Department of Pesticide Regulation
Environmental Monitoring Branch
1001 I Street
Sacramento, California 95812

Meteorological data processing for ISCST3 and AERMOD

Yuzhou Luo, Ph.D.
Research Scientist IV

10/11/2017

1 Introduction

This report presents a methodology to prepare meteorological input data for the Department of Pesticide Regulation's (DPR) Air Program. The methodology is developed to standardize and facilitate data processing, and provide consistent meteorological data in California for air dispersion models (ISCST3 and AERMOD). It extends the previous efforts in processing CIMIS data for ISCST3 (Johnson and Vidrio, 2014; Vidrio and Johnson, 2014). A computer program ("MetProc") is also developed to implement the proposed methodology.

Major components and features of this methodology are summarized as follows:

- Integrate multiple programs (AERMET, AERMINUTE, AERSURFACE, and PCRAMMET) into one interface.
- Automate meteorological data downloading and processing.
- Incorporate surface data from multiple sources (NOAA/NCEI, CIMIS, and DPR).
- Ensure consistent data processing for ISCST3 and AERMOD.
- Process CIMIS data in XML format (so the program will work even if CIMIS changes its output format in the future).
- Use ISD data in full archival format (TD-3505) for ISCST3 model (surface data in SAMSON format are not updated any more).
- Incorporate the stability classification with "σ_A method" in addition to Turner's methodd for the CIMIS stations not collecting net radiation.
- Calculate mixing heights for ISCST3 (mixing height in SCRAM format was discontinued since 1992), and provide the option to set mixing heights to a predefined value (i.e., 320 m).
- Provide advanced options on site characterization for modeling scenario development.

2 Data sources

2.1 NOAA (National Oceanic and Atmospheric Administration)

Surface data are retrieved from Integrated Surface Database (ISD) and Automated Surface Observing System (ASOS), maintained by NOAA/NCEI (National Centers for Environmental

Information, formerly the National Climatic Data Center). Upper air data are obtained from NOAA/ESRL (Earth System Research Laboratory) Radiosonde Database. Web API (Application Program Interface) and/or FTP service are provided by those databases, and integrated in this program for automatic data downloading.

Please note that NOAA data are provided in Greenwich Mean Time (GMT). When they are converted to local standard time (LST) of California, the last 8 hours of the study period will be all missing data. For example, ISD data for the year 2010 will not provide valid values for the hours 17~24 on 2010/12/31. There are two options if those hours are important to your study to: [1] include ISD data for the next year, i.e., combined ISD data for 2010-2011 will cover all hours on 2010/12/31; or [2] include data from CIMIS or DPR, which are reported in LST.

2.2 CIMIS (California Irrigation Management Information System)

CIMIS data in XML format (Extensible Markup Language) and metric units are used in this study. Data retrieval and analysis from two data sources are developed: [1] CIMIS website (http://www.cimis.water.ca.gov/) for manual data downloading, and [2] CIMIS web API (http://et.water.ca.gov/) for automatic data retrieval.

Automatic retrieval of CIMIS data is developed, but not fully implemented at this time due to the limitations in the current version of the CIMIS Web API (as of September 2017):

1) "Standard deviation of wind direction" is not provided in the Web API.
2) The API does not work if the data query includes a long term period (e.g., a whole month) of continuous missing data, especially for data from wind sensors (DWR, 2016).
3) The API limits the maximum data of 1750 records (i.e., ~72 days of hourly data).

In the previous efforts (Johnson and Vidrio, 2014; Vidrio and Johnson, 2014), CIMIS data was processed in CSV (Comma-separated value) format. If there are changes in the output data, e.g., the order or the number of meteorological variables in a row of the CSV file, data pre-processing has to be changed accordingly. For example, new programs were developed to reformat the downloaded CIMIS data in CSV after CIMIS changed the formatting of their output data in 2014 (Johnson and Vidrio, 2014) and 2016 (Johnson, 2016).

