Data Fusion Tool User's Manual

Software version 1.9

July 2020

Acknowledgements

In collaboration with the U.S. EPA and the Institute for the Environment at the University of North Carolina at Chapel Hill (UNC-IE), the South China University of Technology (SCUT) developed the Data Fusion Tool software. This project was funded by the U.S. EPA under contract EP-D-12-044 to UNC-IE.

Table of Contents

List of Figures

1 Introduction

1.1 Overview of Data Fusion tool

Data Fusion (DF) is an innovative tool to derive spatial and temporal fields based on a series of data fusion methods commonly applied in health and economic benefit assessments (e.g., Environmental Benefits Mapping and Analysis Program – BenMAP) and air quality attainment assessments (e.g., Software for Model Attainment Test – SMAT) tools. The objective of this DF tool is to provide scientists and policymakers a user-friendly framework for creating desired spatial and temporal fields from existing air quality monitoring and model data and evaluating the results of fused spatial and temporal fields. The spatial and temporal field can be annual, quarterly, monthly, or daily distribution of predicted concentrations on the entire domain or sub-domain.

This standalone DF tool currently implements five built-in data fusion techniques:

(1) Voronoi Neighbor Averaging (VNA)

(2) Enhanced Voronoi Averaging (eVNA)

(3) Downscaler (DS)

(4) Weighting VNA (wVNA)

(5) Nearest Site (NS)

The DF tool has an easy-to-use graphical user interface (GUI) to allow users to flexibly create spatial and temporal data fields, conduct cross-validation (including Multi-fold, Leave-one-out, and Leave-group-out cross-evaluation methods), and provide visualization analysis capabilities (e.g., map, GIS and chart, etc.), as well as contains a built-in long-term ambient monitoring network database for $PM_{2.5}$ and its component species and ozone.

The DF tool includes five key modules currently:

(1) Data input module

(2) Data fusion technique module

(3) Spatial and temporal option module

(4) Cross validation module

(5) Data viewer module

The detailed description of each module can refer to later chapters.

This document provides instructions on how to use DF to conduct data fusion analysis for $PM_{2.5}$ and O_3 with the following chapters:

- Chapter 1. DF introduction and installation

- Chapter 2. Functional Framework of Data Fusion
- Chapter 3. Quick Start
- Chapter 4. $PM_{2.5}$ Data Fusion Analysis Details
- Chapter 5. Ozone Data Fusion Analysis Details
- Chapter 6. Data Viewer for analyzing results

1.2 Computer Requirements

Data Fusion requires a computer with:

a) Net Framework Version 4.0 or higher.

b) 32-bit or 64-bit Windows 7/Windows 8/Windows 10.

c) 2 GB RAM or greater.

d) 6 GB free disk space or greater.

1.3 Install/Uninstall Data Fusion

1.3.1 Install Data Fusion

➢Download Data Fusion Software Package from the ABaCAS website or Google Drive. This tool is available at the following links:

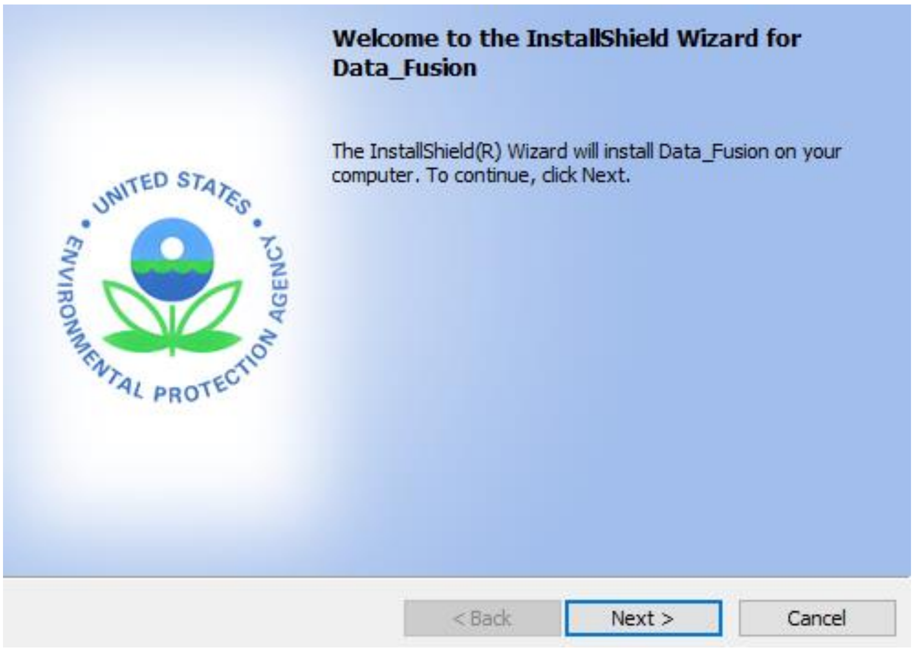
(1) ABaCAS website:

http://abacas.see.scut.edu.cn/tools.

(2) Google Drive:

https://drive.google.com/open?id=1Xl3VqtlRXeBt_FrfHpuCZYumxjqMR9hL

➢Double click the package (e.g., Data Fusion v1.9 setup.exe) to install the program, it will appear the following window (Fig. 1).

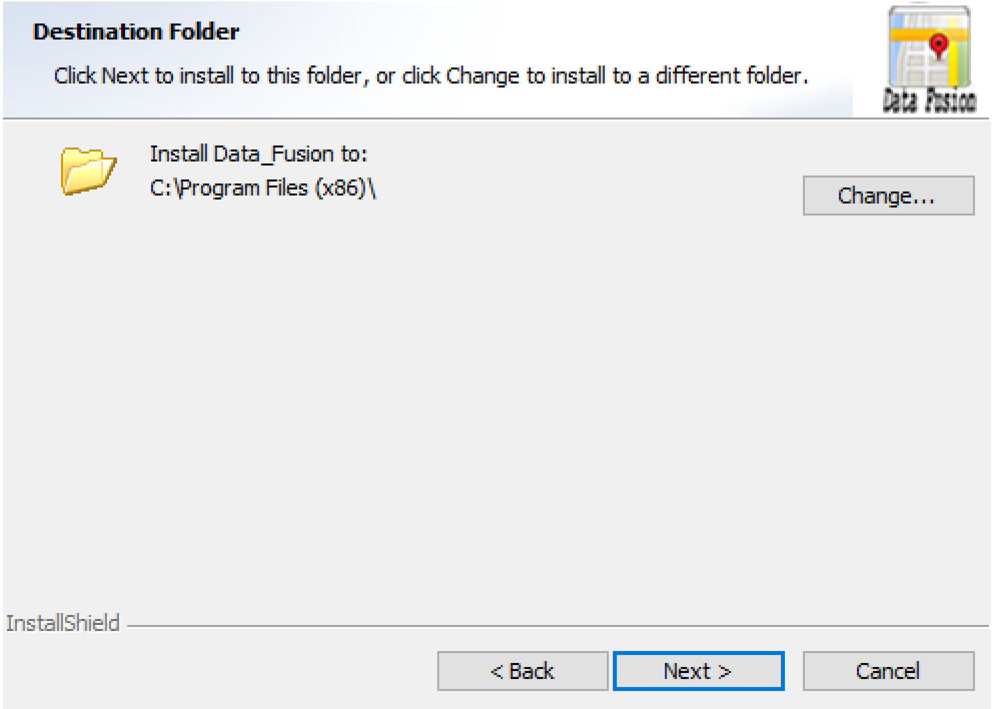


Welcome to the InstallShield Wizard for Data_Fusion

The InstallShield(R) Wizard will install Data_Fusion on your computer. To continue, click Next.

Fig. 1 Setup Window

➢Click the "**Next**" button, users can customize the destination folder to install as shown in Fig. 2.



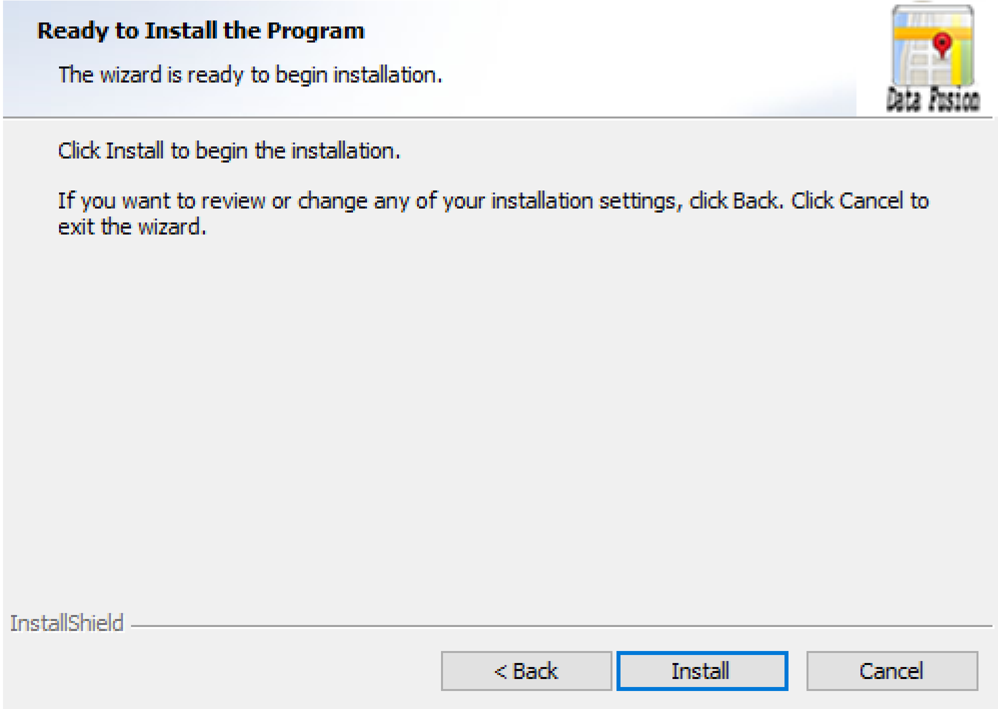
Destination Folder

Click Next to install to this folder, or click Change to install to a different folder.

Install Data_Fusion to:
C:\Program Files (x86)\

Fig. 2 Choose Install Path

➢Click the "**Next**" button, it will show the "Ready to Install the Program" window as shown in Fig. 3.



Ready to Install the Program

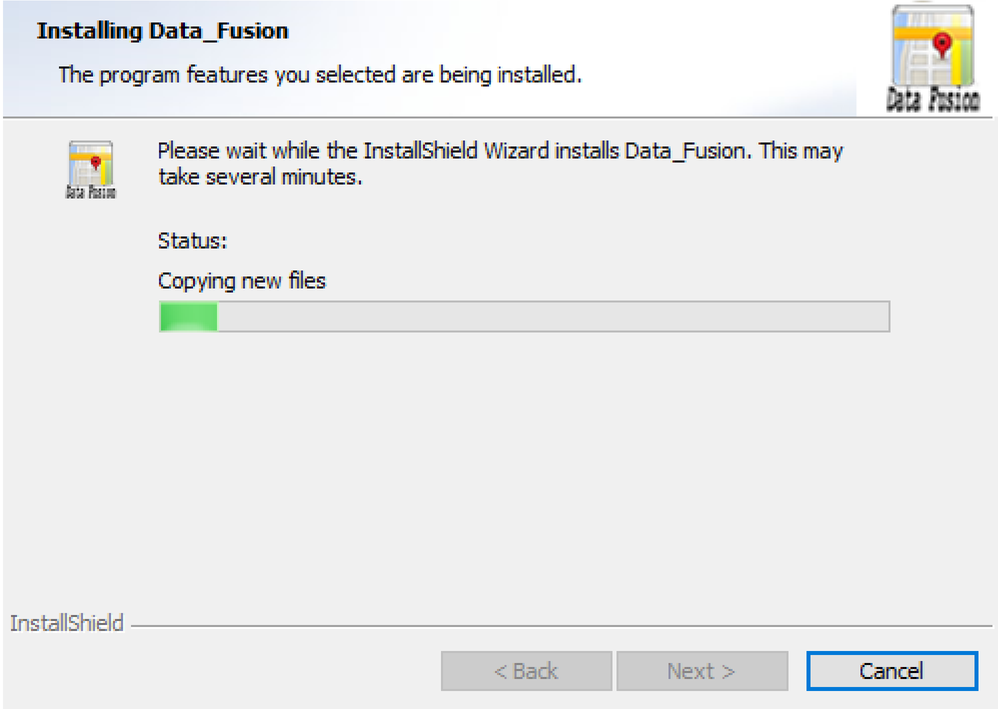
The wizard is ready to begin installation.

Click Install to begin the installation.

If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.

Fig. 3 Ready to Install

➢Click the "**Install**" button, then the Data Fusion will start installing.



Installing Data_Fusion

The program features you selected are being installed.

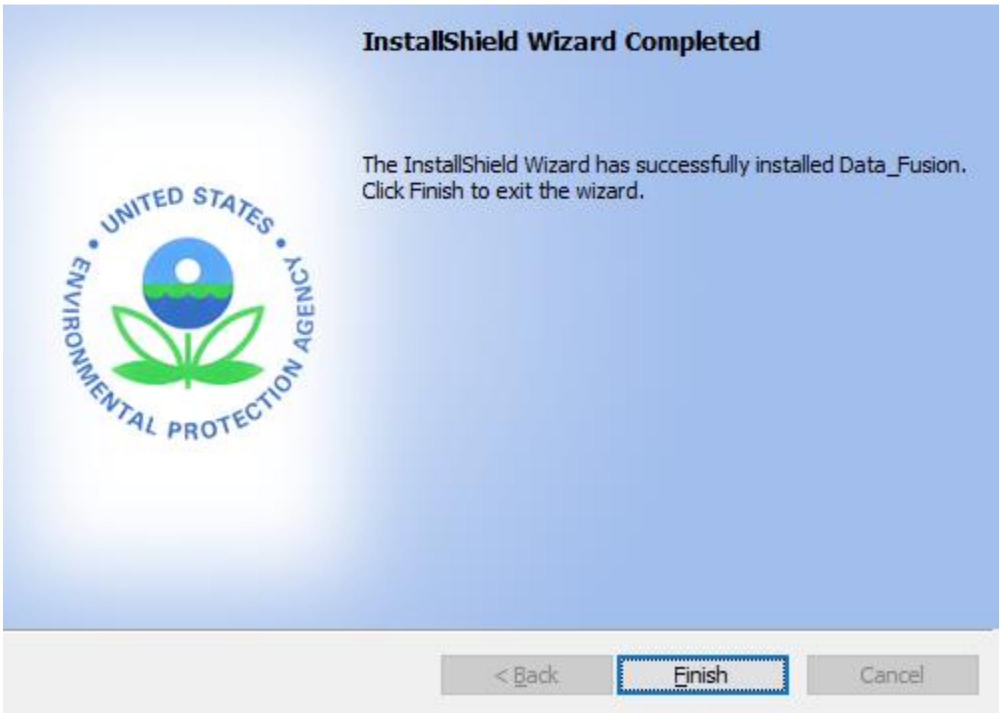
Please wait while the InstallShield Wizard installs Data_Fusion. This may take several minutes.

Status:

Copying new files

Fig. 4 Installation Process

➢Click the "**Finish**" button and complete installation.



InstallShield Wizard Completed

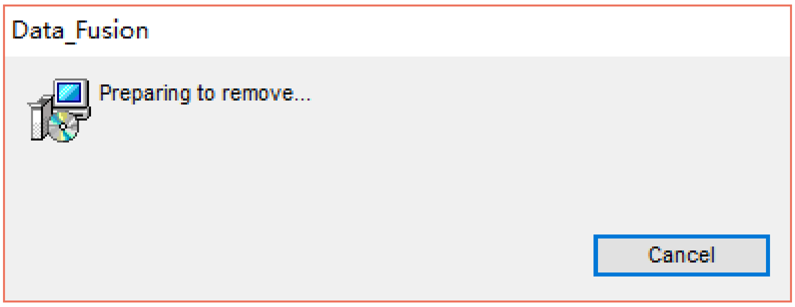
The InstallShield Wizard has successfully installed Data_Fusion. Click Finish to exit the wizard.

Fig. 5 Complete Installation

1.3.2 Uninstall Data Fusion

➢Go to Control Panel.

➢Select Data Fusion and click "**Change/Remove**", the following window will appear.



Data_Fusion

Preparing to remove...

Fig. 6 Uninstall DF

After a few seconds, the uninstallation process will finish.

1.4 Contacts for Comments and Questions

For comments and questions, please contact:

(1) Prof. Yun (Dustin) Zhu at South China University of Technology, Environmental Simulation and Information Laboratory via email at zhuyun@scut.edu.cn;

(2) The Center for Community Modeling and Analyses System (CMAS) at the University of North Carolina at Chapel Hill via email at cmas@unc.edu.

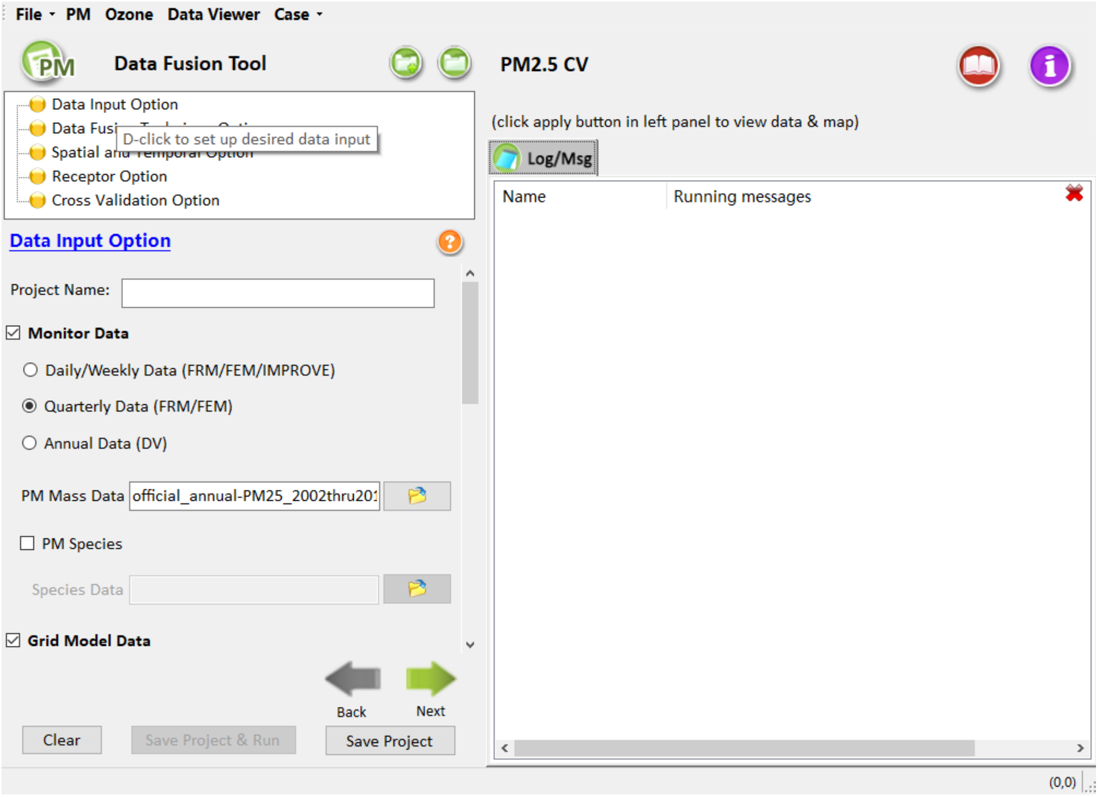
2 Quick Start

This chapter provides the steps required to run DF for PM and Ozone analyses. The Quick Start will use EPA’s data set to demonstrate how to run each of the DF analysis modules. These steps will use the default settings and do not describe the configuration settings for each analysis. For details of the configuration settings for the individual DF modules, refer to the User’s Guide chapter for each module.

2.1 Start the Data Fusion

Double click the Data Fusion icon on the desktop or the executable file (Data Fusion.exe) under the installation path to start the program.

➢ The Start Page window will appear on your screen after starting the program, as shown in Fig. 7.



Project Name:

☑ Monitor Data

○ Daily/Weekly Data (FRM/FEM/IMPROVE)

● Quarterly Data (FRM/FEM)

○ Annual Data (DV)

PM Mass Data: official_annual-PM25_2002thru20:

☐ PM Species

Species Data

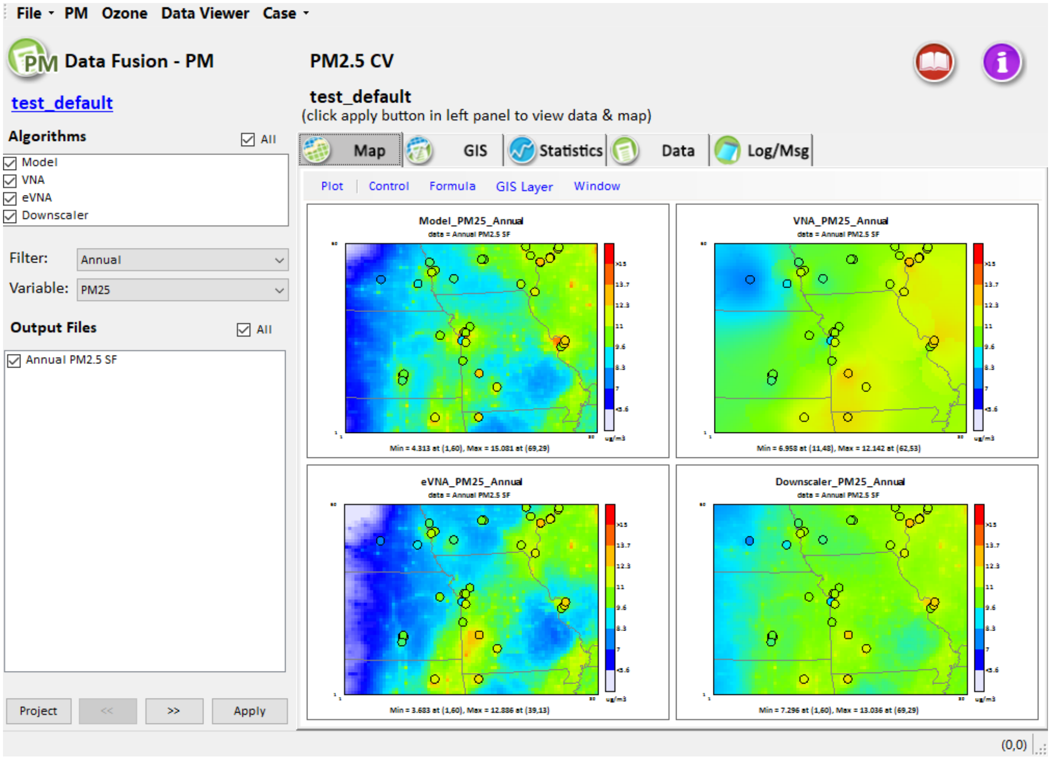
☑ Grid Model Data

Fig. 7 Start Page of Data Fusion

2.2 Load existing project

➢Click the **Open** in the **File** pull-down menu item on the start page, select the project file (*.proj) of interest to open, then it can be loaded into the DF Data Viewer to probe the results. All project files are stored in your computer's document directory (C:\Users\ ... \Documents\My DataFusion Files\Result\Project) by default. Users can select a default project named **test_default**.

