 25%, 50%, 75% and 95% percentiles of the data, color coded by percentile.	single network; single species; single simulation
<code>run_scatterplot_single.csh</code>	<code>scatterplot_single.input</code>	Creates a scatter plot for a single network that includes more specific statistics than <code>run_scatterplot.csh</code> .	single network; single species; multiple simulations
<code>run_scatterplot_skill.csh</code>	<code>scatterplot_skill.input</code>	Creates a forecast skill scatter plot. The script is designed to work specifically with O <sub>3</sub> .	all AQS networks; O <sub>3</sub> ; single simulation
<code>run_soccerplot.csh</code>	<code>soccerplot.input</code>	Creates a soccerplot for one or more species over one or more networks. Criteria and goal lines are plotted in such a way as to form a “soccer goal” on the plot area. Two statistics are then plotted: Bias [ <b>NMB</b> (normalized mean), <b>FB</b> (fractional), or <b>NMdnB</b> (normalized median)] on the x-axis and Error [ <b>NME</b> (normalized mean), <b>FE</b> (fractional), or <b>NMdnE</b> (normalized median)] on the y-axis. The better the performance of the model, the closer the plotted points will fall within the “goal” lines.	multiple network; multiple species; multiple simulations

C-shell Script	Input	Description	Scope
<code>run_stacked_barplot.csh</code>	<code>stacked_barplot.input</code>	Data are averaged (mean or median) for SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , EC, OC, and PM <sub>2.5</sub> for the model and observed values. Averages are then plotted on a stacked bar plot, along with the percent of the total PM <sub>2.5</sub> that each species constitutes.	STN or SEARCH; species predefined; multiple simulations
<code>run_stacked_barplot_AE6.csh</code>	<code>stacked_barplot_AE6.input</code>	Designed specifically for CMAQ AE6 which includes the trace metal species. Data are averaged (mean or median) for SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , EC, OC, Soil, NCOM, Other and PM <sub>2.5</sub> for the model and observed values. Averages are then plotted on a stacked bar plot, along with the percent of the total PM <sub>2.5</sub> that each species constitutes.	CSN; species predefined; multiple simulations
<code>run_stacked_barplot_soil.csh</code>	<code>stacked_barplot_soil.input</code>	Designed specifically for CMAQ AE6 which includes the soil species. Data are averaged (mean or median) for Fe, Si, Ca, Ti, Mg, Mn, K and Al for the model and observed values from CSN or IMPROVE. Averages are then plotted on a stacked bar plot, along with the percent of the total soil concentration that each species constitutes.	CSN or IMPROVE; species predefined; one or two model simulations
<code>run_stacked_barplot_soil_multi.csh</code>	<code>stacked_barplot_soil_multi.input</code>	Designed specifically for CMAQ AE6 which includes the soil species. Data are averaged (mean or median) for Fe, Si, Ca, Ti, Mg, Mn, K and Al for the model and observed values from multiple networks. Averages are then plotted on a stacked bar plot, along with the percent of the total soil concentration that each species constitutes.	CSN and IMPROVE; species predefined; single simulation.



C-shell Script	Input	Description	Scope
<code>run_stats_plots.csh</code>	<code>stats_plots.input</code>	Generates a series of spatial plots of <b>NMB</b> , <b>NME</b> , <b>FB</b> , <b>FE</b> , and <b>Correlation</b> . CSV files with additional domain- and site-specific statistics are also included.	multiple networks; single species; single simulation
<code>run_timeseries.csh</code>	<code>timeseries.input</code>	Creates a time series plot. With multiple sites; the sites are time averaged to create a single plot. Also plots the bias and error between the obs and model.	single network; single species; multiple simulations
<code>run_timeseries_multi.csh</code>	<code>timeseries_multi.input</code>	Creates a time series plot for multiple networks. With multiple sites; the sites are time averaged to create a single plot. Also plots the bias and error between the obs and model.	multiple networks; single species; single simulation

### 7.3 Creating a New Analysis Project

Creating a new analysis project requires the same basic steps for both the MET and the AQ models. When you create your own analysis projects, we recommend that you utilize the structure of naming your directories after your projects (as discussed in Sections 4.1, 6.5, and 6.6). To run analyses, you must already have populated the database with the new project (see Sections 6.5 and 6.6). To create a new analysis project, follow these basic steps:

1. Copy the appropriate example project to a new directory.
2. Rename it after your new project (use the *exact* project name, as many scripts use the project name to navigate directories).
3. Change the appropriate variables in the project C-shell scripts.
4. Run the new analysis scripts.

For example, if we were creating a new WRF project called “**wrfNC2007**”, we would use

```
$ cd $AMETBASE/scripts_analysis
$ cp -r wrfExample wrfNC2007
$ cd wrfNC2007
```

In each of the C-shell scripts you want to run, make sure to change the **AMET\_PROJECT** to **wrfNC2007**. You will also likely change dates and custom titles in many of the scripts.

Run the desired analysis scripts from your new project directory. Several new capabilities have been added to the analyses. Appendix B details the updates in the AMET v1.2 analyses relative to AMET v1.1.

## 8. CMAS Support for AMET

We have added AMET to Bugzilla, the CMAS bug-tracking and support request software system. You are encouraged to contact CMAS via Bugzilla if you have bugs to report, or if you would like assistance with a specific component of AMET. The Bugzilla site for AMET is [http://bugz.unc.edu/enter\\_bug.cgi?product=AMET](http://bugz.unc.edu/enter_bug.cgi?product=AMET). If you have never accessed this site before, you will need to request a user account by from the CMAS administrator at [https://www.cmascenter.org/register/create\\_account.cfm](https://www.cmascenter.org/register/create_account.cfm). We have created the following subsections on the Bugzilla AMET page:

- AQ Analyses
- AQ Database
- Installation
- Met Analyses
- Met Database
- Other

## Reference

Appel, K.W., Gilliam, R.C., Davis, N., Zubrow, A., and Howard, S.C.: Overview of the Atmospheric Model Evaluation Tool (AMET) v1.1 for evaluating meteorological and air quality models, *Environ. Modell. Softw.*, 26, 4, 434-443, 2011.

## **Appendix A:**

### **Configuration and Input Files**

## A.1 Perl Configuration File

This is the configuration file for all Perl scripts used in database population—for example, `$AMETBASE/configure/amet-config.pl`.