In this study, CIMIS data in XML format is used according to the following considerations:

- CIMIS data in XML is self-explanatory and independent of the order and number of reported variables. The variable name, QC flag, and numerical value are presented in the same row of XML, e.g., the value of 5.4 m/s for the variable "air_temp" (air temperature) in Figure 1, so that a computer program can read the data without a predefined data format as we have to prepare for CSV files.

```
<unit>Metric</unit>
<station station_nbr="92" station_name="Kesterson" region="San Joaquin Valley">
    <date val="1/1/2010" hour="0100">
        <eto qc="N"></eto>
        <precip qc=" ">0.00</precip>
        <solar_radiation qc=" ">0</solar_radiation>
        <net_radiation qc="N"></net_radiation>
        <vapor_pressure qc=" ">0.9</vapor_pressure>
        <air_temp qc=" ">5.4</air_temp>
        <rel_hum qc=" ">99</rel_hum>
        <dew_point qc=" ">5.3</dew_point>
        <wind_speed qc=" ">1.5</wind_speed>
        <res_wind_speed qc=" ">1.5</res_wind_speed>
        <wind_dir qc=" ">124</wind_dir>
        <wind_dir_sd qc=" ">13</wind_dir_sd>
        <soil_temp qc=" ">9.6</soil_temp>
        <pm_eto qc="N"></pm_eto>
        <pm_etr qc="N"></pm_etr>
        <exp_1 qc=" ">99.0</exp_1>
        <exp_2 qc=" ">12.8</exp_2>
    </date>
```

Figure 1. Demonstration of CIMIS data in XML format. The example here shows 1-hour (January 1, 2010, 1AM) data of all variables in metric units at the CIMIS station #92 "Kesterson".

- Missing data: in CSV, missing data is indicated by special symbols, e.g., "M", "--","####", and/or unanticipated forms (Johnson and Vidrio, 2014). In XML, missing data is consistently presented as an empty entry. For example, data for net radiation is missing for the hour "January 1, 2010, 1AM" at the CIMIS station #92 "Kesterson" (Figure 1).
- XML is the only format supported by both CIMIS website and CIMIS web API.

In addition, most of the modern computer languages provide specialized tools to handle XML files (e.g., "System.Xml" namespace in .Net Framework by Microsoft) and continuous support for future XML format change, if applicable. In summary, the use of CIMIS data in XML will ensure long-term operability and validity of the program developed in this study.

2.3 DPR-measured weather data

DPR data is saved in CSV format, and presented in the following order and units:

- Column 1: Data logger assigned tag number
- Column 2: Year
- Column 3: Julian day
- Column 4: Hour and minute of measurement
- Column 5: Wind speed (mph)
- Column 6: Wind direction (degree)
- Column 7: Relative humidity (%)
- Column 8: Air temperature (F)
- Column 9: Solar radiation (W/m^2)
- Column 10: Barometric pressure (mbar)

- Columns 11-13: Other data for internal use only

3 Stability (ISCST3 only)

In this program, Turner's method is the default method for stability calculation. DPR-modified σ_A method is also provided as an option if CIMIS data is used. USEPA recommends Turner's method for stability calculation, and accepts alternative methods including "σ_A method" (USEPA, 2000).

"Turner's method" (Turner, 1964) determines stability from wind speed, total cloud cover and ceiling height. By virtue of its historic precedence and widespread use, USEPA considers Turner's method to be the benchmark procedure for determining stability (USEPA, 2000). It has been integrated in the USEPA's PCRAMMET, a meteorological processor for ISCST3 (USEPA, 1999). The program uses the same algorithm for Turner's method as implemented in the subroutine "STABIT" of the PCRAMMET.

"σ_A method" in the previous meteorological data preparation (Johnson and Vidrio, 2014) was modified and integrated in this program. Previously, the determination of day and night was based on the net radiation: negative net radiation was defined as night and positive was defined as day (refer to the program "CIMPRO2v5", subroutine "DOSTAB" for more technical details). In this study, day and night are determined based on the sunrise and sunset times calculated daily at the corresponding CIMIS station location. The calculation follows the NOAA approach (ESRL, 2016a). This change is related to the following two considerations:

- Some CIMIS stations do not collect net radiation, indicated by QC flag of "N" (http://wwwcimis.water.ca.gov/Resources.aspx). For example, net radiation is not collected at the CIMIS station #92 "Kesterson" (Figure 1). By not using net radiation, this program can be applied to more CIMIS stations.
- Sunrise and sunset times have already been used in the previous study (Johnson and Vidrio, 2014) to determine day and night when the data of net radiation is questionable. A predefined dataset of monthly sunrise and sunset times for Redding and San Diego was used. This study extends this effort by calculating daily sunrise and sunset times for each station.