➢After opening a project, the program will jump to the data viewer module, as shown in Fig. 8.



test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All
☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual
Variable: PM25

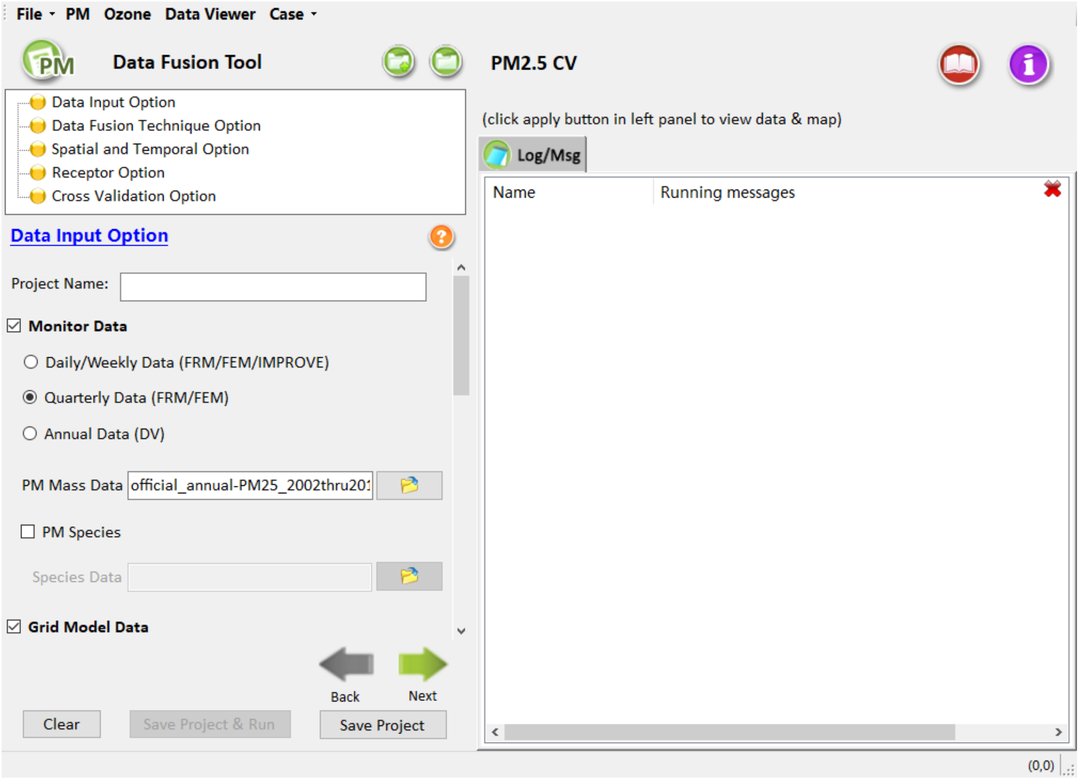
Output Files ☑ All
☑ Annual PM2.5 SF

Fig. 8 Open a previous project

➢After the selected project is loaded, users are also allowed to view the configuration options of this project by clicking the project name with a blue hyperlinked text in the top of output files. Analyzing existing projects can refer to the details of chapter 6 below. Users are allowed to return to the Map view by clicking the **Viewer** tab.

2.3 Create a new project

➢ Click the **New Project** in the **File** pull-down menu item on the start page, the program will reset all settings and jump to the parameter configuration module, as shown in Fig. 9.



Data Input Option

Project Name:

☑ Monitor Data

○ Daily/Weekly Data (FRM/FEM/IMPROVE)

● Quarterly Data (FRM/FEM)

○ Annual Data (DV)

PM Mass Data official_annual-PM25_2002thru20:

☐ PM Species

Species Data

☑ Grid Model Data

(click apply button in left panel to view data & map)

Name	Running messages

Fig. 9 Create a new project

➢ Configure the parameters of each step, then save and run the new project can refer to the details of Chapter 4 (PM$_{2.5}$ data fusion analysis) and Chapter 5 (Ozone data fusion analysis).

2.4 PM$_{2.5}$ Data Fusion Analysis Quick Start

The steps below describe how to use DF to conduct data fusion for PM$_{2.5}$ with monitor data, model data and other related data, and verify the data fusion results. More details refer to Chapter 4.

Step 1. Click **PM** menu on the top of Data Fusion Tool Home Page to launch the PM Analysis module window.

Step 2. The **Data Input Option** window display first. This windows sets the input files of PM$_{2.5}$ moritor data , PM$_{2.5}$ grid model data and other data. Data Fusion Analysis calculates the data fusion result based on the input files.

Use the default settings in the **Data Input Option** window.

➢ Click on the **Data Input Option** hyperlink to display an electronic version of the User’s Manual for this window.

Monitor Data: sets the input file of $PM_{2.5}$ monitor data. There are three time type options: **Daily/Weekly Data**, **Quarterly Data** and **Annual Data** (The default option is **Quarterly Data**).

- **PM Mass Data:** sets the detailed data of $PM_{2.5}$ monitor data.
- **PM Species:** This option is unchecked. (If checked, it will require to input a file of $PM_{2.5}$ species data).
- **Grid Model Data:** sets the data format, base case model data and grid description file.
- **Data Format:** sets the data format of the base year model data. *.csv and IO/API format data are supported.
- **Base Case:** sets the detailed data of $PM_{2.5}$ model data at base year.
- **Grid description:** There are two grid definition file type options: a CMAQ grid description file and a Grid Shape File(*.shp).

Other Data: This option is unchecked. (If checked, it will require to input a file of Diffusion model file data or AOD data).

➢ Click the **Next** arrow at the bottom right of the **Data Input Option** window to proceed to the next step.

Step 3. The **Data Fusion Technique Option** window sets the data fusion algorithms to calculate the data fusion result.

Use the default settings in the **Data Input Technique Option** window.

➢ Click on the **Data Fusion Technique Option** hyperlink to display an electronic version of the User's Manual for this window.

Algorithm: There are five algorithms options: VNA, eVNA ,Downscaler, wVNA and Nearest Site. The default choice are VNA, eVNA and Downscaler.

➢ Click the **Next** arrow at the bottom right of the **Data Input Technique Option** window to proceed to the next step.

Step 4. The **Spatial and Temporal Option** window customizes the space and time separately to comparing the simulation results of entire domain or partial domain and specified time ranges.

Use the default settings in the **Spatial and Temporal Option** window.

- Click on the **Spatial and Temporal Option** hyperlink to display an electronic version of the User's Manual for this window.

Spatial and Temporal Option: select **Entire Domain** or **Partial Domain**. The **Partial Domian** sets the grid cells by customizing the starting and ending grid cell in the entire domain.

Temporal Data Fusion Option: includes the time range of data fusion output results, which are divided into annual average, quarterly average, monthly average and daily average.

- Click the **Next** arrow at the bottom right of the **Spatial and Temporal Option** window to proceed to the next step.

Step 5. The **Receptor Option** window sets the GIS shape file. This GIS shape file is used to calculate average concentration in specified regions (e.g., states, counties).

Use the default settings in the **Receptor Option** window.

- Click on the **Receptor Option** hyperlink to display an electronic version of the User's Manual for this window.
- Click the **Next** arrow at the bottom right of the **Receptor Option** window to proceed to the next step.

Step 6. The **Cross Validation Option** window sets cross-validation method to conduct scientific and objective analysis of the validity and stability of data fusion results.

Use the default settings in the **Cross Validation Option** window.

- Click on the **Cross Validation Option** hyperlink to display an electronic version of the User's Manual for this window.

Cross Validation: This option is unchecked. (If checked, it requires to select a kind of cross-validation method).

- **Multi-fold:** selects the monitoring data file and groups it, and finally verifies the data fusion result according to the result set synthesized by each group. The result will intuitively show the validity and stability of the data fusion result.
- **Leave-one-out:** eliminates different monitoring value points selected by users one by one, compares and analyzes data fusion results in the absence of different monitoring points, and then finds and removes abnormal data values.

- **Leave-group-out:** selects the group data sets to compare and analyze data fusion results in the absence of different monitoring points, and then find and removes abnormal data values.

➢ Click the **Next** arrow at the bottom right of the **Cross Validation Option** window to complete the PM Analysis configuration and run the DF project.

➢ Click **Save & Run Project** in the pop-up window. Click Save to create a project (*.proj) file for this tutorial exercise.

The PM Analysis will complete after a few minutes and the Data Fusion Tool Data Viewer will present the results of this analysis. See Chapter 6 for details on how to use the Data Viewer to analyze the results.

2.5 Ozone Data Fusion Analysis Quick Start

The steps below describe how to use DF to conduct data fusion operations on Ozone with ozone monitor data, model data and other related data, and verify the data fusion results. More details refer to Chapter 5.

Step 1. Click **Ozone** menu on the top of Data Fusion Tool Home Page to launch the PM Analysis module window.

Step 2. The **Data Input Option** window display first. This windows sets the input files of Ozone moritor data, grid model data and other data. Data Fusion Analysis calculates the data fusion result based on the input files.

Use the default settings in the **Data Input Option** window.

➢ Click on the **Data Input Option** hyperlink to display an electronic version of the User's Manual for this window.

Monitor Data: set the input file of Ozone monitor data. There are two time type options: **Annual** and **Daily**(The default option is **Annual**).

- **Selcet Ozone Design Values (DV) Year:** sets the **Start Year** and **End Year** corresponding to the monitor data year.

Grid Model Data: sets the data format, base case model data and grid description file.

- **Data Format:** sets the data format of the base year model data. *.csv and IO/API format data are supported. By default, annual monitor data corresponds to model data in CSV format. Daily monitor data corresponds to model data in IO/API format

- **Base Case:** sets the detailed data of Ozone model data at base year.
- **Control Case:** This option is unchecked (If checked, it requires to select a file including the detailed data of Ozone model data at control year).
- **Grid description:** There are two grid definition file type options: CMAQ grid description file and Grid Shape File(*.shp).

Other Data: This option is unchecked. (If checked, it requires to input a file of Diffusion model file data or AOD data).

- ➢ Click the **Next** arrow at the bottom right of the **Data Input Option** window to proceed to the next step.

<u>Step 3.</u> The **Data Fusion Technique Option** window sets the data fusion algorithms to calculate the data fusion result.

Use the default settings in the **Data Input Technique Option** window.

- ➢ Click on the **Data Fusion Technique Option** hyperlink to display an electronic version of the User's Manual for this window.

Algorithm: There are five algorithms option: VNA, eVNA ,Downscaler, wVNA and Nearest Site. The default choice are VNA, eVNA and Downscaler.

- ➢ Click the **Next** arrow at the bottom right of the **Data Input Technique Option** window to proceed to the next step.

<u>Step 4.</u> The **Spatial and Temporal Option** window customizes the space and time separately to comparing the simulation results of entire domain or partial domain and specified time ranges.

Use the default settings in the **Spatial and Temporal Option** window.

- ➢ Click on the **Spatial and Temporal Option** hyperlink to display an electronic version of the User's Manual for this window.

Spatial and Temporal Option: select **Entire Domain** or **Partial Domain**. The **Partial Domian** sets the grid cells by customizing the starting and ending grid cell in the entire domain.

Temporal Data Fusion Option: includes the time range of data fusion output results, which are divided into annual average, quarterly average, monthly average and daily average.

- ➢ Click the **Next** arrow at the bottom right of the **Spatial and Temporal Option** window to proceed to the next step.

Step 5. The **Receptor Option** window sets the GIS shape file. This GIS shape file is used to calculate average concentration in specified regions (e.g., states, counties).

Use the default settings in the **Receptor Option** window.

- ➢ Click on the **Receptor Option** hyperlink to display an electronic version of the User's Manual for this window.
- ➢ Click the **Next** arrow at the bottom right of the **Receptor Option** window to proceed to the next step.

Step 6. The **Cross Validation Option** window sets cross-validation method to conduct scientific and objective analysis of the validity and stability of data fusion results.

Use the default settings in the **Cross Validation Option** window.

- ➢ Click on the **Cross Validation Option** hyperlink to display an electronic version of the User's Manual for this window.

Cross Validation: This option is unchecked. (If checked, it requires to select a kind of cross-validation method).

- **Multi-fold:** selects the monitoring data file and groups it, and finally verifies the data fusion result according to the result set synthesized by each group. The result will intuitively show the validity and stability of the data fusion result
- **Leave-one-out:** eliminates different monitoring value points selected by users one by one, compares and analyzes data fusion results in the absence of different monitoring points, and then finds and removes abnormal data values.
- **Leave-group-out:** selects the group data sets to compare and analyze data fusion results in the absence of different monitoring points, and then find and removes abnormal data values.
- Click the **Next** arrow at the bottom right of the **Cross Validation Option** window to complete the Ozone Analysis configuration and run the DF project.

- Click **Save & Run Project** in the pop-up window. Click Save to create a project (*.proj) file for this tutorial exercise.

The Ozone Analysis will complete after a few minutes and the Data Fusion Tool Data Viewer will present the results of this analysis. See Chapter 6 for details on how to use the Data Viewer to analyze the results.

3 Functional Framework of Data Fusion

Fig. 10 shows the functional modules of Data Fusion, including five modules:

(1) Data input module

(2) Data fusion technique module

(3) Spatial and temporal option module

(4) Cross validation module

(5) Data viewer module

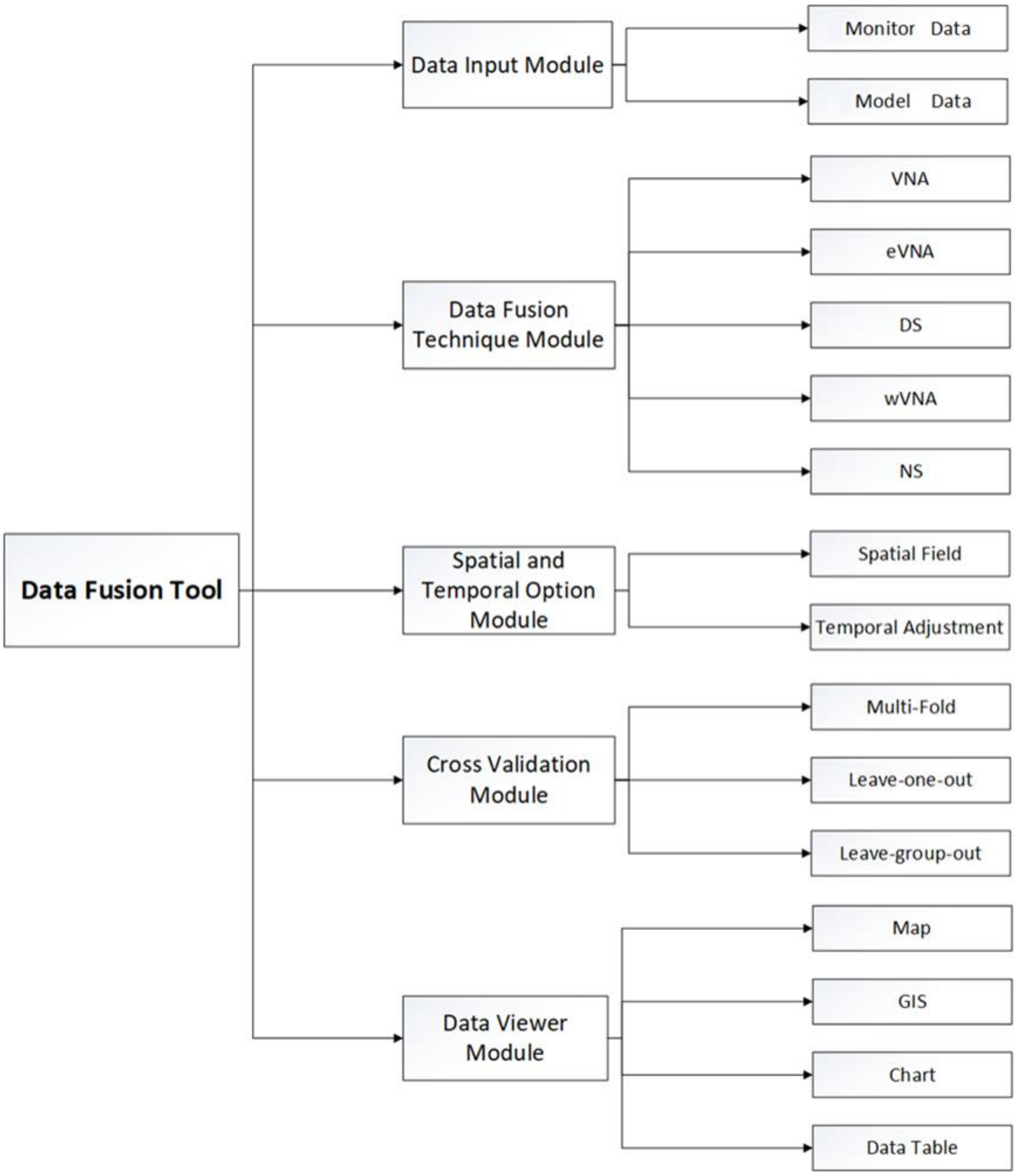


Fig. 10 The functional framework of Data Fusion

➢ **Data input module:** The **Data input module** loads the monitor data and the model data to use for the designed data fusion calculations.

➢ **Data fusion technique module:** The **Data fusion technique module** selects the data fusion methods used for calculations. Five algorithms (VNA, eVNA, DS, wVNA and NS) are provided in this module, and details of these methods can be found in **Appendix I**.

➢ **Spatial and temporal option module:** The **Spatial and temporal option module** sets the spatial field and period used for the designed data fusion calculations.

➢ **Cross validation module:** The **Cross-validation module** evaluates the performance of each data fusion technique and analyzes the results at monitor points under the absence of different scenarios. The results are evaluated by the cross-validation methods, including **Multi-fold**, **Leave-one-out,** and **Leave-group-out** methods listed below:

● **Multi-fold Cross Validation:** Multi-fold cross validation assumes that all monitoring locations are randomly or sequentially divided into K groups across regions. The set of monitors in one group is excluded and then the interpolating algorithm is performed on the remaining monitors. Each algorithm interpolated from the training set is used to make predictions at those excluded monitoring locations. This process is repeated K times, until all groups served as the test set once, thereby creating out-of-sample predictions for all monitoring locations. For example, the program uses a 10-fold CV by default. 10-fold cross validation (10-fold CV) means that all monitoring sites were randomly divided into groups, one of which contains 90% of the sites and the other group contains the remaining 10%. Each algorithm was then applied using data from 90% of the monitoring sites to predict $PM_{2.5}$ at the remaining 10%. This process was repeated for all groups so that predictions could be evaluated against measurements at all sites. The spatial and temporal correlation (e.g., R^2) between predicted $PM_{2.5}$ and monitored $PM_{2.5}$ are calculated and showed in the **Statistic** tab page of the **Data Viewer** Module.

● **Leave-one-out Cross Validation:** Leave-one-out is the most extreme case of K-Fold. The training subset is composed of all samples but one, which is used for validation. The procedure is also repeated K times.

● **Leave-group-out Cross Validation:** Leave-group-out is a clustered cross-validation approach, which splits the points by space used a k-means algorithm. This approach is an unsupervised learning technique requiring a pre-specified number of clusters based on proximity to the nearest mean (defined spatially), effectively grouping points into Voronoi polygons given by those means. The distance metric used to define space is the conventional Euclidean distance. Cross-validation groups are contained within regions under this approach, with a predetermined number of pf clusters per region based on the selected **Radius of Selected Monitors Withheld**.

4 $PM_{2.5}$ Data Fusion Analysis Details

Details of the $PM_{2.5}$ Analysis options in the DF are provided in this Chapter. A default project is used to demonstrate the options for this data fusion analysis.

The $PM_{2.5}$ data fusion analysis in DF is organized into five steps. The steps include the data input and configuration options for interpolation. The following configuration steps correspond to different DF windows are described in detail in this chapter:

➢**Data Input Option:** Specify related data needed to be interpolated in the DF, including the monitor data, species monitor data, model data, and other data.

➢**Data Fusion Technique Option:** Select those algorithms of interest, it is also supported to select multiple algorithms at a time.

➢**Spatial and Temporal Option:** Customize the selection of spatial and temporal for analysis of the model data.

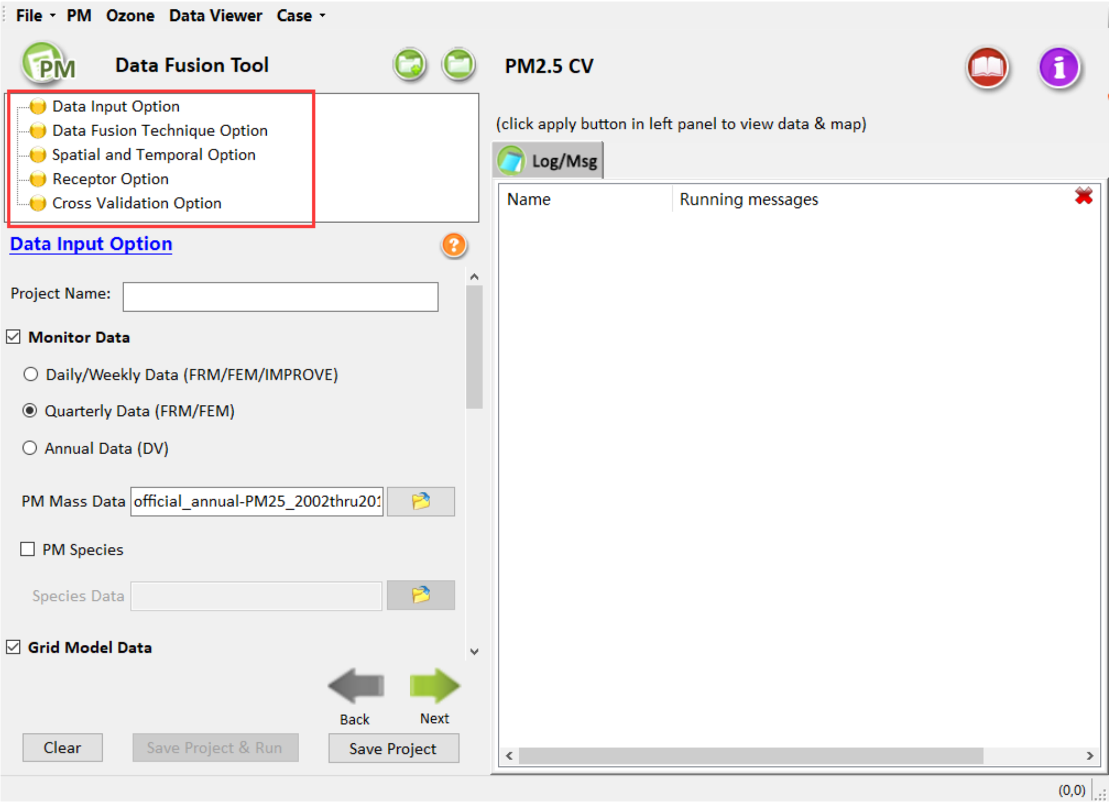
➢**Receptor Option:** Select the corresponding GIS shapefile to calculate the aggregation average concentration of pollutants from Gird to region (e.g., state, county, city) level.

➢**Cross Validation Option:** Select cross validation methods to validate the data fusion results and analyze the results at monitor points under the absence of different scenarios.

To conduct a $PM_{2.5}$ analysis, click the **PM** menu from the main menu on the DF Home Page. Fig. 11 shows the initial window for PM Analysis. The box in the upper left of the window (highlighted in red in Fig. 11) lists the configuration steps of the PM Analysis. Each step listed in this box has a different set of configuration options for PM Analysis. Once a step is successfully configured, the yellow buttons in the box will change from yellow to green. In general, the configuration steps must be followed in order, from top to bottom, as they are listed in the box. Previously completed steps may be accessed and modified by double-clicking on the step name in the box. Once the configuration for a step is complete, users can move to the next step by either clicking on it or by selecting the **Next** button (shown in the blue box in Fig. 11).

A previous project can be loaded by clicking the **Open** button in the **File** drop-down menu item, and a new project may be initiated at any time in the PM Analysis window by clicking **PM** menu on the top of DF start page or selecting **New Project** button from the **File** drop-down menu item.

Each of the PM Analysis configuration steps is described in the following sections.



(click apply button in left panel to view data & map)

Project Name:

☑ Monitor Data

○ Daily/Weekly Data (FRM/FEM/IMPROVE)

● Quarterly Data (FRM/FEM)

○ Annual Data (DV)

PM Mass Data official_annual-PM25_2002thru20:

☐ PM Species

Species Data

☑ Grid Model Data

Fig. 11 Start Page of Creating new $PM_{2.5}$ data fusion

4.1 Data Input Option

Data Input Option is the first step in $PM_{2.5}$ data fusion configuration options. The interface is shown in Fig. 12, including configuration options listed below:

➢ **Project Name:** Text string to identify this analysis and set the name of the project file.