**Table A-1. amet-config.pl**

Variable	Description
<code>\$amet_base</code>	Top of AMET directory tree. It is recommended that you keep this as the value of the environmental variable <code>\$AMETBASE</code> . If you use something else, the provided scripts will need significant modification.
<code>\$mysql_server</code>	MySQL server location. Examples are “localhost” for the same machine as AMET, or “rama.cempd.unc.edu” for a server on a remote host called <b>rama</b> .
<code>\$root_login</code>	MySQL user for adding data to the database and performing queries. “ <b>ametsecure</b> ” is recommended. This user will be created in the database setup. <i>Note:</i> To increase system security, users may want to restrict this configuration file to read and write only by user.
<code>\$root_pass</code>	Password for “ <b>ametsecure</b> ”, or your <code>\$root_login</code> (if changed from “ <b>ametsecure</b> ”). This user will be created in the database setup.
<code>\$dbase</code>	The name of the MySQL database created for all AMET projects. This variable is set in the AMET run scripts, allowing users to easily switch between different databases.
<code>\$obs_data_dir</code>	Full path to AQ observation data.
<code>\$perl_lib</code>	Location of additional Perl libraries. In our example, we added additional libraries through CPAN to “ <b>/usr/local/lib/perl</b> ”. Below <code>\$perl_lib</code> is the version number directory and below that are the additional modules.
<code>\$R_dir</code>	Base directory for R installation.
<code>\$R_exe</code>	Full path to R executable.
<code>\$R_lib</code>	All of the R library paths that should be searched. In our example, “ <b>/usr/local/pkg/R/packages:\$R_dir/lib/R/site-library</b> ”, we installed additional R libraries under <b>/usr/local/pkg/R/packages</b> .
<code>\$EXEC_aqs</code>	Full path to <b>cmp_airs</b> executable.

Variable	Description
<code>\$EXEC_cast</code>	Full path to <b>cmp_castnet</b> executable.
<code>\$EXEC_all</code>	Full path to <b>sitecmp</b> executable.
<code>\$fslftp_madis</code> <sup>1</sup>	MADIS ftp site.
<code>\$login_madis</code> <sup>1</sup>	MADIS-provided login.
<code>\$pass_madis</code> <sup>1</sup>	MADIS-provided password.
<code>\$fslftp_ncep</code> <sup>1</sup>	NCEP ftp site).
<code>\$login_ncep</code> <sup>1</sup>	NCEP login; most likely “ <b>anonymous</b> ”.
<code>\$pass_ncep</code> <sup>1</sup>	Anonymous password for NCEP, usually your email address. Note the necessary “\@”.
<code>\$amet_verbose</code>	Verbose <b>stdout</b> . Set to “ <b>yes</b> ” for more verbose, “ <b>no</b> ” for less verbose.

<sup>1</sup> Needed only if using the MET side of AMET

## A.2 R Configuration File

This is the configuration file for all R-based analysis scripts. Note that the values here should likely match the equivalent values in the Perl configuration file—for example, `$AMETBASE/configure/amet-config.R`.

Table A-2. `amet-config.R`

Variable	Description
<b>server</b>	MySQL server location. Examples are “ <b>localhost</b> ” if installed on the same machine as AMET, or “ <b>rama.cempd.unc.edu</b> ” for a server on a remote host called <b>rama</b> .
<b>login</b>	MySQL user for adding data to the database and performing queries. “ <b>ametsecure</b> ” is recommended. This user will be created in the database setup. <i>Note:</i> To increase system security, users may want to restrict this configuration file to read and write only by user.

Variable	Description
<b>passwd</b>	Password for “ <b>ametsecure</b> ”, or your login (if changed from “ <b>ametsecure</b> ”). This user will be created in the database setup. To provide additional security, AMET is shipped with permissions that allow this file to be read only by the user. <i>Note:</i> The R and Perl configuration files should have matching passwords and logins.
<b>dbase</b>	The name of the MySQL database created for all AMET projects.. This variable is set in the AMET run scripts, allowing users to easily switch between different databases.
<b>maxrec</b>	Maximum number of records to extract from the database for any one query. No maximum = <b>-1</b> .
<b>newLibPath</b>	Should include all R library paths where R packages were installed. In our installation, they are: <code>c (“/usr/local/pkg/Rpackages”,oldLibPath)</code> , since we installed additional R libraries under <code>/usr/local/pkg/Rpackages</code> . Note that the <code>oldLibPath</code> needs to be included in this variable.

### A.3 MET Project Setup Input File

This is the MET input file for all new projects, and sets up an empty project—for example, `$AMETBASE/scripts_db/wrfExample/setup_project.input`.

Table A-3. MET `setup_project.input`

Variable	Description
<b>\$run_id</b>	Project name. Must be unique across both MET and AQ project names.
<b>\$model</b>	MET model. Examples: “ <b>mm5</b> ”, “ <b>wrf</b> ”.
<b>\$login</b>	Your username. Identifier used to populate <b>user</b> column in <b>project_log</b> table.
<b>\$email</b>	Your email. Note: use the backslash character “\” to escape the “@” character for Perl.
<b>\$description</b>	Brief project description.

## A.4 AQ Project Setup Input File

This is the AQ input file for all new projects, and sets up an empty project—for example, `$AMETBASE/scripts_db/aqExample/setup_project.input`.

Table A-4. AQ `setup_project.input`

Variable	Description
<code>\$run_id</code>	Project name. Must be unique across both MET and AQ project names.
<code>\$model</code>	AQ model. Example: “ <b>CMAQ</b> ”.
<code>\$login</code>	Your username. Identifier used to populate <b>user</b> column in <b>aq_project_log</b> table.
<code>\$email</code>	Your email. Note: use the backslash character “\” to escape the “@” character for Perl.
<code>\$description</code>	Brief project description.
<code>\$delete_id</code>	Delete project if project already exists: “ <b>y</b> ” – yes, delete project, with prompt “ <b>n</b> ” – no, do not delete project
<code>\$exit_exists</code>	Returned exit status of <b>AQ_create_project.pl</b> if project already exists: 1 – return error 0 – return success

## A.5 MET Project Populate Input Files

This is the MET input file for populating new projects—for example, `$AMETBASE/scripts_db/wrfExample/populate_project.input`.

Table A-5. MET `populate_project.input`

Variable	Description
<code>\$pid</code>	Project name. Must be unique across both MET and AQ project names.
<code>\$model</code>	MET model. Examples: “ <b>mm5</b> ”, “ <b>wrf</b> ”, “ <b>mcip</b> ”, or “ <b>eta</b> ”.