4 Mixing heights

The mixing height (z_i) in the convective boundary layer depends on both mechanical and convective processes and is assumed to be the larger of a mechanical mixing height (z_{im}) and a convective mixing height (z_{ic}). In the stable boundary layer, z_i results exclusively from mechanical (or shear-induced) turbulence and therefore is identically equal to z_{im}.

AERMET is used by this program to calculate z_{im} and z_{ic}: z_{im} is calculated for all hours, while z_{ic} is only reported for unstable atmosphere and flagged as -999 for stable hours. AERMOD will read both parameters z_{im} and z_{ic}, and calculate z_i as a built-in function.

For ISCST3 which requires z_i, two options are provided in this program to: [1] set z_i to 320 m for all hours, as justified by DPR (Johnson and Vidrio, 2014), or [2] calculate hourly z_i=max(z_{im}, z_{ic}) and incorporate the results into the meteorological input files.

5 Data processing

This section summarizes the meteorological data analysis, and associated options and assumptions implemented in this program:

- Surface data may be from multiple sources, and the priority order is set as: DPR data > CIMIS data > NOAA data. For each variable in each hour, if a valid data value is available from a source with higher priority, the value will be used in the data processing. On a certain hour, for example, if both CIMIS and NOAA report valid wind speed data, then the CIMIS value will be used.
- Due to limited data availability, upper air data are pre-downloaded for three stations of NKX (Miramar Naval Air Station), OAK (Oakland International Airport), and VBG (Vandenberg) (ESRL, 2016b). OAK is set as the default station. Data for 2005-2015 in FSL Radiosonde data format are prepared for the three stations.
- National Land Cover Dataset (NLCD) 1992 for California is pre-downloaded from USGS (USGS, 2016). This data and format are required by the current version of AERSURFACE (USEPA, 2017).
- Predefined parameter values (users cannot change those built-in values):
    - AERMET stage 1: threshold for wind speed is set as 0.5 m/s.
    - AERSURFACE: study radius for surface roughness = 1 km; number of sectors =12; temporal variation = monthly.
    - PCRAMMET: deposition = no.
- With user-specified data in the program interface (i.e., study period, weather stations, and data processing options), the program will first retrieve basic information from the internet for the selected stations, such as coordinates, elevation, and data availability for ASOS and IFWS (Ice Free Wind Sensor). Those characteristics will be in the subsequent data downloading and processing.
- For the ISD (Integrated Surface Database) station, specified as a five-digit WBAN (Weather-Bureau-Army-Navy) identifier by the user, the program will download hourly surface data (in full archival format, or TD-3505) from NOAA/NCEI (NCEI, 2016). Users are encouraged to check the ISD inventory (ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-inventory.txt) for data availability for the selected station and period.
- If the option “use ASOS data if available” is checked, the program will check if the 1-minute ASOS (Automated Surface Observing System) data is available for the selected WBAN station. If so, the program will download the 1-minute ASOS data (“page-1” data, in TD6405 format) from NOAA/NCEI (NCEI, 2016). This also activates AERMINUTE processing.
- Onsite data from CIMIS is an option in the program. If it’s selected, the program will ask the user to download the data (in XML format, by year) from the CIMIS website for the selected station and period. The program will combine and format the yearly on-site data into one data file.