➢ **Monitor Data:** DF requires monitoring data in the form of a text file. DF uses both "**Quarterly Data**" and "**Daily/Weekly Data**" as well as $PM_{2.5}$ species data in its calculations.

Quarterly Data. The EPA-approved quarterly average $PM_{2.5}$ FRM data that have been used to calculate $PM_{2.5}$ design values. These data are commonly used in the attainment test tool (e.g., SMAT-CE). The default data file in DF was created by EPA OAQPS. In most cases, the data should not be altered, however, in some cases (e.g. sensitivity analysis) there may be a need to add or remove data.

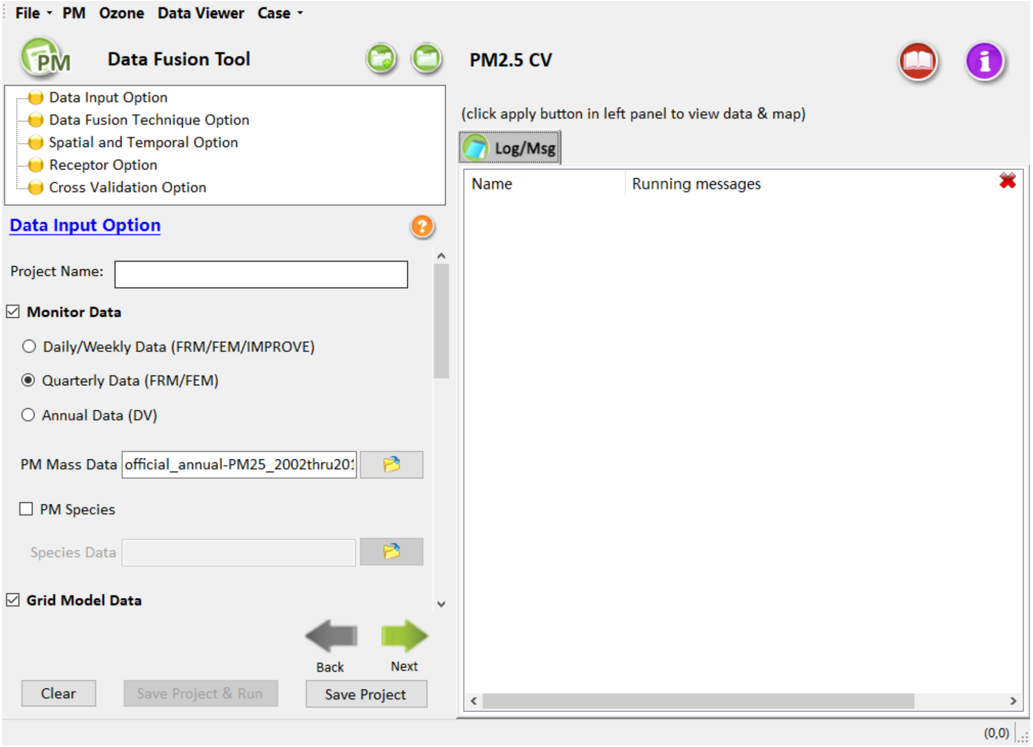
Daily/Weekly Data. $PM_{2.5}$ data that are needed to calculate species fractions. These data are used in combination with the Species Monitor Data File.

A detailed description of the formats of the input data files used in this step is available in Section 2.4 of the User's Manual of SMAT-CE.

➢**Grid Model Data:** These are gridded model output from models such as CMAQ or

CAMx. Users can choose either daily model data input or quarterly model data input (which is just a quarterly average of the daily model data). Either will work for the $PM_{2.5}$ Analysis. The default setting is daily average data. It supports two kinds of formats: one is the simple comma-separated text format (*.csv) and the other is IOAPI format. Note that the model data year should be consistent with the base year of monitor data.

➢**Other Data:** Other data mainly include dispersion model data, related geographic information data, and Aerosol Optical Depth (AOD) data. AOD data is mainly used for satellite data fusion. Users can choose according to their requirements. The program does not use it by default.



Project Name:

☑ Monitor Data

○ Daily/Weekly Data (FRM/FEM/IMPROVE)

● Quarterly Data (FRM/FEM)

○ Annual Data (DV)

PM Mass Data: official_annual-PM25_2002thru20:

☐ PM Species

Species Data

☑ Grid Model Data

Fig. 12 Data Input Option ($PM_{2.5}$)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Data Fusion Technique Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

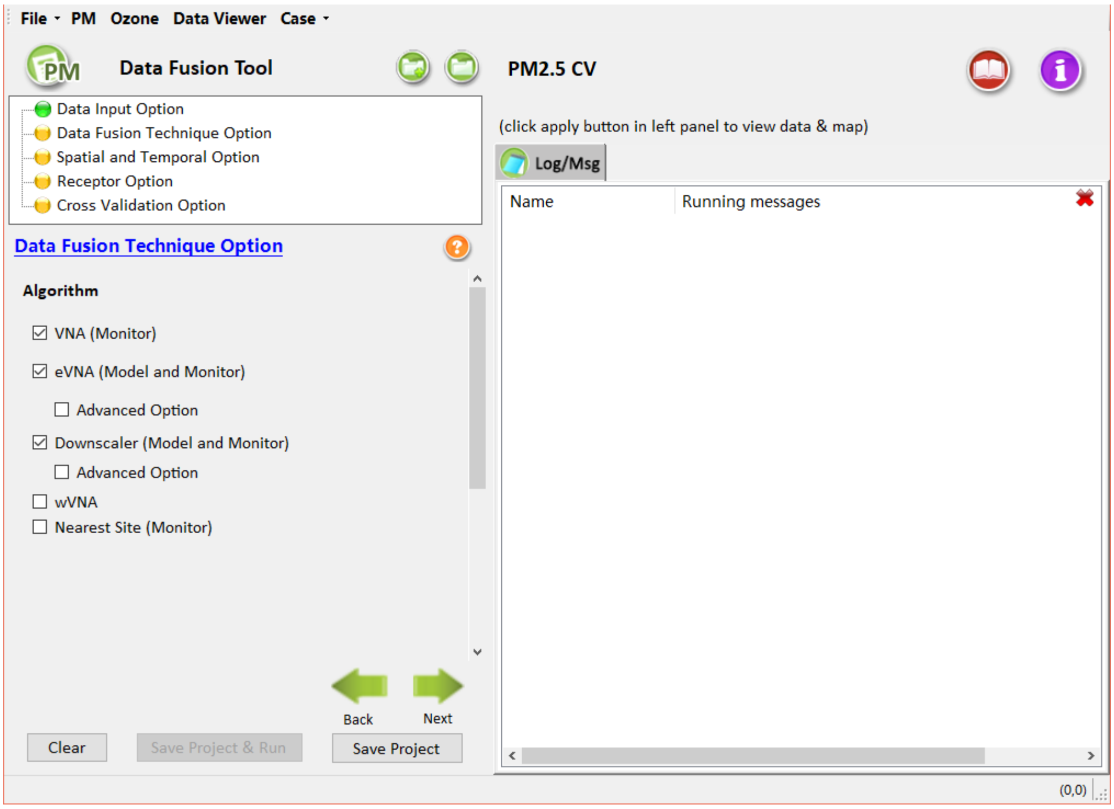
4.2 Data Fusion Technique Option

Data Fusion Technique Option includes five data fusion algorithms. Users can choose one or more algorithms of interest to calculate at a time. The interface is shown in Fig. 13, including five algorithms of VNA, eVNA, DS, wVNA, and NS. The preferred default choices are VNA, eVNA, and Downscaler. The calculation principle of these five algorithms are described in **Appendix I**.

➢Users can check the **Advanced options** under the **eVNA** algorithm option to set the Inverse Distance Weighting Power (the default is 2) and the Gradient-adjusted threshold (GAT, the default is 3), as shown in Fig. 14. Inverse Distance Weighting Power means that the value is the sum of the coefficient of the distance from the monitor point to the interpolation point. Gradient-adjusted threshold means that the value in the model grid cell will be replaced with the threshold value (= GAT × $C_{Monitor}$) if the ratio of modeled concentration to monitored concentration in a target grid cell is larger than the Gradient-adjusted threshold.

➢Users can check the **Advanced options** under the **Downscaler** algorithm option to set the number of iterations and related parameters to reduce the computational load, and improve the calculation effectiveness, as shown in Fig. 15.

➢Users can check the **Advanced options** under the **wVNA** algorithm option to set the value of weight between VNA and eVNA, as shown in Fig. 16.



Algorithm

☑ VNA (Monitor)

☑ eVNA (Model and Monitor)

☐ Advanced Option

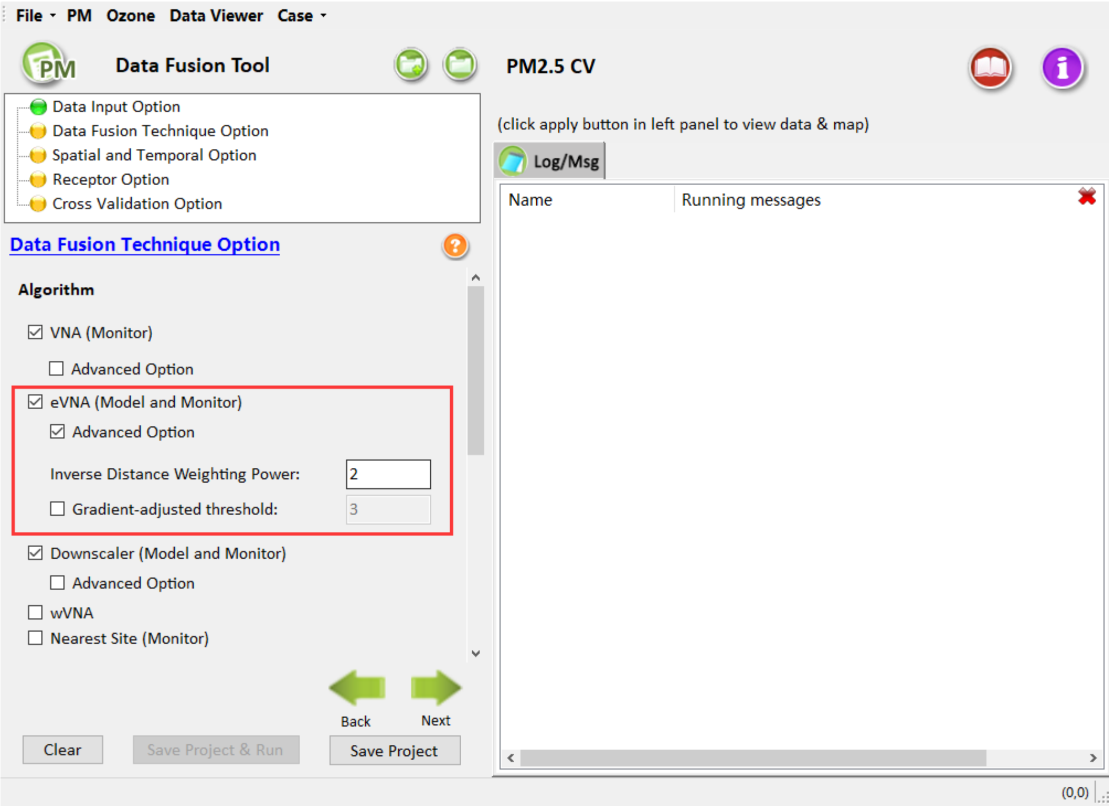
☑ Downscaler (Model and Monitor)

☐ Advanced Option

☐ wVNA

☐ Nearest Site (Monitor)

Fig. 13 Data Fusion Technique Option ($PM_{2.5}$)



(click apply button in left panel to view data & map)

Name	Running messages

Data Fusion Technique Option

Algorithm

☑ VNA (Monitor)

☐ Advanced Option

☑ eVNA (Model and Monitor)

☑ Advanced Option

Inverse Distance Weighting Power: 2

☐ Gradient-adjusted threshold: 3

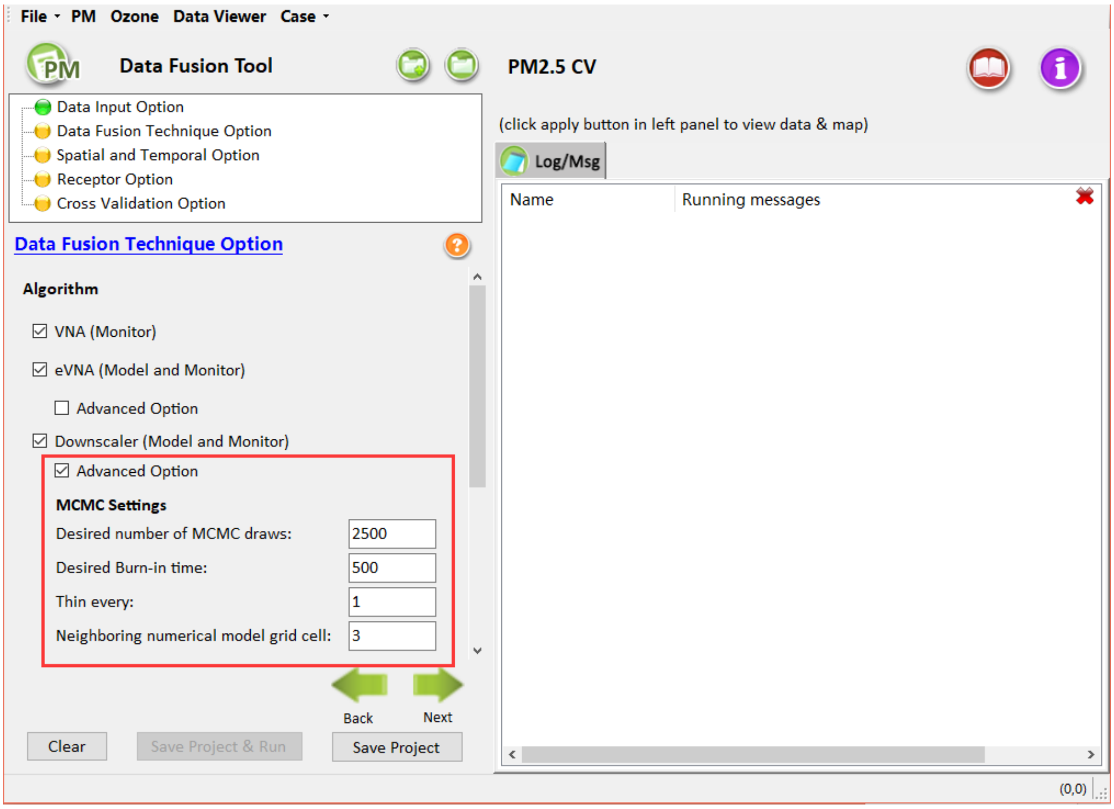
☑ Downscaler (Model and Monitor)

☐ Advanced Option

☐ wVNA

☐ Nearest Site (Monitor)

Fig. 14 Advanced Option in eVNA ($PM_{2.5}$)



(click apply button in left panel to view data & map)

Name	Running messages

Data Fusion Technique Option

Algorithm

☑ VNA (Monitor)

☑ eVNA (Model and Monitor)

☐ Advanced Option

☑ Downscaler (Model and Monitor)

☑ Advanced Option

MCMC Settings

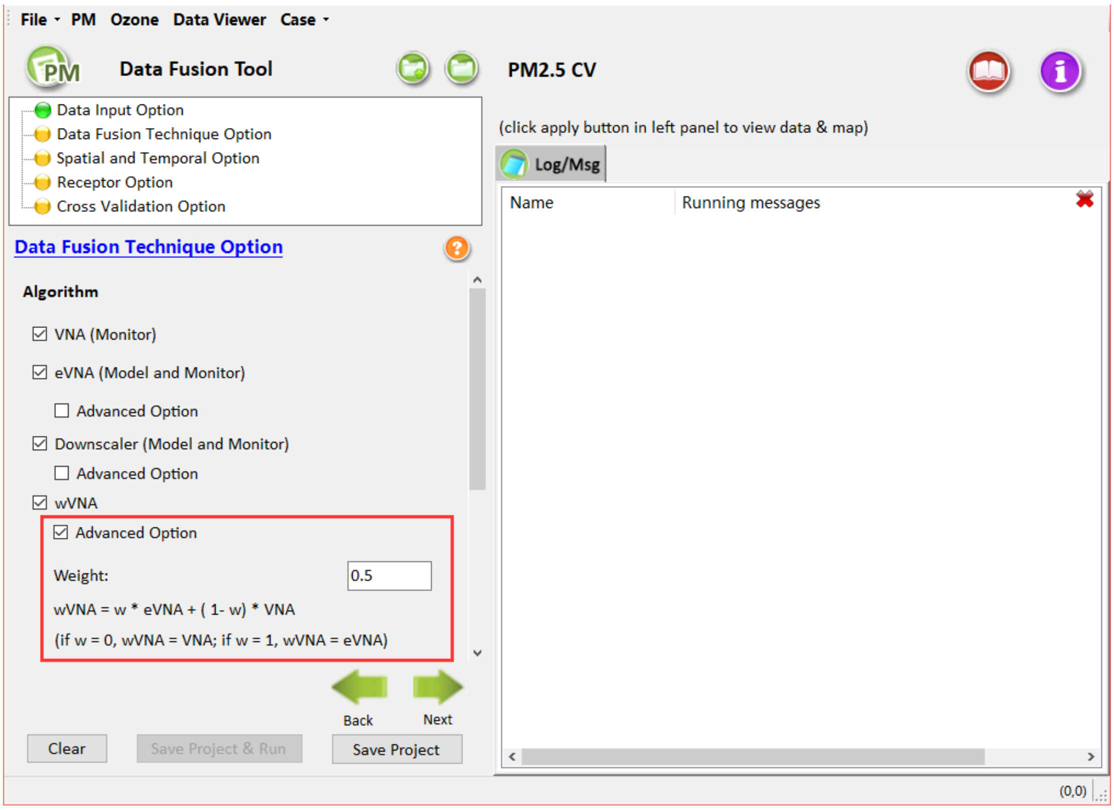
Desired number of MCMC draws: 2500

Desired Burn-in time: 500

Thin every: 1

Neighboring numerical model grid cell: 3

Fig. 15 Advanced Option in Downscaler ($PM_{2.5}$)



Data Fusion Technique Option

(click apply button in left panel to view data & map)

Algorithm

☑ VNA (Monitor)

☑ eVNA (Model and Monitor)

☐ Advanced Option

☑ Downscaler (Model and Monitor)

☐ Advanced Option

☑ wVNA

☑ Advanced Option

Weight: 0.5

wVNA = w * eVNA + (1- w) * VNA

(if w = 0, wVNA = VNA; if w = 1, wVNA = eVNA)

Name	Running messages

Fig. 16 Advanced Option in wVNA ($PM_{2.5}$)

The default configuration is that the VNA, eVNA, and Downscalar. algorithm options are selected. Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Spatial and Temporal Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

4.3 Spatial and Temporal Option

The interface of **Spatial and Temporal Option** is shown in Fig. 17, including Spatial Data Fusion Option and Temporal Data Fusion Option. Users can specify the spatial and temporal options.

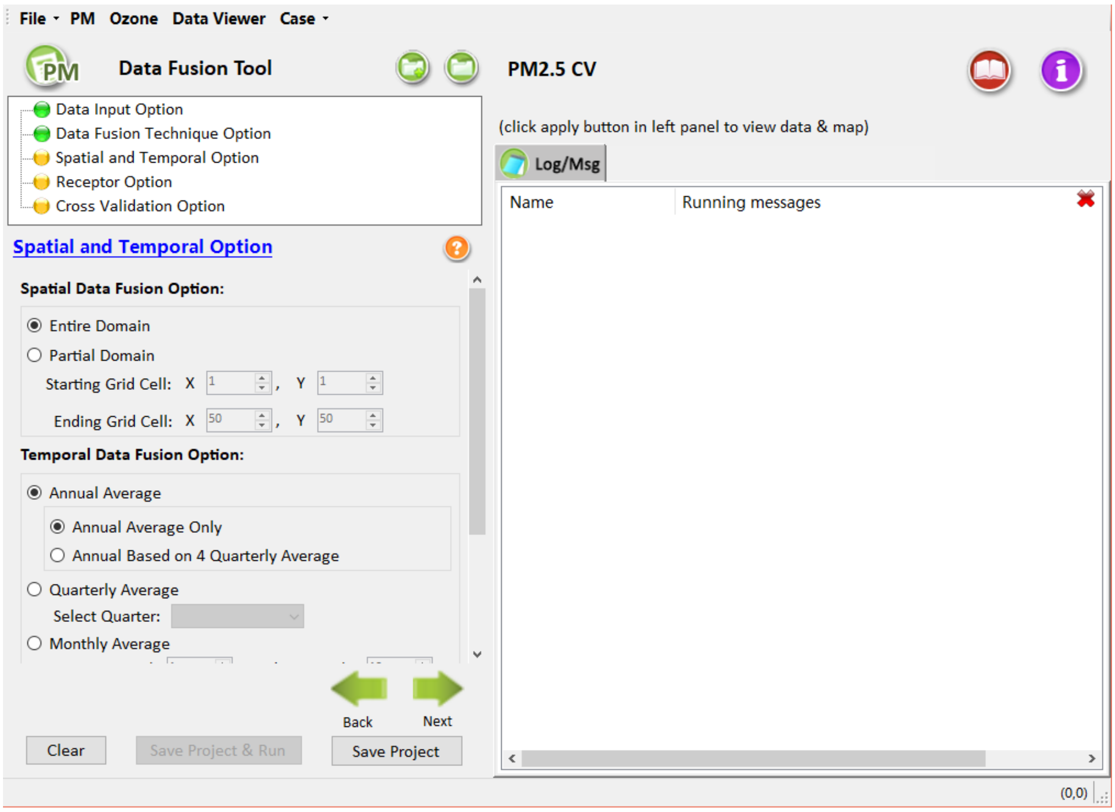
➢**Spatial Data Fusion Option:** Set the number of the model grid cells used in the calculations to interpolate from the monitor data, including the **Entire Domain** and customized **Partial Domain**. For the **Entire Domain**, DF will output the results for all specified monitors within the domain. For the **Partial Domain**, DF will create a spatial filed that matches the size of the gridded model domain according to the **Starting Grid Cell** and **Ending Grid Cell.**

➢ **Temporal Data Fusion Option:** Set the period used in the calculations to interpolate from the monitor data. It is divided into the **Annual Average**, **Quarterly Average**,

Monthly Average, and **Daily Average**.

When selecting the **Daily/Weekly Data** option at the step of **Data Input Option**(chapter 4.1)**,** all options of **Temporal Data Fusion Option** will be available. When selecting the **Quarterly Data** option at the step of **Data Input Option, Monthly Average** and **Daily Average** will be unavailable. When selecting the **Annual Data** option at the step of **Data Input Option,** only the Annual Average will be available and other options will be greyed out.

DF will calculate the **Annual Average**, **Quarterly Average**, **Monthly Average**, and **Daily Average** from the input DF formatted daily average gridded model files or quarterly average files and optionally output the annual, quarterly, monthly, and daily average concentrations into text files (CSV files) respectively.