Variable	Description
<code>\$obs_format</code>	Observation format: “ <b>madis</b> ” – MADIS observations “ <b>td1</b> ” – preprocessed nonstandard obs
<code>\$model_data_dir</code>	Directory of model output.
<code>\$mcip_grid_file</code>	Path to MCIP <b>GRIDCRO2D</b> file.
<code>\$obs_dir</code>	Top of observation directory. For MADIS, this is the equivalent of <b>\$TOP</b> and <b>\$MADIS_DATA</b> . For typical setup, this will be <b>\$AMETBASE/obs/MET</b> .
<code>\$tmp_dir</code>	Output directory for temporary MADIS files and MM5 netCDF files.
<code>\$interp_method</code>	Interpolation technique from model grid to observation location: <b>0</b> – bilinear <b>1</b> – nearest neighbor
<code>\$eval_class</code>	Observation types to include: “ <b>surface</b> ” – surface observations only “ <b>surfrad</b> ” – SurfRAD radiation only “ <b>profiler</b> ” – wind profiler only “ <b>raob</b> ” – rawinsonde only “ <b>acars</b> ” – aircraft only (MADIS acars profile data) “ <b>vadwind</b> ” – Doppler VAD wind (Not fully supported as VAD obs need conversion from Prepbufr to MADIS-like profiler format) “ <b>all</b> ” – all standard observation types (surface, profiler and raob)
<code>\$diagnose_sfc</code>	Diagnose surface model: <b>1</b> – MM5 with PX surface model <b>0</b> – WRF or MM5 without PX surface model Note that if you use <b>0</b> for MM5, you need to change <b>\$match_config</b> to non-PX configuration.
<code>\$output_int</code>	Model output time step in hours.
<code>\$eval_int</code>	Observations time step in hours.



Variable	Description
<b>\$forecast</b>	Model is a forecast, so observations may not be available for all hours: 1 – yes (forecast hour is set to time index in model output × output interval) 0 – no (forecast hour is set to zero)
<b>\$auto_ftp</b>	Automatically ftp observations from MADIS: 1 – yes 0 – no
<b>\$auto_unzip</b>	Automatically unzip MADIS data stored in gzip format: 1 – yes 0 – no
<b>\$auto_zip</b>	Automatically zip MADIS data in gzip format for storage: 1 – yes 0 – no
<b>\$time_index_start</b>	Starting index of model output. This can be used to skip a spin-up period. The starting point is index + 1. Example values: -1 – start at 0th hour, i.e., include all time steps 0 – skip first time step, start at hour 1 11 – start at hour 12
<b>\$process_npa</b>	Match model precip with National Precipitation Analysis (not supported at this time): 1 – yes 0 – no
<b>\$real_time</b>	Real-time run, use current system date as time range in scripts: 1 – yes 0 – no

Variable	Description
<code>\$file_date</code>	For MM5 data, model data need to be in netCDF format or converted to netCDF. If they have been converted, they need to use the naming convention <code>mmout.YYYY.MM.DD-NHRS</code> . For example, <code>mmout.2001.04.01.00-48</code> is a model run starting at 4/1/2001 00 UTC and running for 48 hours. MM5 has already been converted to netCDF:  1 – yes 0 – no
<code>\$rm_output_file</code>	Remove model output file after matching process:  1 – yes 0 – no
<code>\$match_config</code>	Full path to configuration file for matching MADIS variables to model variables. Note that if you are using MM5 with a non-PX surface model, you will have to use <code>MET_matching_nonPX.conf</code> and change the <code>\$diagnose_sfc</code> value to 0. All MCIP output should use the MCIP configuration file.

## A.6 AQ Project Script Options

This is the AQ populate database script for populating new projects—for example, `$AMETBASE/scripts_db/aqExample/aqProject.csh`. There have been several changes to the script to incorporate the options that were previously selected via the `populate_project.input` file, which is no longer required in the AQ side of AMET.

Table A-6. AQ `aqProject.csh`

Variable	Description
<code>AMETBASE</code>	Base directory where AMET is installed.
<code>AMET_PROJECT</code>	Unique project name for AQ and MET projects.
<code>AMET_DATABASE</code>	MySQL database to use (can simply use default)
<code>AMET_OBS</code>	Directory with observation data and site files

Variable	Description
<b>AMET_OUT</b>	Directory where output will be written
<b>WRITE_SITEX</b>	Option to write the <b>sitcmp</b> scripts or not. Leave as T unless you have already created the <b>sitcmp</b> files and do not want to write new ones.
<b>RUN_SITEX</b>	Option to run (or not run) the <b>sitcmp</b> scripts. Leave as T unless you have already run the <b>sitcmp</b> scripts and do not want to run them again (and create new output files).
<b>LOAD_SITEX</b>	Option to load the output from the <b>sitcmp</b> scripts into the database. Leave as T unless you do not want to load the data into the database.
<b>AMET_SPECIES_FILE</b>	Flag to include the new trace metal (soil) species that are included in the CMAQv5.0 AERO6 module. Setting this option to T will process additional trace metal and soil species for several of the networks. By default, this option is set to F.
<b>INC_AERO6_SPECIES</b>	Flag to include the new trace metal (soil) species that are included in the CMAQv5.0 AERO6 module. Setting this option to T will process additional trace metal and soil species for several of the networks. By default, this option is set to F.
<b>INC_CUTOFF</b>	Flag to process additional model species using a sharp 2.5 $\mu\text{m}$ cutoff diameter for several of the network PM species. This requires you to have included this species in the combine file. If they are not available in your combine file, leave this option as F. By default, this option is set to F.
<b>TIME_SHIFT</b>	This option allows for temporal shifting in the <b>sitcmp</b> matching script. This is typically used if you are using ACONC files and have time-shifted those files to match the input MCIP files. By default this option is set to 0, which indicates no time shifting will take place. Unless you have time-shifted one of the CMAQ output files (typically ACONC files), leave this option set to 0.
<b>START_DATE</b>	Start date of model-observation matching.
<b>END_DATE</b>	End date of model-observation matching.