- Onsite data from DPR is an option in this program. If it's selected, the program will ask the user to locate the data file. Some pre-processing may be required in order to sort the data records in ascending order by the date/time of measurements, and format the columns according to the requirements in Section 2.3. With user-provided DPR data:
    - If CIMIS data is not used, the program will generate an onsite data file from DPR data;
    - If CIMIS data is also used, the program will modify the previously generated on-site data (see the last step) by DPR data.
- With all required meteorological data downloaded or provided, the program will call AERMINUTE (if applicable) and AERSURFACE to prepare additional input data for AERMET.
- The program will call AERMET by providing control files for its stage-1, -2, and -3 processing, respectively.
- After AERMET processing, the program will extract the following variables from the intermediate and final outputs of AERMET: total cloud cover, ambient temperature, wind direction (and its standard deviation, if CIMIS data is used), wind speed, and ceiling height.
- Based on the user-selected option for stability determination for ISCST3, the program will either [1] call PCRAMMET for Turner's method, or [2] use the new algorithm developed in the program for DPR-modified σ_A method (see section 3).
- The program will compile an hourly meteorological data file in ISCST3 format, with variables of wind direction (in the form of wind vector), wind speed, ambient temperature, stability class, and mixing height. Wind direction, wind speed and ambient temperature are the same as the AERMET outputs; stability class is determined from the previous step; and mixing height is taken from AERMET outputs or set as 320 m according to user's decision (see section 4).
- Finally, three files are generated for the user-specified location and period: "SFC" as the boundary layer parameter file (used by AERMOD), "PFL" as the profile file (AERMOD), and "MET" as the ISCST3 input file. The "RP1" file summarizes the quality assessments.

6 User's guide and tutorial

6.1 Program overview

The program consists of an executable file ("MetProc.exe") and a folder with supporting data ("MetCache"). The latest USEPA meteorological data processors (as of this study) are used in the program, including AERMET v16216, AERMINUTE v15272, AERSURFACE v13016 (USEPA, 2017).