(click apply button in left panel to view data & map)

Name	Running messages

Spatial and Temporal Option

Spatial Data Fusion Option:

● Entire Domain

○ Partial Domain

Starting Grid Cell: X 1 , Y 1

Ending Grid Cell: X 50 , Y 50

Temporal Data Fusion Option:

● Annual Average

● Annual Average Only

○ Annual Based on 4 Quarterly Average

○ Quarterly Average

Select Quarter:

○ Monthly Average

Fig. 17 Spatial and Temporal Option ($PM_{2.5}$)

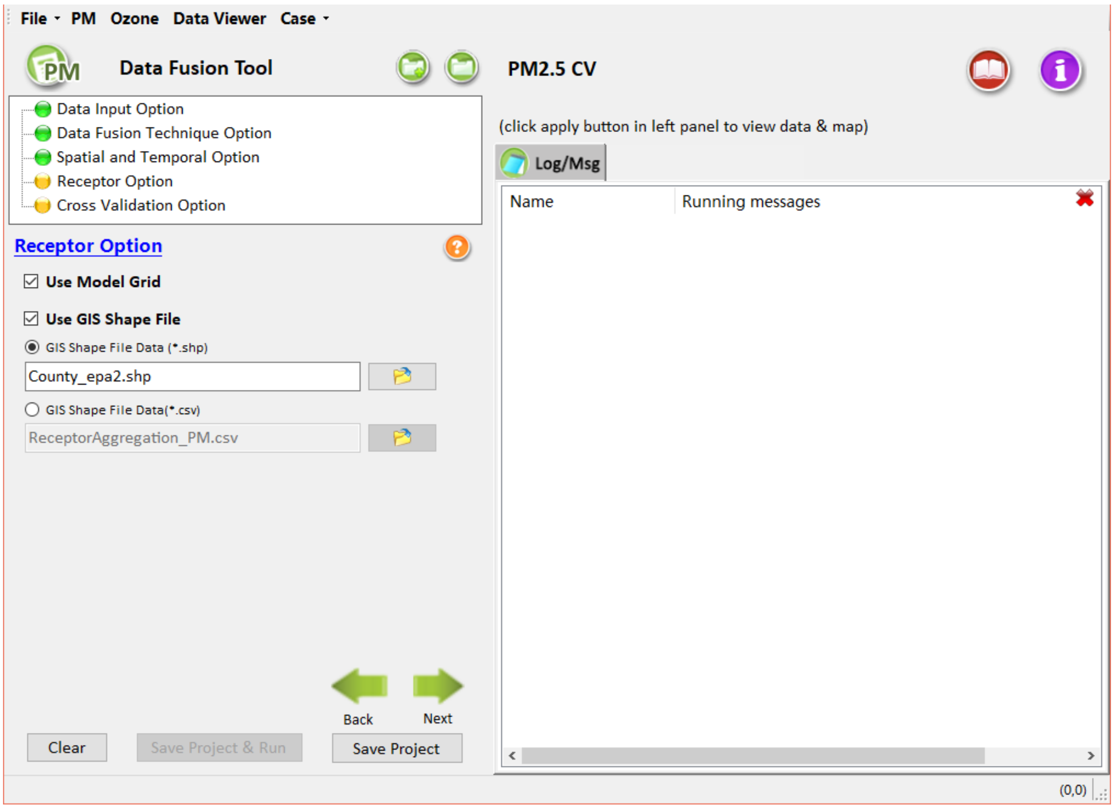
Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Receptor Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

4.4 Receptor Option

The interface of **Receptor Option** is shown in Fig. 18, including **Use Model Grid** and **Use GIS Shape File** options.

➢ **Use Model Grid:** The model grid file in the area of data fusion (checked in default).

➢ **Use GIS Shape File:** The average concentration of pollutants in each region, which can be aggregated by selecting the GIS-related file of the analysis area in the interface of Receptor Option. The format of the input file can be either *.shp or *.csv.



Receptor Option

☑ Use Model Grid

☑ Use GIS Shape File

● GIS Shape File Data (*.shp)

County_epa2.shp

○ GIS Shape File Data(*.csv)

ReceptorAggregation_PM.csv

(click apply button in left panel to view data & map)

Fig. 18 Receptor Option ($PM_{2.5}$)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Cross Validation Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

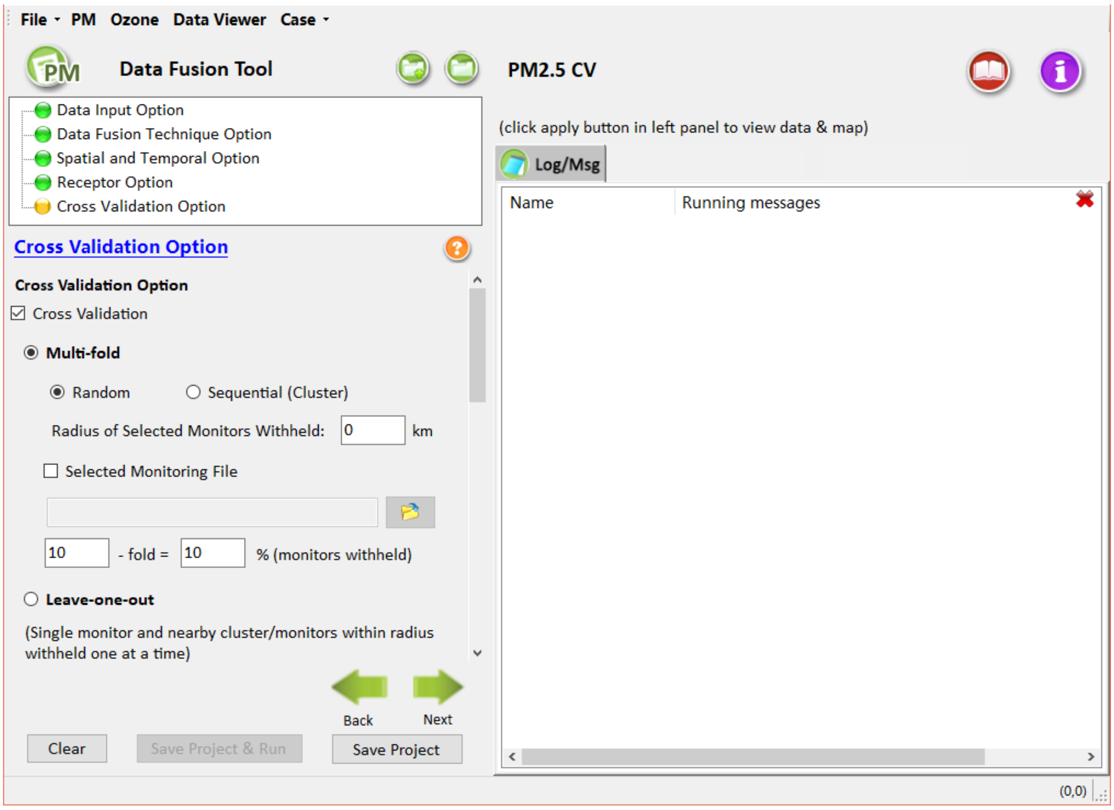
4.5 Cross Validation Option

The data fusion results can be evaluated by the following three cross-validation methods: **Multi-fold**, **Leave-one-out,** and **Leave-group-out** methods (Fig. 19). Users can select a method of interest at a time. Users can also set a **Radius of Selected Monitors Withheld** to implement the Cluster analysis for leaving-out the cluster of monitors within a specific radius of selected monitors. The calculation principle of these five algorithms is described in **Appendix I**.

➢ Under the **Multi-fold** cross validation option (checked in default), users can set the number of groups and grouping types (**Random** and **Sequential (Cluster)**), as shown in Fig. 19. **Random** means all monitor sites are randomly divided into specified groups, while **Sequential** means all monitor sites are grouped by the order of monitors. Users can also set the **Radius of Selected Monitors Withheld** to implement a radius of withholding for the random & sequential Multi-Fold cross validation. The **Radius of Selected Monitors Withheld** means to remove nearby clusters/monitors with radius withheld at a time. For example, the radius = 0 km means just removing selected monitor only at a time, while radius = 1 km means removing a cluster of monitors within 1km for each selected monitor.

➢ Under the **leave-one-out** cross validation option, users can select the monitor sites to leave out by importing it into the right panel, as shown in Fig. 20. The option means that a single monitor or nearby cluster/monitors within the radius will be removed from the selected monitors at a time.

➢ Under **the leave-group-out** cross validation option, users can select the group data sets to be verified and import it into the right panel, as shown in Fig. 21. Users can also set the **Radius of Selected Monitors Withheld** to leave a group of individual monitors and nearby clusters/monitors with radius withheld at the same time.



PM2.5 CV

(click apply button in left panel to view data & map)

Name | Running messages

Cross Validation Option

Cross Validation Option

☑ Cross Validation

● Multi-fold

● Random ○ Sequential (Cluster)

Radius of Selected Monitors Withheld: 0 km

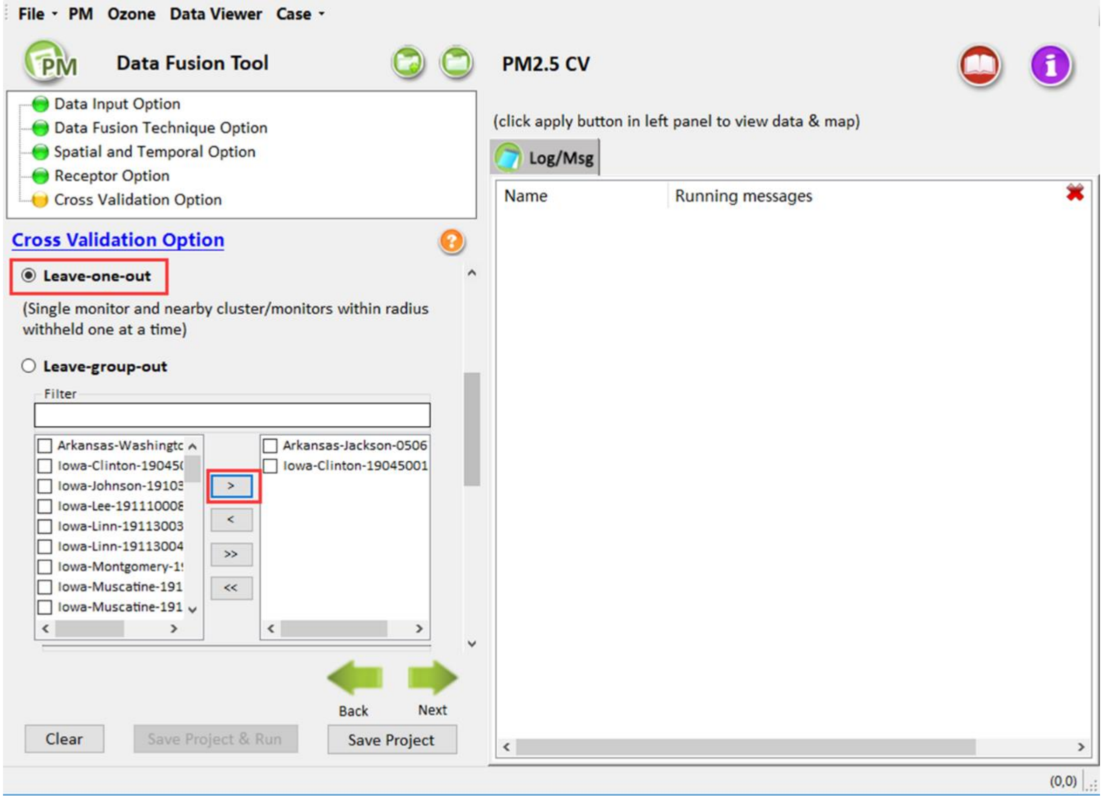
☐ Selected Monitoring File

10 - fold = 10 % (monitors withheld)

○ Leave-one-out

(Single monitor and nearby cluster/monitors within radius withheld one at a time)

Fig. 19 Cross Validation Option ($PM_{2.5}$)



PM2.5 CV

(click apply button in left panel to view data & map)

Name | Running messages

Cross Validation Option

● Leave-one-out

(Single monitor and nearby cluster/monitors within radius withheld one at a time)

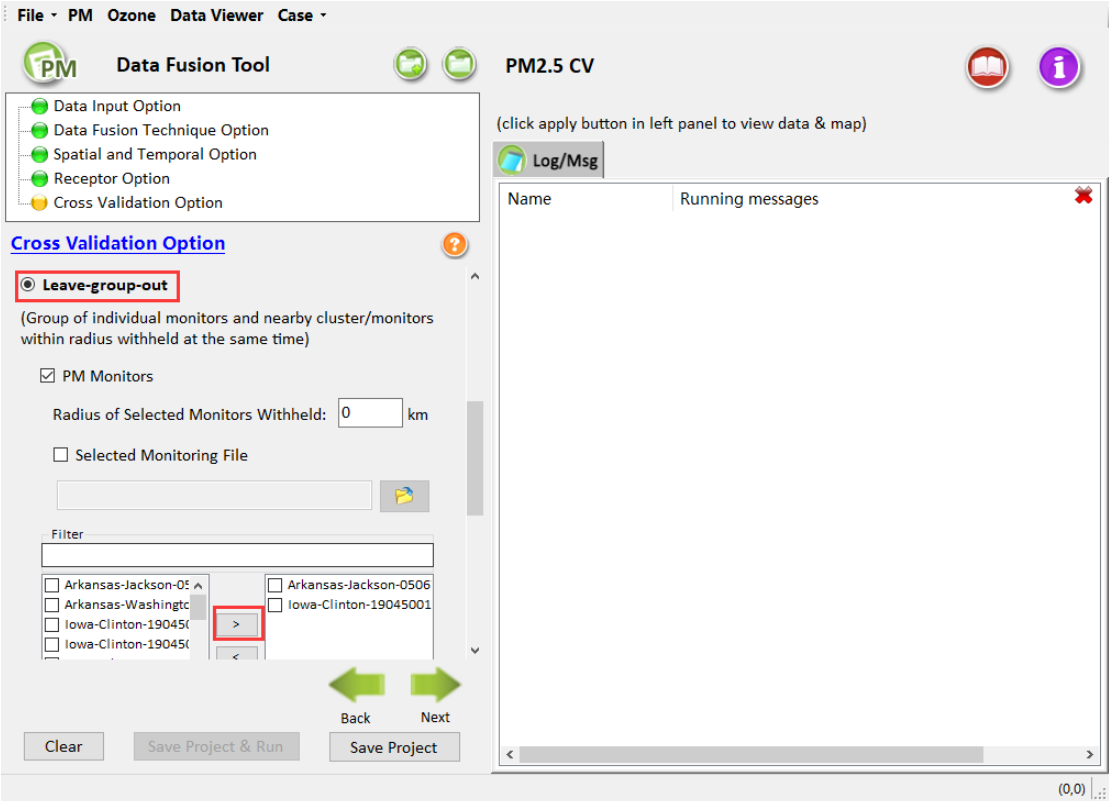
○ Leave-group-out

Filter

☐ Arkansas-Washingtc
☐ Iowa-Clinton-19045(
☐ Iowa-Johnson-19103
☐ Iowa-Lee-19111000{
☐ Iowa-Linn-19113003
☐ Iowa-Linn-19113004
☐ Iowa-Montgomery-1!
☐ Iowa-Muscatine-191
☐ Iowa-Muscatine-191

☐ Arkansas-Jackson-0506
☐ Iowa-Clinton-19045001

Fig. 20 Leave-one-out Cross Validation



Data Input Option
Data Fusion Technique Option
Spatial and Temporal Option
Receptor Option
Cross Validation Option

Cross Validation Option

● Leave-group-out

(Group of individual monitors and nearby cluster/monitors within radius withheld at the same time)

☑ PM Monitors

Radius of Selected Monitors Withheld: 0 km

□ Selected Monitoring File

Filter

□ Arkansas-Jackson-05
□ Arkansas-Washingtc
□ Iowa-Clinton-19045(
□ Iowa-Clinton-19045(

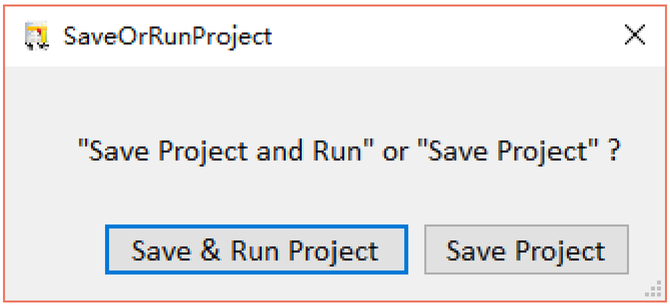
□ Arkansas-Jackson-0506
□ Iowa-Clinton-19045001

(click apply button in left panel to view data & map)

Name	Running messages

Fig. 21 Leave-group-out Cross Validation

Users can keep default options and configuration, or change options and configuration according to their requirements. After completing this step, all configuration options for creating new $PM_{2.5}$ data fusion have been completed. Click the **Next** button to **Save Project**, or **Save & Run Project**, as shown in Fig. 22. Users can also click the **Save Project & Run** or **Save Project** button at the left bottom of the interface to save the project.



"Save Project and Run" or "Save Project" ?

Fig. 22 Save & Run Project

5 Ozone Data Fusion Analysis Details

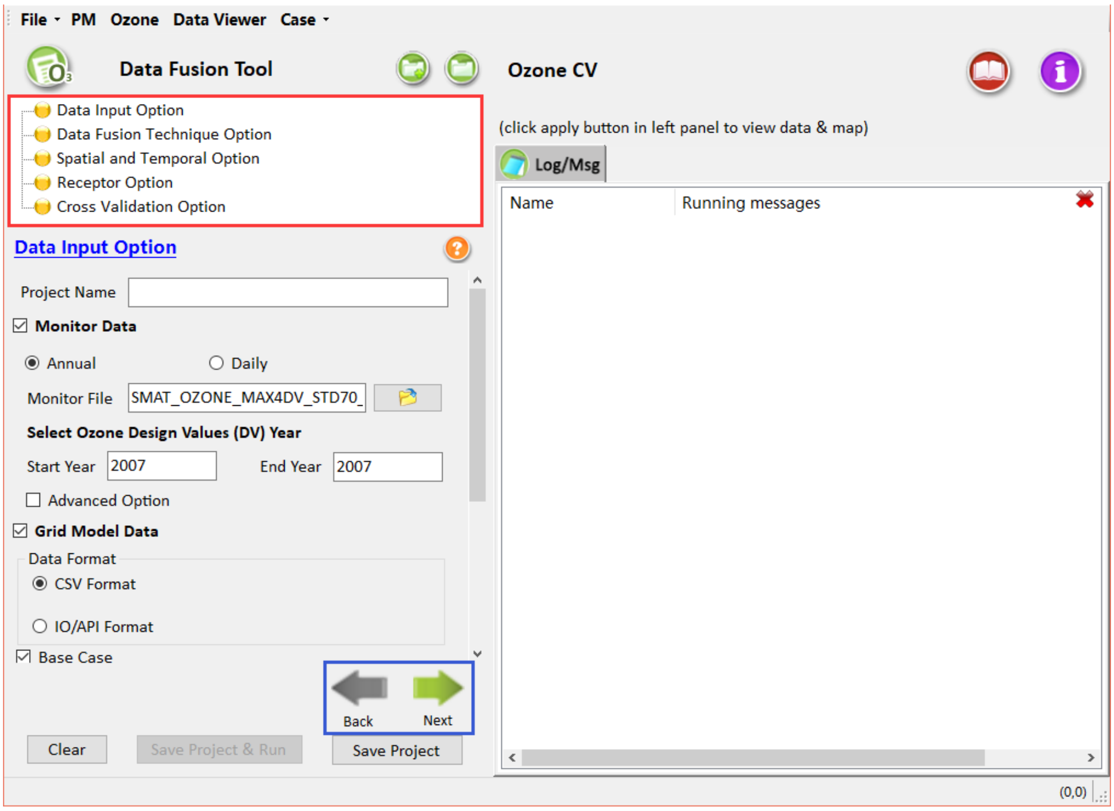
Details of the Ozone Analysis options in the DF are provided in this Chapter. A default project is used to demonstrate the options for this data fusion analysis.

The steps of creating new Ozone data fusion analysis are similar to the steps of creating new $PM_{2.5}$ data fusion analysis.

To conduct an Ozone analysis, click the **Ozone** menu from the main menu on the DF Start Page. Fig. 23 shows the initial window that is displayed when the Ozone Analysis is selected. The box in the upper left of the window (highlighted in red in Fig. 23) lists the configuration steps of the Ozone Analysis. Each step listed in this box has a different set of configuration options that are displayed in the Ozone Analysis window. Once each step is successfully configured, the yellow buttons in the box will change from yellow to green. In general, the configuration steps must be followed in order, from top to bottom, as they are listed in the box. Previously completed steps may be accessed and modified by double-clicking on the step name in the box. Once the configuration for a step is complete, users can move to the next step by either clicking on it or by selecting the Next button (shown in the blue box in Fig. 23).

A previous project can be loaded by clicking the **Open** button in the **File** drop-down menu item, and a new project may be initiated at any time in the Ozone Analysis window by selecting **Ozone** menu on the top of DF start page or clicking **New Project** button from the **File** drop-down menu item.

Each of the Ozone Analysis configuration steps is described in the following sections.



(click apply button in left panel to view data & map)

Data Input Option

Project Name

☑ Monitor Data

● Annual ○ Daily

Monitor File SMAT_OZONE_MAX4DV_STD70_

Select Ozone Design Values (DV) Year

Start Year 2007 End Year 2007

□ Advanced Option

☑ Grid Model Data

Data Format

● CSV Format

○ IO/API Format

☑ Base Case

Name	Running messages

(0,0)

Fig. 23 Start Page of Creating new Ozone data fusion

5.1 Data Input Option

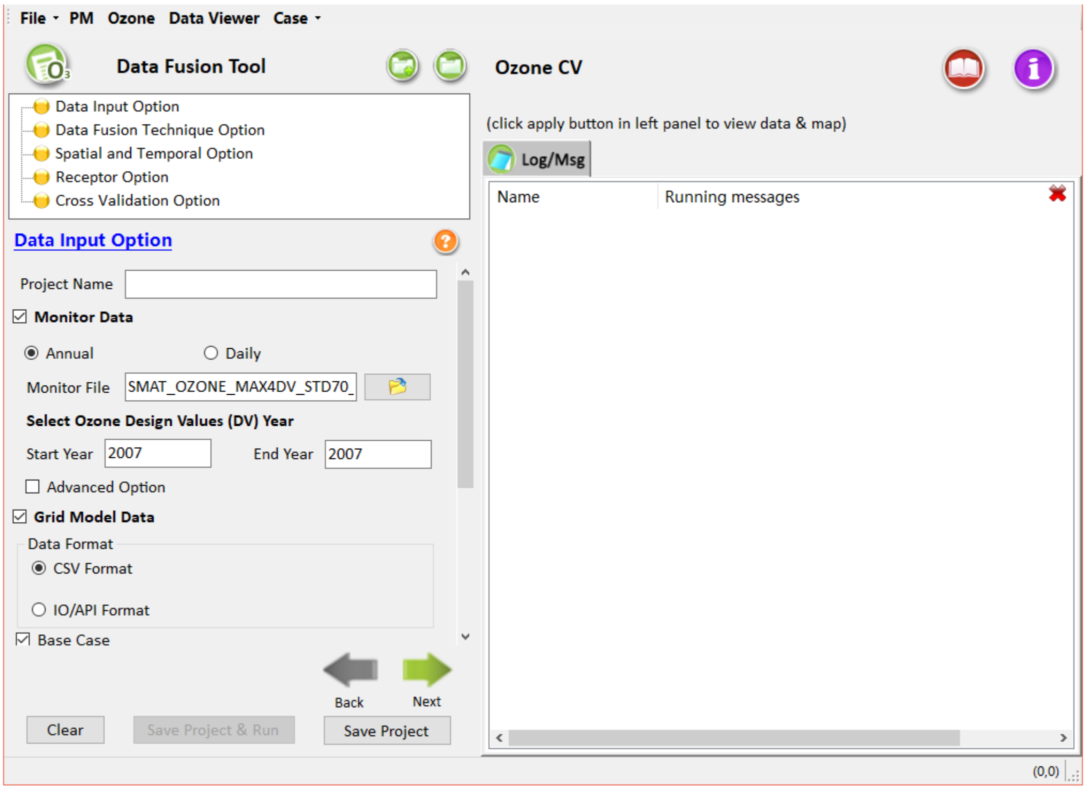
Data Input Option is the first step in Ozone data fusion configuration options. The interface is shown in Fig. 24, including configuration options listed below:

➢ **Project Name:** Text string to identify this analysis and set the name of the project file.

➢ **Monitor Data:** DF requires monitoring data in the form of a text file. DF uses both "**Annual**" Data and "**Daily**" Data in its calculations. The start and end years should be changed to applicable periods, depending on the base modeling year. Users can set the additional **Advanced Option** to specify the area of grid cells used in calculations.