Variable	Description
<b>CONC_FILE_1</b>	List of model output file(s). Up to 10 model concentration files, numbered sequentially, starting at 1
<b>DEP_FILE_1</b>	List of model output file(s). Up to 10 model wet deposition files, numbered sequentially, starting at 1.
<b>CASTNET</b>	Example of monitoring network flag (CASTNET). Each network can be either included or excluded: <b>"T"</b> / <b>"t"</b> / <b>"Y"</b> / <b>"y"</b> – include monitoring network's observations <b>"F"</b> / <b>"f"</b> / <b>"N"</b> / <b>"n"</b> – exclude monitoring network's observations
<b>O3_UNITS</b>	Concentration units for model O <sub>3</sub> : <b>"ppb"</b> or <b>"ppm"</b>
<b>PRECIP_UNITS</b>	Precipitation units in model: <b>"cm"</b> or <b>"mm"</b>
<b>\$user_def_spec</b>	Additional or different definition for chemical species: <b>"y"</b> – user is providing new definitions <b>"n"</b> – user is not providing new definitions

## A.7 MET Analysis Input Files

The MET analysis input files are found in \$AMET/**scripts\_analysis/wrfExample**. The following is a partial list of variables. Not all of these variables are available in every input file.

Table A-7. MET analysis input variables

Variable	Description
<b>acars.only</b>	Either plot only ACARS airport sites on site location plot, or include all profile stations: <b>TRUE</b> – only ACARS sites <b>FALSE</b> – All sites
<b>ametp</b>	Flag to generate figures.
<b>aq_database</b>	Name of air quality database. This is typically the same as the meteorological database (e.g., " <b>amet</b> ").

Variable	Description
<b>aq_network</b>	Air quality network to be plotted. Default from <b>csb</b> script.
<b>aq_project</b>	Air quality project to be used. Default from <b>csb</b> script.
<b>aq_site_table</b>	Table in AMET database that contains metadata of air quality sites.
<b>aq_species_col</b>	Air quality variables to be included in analysis. Default value is from <b>csb</b> script.
<b>checksave</b>	Check to see if R datafile exists (i.e., if savefile is activated R datafile is saved, so there is no need to query database).
<b>colp</b>	Various color specifications for model evaluation metrics.
<b>colpRH</b>	Color specifications for model evaluation metrics involving relative humidity.
<b>colpT</b>	Color specifications for model evaluation metrics involving 2-m temperature.
<b>colpWS</b>	Color specifications for model evaluation metrics involving 10-m wind.
<b>convert</b>	ImageMagick command. Used to crop margins of images. This variable has been deprecated.
<b>date</b>	Either one date used as both starting and ending date (default from <b>csb</b> script), or a vector containing starting and ending dates. Format: <b>YYYYMMDD</b>
<b>datee</b>	In <b>met_aq_coupled.input</b> , used to select ending date of analysis. Default uses value from <b>csb</b> script. In <b>summary.input</b> , used in the plot header to show the ending date used in the analysis. Does not change dates used in analysis. Format: <b>YYYYMMDD</b>

Variable	Description
<b>dates</b>	<p>In <b>met_aq_coupled.input</b>, used to select starting date of analysis. Default uses value from <b>csb</b> script.</p> <p>In <b>summary.input</b>, used in the plot header to show the starting date used in the analysis. Does not change dates used in analysis.</p> <p>Format: <b>YYYYMMDD</b></p>
<b>Daydelay</b>	Used in real-time mode to lag statistics by <b>daydelay</b> days.
<b>De</b>	Ending day of time series analysis. Default uses value from <b>csb</b> script.
<b>Diurnal</b>	Flag to partition and plot statistics as a function of time of day.
<b>Ds</b>	Starting day of time series analysis. Default uses value from <b>csb</b> script.
<b>elev</b>	Elevation criterion.
<b>extra</b>	Additional SQL criterion to use in database query.
<b>extra2</b>	Additional SQL criterion to use in database query. This is for the second site in time series plot.
<b>fcasthr</b>	Forecast hour criteria used to isolate a particular segment of forecast data.
<b>figdir</b>	Directory where figures will be output. Default value comes from the <b>csb</b> script.
<b>figid_sub</b>	Secondary figure label.
<b>figure</b>	Full figure path and name.
<b>fixed.legend</b>	<p><b>FALSE</b> – Use default legend</p> <p><b>TRUE</b> – Use custom legend from <b>legend.interval</b></p>
<b>he</b>	Ending hour of analysis.
<b>histplot</b>	Flag to plot histogram of statistics.

Variable	Description
<b>hs</b>	Starting hour of analysis.
<b>imageplot</b>	Include contours on profile plots: <b>TRUE</b> – Contour and shade plot <b>FALSE</b> – Only shade plot
<b>landuse</b>	Land use classification criteria for MySQL query.
<b>lat</b>	Latitude specification for MySQL query.
<b>layer</b>	Layer specification for profiler statistics.
<b>layerlab</b>	Label for specified layer.
<b>layerunit</b>	Units for layer statistics.
<b>layervar</b>	Variable for layer statistics.
<b>legend.div</b>	Number of legend intervals.
<b>legend.interval</b>	Specific legend intervals.
<b>legend.interval.bias</b>	Legend intervals for mean bias.
<b>level</b>	Level description.
<b>levsRH</b>	Levels for relative humidity statistics.
<b>levsT</b>	Levels for temperature statistics.
<b>levsWS</b>	Levels for wind speed statistics.
<b>lon</b>	Longitude specification for query.
<b>LT.offset</b>	Maximum and minimum UTC to local time offset in domain.
<b>max.dist</b>	Maximum distance allowed between AQ and MET sites in km.
<b>maxrec</b>	Maximum number of records to allow from database query (set to <b>-1</b> for unlimited).

Variable	Description
<b>me</b>	Ending month of analysis; default comes from <b>cs</b> run script.
<b>met_database</b>	MySQL database name that holds meteorological project.
<b>met_network</b>	Meteorological network.
<b>met_project</b>	Meteorological project name to be used in analysis. Default uses value from <b>cs</b> script.
<b>met_site_table</b>	Name of the table that contains meteorological site information.
<b>met_variable_col</b>	Modeled and observational variables of interest. Should be column names from database.
<b>model</b>	AMET project name.
<b>model1</b>	Primary AMET project for time series plot.
<b>model2</b>	Secondary AMET project for time series plot.
<b>ms</b>	Starting month of analysis; default comes from <b>cs</b> run script.
<b>obnetwork</b>	Network used in analysis. This value is used only in the plot header, and does not affect the analysis.
<b>obtime</b>	Times used in analysis. This value is used only in the plot header, and does not affect the analysis.
<b>pheight</b>	Plot height in pixels.
<b>pid</b>	A separate identification that is attached to the output. Can be used to distinguish between different output subsets for the same project. Default value is input from the <b>cs</b> script.
<b>player</b>	Logical to plot layer statistics.
<b>plotfmt</b>	File type of output. Default is taken from the <b>cs</b> script. Acceptable Values: <b>png</b> , <b>pdf</b> , <b>jpg</b> , or <b>eps</b>
<b>plotSingleProfile</b>	Logical (T or F) for generating hourly model-obs wind vector and potential temperature profile plots