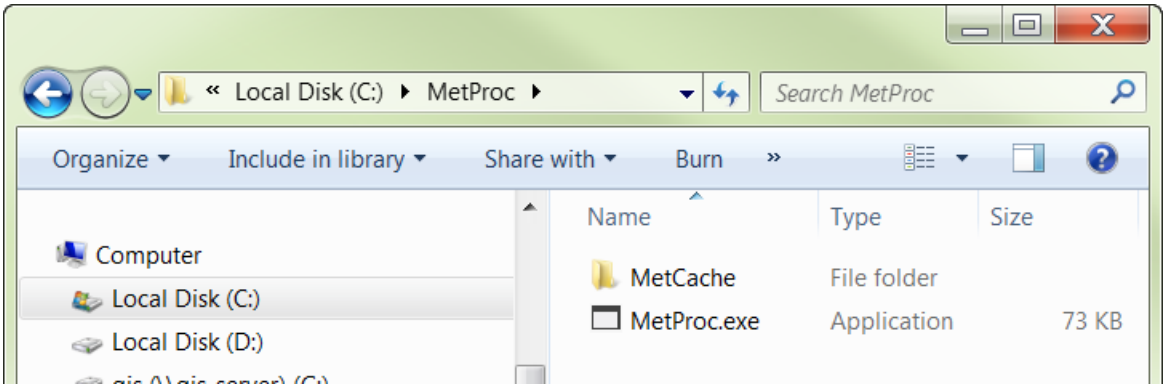


Figure 2. File structure of the program

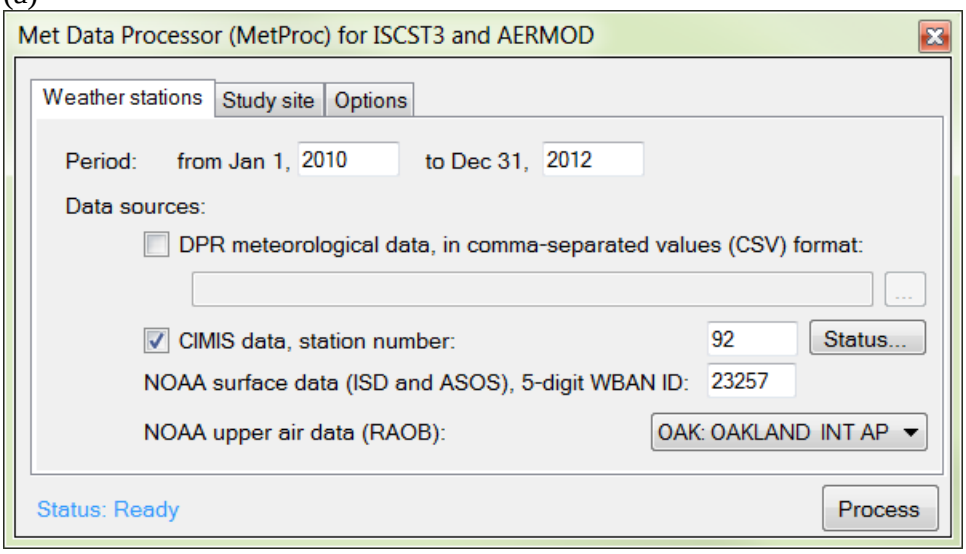
6.2 Configure study domain

As a case study for demonstration purpose, this program will be used to generate meteorological input data with the following configurations:

- Period: January 1, 2010 ~ December 31, 2012
- WBAN station ID: 23257
- Upper air data: OAK (Oakland)
- CIMIS station number: 92 (Kesterson, Merced)
- Site location: latitude = 37.4048 and longitude = -120.8680
- Other options: default values

Double click the executable file "MetProc.exe" to start the program, and specify the above settings to the program interface (Figure 3).

(a)



(b)

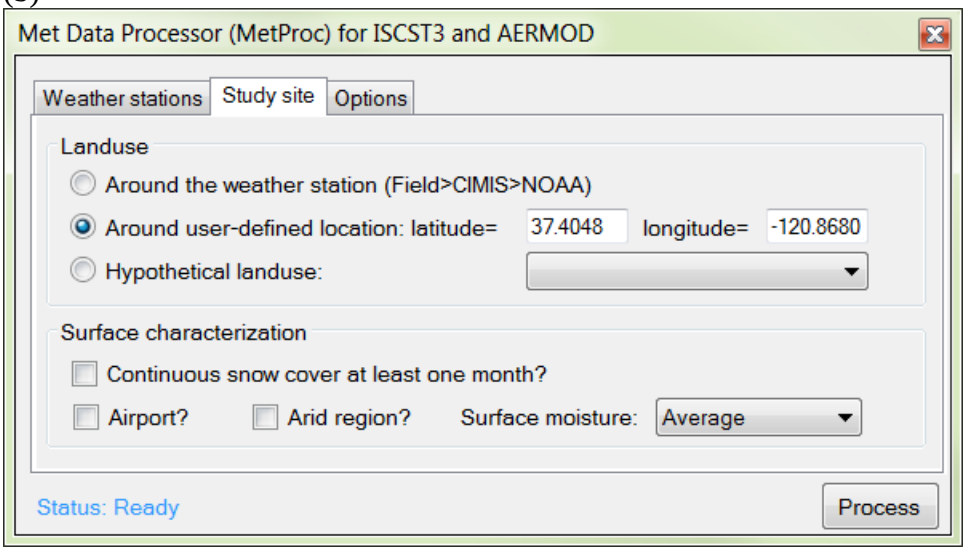


Figure 3. Settings in the case study for (a) weather stations and (b) study site.

The above settings for the case study are consistent with the 1,3-D monitoring in Merced during 10/14/2010 to 1/16/2012 (Rotondaro and van Wesenbeeck, 2012). The same CIMIS station was used in the previous modeling efforts (Johnson, 2014). There are several ISD stations around the study area. WBAN 23257 is also an ASOS station, so it's used here to demonstrate the program's capability to handle 1-minute ASOS data.

6.3 Download CIMIS data

This step is required if the option for "CIMIS data" is selected in the program (Figure 3). Due to the current limitations in CIMIS Web API (see section 2.2), CIMIS data should be manually downloaded, by year and in XML format.

The CIMIS webpage (http://www.cimis.water.ca.gov/WSNReportCriteria.aspx) requires an account. , which can be created at http://www.cimis.water.ca.gov/Auth/Register.aspx), CIMIS data for the first year of the study period for the case study (i.e., 2010) can be downloaded by following the instructions below (Figure 4):

1) For the "report style and date range", select "an hourly" "XML Report" in "Metric Units" from "1/1/2010" to "12/31/2010".
2) Select the CIMIS station from the site list, station 92 (Kesterson) in this case.
3) Check the option to "Select Sensors"; the sensor list will show up.
4) Select all sensors in the list (select the first sensor "ETo", then scroll down to the last sensor "Experimental 2" and select it by SHIFT + left click. In this way all sensors will be selected.)