➢ **Grid Model Data:** These are gridded model output from models such as CMAQ or CAMx. Users can choose either daily model data input or annual model data input (which is just an annual average of the daily model data). Either will work for the Ozone Analysis. The default setting is the annual average data. It supports two kinds of formats: one is the simple comma-separated text format (*.csv) and the other is IOAPI format. The model value and the monitor value can be matched automatically according to the selected daily monitor data, as shown in Fig. 25.

➢ **Other Data:** Other data mainly include dispersion data and related geographic information data. Users can choose according to their requirements. The program does not use it by default.



(click apply button in left panel to view data & map)

Name	Running messages

Project Name

☑ **Monitor Data**

● Annual ○ Daily

Monitor File SMAT_OZONE_MAX4DV_STD70_

Select Ozone Design Values (DV) Year

Start Year 2007 End Year 2007

☐ Advanced Option

☑ **Grid Model Data**

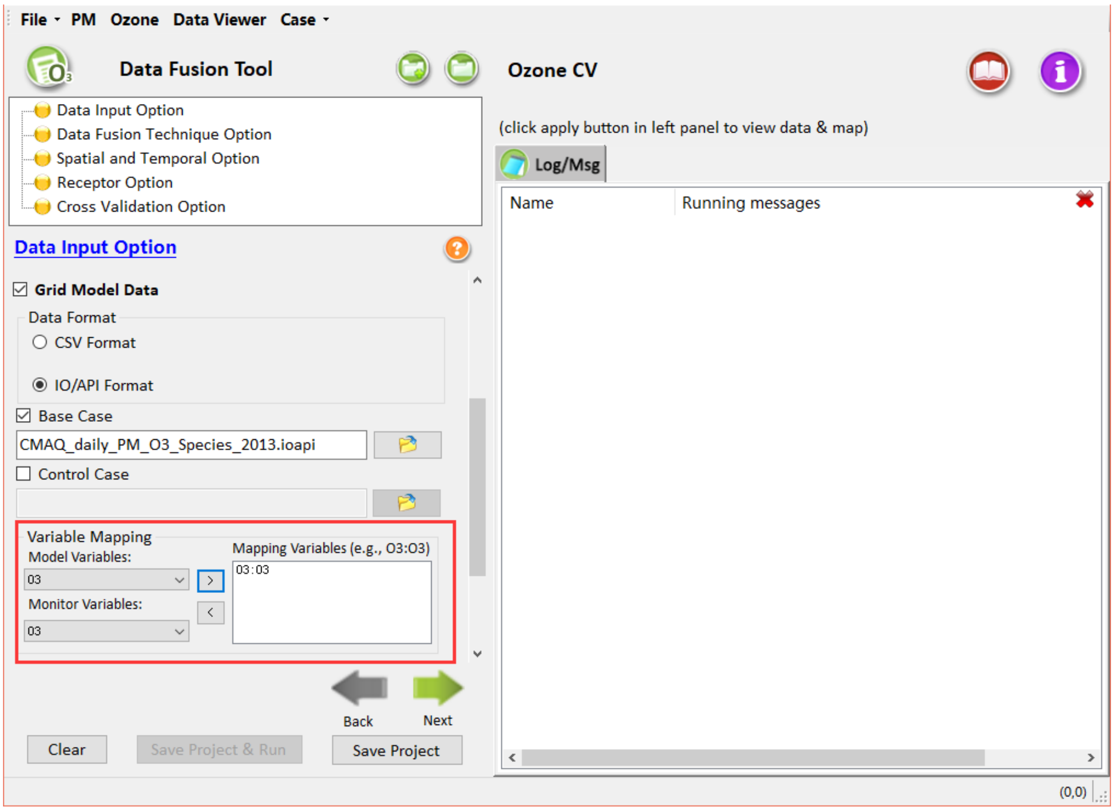
Data Format

● CSV Format

○ IO/API Format

☑ Base Case

Fig. 24 Data Input Option (Ozone)



(click apply button in left panel to view data & map)

Name	Running messages

☑ **Grid Model Data**

Data Format

○ CSV Format

● IO/API Format

☑ Base Case

CMAQ_daily_PM_O3_Species_2013.ioapi

☐ Control Case

Variable Mapping

Model Variables: O3

Monitor Variables: O3

Mapping Variables (e.g., O3:O3)

O3:O3

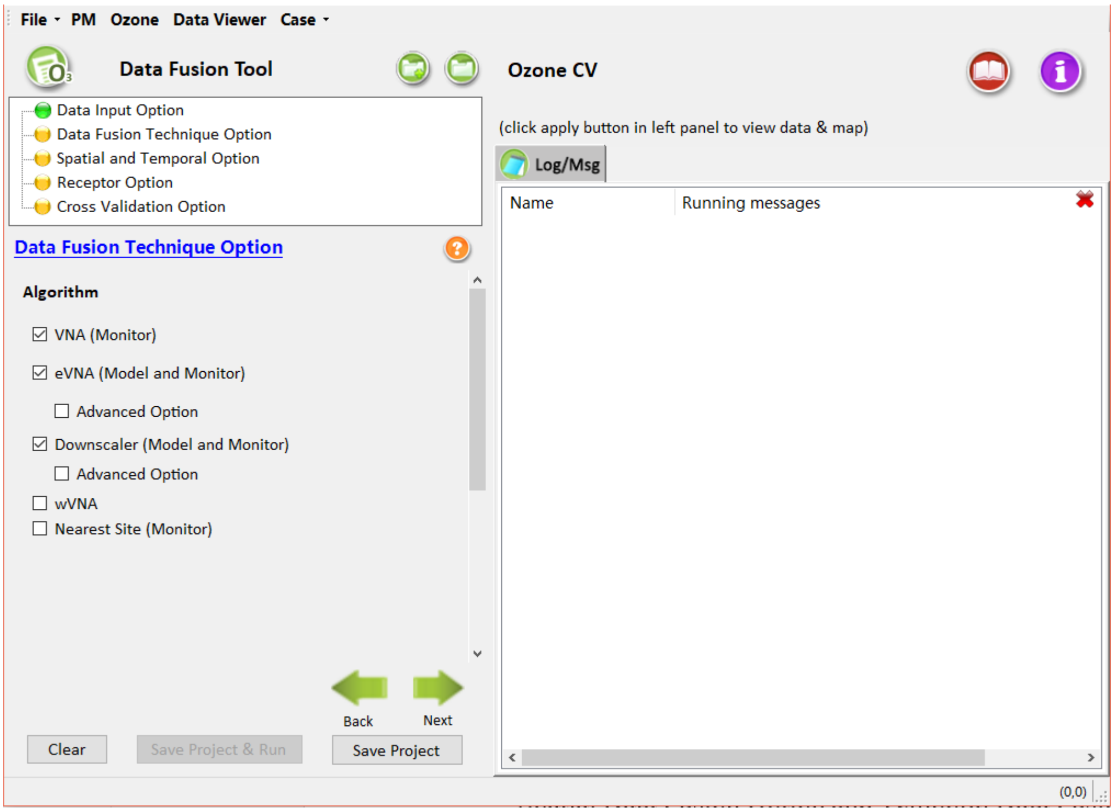
Fig. 25 Variable Mapping (Ozone)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the

Next button or double-click the node of **Data Fusion Technique Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

5.2 Data Fusion Technique Option

The interface design of the **Data Fusion Technique Option** for Ozone data fusion analysis is similar to that of $PM_{2.5}$ data fusion analysis, as shown in Fig. 26. The preferred default choices are VNA, eVNA, and Downscaler. Users can refer to chapter 4.2 above.



(click apply button in left panel to view data & map)

Name	Running messages

Data Fusion Technique Option

Algorithm

☑ VNA (Monitor)

☑ eVNA (Model and Monitor)

☐ Advanced Option

☑ Downscaler (Model and Monitor)

☐ Advanced Option

☐ wVNA

☐ Nearest Site (Monitor)

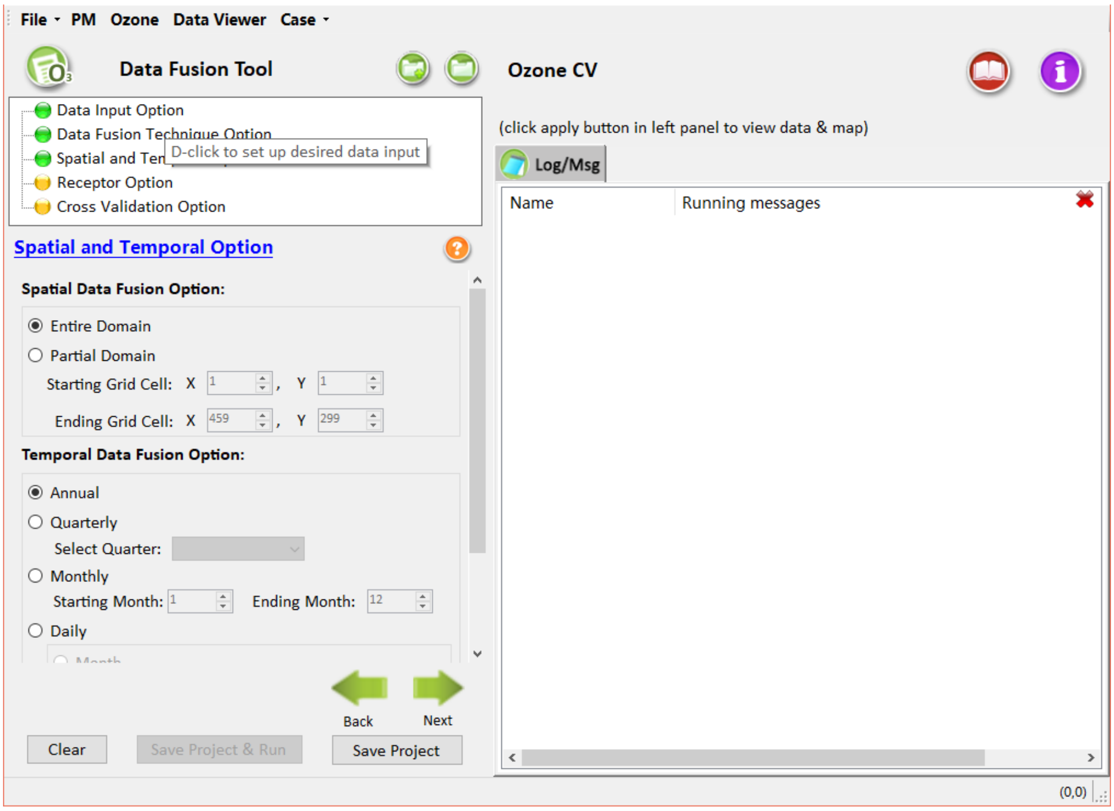
Fig. 26 Data Fusion Technique Option (Ozone)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Spatial and Temporal Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

5.3 Spatial and Temporal Option

The interface of Ozone **Spatial and Temporal Option** is shown in Fig. 27, including Spatial Data Fusion Option and Temporal Data Fusion Option. Users can specify the spatial and temporal options.

- **Spatial Data Fusion Option:** Set the number of model grid cells used in the calculations to interpolate from the monitor data, including the **Entire Domain** and customized **Partial Domain**. For the **Entire Domain**, DF will output the results for all specified monitors within the domain. For the **Partial Domain**, DF will create a spatial filed that matches the size of the gridded model domain according to the **Starting Grid Cell** and **Ending Grid Cell.**
- **Temporal Data Fusion Option:** Set the period to use in the calculations to interpolate from the monitor data. It is divided into the **Annual Average**, **Quarterly Average**, **Monthly Average**, and **Daily Average**. Note that the **Annual Average** option for Ozone data fusion analysis **is not available for annual average calculations based on quarterly values,** as shown in Fig. 28.



Spatial Data Fusion Option:

● Entire Domain

○ Partial Domain

Starting Grid Cell: X 1 , Y 1

Ending Grid Cell: X 459 , Y 299

Temporal Data Fusion Option:

● Annual

○ Quarterly

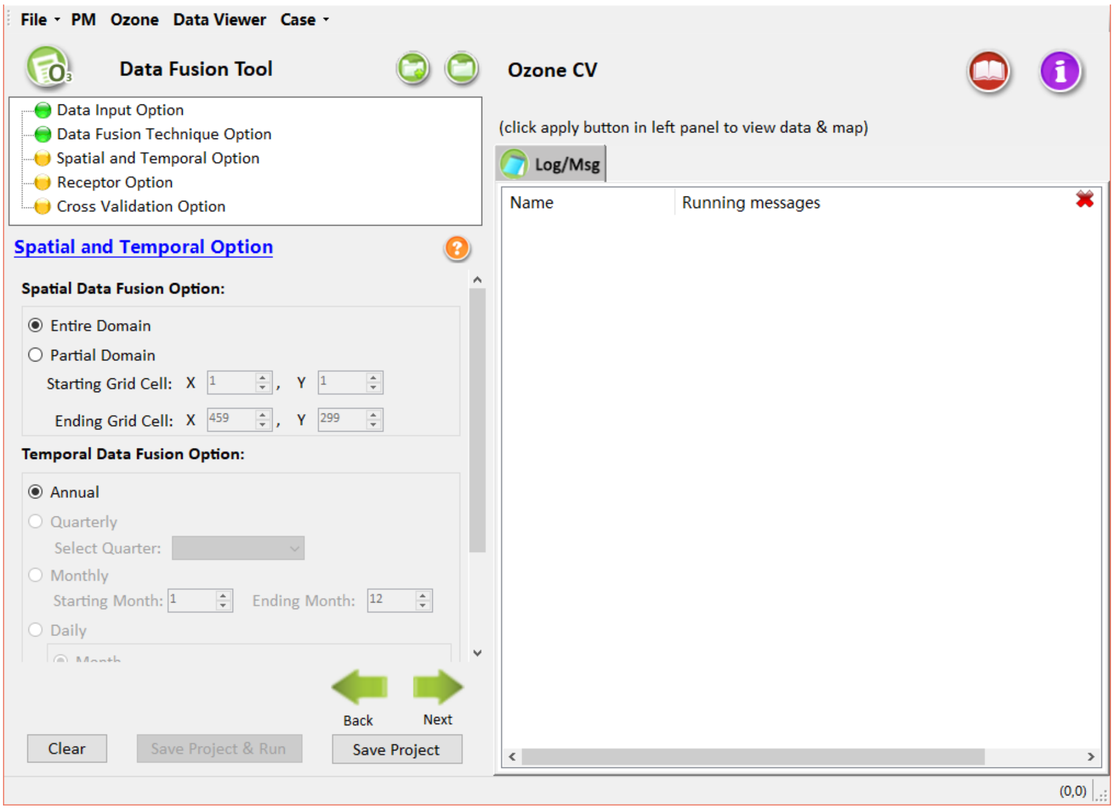
Select Quarter:

○ Monthly

Starting Month: 1 Ending Month: 12

○ Daily

Fig. 27 Spatial and Temporal Option (Ozone)



- Data Input Option
- Data Fusion Technique Option
- Spatial and Temporal Option
- Receptor Option
- Cross Validation Option

(click apply button in left panel to view data & map)

Spatial and Temporal Option

Spatial Data Fusion Option:

● Entire Domain

○ Partial Domain

Starting Grid Cell: X 1 , Y 1

Ending Grid Cell: X 459 , Y 299

Temporal Data Fusion Option:

● Annual

○ Quarterly

Select Quarter:

○ Monthly

Starting Month: 1 Ending Month: 12

○ Daily

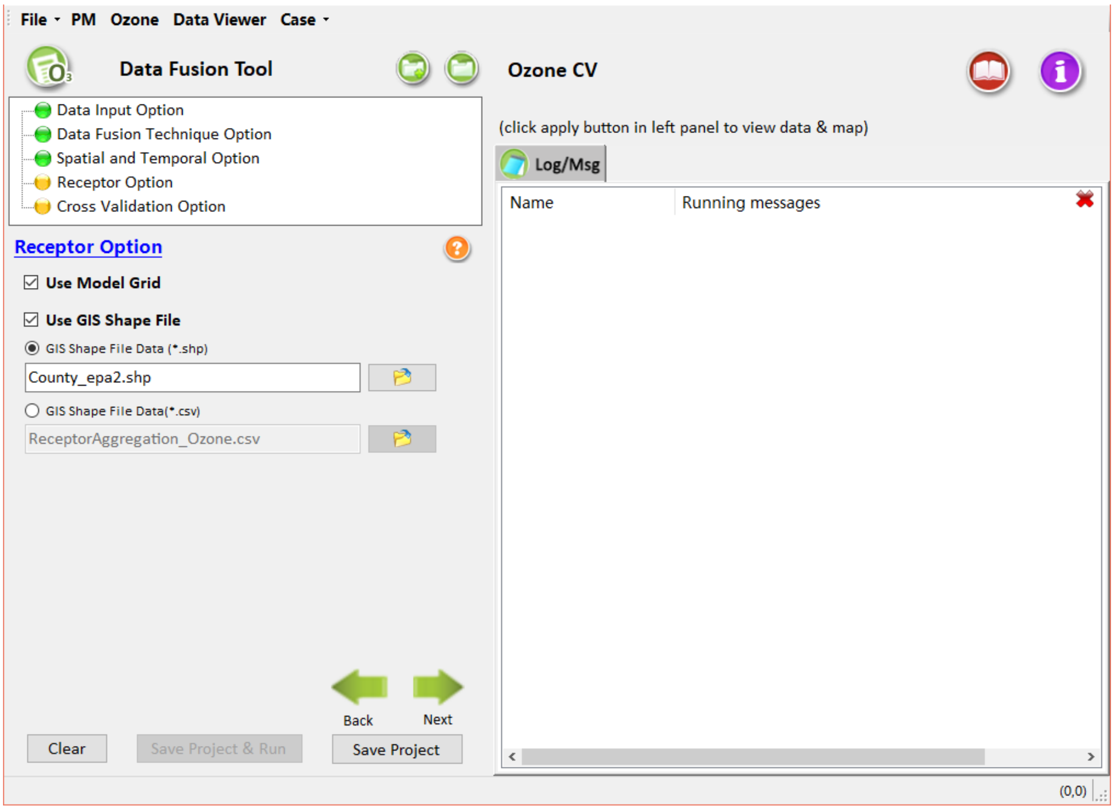
Month

Fig. 28 Spatial and Temporal Option - Only Annual Average (Ozone)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Receptor Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step is ready.

5.4 Receptor Option

The interface design of **Receptor Option** for Ozone data fusion analysis is similar to that of $PM_{2.5}$ data fusion analysis, as shown in Fig. 29. Users can refer to chapter 4.4.



(click apply button in left panel to view data & map)

Receptor Option

☑ Use Model Grid

☑ Use GIS Shape File

● GIS Shape File Data (*.shp)

County_epa2.shp

○ GIS Shape File Data(*.csv)

ReceptorAggregation_Ozone.csv

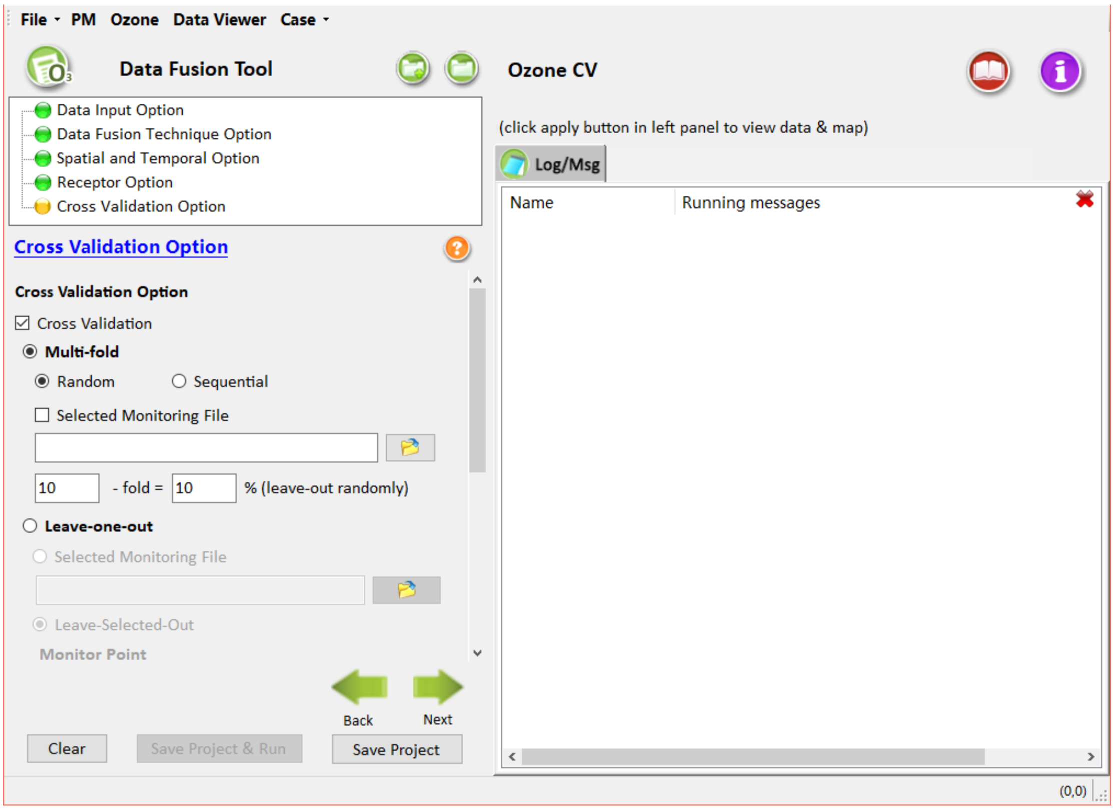
Name	Running messages

Fig. 29 Receptor Option (Ozone)

Users can keep default options and configuration, or change options and configuration according to their requirements. After setting all the options, click the **Next** button or double-click the node of **Cross Validation Option** in the upper left corner to proceed to the next step. The color of this node icon changes from yellow to green, indicating that this step has completed successfully.

5.5 Cross Validation Option

The interface design of **Cross Validation Option** for Ozone data fusion analysis is similar to that of $PM_{2.5}$ data fusion analysis (except that the functions of **Leave-group-out** and **cluster analysis** are still under development), as shown in Fig. 30. Users can refer to chapter 4.5.



(click apply button in left panel to view data & map)

Cross Validation Option

Cross Validation Option

☑ Cross Validation

● Multi-fold

● Random ○ Sequential

☐ Selected Monitoring File

10 - fold = 10 % (leave-out randomly)

○ Leave-one-out

○ Selected Monitoring File

● Leave-Selected-Out

Monitor Point

Fig. 30 Cross Validation Option (Ozone)

Users can keep default options and configuration, or change options and configuration according to their requirements. After completing this step, all configuration options for creating new Ozone data fusion have been completed, the steps of **Save & Run Project** are the same as that of saving $PM_{2.5}$ data fusion results.

6 Data Viewer for analyzing results

The main function of the Data Viewer module is to display the data fusion results in multiple ways. The Data Viewer module can be loaded in the following ways:

- Click the **File** drop-down menu to open an existing project, then the interface of the **Data Viewer** of this project will be exhibited automatically.
- Click the **Data Viewer** menu item to enter the **Data Viewer** interface directly, then click the **Project** button at the bottom left to open the related project.
- Select **Save & Run Project** after creating a new $PM_{2.5}$ or Ozone data fusion analysis, then the data viewer interface of the project will be loaded automatically after the run.

The interface of the Data Viewer module is mainly divided into two parts: the **Output File area** and the **Data Viewer area**, as shown in Fig. 31. The left-hand pane of the interface is the output files area, listing the output files created during the configuration of the DF analysis. Which types of output files are available is dependent on the type of analysis ($PM_{2.5}$, Ozone) and the output choices set in the Configuration File. Clicking on any of the output files will load the data into the graphical and tabular analysis module of Data Viewer. The right-hand pane of the interface is the data viewer area, displaying the data fusion analysis results of the example testing case.