Variable	Description
<b>plotSiteMap</b>	Logical (T or F) for generating a site location plot that can aid in identifying site IDs and locations.
<b>plotsize</b>	Scale factor to increase or decrease the size of plots. <b>1</b> = 541 x 700 pixel ( <b>png</b> ) or 8.5 x 11 inch ( <b>pdf</b> )
<b>processprof</b>	Logical to generate profile comparisons.
<b>prof</b>	Logical ( <b>T</b> or <b>F</b> ) for plotting raob-model profile comparison.
<b>proflim</b>	Lower and upper limit of profile plot.
<b>project</b>	AMET project name to be used in analysis. Default value is taken from <b>csH</b> run script.
<b>pwidth</b>	Width of plot.
<b>qcQ</b>	Quality control limits of moisture data. ( <i>Note:</i> All data outside of this range are not considered.)
<b>qcT</b>	Quality control limits of temperature data.
<b>qcWS</b>	Quality control limits of wind speed data.
<b>query</b>	MySQL query.
<b>queryID</b>	A separate identification that is attached to the output. Can be used to distinguish different output subsets for the same project. Default value is transferred from <b>pid</b> .
<b>querystr</b>	Additional SQL criteria that can be used to subset the data used by the analysis. Default value is transferred from the run script.
<b>realtime</b>	Run script in real-time automated mode.
<b>savedir</b>	Directory in which plots and other output will be saved. Default value is input from <b>csH</b> script.
<b>savefile</b>	Logical to generate an R data file that contains the data used in the statistics plots.
<b>saveid</b>	Name of R data file.

Variable	Description
<b>scex</b>	Scale factor for statistics text size.
<b>shadeplot</b>	Flag to plot shaded statistics plot in addition to point statistics plot.
<b>sres</b>	Resolution of shaded plot in degrees.
<b>statid</b>	In <b>timeseries_plot</b> , the site ID to be used in the analysis. Default value is taken from the <b>cs</b> script. In <b>summary_plot</b> , the station label to be used in labeling the plot.
<b>symb</b>	Symbol shape to be used in plots. See R documentation for shape numbers.
<b>symbo</b>	Plot symbol.
<b>symsiz</b>	Scale factor to adjust size of symbols on plots. <b>0.5</b> is very small while <b>1.5</b> is large.
<b>syncond</b>	This variable has been deprecated.
<b>t.test.flag</b>	Logical to apply statistical significance test to the spatial statistics. If it is applied and the model and observation data are not different statistically, the values are not plotted.
<b>textout</b>	Logical to write text output of statistics and underlying data.
<b>textstats</b>	Logical to write text file of statistics.
<b>thresh</b>	Used in spatial surface to identify the minimum number of data points required at a particular site to compute the statistics.
<b>time.of.day.utc</b>	Range of time (UTC) to isolate met and AQ data (e.g., compare average temperature and average PM for hours between 6 and 12 UTC).
<b>tserieslen</b>	Length in days of time series if real-time mode is activated.
<b>uniquepnum</b>	Unique plot number (random).

Variable	Description
<b>wantfigs</b>	Flag to generate figures.
<b>wantsave</b>	Flag to save station statistics data in R data file.
<b>wdweightws</b>	Logical to weight wind direction statistics by the wind speed (e.g., if wind speed falls below 3 m/s, the difference between model and observed wind direction is mitigated).
<b>ye</b>	Ending year of analysis default; comes from the <b>cs</b> script.
<b>ys</b>	Starting year of analysis default; comes from the <b>cs</b> script.
<b>zlims</b>	Specification of lower and upper vertical level of profile.

## A.8 AQ Network Input File

In addition to the analysis input files that are script specific, AMET makes use of the input file **\$AMET/scripts\_analysis/aqExample/Network.input**. This file allows all of the network-specific processing to be handled in one location, and allows for easier addition of new networks into the analysis scripts. The following are the settings used in this file.

Table A-8. AQ network input file settings

Variable	Description
<b>inc_airbase_hourly</b>	Include AirBase hourly “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_airbase_daily</b>	Include AirBase daily “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_airmon</b>	Include AIRMoN “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_admn</b>	Include ADMN “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_aganet</b>	Include AGANET “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_aurn_hourly</b>	Include AURN hourly “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_aurn_daily</b>	Include AURN daily “network”: “ <b>y</b> ” or “ <b>n</b> ”.
<b>inc_aqs_daily_o3</b>	Include AQS daily ozone values (i.e. 1hr, 8hr, and 9cell maxes): “ <b>y</b> ” or “ <b>n</b> ”.

Variable	Description
<code>inc_aqs_daily_pm</code>	Include AQS daily PM “ <b>network</b> ” (i.e. PM2.5 and PM10): “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_aqs_hourly</code>	Include AQS hourly “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_castnet</code>	Include CASTNET “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_castnet_hr</code>	Include CASTNET hourly O <sub>3</sub> “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_emep_hourly</code>	Include EMEP hourly “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_emep_daily</code>	Include EMEP daily “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_improve</code>	Include IMPROVE “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_nadp</code>	Include NADP “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_namn</code>	Include NAMN “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_naps</code>	Include NAPS “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_search</code>	Include SEARCH “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_search_daily</code>	Include SEARCH daily “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>inc_csn</code>	Include CSN “ <b>network</b> ”: “ <b>y</b> ” or “ <b>n</b> ”.
<code>network_label</code>	Label of networks to be include in analysis.
<code>network_names</code>	Names of networks to be included in analysis.
<code>total_networks</code>	Total number of networks in analysis.

## A.9 AQ Analysis Input Variables

The AQ analysis input files are found in **\$AMET/scripts\_analysis/aqExample**. The following is a partial list of variables in the AQ analysis input files. Not all of these variables are available in every input file.