   Please note that only some of the downloaded variables will be actually used in the data processing. Selecting all data simplifies the process and minimizes potential mistakes during the selection of individual items. Moreover, the complete dataset for the specific station and year will be retrieved in this way, so we don't have to re-download the data when additional variables are needed in the future. The data not used currently may be required for the program development or for other air dispersion models.

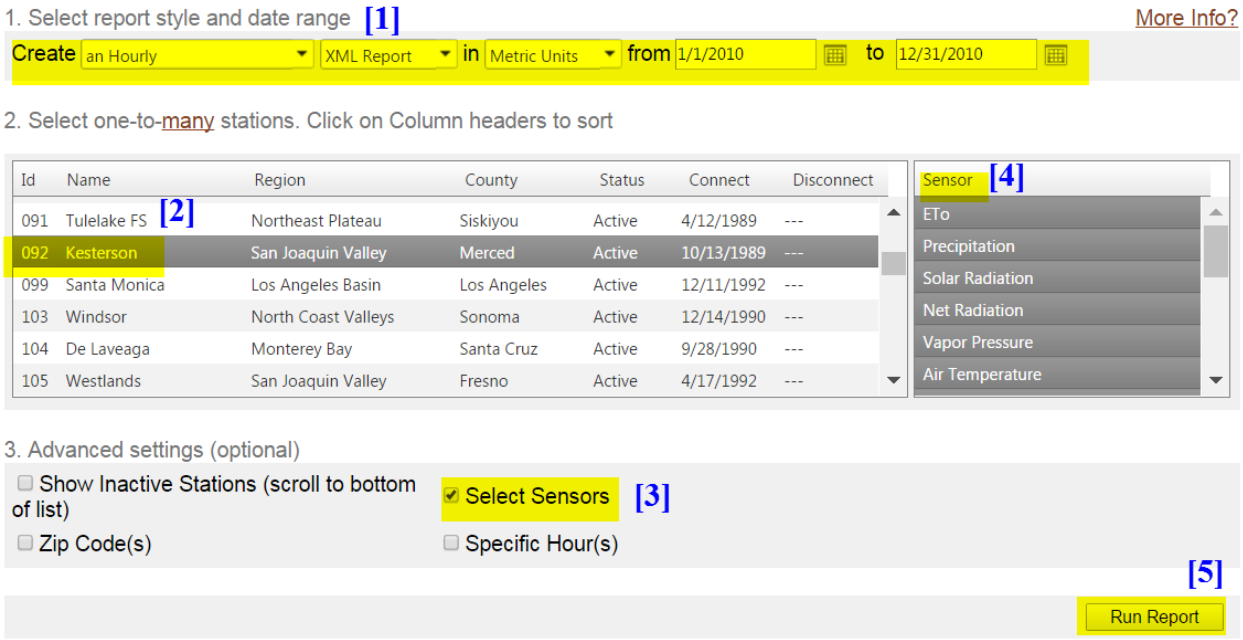


Figure 4. CIMIS data downloading of hourly data for a whole year (station 92 and year 2010 as an example)

5) "Run report". A "Save As" dialog box will show up with a default file name of "hourly.xml". Rename it according to the station and year, e.g., "CIMIS92_2010.xml", and save it to the [MetCache] folder.

Please note that you can save the data to any location with any name, but during data processing the program will copy them to the [MetCache] folder and rename them in the format "CIMIS[ID]_[YYYY].xml".

The above 5-step procedure can be repeated for each year of the study period. Finally, required CIMIS data are downloaded as three files for the case study, e.g., "CIMIS92_2010.xml", CIMIS92_2011.xml", and "CIMIS92_2012.xml".

6.4 Run the program

After downloading data from CIMIS, go back to the program GUI and click the "Process" button to run the program (Figure 3). The processing status will be reported as "Downloading", "Processing", or "Ready" in the lower-left corner of the GUI. Data downloading from NOAA may take several minutes for each year. Multiple MS-DOS prompt windows may appear during data processing. Please wait until the program completes all operations (Figure 5).