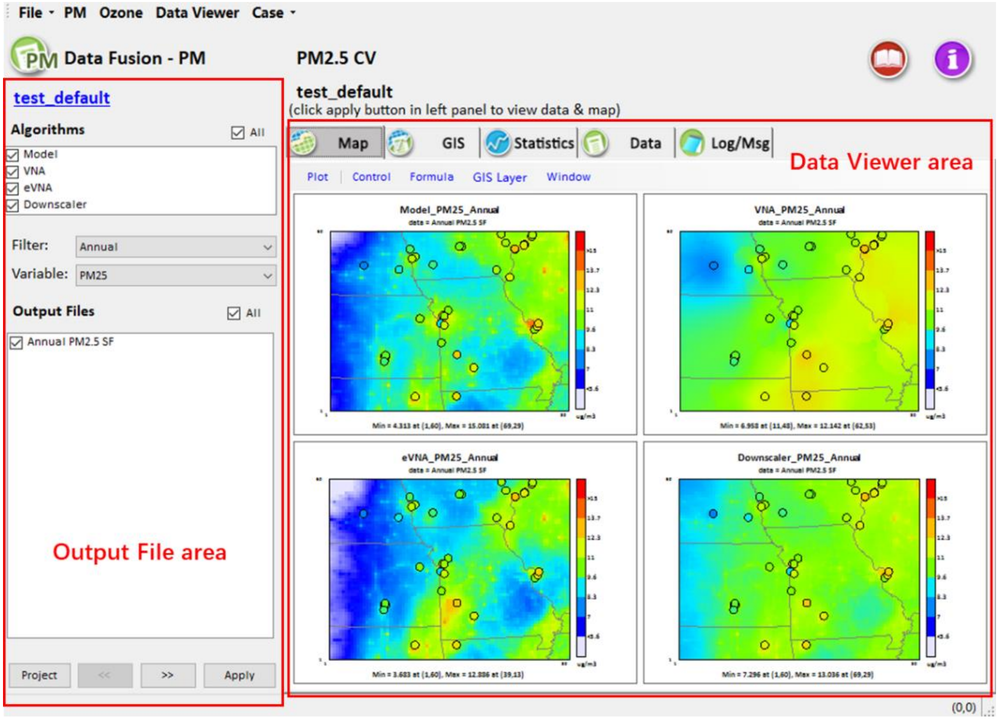


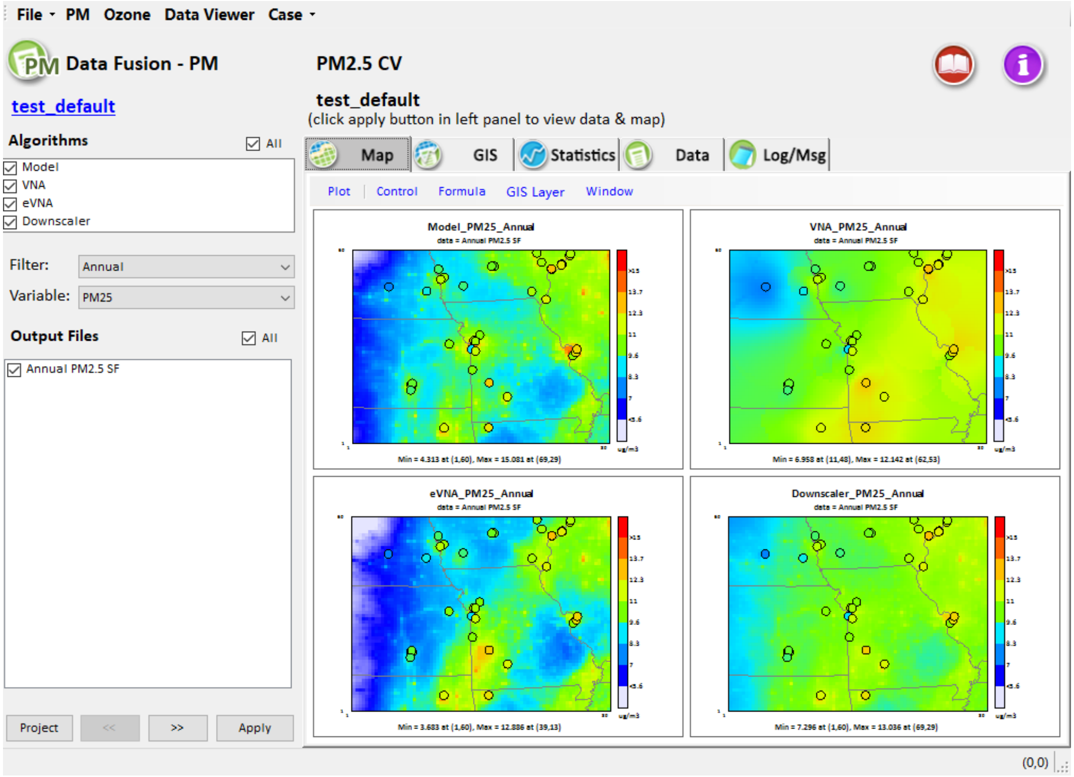
Fig. 31 The interface of the Data Viewer

The results can be displayed in the following ways:

- **Map Viewer:** Spatial tile plots of model output data used for estimating relative response factors.
- **GIS Viewer:** Open-source geographic information system (GIS) graphical interface for displaying point analysis results of the DF at monitor locations. Multiple data layers can be added to the map and different attributes within each layer can be plotting in the GIS tab.
- **Statistics Viewer:** Display of the DF input and output data through the bar chart.
- **Data Viewer:** Display of the DF input and output data through the table.
- **Log/Msg Viewer:** Logging information generated during the currently loaded project.

6.1 Map Viewer

The Map Viewer window displays the results of model results at monitor locations in the modeling domain. Example results available to view in the Map window include the spatial and temporal data fusion results interpolated by the algorithms used in the calculations. The configuration options available in the Map window include zooming in/out the map; clicking on a monitoring site in the modeling domain to view the site information; specifying the color legend; checking to display the location and concentration of monitor sites; selecting background areas; using the formula function developed for common mathematical calculations between different data fusion results; saving the results as an image or CSV file. The interface of the Map for $PM_{2.5}$ or Ozone is shown in Fig. 32 and the interface of the Map for AOD are shown in Fig. 33.


Data Fusion - PM

PM2.5 CV

test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All
☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ☑ All
☑ Annual PM2.5 SF

Min = 4.313 at (1,60), Max = 15.081 at (69,29)

Min = 6.958 at (11,48), Max = 12.142 at (62,53)

Min = 3.683 at (1,60), Max = 12.886 at (39,13)

Min = 7.296 at (1,60), Max = 13.036 at (69,29)

(0,0)

Fig. 32 The interface of the Map Viewer ($PM_{2.5}$ or Ozone)

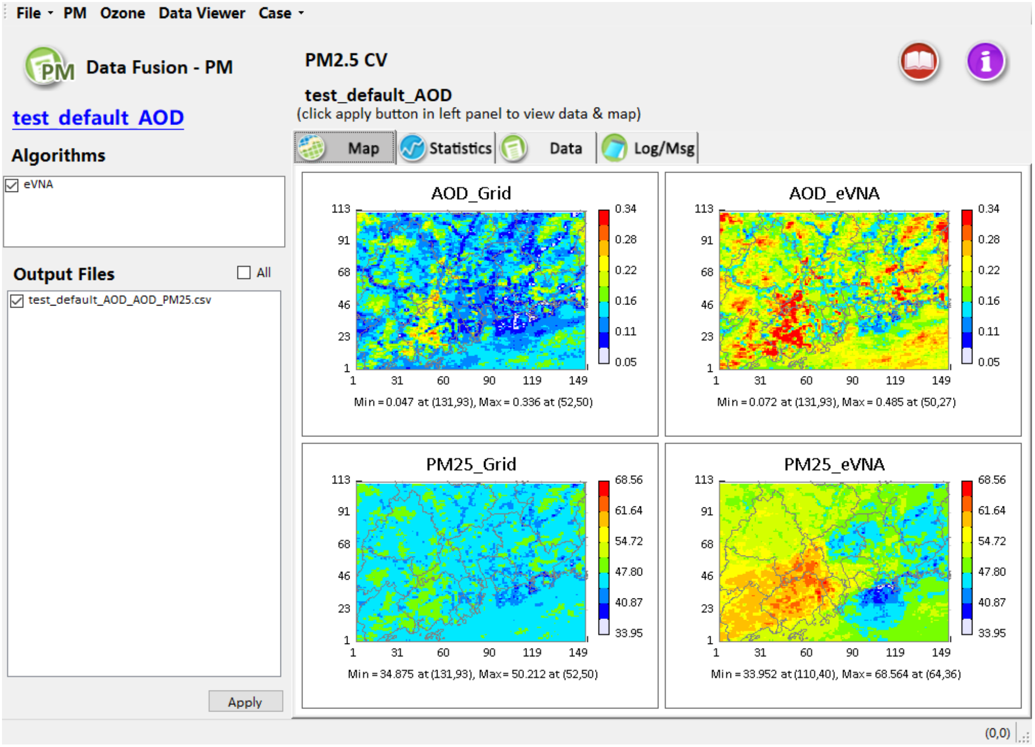

File PM Ozone Data Viewer Case
Data Fusion - PM
test_default_AOD
Algorithms
☑ eVNA
Output Files
□ All
☑ test_default_AOD_AOD_PM25.csv
Apply
PM2.5 CV
test_default_AOD
(click apply button in left panel to view data & map)
Map
Statistics
Data
Log/Msg
AOD_Grid
Min = 0.047 at (131,93), Max = 0.336 at (52,50)
AOD_eVNA
Min = 0.072 at (131,93), Max = 0.485 at (50,27)
PM25_Grid
Min = 34.875 at (131,93), Max = 50.212 at (52,50)
PM25_eVNA
Min = 33.952 at (110,40), Max = 68.564 at (64,36)
0.34 0.28 0.22 0.16 0.11 0.05
68.56 61.64 54.72 47.80 40.87 33.95
113 91 68 46 23 1
1 31 60 90 119 149
(0,0)

Fig. 33 The interface of the Map Viewer (AOD)

➢ To examine the results from a completed DF analysis, select one of the algorithms in the left-hand pane of the Data Viewer window to display the resulting data on the DF map for analysis (Fig. 34); select one of the periods in the **Filter** drop-down boxes and one of the pollutants in the **Variable** drop-down boxes for analysis (the default case calculates the average value for $PM_{2.5}$ only, the cases for analysis of other variables within different periods can perform this operation); select one of the output files in the **Output Files** pane to show the file data on the map for analysis (the default case contains only one output file, the analysis cases with multiple output files can perform this operation). Click the **Apply** button to display the selected analysis choices on the Map of the DF.

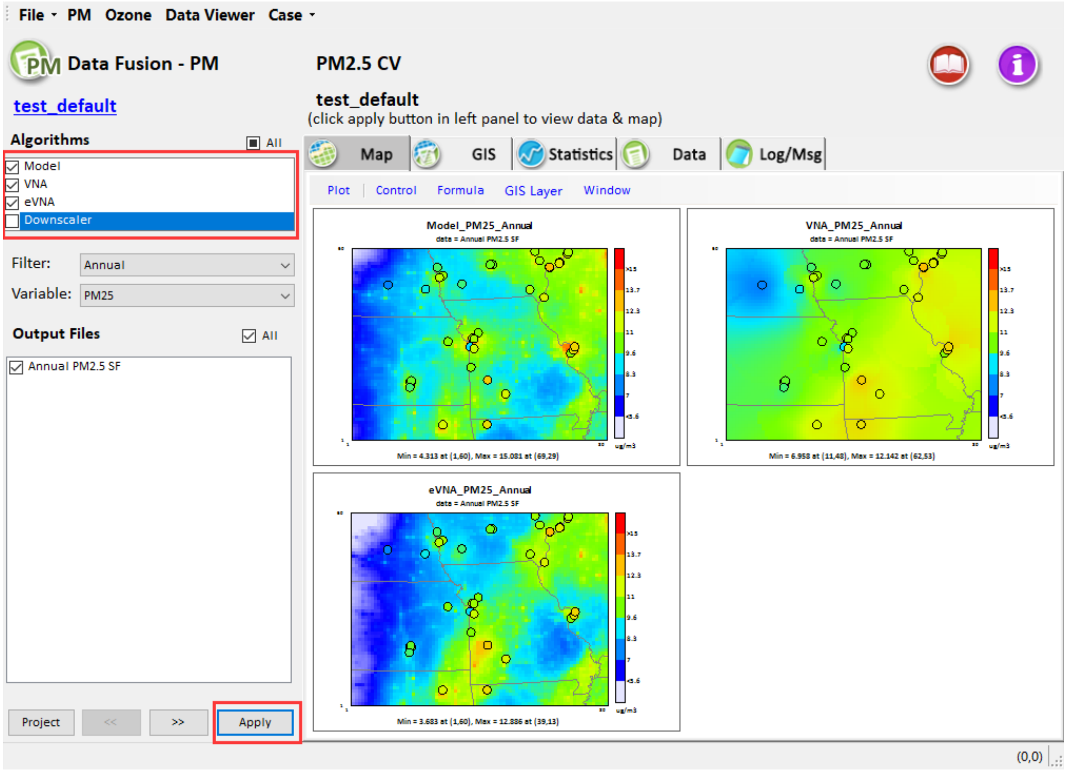
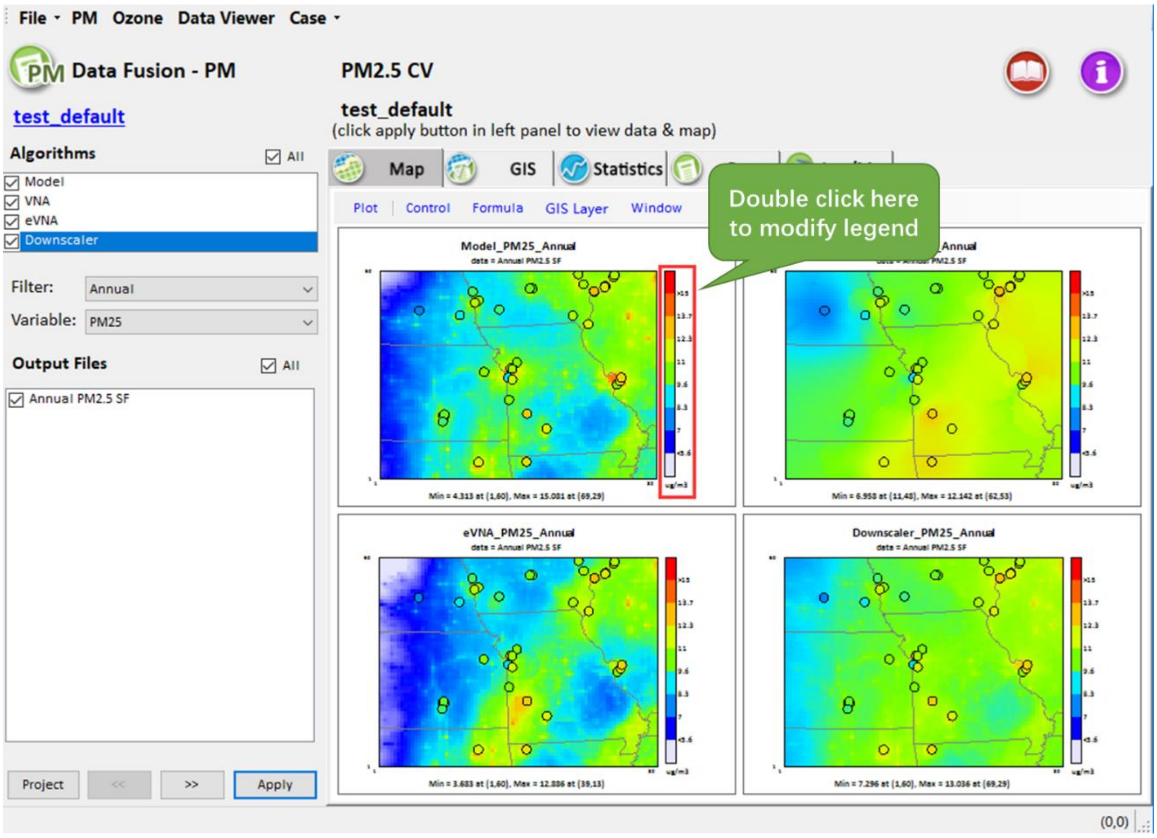


Fig. 34 Show map of selected algorithms

➢ As shown in Fig. 35, this option adjusts the map legend levels. Double-click the color bar on the right -hand of the map to display the legend controls. The legend controls can set the minimum and maximum values of the legend and the individual colors on the scale, as is shown in Fig. 36.


File PM Ozone Data Viewer Case
Data Fusion - PM
test_default
Algorithms
Model
VNA
eVNA
Downscaler
Output Files
Project
<<
>>
Apply
Map
GIS
Statistics
Plot
Control
Formula
GIS Layer
Window
Double click here to modify legend
Model_PM25_Annual
eVNA_PM25_Annual
Downscaler_PM25_Annual
data = Annual PM2.5 SF
ug/m3
Min = 4.313 at (1,60), Max = 15.081 at (69,29)
Min = 6.958 at (11,48), Max = 12.142 at (62,53)
Min = 3.683 at (1,60), Max = 12.886 at (39,13)
Min = 7.296 at (1,60), Max = 13.036 at (69,29)
(0,0)

PM2.5 CV

test_default

(click apply button in left panel to view data & map)

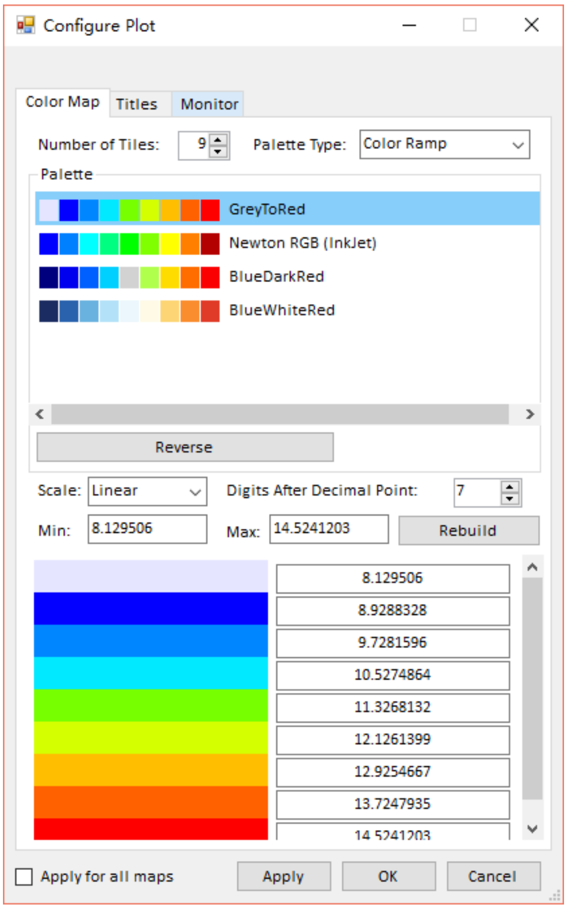
☑ All

Filter: Annual

Variable: PM25

☑ Annual PM2.5 SF

Fig. 35 Double click to change legend



Number of Tiles: 9

Palette Type: Color Ramp

Scale: Linear

Digits After Decimal Point: 7

Min: 8.129506

Max: 14.5241203

8.129506

8.9288328

9.7281596

10.5274864

11.3268132

12.1261399

12.9254667

13.7247935

14.5241203

□ Apply for all maps

Fig. 36 Configure Plot

➢ As shown in Fig. 37 and Fig. 38, these options control zooming in/out, exporting the map image, and overlaying monitor data on the map image.