**Table A-9. AQ analysis input variables**

Variable	Description
<b>abs_error_max</b>	max value on spatial plots for error (absolute value): “ <b>NULL</b> ” – script defined limit
<b>abs_range_max</b>	max value on spatial plots for model/ob value (absolute value): “ <b>NULL</b> ” – script defined limit
<b>abs_range_min</b>	min value on spatial plots for model/ob value (absolute value): “ <b>NULL</b> ” – script defined limit
<b>add_query</b>	Additional query syntax to add to the MySQL query.
<b>aq_database</b>	AQ MySQL database. Most likely “ <b>amet</b> ”.
<b>aq_network</b>	AQ monitoring network.
<b>aq_project</b>	AQ project name.
<b>aq_site_table</b>	AQ monitoring site MySQL table. Most likely “ <b>site_metadata</b> ”.
<b>aq_species_col</b>	AQ variables, column names, from AQ project table.
<b>averaging</b>	Average across time: “ <b>n</b> ” – no averaging (default) “ <b>s</b> ” – seasonal averaging (DJF; MAM; JJA; SON) “ <b>m</b> ” – monthly averaging “ <b>a</b> ” – entire time period averaging
<b>bias_range_max</b>	Bias range max limit: “ <b>NULL</b> ” – script-defined limit
<b>bias_range_min</b>	Bias range min limit: “ <b>NULL</b> ” – script-defined limit
<b>Bldoverlay_exe</b>	The location of the <b>bldoverlay</b> Fortran executable. Most likely <b>\$AMETBASE/bin/bldoverlay</b> . (AQ only)
<b>conf_line</b>	Add confidence lines to scatterplots: “ <b>y</b> ” or “ <b>n</b> ”.

Variable	Description
<b>coverage_limit</b>	% necessary for data completeness (e.g., <b>75</b> means 75% data completeness).
<b>custom_title</b>	Custom title for plots: “” – no custom title
<b>datee</b>	End date of query, <b>YYYYMMDD</b> format ( <b>met_aq_coupled.input</b> only).
<b>dates</b>	Start date of query, <b>YYYYMMDD</b> format ( <b>met_aq_coupled.input</b> only).
<b>density_zlim</b>	Maximum values for density scatter plot: “ <b>NULL</b> ” – script-defined limit
<b>diff_range_max</b>	Difference range max limit: “ <b>NULL</b> ” – script defined limit
<b>diff_range_min</b>	Difference range min limit: “ <b>NULL</b> ” – script defined limit
<b>end_date</b>	End date of query, <b>YYYYMMDD</b> format.
<b>end_hour</b>	End hour of query, <b>HH</b> format.
<b>error_range_max</b>	Error range max limit: “ <b>NULL</b> ” – script defined limit
<b>figdir</b>	Output directory for plots.
<b>fixed.legend</b>	Fixed legend intervals ( <b>met_aq_coupled.input</b> ): “ <b>F</b> ” – False, use full range of data “ <b>T</b> ” – True, specify custom legend through <b>legend.interval</b>
<b>greyscale</b>	Plot spatial plot in greyscale instead of color scale: “ <b>y</b> ” or “ <b>n</b> ”. Currently only applies to run_plot_spatial.csh script.
<b>inc_counties</b>	Option to include county lines on spatial plots: “ <b>y</b> ” or “ <b>n</b> ”.

Variable	Description
<code>inc_ranges</code>	Included quartile <b>ranges</b> on box plots.
<code>legend.div</code>	Number of levels if a fixed interval is not specified.
<code>legend.interval</code>	Custom legend interval ( <b>met_aq_coupled.input</b> ).
<code>legend.interval.bias</code>	Custom legend bias interval ( <b>met_aq_coupled.input</b> ).
<code>LT.offset</code>	Maximum and minimum UTC to LT offset in domain.
<code>max.dist</code>	Maximum distance allowed between MET and AQ “buddy” sites. These “buddy” sites are used to compare MET and AQ results in <b>met_aq_coupled</b> .
<code>median</code>	Statistical averaging method to use for stacked barplot: <b>TRUE</b> – median <b>FALSE</b> – mean
<code>met_database</code>	MET MySQL database. Most likely “ <b>amet</b> ”.
<code>met_network</code>	MET monitoring network.
<code>met_project</code>	MET project name.
<code>met_site_table</code>	MET monitoring site MySQL table. Most likely “ <b>stations</b> ”.
<code>met_variable_col</code>	MET variables, column names, from AQ project table.
<code>num_density_bins</code>	Number of bins to use in density scatter plot: “ <b>NULL</b> ” – script-defined limit
<code>num_ints</code>	The number of color intervals to use for spatial plots. The script will ultimately determine the number of intervals, but <b>num_ints</b> can be set to increase or decrease the number of intervals.

Variable	Description
<b>obs_per_day_limit</b>	Set the minimum number of observations required for single to be included in the analysis. Currently applies to time series plots only. Useful if several days have only a few observations that skew the plot formatting. Default value typically set to 10, but any numeric number can be used.
<b>overlay_opt</b>	PAVE overlay times: <b>1</b> – hourly <b>2</b> – daily <b>3</b> – 1-hr daily max <b>4</b> – 8-hr daily max
<b>perc_error_max</b>	max value on spatial plots for error (percent value): <b>“NULL”</b> – script defined limit
<b>perc_range_max</b>	max value on spatial plots for model/ob value (percent value): <b>“NULL”</b> – script defined limit
<b>perc_range_min</b>	min value on spatial plots for model/ob value (percent value): <b>“NULL”</b> – script defined limit
<b>pid</b>	Project name; must be unique across both MET and AQ.
<b>plot_colors</b>	Scatter plot symbol colors for primary simulation.
<b>plot_colors2</b>	Scatter plot symbol colors for secondary simulation.
<b>plotfmt</b>	Plot format, output type: <b>“PDF”</b> – pdf format <b>“PNG”</b> – png and pdf formats
<b>plotsize</b>	Scale factor to increase or decrease the size of a 541 x 700 pixel ( <b>png</b> ) or 8.5 x 11 inch ( <b>pdf</b> ) plot.
<b>query</b>	MySQL query to select data from database. In most cases, this is only part of the query. The complete query is constructed in the corresponding R script.



Variable	Description
<b>remove_mean</b>	Remove model/ob mean value from average value: “ <b>y</b> ” or “ <b>n</b> ”. default = “ <b>n</b> ”.
<b>remove_negatives</b>	Remove negative values: “ <b>y</b> ” or “ <b>n</b> ”. default = “ <b>y</b> ”.
<b>remove_other</b>	Remove “PM other” category from stacked bar plot analysis.
<b>rmse_range_max</b>	<b>RMS Error</b> range max limit: “ <b>NULL</b> ” – script defined limit
<b>rpo</b>	Subset by Regional Planning Organization (pre-defined subset of states). Acceptable values are: “VISTAS”, “MANE-VU”, “LADCO”, “CENRAP”, “WRAP” and “None” (default).
<b>run_info_text</b>	Include model run info as additional text to plots: “ <b>y</b> ” or “ <b>n</b> ”
<b>run_name1</b>	Project name; must be unique across AQ and MET.
<b>run_name*</b>	Additional project names (allowed only in some scripts). * can typically range from 2-6.
<b>site</b>	Plot label for when you are including only certain sites. Note that you will need to use an additional query to actually subset the data to these sites.
<b>soccerplot_opt</b>	Flag for soccer and bugle plot options: <b>1</b> – normalized mean bias/error <b>2</b> – fractional bias/error
<b>species</b>	Chemical species to plot.
<b>start_date</b>	Start date of query, <b>YYYYMMDD</b> format.
<b>start_hour</b>	Start hour of query, <b>HH</b> format.
<b>state</b>	Plot label for indicating certain states. Note that you will need to use an additional query to actually subset the data to these states.