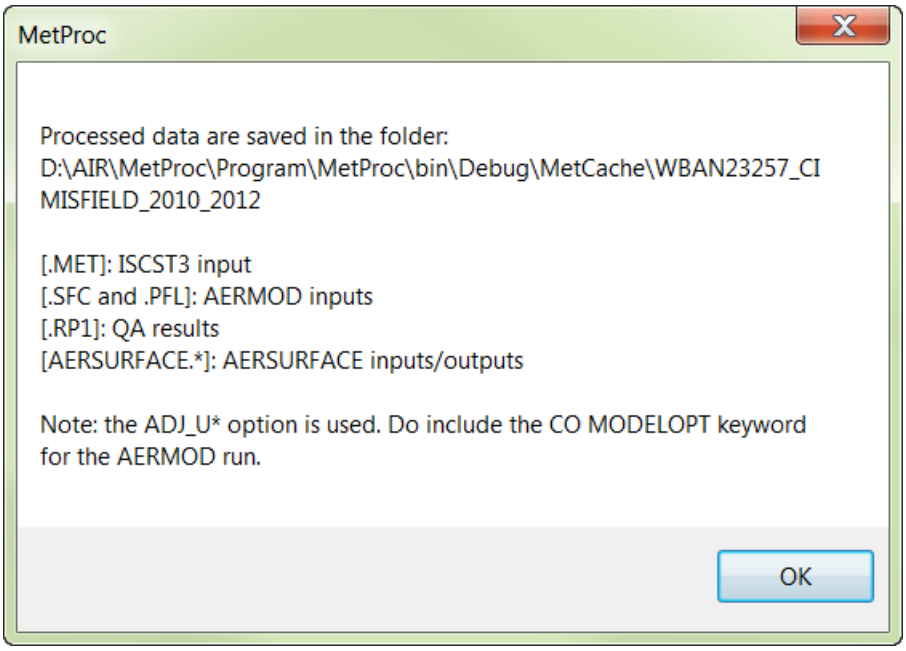


Figure 5. A message showing that the data processing has been finished

7 Verification

Outputs of this methodology are verified by comparing them with the processed AERMOD-ready meteorological input data available online. There are four agencies providing such data for California:

- California Air Resources Board or ARB (https://www.arb.ca.gov/toxics/harp/metfiles2.htm)
- San Joaquin Valley Unified Air Pollution Control District (http://www.valleyair.org/busind/pto/Tox_Resources/AirQualityMonitoring.htm)
- Santa Barbara County Air Pollution Control District (http://www.ourair.org/metdata/)

- South Coast Air Quality Management District (http://www.aqmd.gov/home/library/air-quality-data-studies/meteorological-data/data-for-aermod)

One-year data for one site is randomly selected from each agency (Table 1). MetProc is configured accordingly and its results are compared to the data posted by the agencies. Note that MetProc uses the latest version of AERMET (16216) while the posted data were generated with old versions (15181 for San Joaquin Valley and 14134 for others).

Table 1. Selected station/data and configurations for model verification. Highlighted are settings different to the program in this study

Dataset ID	Agency	Upper air station	Surface station	Year	AERMET version	Use ASOS data
1	California Air Resources Board	OAK	23230 (Oakland)	2010	14134	No
2	San Joaquin Valley Unified Air Pollution Control District	OAK	23237 (Stockton)	2010	15181	Yes
3	Santa Barbara County Air Pollution Control District	VBG	23190 (Santa Barbara)	2010	14134	Yes
4	South Coast Air Quality Management District	NKX	Unspecified onsite data (Riverside)	2010	14134	unknown

Note: Only dataset 2 is reported with surface characteristics: airport site; moisture condition = wet (for 2010 in San Joaquin Valley). For other datasets, no relevant information is available, and the same options (airport site; moisture condition = wet) are assumed in the verification.