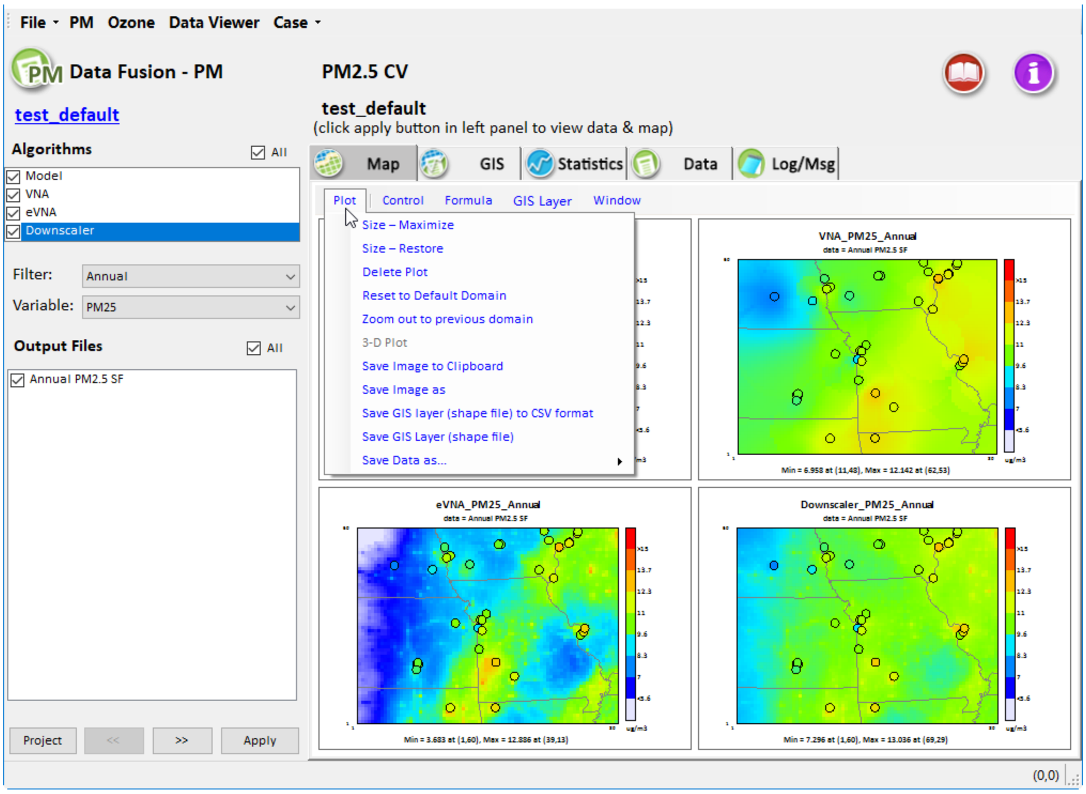


Fig. 37 Click the Plot menu item to manage map

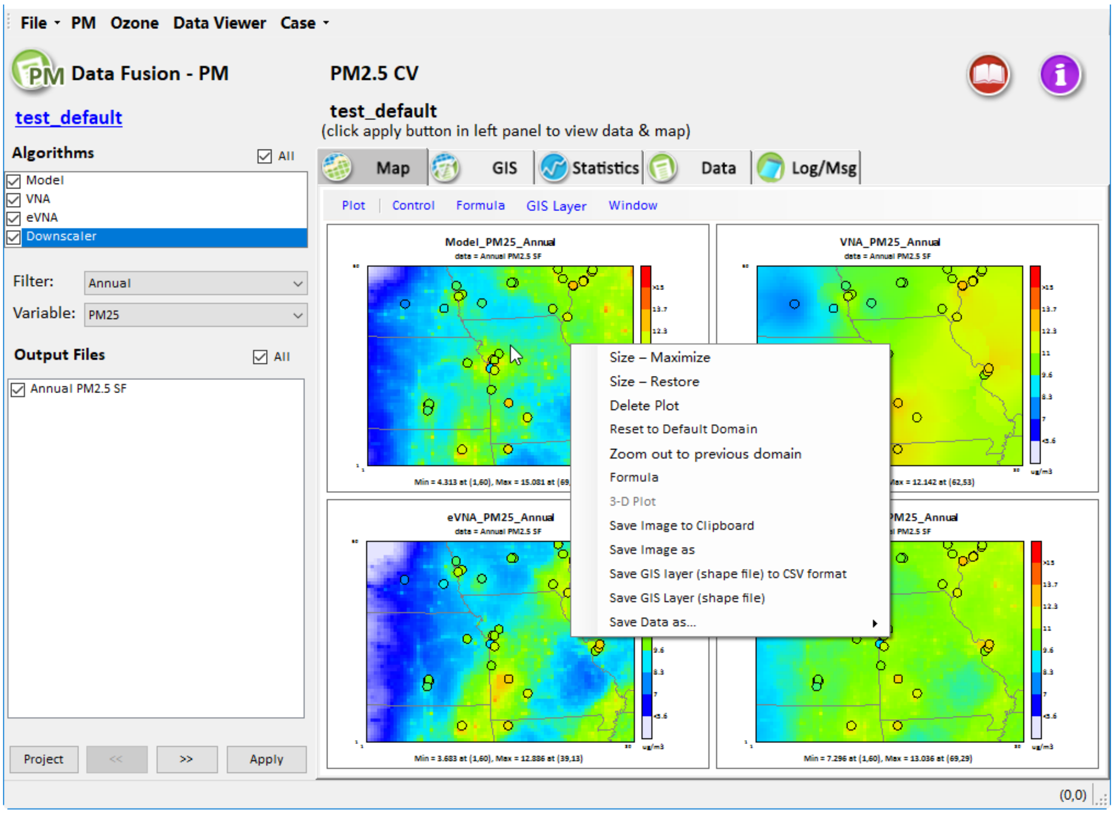
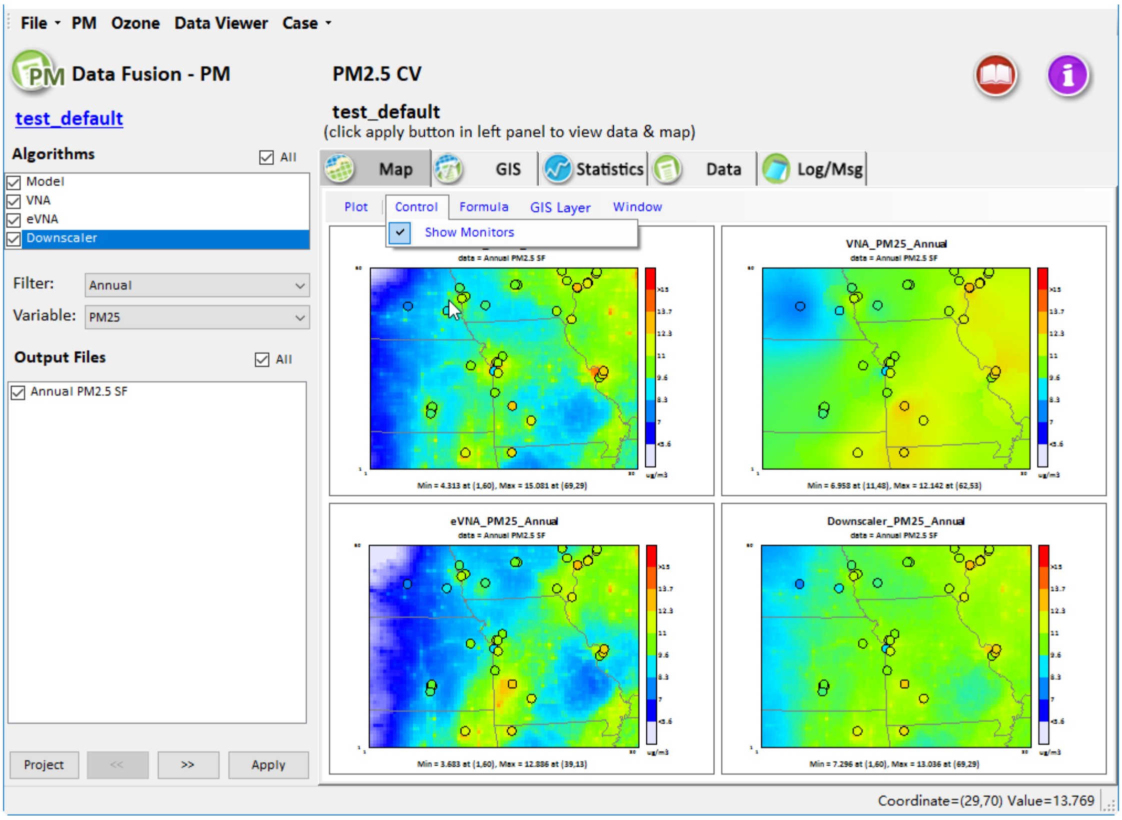


Fig. 38 Right-click the map

➢ As shown in Fig. 39, this option controls the monitors on the DF map, the program checks **show monitors** here in default.



PM2.5 CV

test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ☑ All

☑ Annual PM2.5 SF

☑ Show Monitors

data = Annual PM2.5 SF
Min = 4.313 at (1,60), Max = 15.081 at (69,29)

VNA_PM25_Annual
data = Annual PM2.5 SF
Min = 6.958 at (11,48), Max = 12.142 at (62,53)

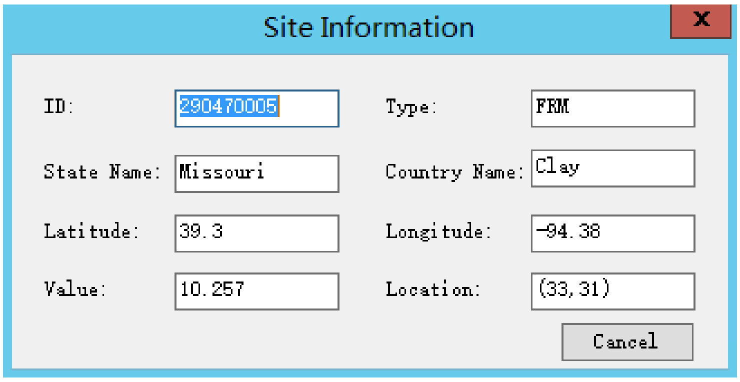
eVNA_PM25_Annual
data = Annual PM2.5 SF
Min = 3.683 at (1,60), Max = 12.886 at (39,13)

Downscaler_PM25_Annual
data = Annual PM2.5 SF
Min = 7.296 at (1,60), Max = 13.036 at (69,29)

Coordinate=(29,70) Value=13.769

Fig. 39 Show Monitors

➢ As shown in Fig. 40, the site information (location, latitude, longitude, country name, etc.) can be viewed by double-clicking on the monitor location on the map of the DF.

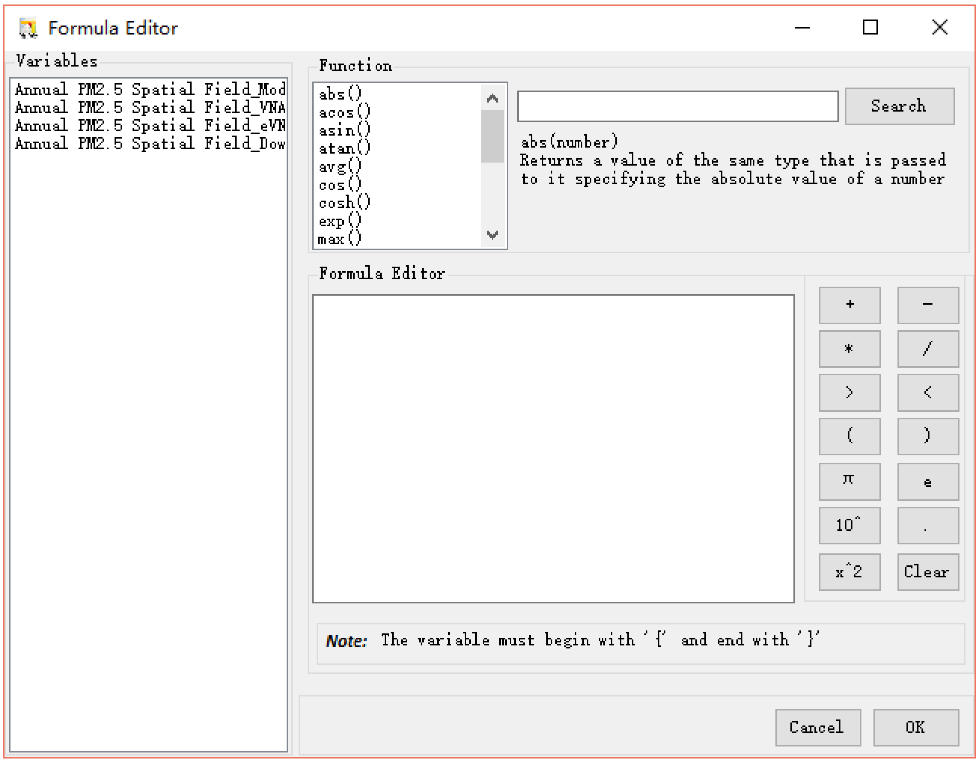


Site Information

ID:	290470005	Type:	FRM
State Name:	Missouri	Country Name:	Clay
Latitude:	39.3	Longitude:	-94.38
Value:	10.257	Location:	(33,31)

Fig. 40 Show site information

➢ As shown in Fig. 41, this option is developed for the common mathematical calculations between multiple data fusion results through the formula function.



Annual PM2.5 Spatial Field_Mod
Annual PM2.5 Spatial Field_VNA
Annual PM2.5 Spatial Field_eVN
Annual PM2.5 Spatial Field_Dow

abs()
acos()
asin()
atan()
avg()
cos()
cosh()
exp()
max()

abs(number)
Returns a value of the same type that is passed to it specifying the absolute value of a number

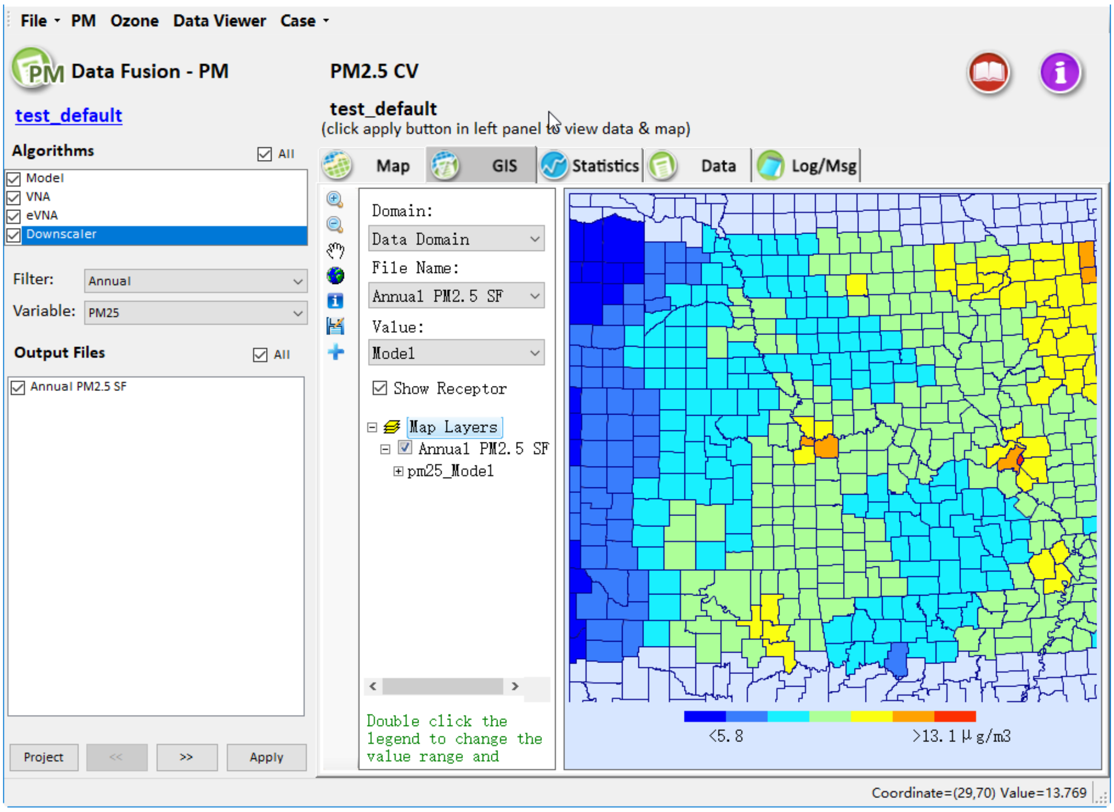
Note: The variable must begin with '{' and end with '}'

Fig. 41 Formula Editor

➢ Click the **Window** menu item with the blue font above the DF Map Viewer window, then choose **Close** to close the window.

6.2 GIS Viewer

The GIS window displays the results in the selected modeling domain. Example results available to view in the Map window include the spatial and temporal data fusion results interpolated by the algorithms used in the calculations. The configuration options available in the Map window include zooming in/out the map; clicking the information button to view the site information (location, latitude, longitude, concentration, country name, etc.) of each monitor location; saving layer files and add layer files when selected Corresponding regional geographic information files in the parameter configuration module before. The interface of the DF GIS Viewer is shown in Fig. 42.



PM2.5 CV

test_default

(click apply button in left panel to view data & map)

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ☑ All

☑ Annual PM2.5 SF

Domain: Data Domain

File Name: Annual PM2.5 SF

Value: Model

☑ Show Receptor

Map Layers
☑ Annual PM2.5 SF
pm25_Model

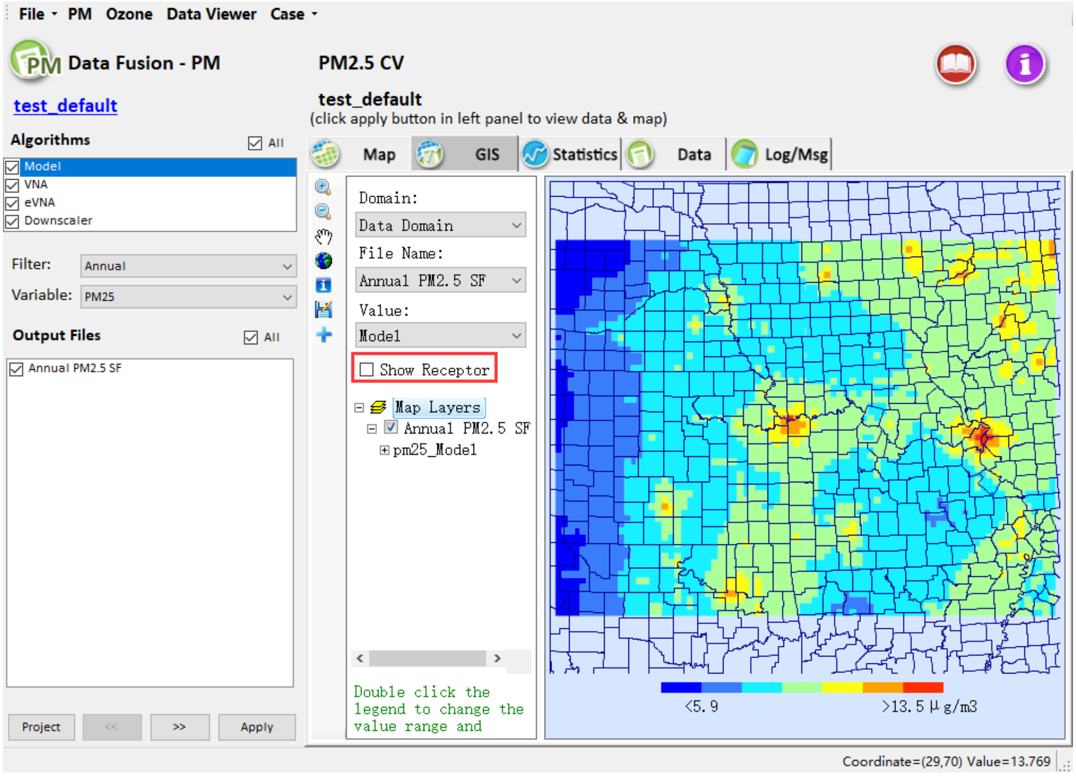
Double click the legend to change the value range and

Coordinate=(29,70) Value=13.769

Fig. 42 The interface of the GIS Viewer

➢ To examine the results from a completed DF analysis, select one of the algorithms in the left-hand pane of the Data Viewer window to display the result data on the DF GIS for analysis; select one of the periods in the **Filter** drop-down boxes and one of the pollutants in the **Variable** drop-down boxes for analysis (the default case calculates the average value for $PM_{2.5}$ only, the cases for analysis of other variables within different periods can implement this operation); select one of the output files in the **Output Files** pane to show the file data on the GIS for analysis (the default case contains only one output file, the analysis cases with multiple output files can implement this operation). Click the **Apply** button to display the selected analysis choices through GIS on the DF.

➢ The GIS Viewer shows the aggregation average concentration of the selected modeling domain in default, as shown in Fig. 42, unchecking the **Show Receptor** can show the spatial distribution of the pollutants, as shown in Fig. 43.



test_default

(click apply button in left panel to view data & map)

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

☑ Annual PM2.5 SF

Domain: Data Domain

File Name: Annual PM2.5 SF

Value: Model

□ Show Receptor

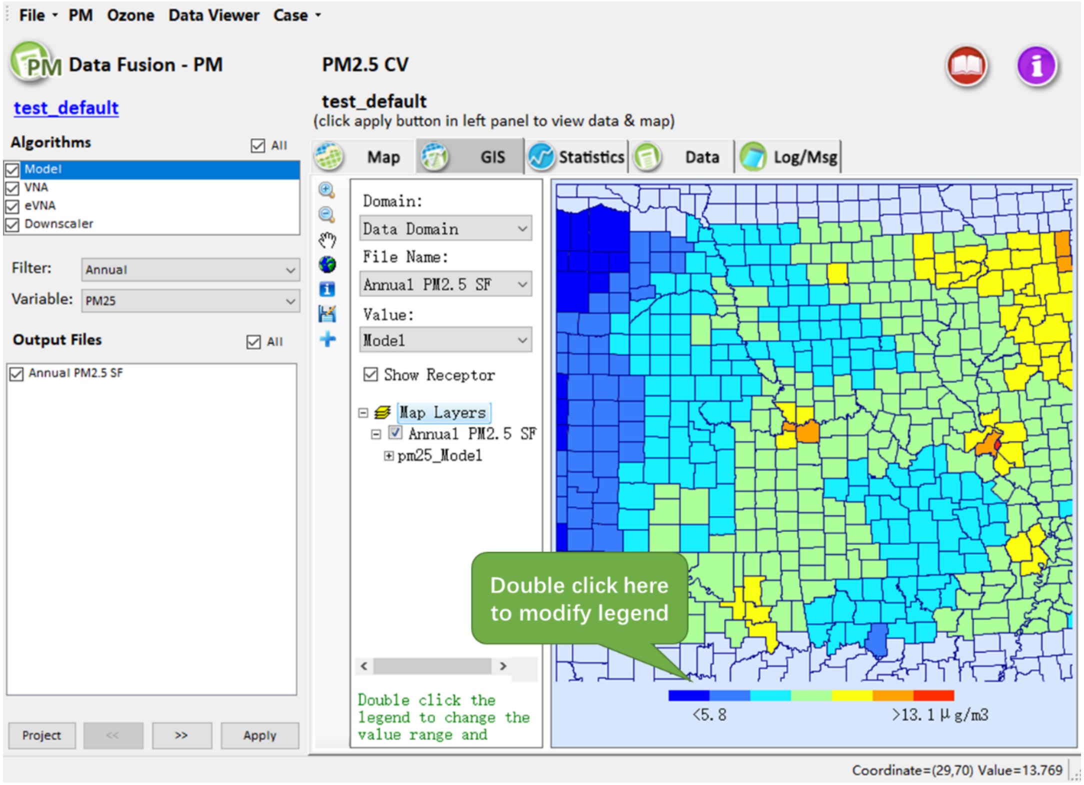
☑ Annual PM2.5 SF

Double click the legend to change the value range and

Coordinate=(29,70) Value=13.769

Fig. 43 GIS Viewer (spatial distribution of pollutants)

➢ As shown in Fig. 44, this option provides a quick adjustment of the GIS map legend levels. Double click the color bar to display the legend controls, as shown in Fig. 45. The legend controls can set the minimum and maximum values of the legend and the individual colors on the scale. More detailed customization of the legend of the map is available by double-clicking a map layer in the Layer Controls section of the GIS window.



PM2.5 CV

test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All
☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual
Variable: PM25

Output Files ☑ All
☑ Annual PM2.5 SF

Domain: Data Domain
File Name: Annual PM2.5 SF
Value: Model
☑ Show Receptor

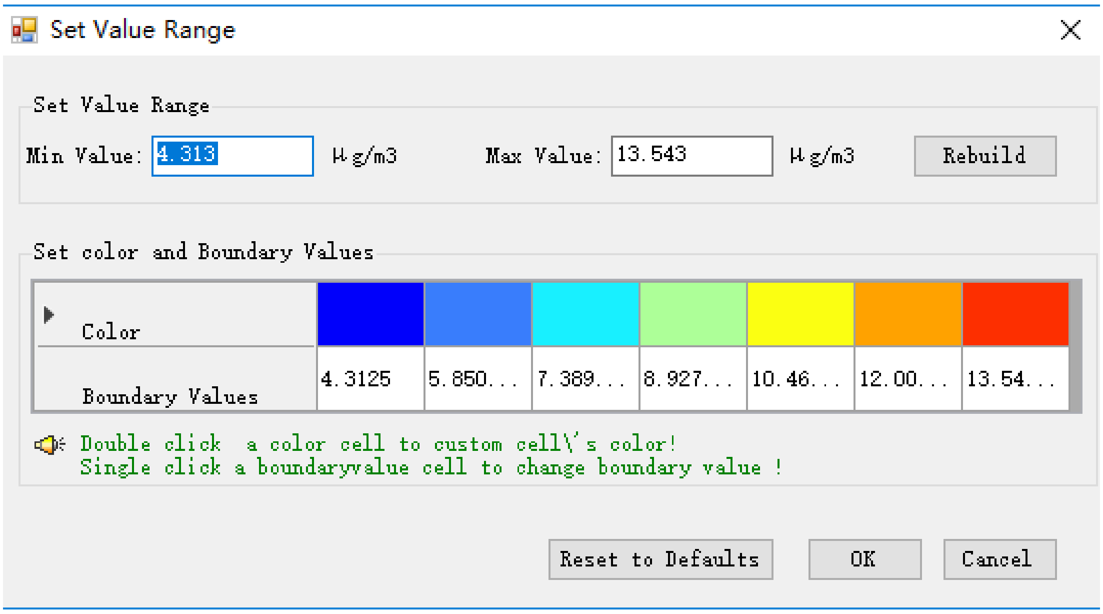
⊟ Map Layers
⊟ ☑ Annual PM2.5 SF
⊞ pm25_Model

Double click the legend to change the value range and

Double click here to modify legend

Coordinate=(29,70) Value=13.769

Fig. 44 Double click to change legend


Set Value Range

Set Value Range

Min Value: 4.313 μg/m3 Max Value: 13.543 μg/m3 Rebuild

Set color and Boundary Values

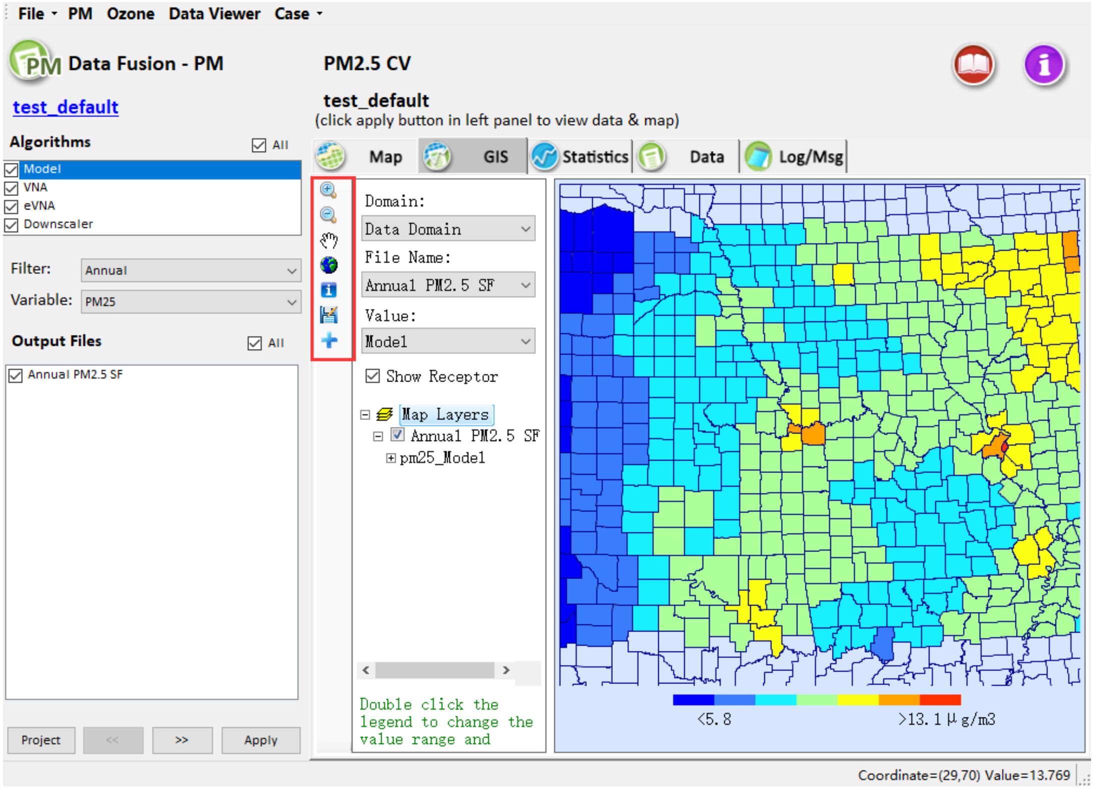
Color	🟦	🟦	🟦	🟩	🟨	🟧	🟥
Boundary Values	4.3125	5.850...	7.389...	8.927...	10.46...	12.00...	13.54...