Variable	Description
<b>stat_file</b>	File containing specific list of stations to analyze. User-defined.
<b>stats_flags</b>	Flags to determine which statistics are included on the <b>run_scatterplot.csh</b> script. Up to five statistics can be included, and are indicated by a 'y'. Unused statistics are left blank. The order of the statistics flags is: <b>index of agreement (IA), correlation (r), RMSE, systematic RMSE, unsystematic RMSE, NMB, NME, Normalized Median Bias, Normalized Median Error, Mean Bias, Mean Error, Median Bias, Median Error, Fractional Bias, Fractional Error</b>
<b>sympb</b>	Plot symbol: <b>15</b> – square <b>19</b> – circle
<b>sympo</b>	Plot symbol.
<b>sympsiz</b>	Plot symbol size: ( <b>0.5</b> very small to <b>1.5</b> large). A value of <b>1</b> is recommended for most applications.
<b>textstats</b>	Produce text statistics file: <b>TRUE</b> or <b>FALSE</b> .
<b>time.of.day.utc</b>	Range of time (UTC) to isolate met and AQ data (e.g., compare average temperature and average PM for hours between 6 and 12 UTC).
<b>use_avg_stats</b>	Use time-averaged statistics: “y” or “n”.
<b>use_median</b>	Use median value instead of mean for stacked bar plots: “y” or “n”.
<b>use_TC</b>	Use total carbon instead of EC and OC species in stacked bar plots: “y” or “n”.
<b>valid_only</b>	Applies to NADP only. Use only valid data based on the validity code in the database: “y” or “n”.

Variable	Description
<b>x_axis_min</b>	x axis min limit: “NULL” – script-defined limit
<b>x_axis_max</b>	x axis max limit: “NULL” – script-defined limit
<b>y_axis_min</b>	y axis min limit: “NULL” – script-defined limit
<b>y_axis_max</b>	y axis max limit: “NULL” – script-defined limit
<b>zeroprecip</b>	Include zero-precipitation obs: “y” or “n”. (typically set to “n”)

## A.10 AQ Variables-Network Mapping for North America

We provide below a mapping between monitoring networks, frequency (of observations), and chemical species that are included in the observed datasets, and that can be used in the model evaluation scripts. The units of all these variables were provided already in Section 4.2.

**Table A-8. AQ networks, frequency of observations, and chemical species for North America**

Network	Frequency	Chemical Species
<b>AQS</b>	hourly	O <sub>3</sub> , SO <sub>2</sub> , NO, NO <sub>2</sub> , NO <sub>x</sub> , CO, PM <sub>2.5</sub>
<b>AQS</b>	daily	PM <sub>2.5</sub> , PM <sub>10</sub>
<b>CASTNET</b>	weekly	SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , HNO <sub>3</sub> , TNO <sub>3</sub> , SO <sub>2</sub>
<b>CASTNET</b>	hourly	O <sub>3</sub> , T, RH, SolRad, Precip, Wind Speed and Direction
<b>IMPROVE</b>	daily (every 3 <sup>rd</sup> day)	SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , PM <sub>2.5</sub> , EC, OC, TC, Cl, Na, Fe, Ca, Si, Ti, Mg, Mn, K, Al
<b>MDN</b>	weekly	Wet deposition and concentration for Hg

Network	Frequency	Chemical Species
<b>NADP</b>	weekly accumulated	Wet deposition and concentration for SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , and precipitation
<b>SEARCH</b>	hourly	NO, NO <sub>2</sub> , SO <sub>2</sub> , O <sub>3</sub> , CO, NO <sub>y</sub> , EC, OC, TC
<b>SEARCH</b>	daily	SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , EC, OC, PM <sub>2.5</sub>
<b>CSN</b>	daily (every 3 <sup>rd</sup> day)	SO <sub>4</sub> , NO <sub>3</sub> , NH <sub>4</sub> , PM <sub>2.5</sub> , EC, OC, TC, Cl, Na, Fe, Ca, Si, Ti, Mg, Mn, K, Al
<b>NAPS</b>	hourly	O <sub>3</sub> , SO <sub>2</sub> , NO, NO <sub>2</sub> , NO <sub>x</sub> , CO, PM <sub>2.5</sub> , PM <sub>10</sub>

## A.11 AQ Variables-Network Mapping for Europe

We provide below a mapping between monitoring networks, frequency (of observations), and chemical species that are included in the observed datasets, and that can be used in the model evaluation scripts. The units of all these variables were provided already in Sections 4.2.10-15.

**Table A-11. AQ networks, frequency of observations, and chemical species for Europe**

Network	Frequency	Chemical Species
<b>ADMN</b>	weekly	SO <sub>x</sub> , NO <sub>y</sub> , NH <sub>x</sub> , Na, Cl, NSS_SO <sub>x</sub>
<b>AGANet</b>	monthly	NO <sub>y</sub> , SO <sub>x</sub> , Na, Cl
<b>AirBase</b>	hourly/daily	O <sub>3</sub> , SO <sub>2</sub> , CO, NO, NO <sub>2</sub> , NO <sub>x</sub> , PM <sub>2.5</sub> , PM <sub>10</sub>
<b>AURN</b>	hourly/daily	NH <sub>x</sub> , NO <sub>y</sub> , SO <sub>x</sub> , CO, SO <sub>2</sub> , O <sub>3</sub> , NO, NO <sub>2</sub>
<b>EMEP</b>	hourly/daily	NH <sub>x</sub> , NO <sub>y</sub> , SO <sub>x</sub> , PM <sub>2.5</sub> , PM <sub>10</sub> , CO, SO <sub>2</sub> , O <sub>3</sub> , NO, NO <sub>2</sub>
<b>NAMN</b>	monthly	NH <sub>x</sub> , NH <sub>3</sub>

## **Appendix B:**

### **What Is New in AMETv1.2**

This appendix highlights updates that have been made in AMET v1.2 relative to AMET v1.1. These were implemented to accommodate new chemical species, and to improve efficiency and flexibility in accessing observational data and in matching observational and model data. New observational data and several new analysis scripts have been added, resulting in changes to some of the original scripts and input files; also, additional queries and options have been added to some existing scripts. The major changes are listed below and noted in the appropriate sections of the main body.