The program developed in this study was conducted with the same configurations as shown in Table 1. The results, referred as "MetProc results", are compared to the downloaded datasets 1-4. Results of the comparison are summarized as follows:

- Dataset 1, Oakland, by California Air Resource Board

  All weather data (wind speed, wind direction, and air temperature) are the same between dataset 1 and MetProc results. Differences are observed for surface parameters, especially roughness length. Dataset 1 has significantly larger values of roughness length. Landuse data and parameters for surface characterization used by ARB were not reported.

- Dataset 2, Stockton, by San Joaquin Valley Unified Air Pollution Control District

  Dataset 2 and MetProc results have exactly the same data values. Surface characterization options are: airport (Y), arid region (N), and moisture (Wet).

- Dataset 3, Santa Barbara, by Santa Barbara County Air Pollution Control District

Dataset 3 and MetProc results have exactly the same data values. Surface characterization options are: airport (Y), arid region (N), and moisture (Average).

- Dataset 4, Riverside, South Coast Air Quality Management District

No comparison is conducted. Dataset 4 was based on onsite data, but the station information and source data were not provided. In addition, the roughness length and Bowen ratio were manually modified (roughness length=0.314; Bowen ratio=1.0 for all hours in Riverside) (SCAQMD, 2016).

Acknowledgements

The author would like to acknowledge Jing Tao, Atac Tuli, and Edgar Vidrio for valuable discussions and critical reviews.

References

DWR (2016). Personal communication with DWR CIMIS Support, 9/19/2016. California Department of Water Resources (DWR), Sacramento, CA.

ESRL (2016a). Solar calculation details (http://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html). National Oceanic & Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL), Boulder, CO.

ESRL (2016b). NOAA/ESRL Radiosonde Database (http://esrl.noaa.gov/raobs/). National Oceanic & Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL), Boulder, CO.

Johnson, B. (2014). Comparison of one-year township monitoring results from Merced to SOFEA simulation results (http://cdpr.ca.gov/docs/emon/pubs/analysmemos.htm). California Department of Pesticide Regulation, Sacramento, CA.

Johnson, B., Vidrio, E. (2014). Part 2: Processing CIMIS Meteorological Files to Produce ISCST3-ready Weather Files: Technical Manual (http://cdpr.ca.gov/docs/emon/pubs/analysmemos.htm). California Department of Pesticide Regulation, Sacramento, CA.

Johnson, B. (2016). Report on Revisions in California Irrigation Management System (CIMIS) Data Processing February 2016 (under review). California Department of Pesticide Regulation, Sacramento, CA.

NCEI (2016). Land-based datasets and products (https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets). National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information (NCEI, formerly the National Climatic Data Center), Asheville, NC.

Rotondaro, A., van Wesenbeeck, I. (2012). Monitoring of Cis- and Trans-1,3-Dichloropropene in air in 9 high 1,3-Dichloropropene use townships Merced County, California. Study ID: 101778. Dow AgroSciences.

SCAQMD (2016). Surface Characteristics of Meteorological Sites Used in AERMET (http://www.aqmd.gov/home/library/air-quality-data-studies/meteorological-data/aermod-table-2). South Coast Air Quality Management District, Diamond Bar, CA.

Turner, D.B., 1964. A Diffusion Model for an Urban Area. J. Applied Meteorology 3, 83-91.
USEPA (1999). PCRAMMET user's guide, EPA-454/B-96-001, revised June 1999 (https://www3.epa.gov/scram001/metobsdata_procaccprogs.htm). U.S. Environmental Protection Agency, Washington, DC.
USEPA (2000). Meteorological Monitoring Guidance for Regulatory Modeling Applications. EPA-454/R-99-005. United States Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.
USEPA (2017). Meteorological Processors and Accessory Programs (https://www3.epa.gov/scram001/metobsdata_procaccprogs.htm). U.S. Environmental Protection Agency, Washington, DC.
USGS (2016). Multi-resolution land characteristics consortium (MRLC) project (http://www.mrlc.gov/). U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD.
Vidrio, E., Johnson, B. (2014). Part 1: Database Creation of ISCST3-ready Files Prepared from CIMIS Meteorological Network Data (http://cdpr.ca.gov/docs/emon/pubs/analysmemos.htm). California Department of Pesticide Regulation, Sacramento, CA.