### MET component updates

- The code underwent major cleaning and reformatting with better headers to track all changes. This includes the model-obs matching Perl codes as well as R and .csh run scripts. See code headers for details of all updates in the scripts provided in the **\$AMETBASE/perl** and **\$AMETBASE/R** directories.
- A bug was fixed on specification of the lower-left latitude and longitude in the perl script **MET\_init\_amet.pl** that has a small, but potential meaningful impact on the model index corresponding to the observation site. The old version used index 1,1, but in Perl the first index starts at 0. Users should test the new code and compare statistics with the old code for differences.
- Subhourly model output and observations such as profiler data can now be handled.
- MySQL time stamp definition was changed from **TIMESTAMP(8)** to **TIMESTAMP** in version 5.0. This required a change in AMET, but should have no impact on users.
- VAD wind comparison capability: A new eval class **vadwind** was added to match WRF-only model outputs with VAD wind observations. This option is in the new code, but not fully supported, as it requires a utility that converts VAD profiles in NCEP PREPBUFR obs files to a MADIS-like text format. Interested users can contact US EPA at [Gilliam.rober@epa.gov](mailto:Gilliam.rober@epa.gov) for guidance. To match model output with VAD wind, set **\$eval\_class= "vadwind"** in **populate\_project.input**.
- Observation matching of SurfRAD shortwave surface radiation was hidden in the old version. This was improved and is now activated in AMET 1.2. Users need to download daily SurfRAD observation text files and place them in **\$AMETBASE/obs/MET/surfrad/YYYY**, where YYYY is the year of the observations. In **populate\_project.input** set **\$eval\_class = "surfrad"**
- Statistics in time series plots were incorrect for wind speed and direction. The main error was in the mean absolute error in wind direction. For wind speed, the index of agreement had a small error related to the masking out of values less than an assigned minimum. These errors have been fixed in the AMETv1.2.

- Time series plotting of MARITIME sites is now possible. Previously the lack of mixing ratio data caused an abort in the plotting.
- Daily barplot creation would abort for days of missing model-obs pairs. In the current code version this failure is averted by skipping these days in the bar plotting.
- Improvement of plot profile statistical analysis:
  - min.sample option has been implemented (=10, corresponding to 10 days of analysis) so that diurnal statistics are ignored for smaller sample sizes.
  - min.WS.error that ignores model-obs pairs when their wind speed difference is greater than this value. This was implemented because bad observations are apparent in the profiler data that are difficult to eradicate. Use with caution and transparency.
  - Mean model PBL height is computed and displayed on the plots of diurnal mean modeled and observed wind speed and wind direction.
  - New plotting capabilities: Hourly average wind profiles (set AMET\_HOURLY\_PROF to T), boxplot profiles of metrics (set AMET\_VDIST\_PROF to T) and histogram of RMSE of wind speed and direction for all times of day and levels(set AMET\_HIST\_PROF T).
- Spatial surface statistics.
  - These have been revised to allow users to set a minimum threshold on the sample size below which to ignore a site.
  - Manual lat-lon plot boundaries can be specified in run script
- Plot model comparison with SurfRAD radiation data using run\_plot\_srad.csh. Plots including diurnal statistics, time series and spatial plots are available at three times daily: early day, midday and late day.

## **AQ component updates**

- Changes to **combine** program:
  - The concentration and deposition species definition files have been updated for CMAQv5.0 AERO6 module species.
- Changes to **AQ\_add\_sites\_dbase.pl** script:
  - New network site lists have been added.
- Changes to **AQ\_create\_project.pl** script:
  - New species have been added.

- Updates to the AQ database population script and inputs:
  - Options previously provided in the **populate\_project.input** file have been assimilated into the **aqProject.csh** script, removing the need for the **populate\_project.input** file.
  - Other options have been added to the **aqProject.csh** script to increase flexibility in the project database population.
  - The option to set which database to use has been moved from the **amet.pl/amet.R** configure files to the various AMET scripts so that users can switch more easily between different AQ databases (this applies to MET scripts as well).
  - Additional networks have been made available in the **aqProjects.csh** script, and several existing ones have been updated.
  - The daily max variables in AQS are now all populated in the database under a single network name (AQS\_Daily). This functionality was previously separated under different “networks” for the AQS ozone variable (e.g., O3\_1hrmax, O3\_8hrmax).
- Updates to the **AQ\_matching.pl** script:
  - Users no longer have to edit or create a network template file for this script. Instead, the Site Compare (**sitecmp**) scripts are generated using network-specific information in this script, and in a new file called **AQ\_species\_list.pl**. This file contains the list of species for each network, which is read by the **AQ\_matching.pl** script to generate a network-specific **sitecmp** script.
  - This greatly simplifies the process of adding or modifying the species processed for each network (now done by modifying the **AQ\_species\_list.pl** file).
  - To add a new network, users would create a new network in the **AQ\_species\_list.pl** following a format similar to that for an existing network, and then add the new network in the **AQ\_matching.pl** script to point to the appropriate network site and obs files.
- Updates to the **AQ\_add\_aq2dbase.pl** script:
  - When adding a new network, the user will no longer need to edit the **AQ\_add\_aq2dbase.pl** script to include the new network, assuming that the new network output file uses the standard Site Compare output format. Previously, the user had to add the specific network name to those specified in the *if* statements in the script.
  - Some obsolete code has been removed from the script.



- The “update project log” portion of the code has been updated to account for changes to the time field in the newer versions of the MySQL database (the new code is designed to be more versatile).
- Updates to the **delete\_AMETdb\_user.pl** script:
  - The script now determines the database from an environment variable instead of from the **configure** file.
- Updates to the AQ R analysis:
  - Numerous new analysis scripts have been added.
  - There are new or updated analysis files (**.csh** and **.input**) for new scripts and additional options.
  - See Table 7-2 in main body for additional information regarding the new or updated scripts.
- Additional query criteria and script options:
  - old script input files may not work as provided due to updates AQ analysis scripts (i.e., they will need updating).
  - See Table 7-2 in main body and Appendix A, Table A-9 for additional information on script input options.