Double click a color cell to custom cell\'s color!
Single click a boundaryvalue cell to change boundary value !

Reset to Defaults OK Cancel

Fig. 45 Set Value Range

➢As shown in Fig. 46 and Fig. 47, these options control zooming in/out and panning the map display, probing monitor values, exporting the map image, saving a shapefile of the results, and adding external shapefile data to the GIS map.



Data Fusion - PM

test_default

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ☑ All

☑ Annual PM2.5 SF

PM2.5 CV

test_default
(click apply button in left panel to view data & map)

Domain: Data Domain

File Name: Annual PM2.5 SF

Value: Model

☑ Show Receptor

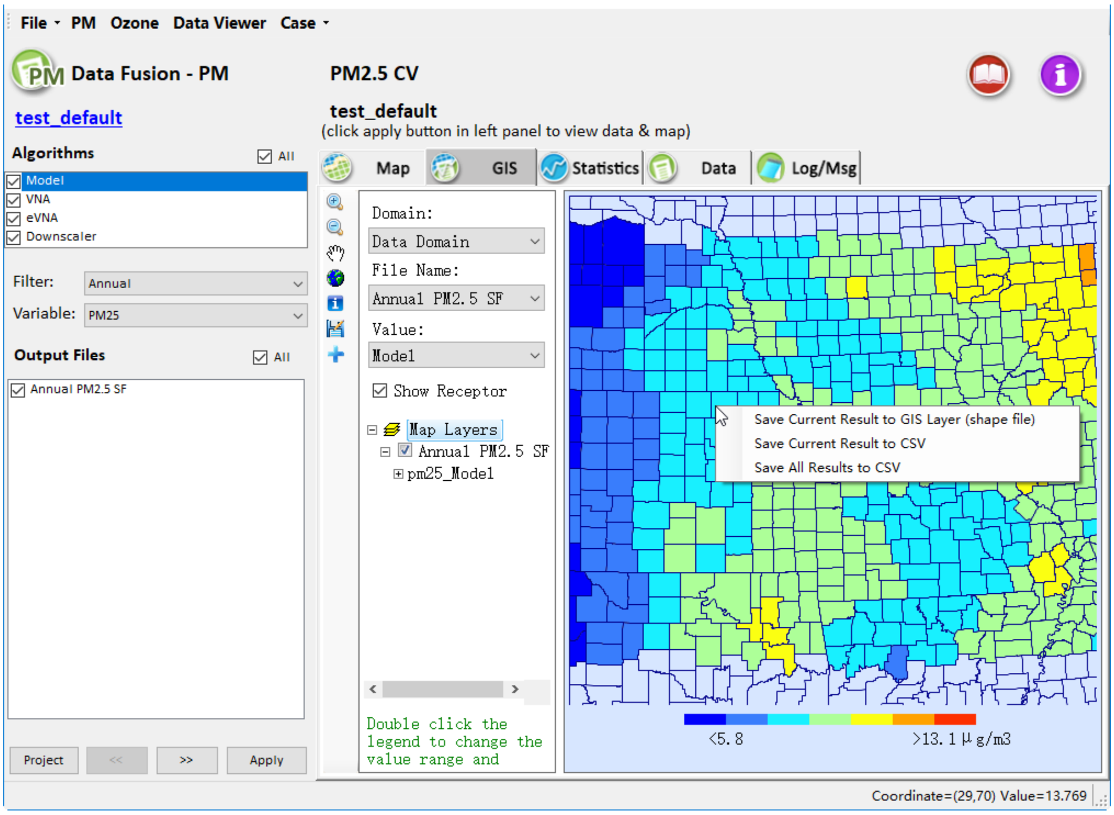
Map Layers
☑ Annual PM2.5 SF
pm25_Model

Double click the legend to change the value range and

<5.8 >13.1 μg/m3

Coordinate=(29,70) Value=13.769

Fig. 46 Click shortcut key to change the GIS map configurations



Data Fusion - PM

test_default

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ☑ All

☑ Annual PM2.5 SF

PM2.5 CV

test_default
(click apply button in left panel to view data & map)

Domain: Data Domain

File Name: Annual PM2.5 SF

Value: Model

☑ Show Receptor

Map Layers
☑ Annual PM2.5 SF
pm25_Model

Double click the legend to change the value range and

<5.8 >13.1 μg/m3

Coordinate=(29,70) Value=13.769

Fig. 47 Right-click to save the GIS Layer

➢As shown in the red box in Fig. 48, these options control the layer attributes of the DF result to display on the map, the ability to quickly zooms to selected sub-regions

(Zoom To) and the boundary layer attributes to display on the map (Domain Selections). The Map Layers section of this area can display different GIS layers, and right-clicking on any of the layers to provide advanced configuration controls for the layer.

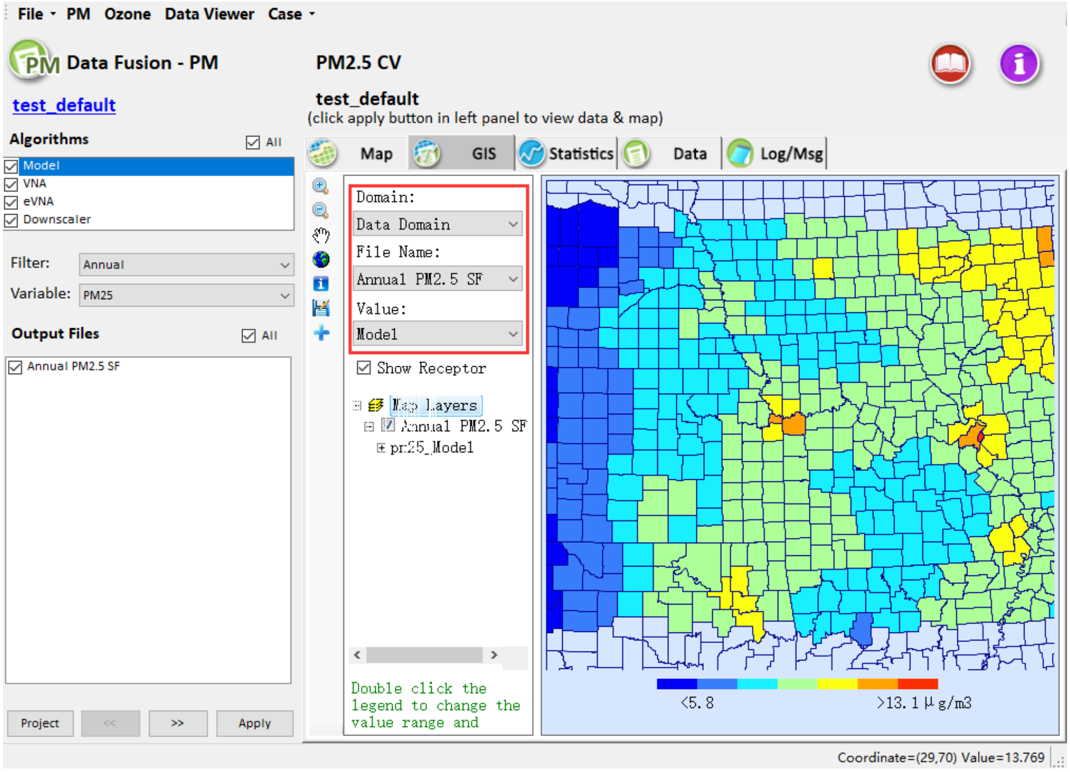


Fig. 48 Options of showing different GIS layers

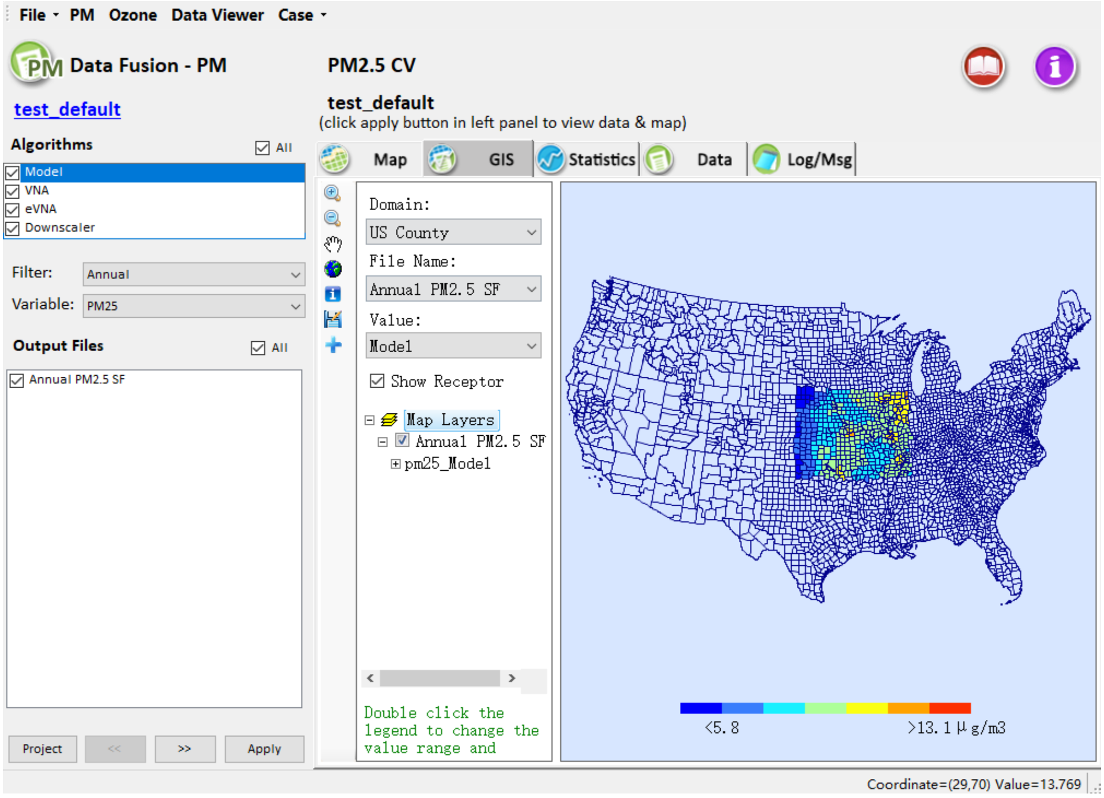


Fig. 49 Show GIS layers of different regions

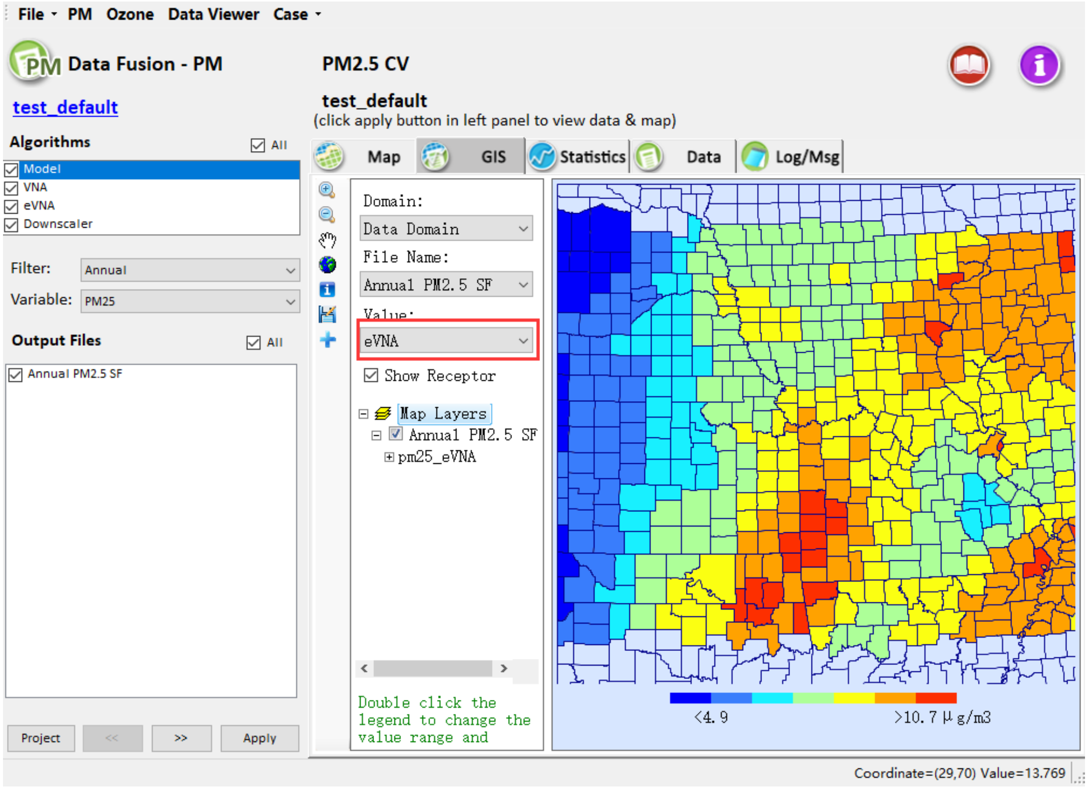
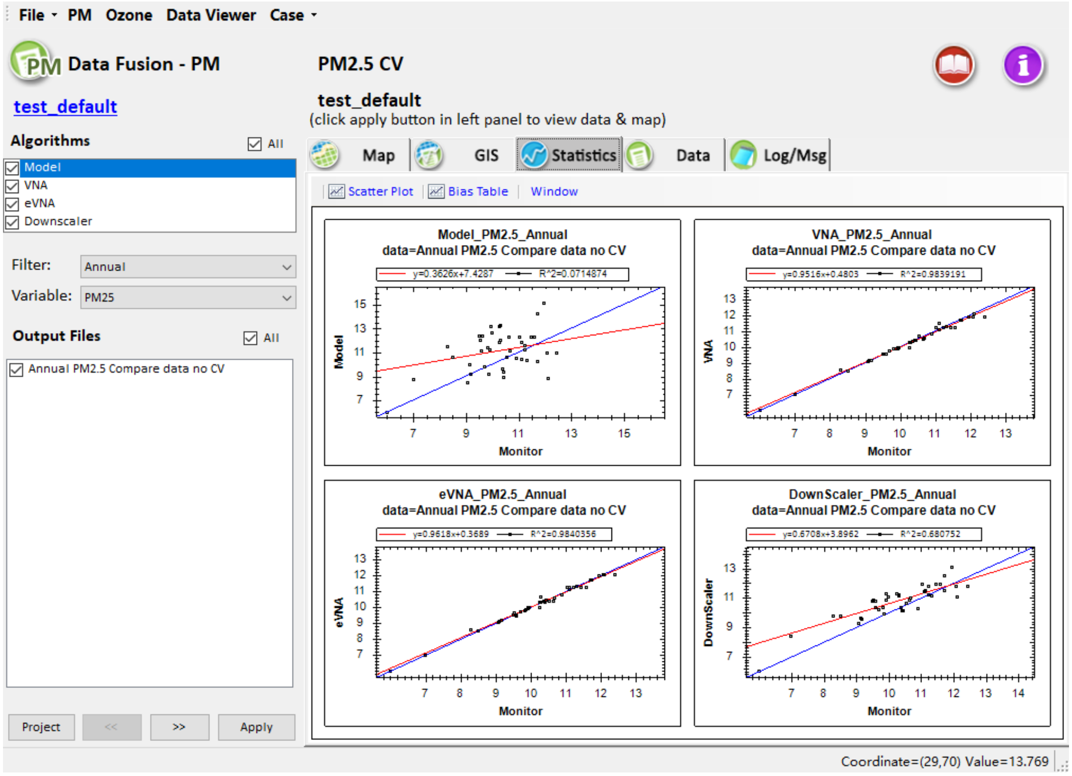


Fig. 50 Show GIS layers of different algorithms

6.3 Chart Viewer

The Chart window displays the DF input and output data as a scatter plot. This type of display is useful for comparing data fusion results interpolated by different algorithms. The Chart view will display the analysis results as a scatter plot on the right-hand panel of the Data Viewer window (Fig. 51).



test_default

test_default
(click apply button in left panel to view data & map)

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

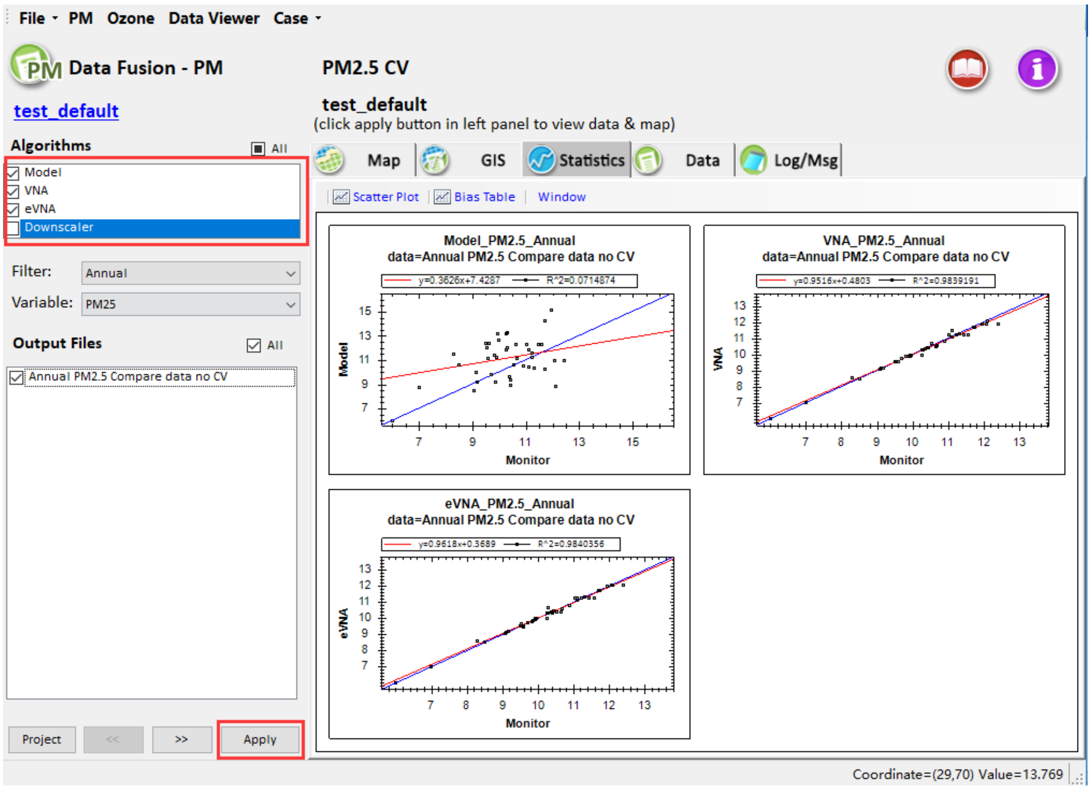
Variable: PM25

☑ Annual PM2.5 Compare data no CV

Coordinate=(29,70) Value=13.769

Fig. 51 The interface of the Chart Viewer

➢To validate the results from a completed DF analysis, select one of the algorithms in the left-hand pane of the Data Viewer window to display the result data on the DF Chart for analysis (Fig. 52); Select one of the periods in the **Filter** drop-down boxes and one of the pollutants in the **Variable** drop-down boxes for analysis (the default case calculates the average value for $PM_{2.5}$ only, the cases for analysis of other variables within different periods can perform this operation); select one of the output files in the **Output Files** pane to show the file data on the Chart for analysis (Fig. 53). Click the **Apply** button to display the selected analysis choices through Chart on the DF.



Data Fusion - PM

PM2.5 CV

test_default

(click apply button in left panel to view data & map)

Algorithms ■ All

☑ Model
☑ VNA
☑ eVNA
□ Downscaler

Filter: Annual

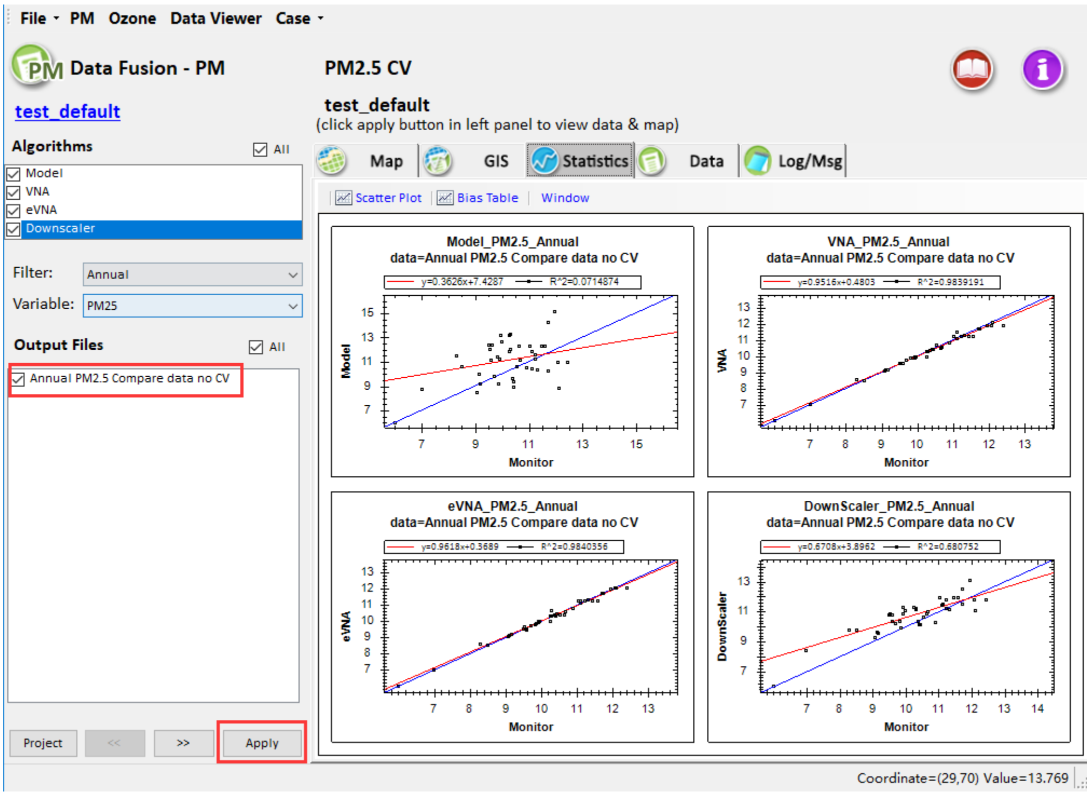
Variable: PM25

Output Files ☑ All

☑ Annual PM2.5 Compare data no CV

Coordinate=(29,70) Value=13.769

Fig. 52 Show chart of selected algorithms



Data Fusion - PM

PM2.5 CV

test_default

(click apply button in left panel to view data & map)

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

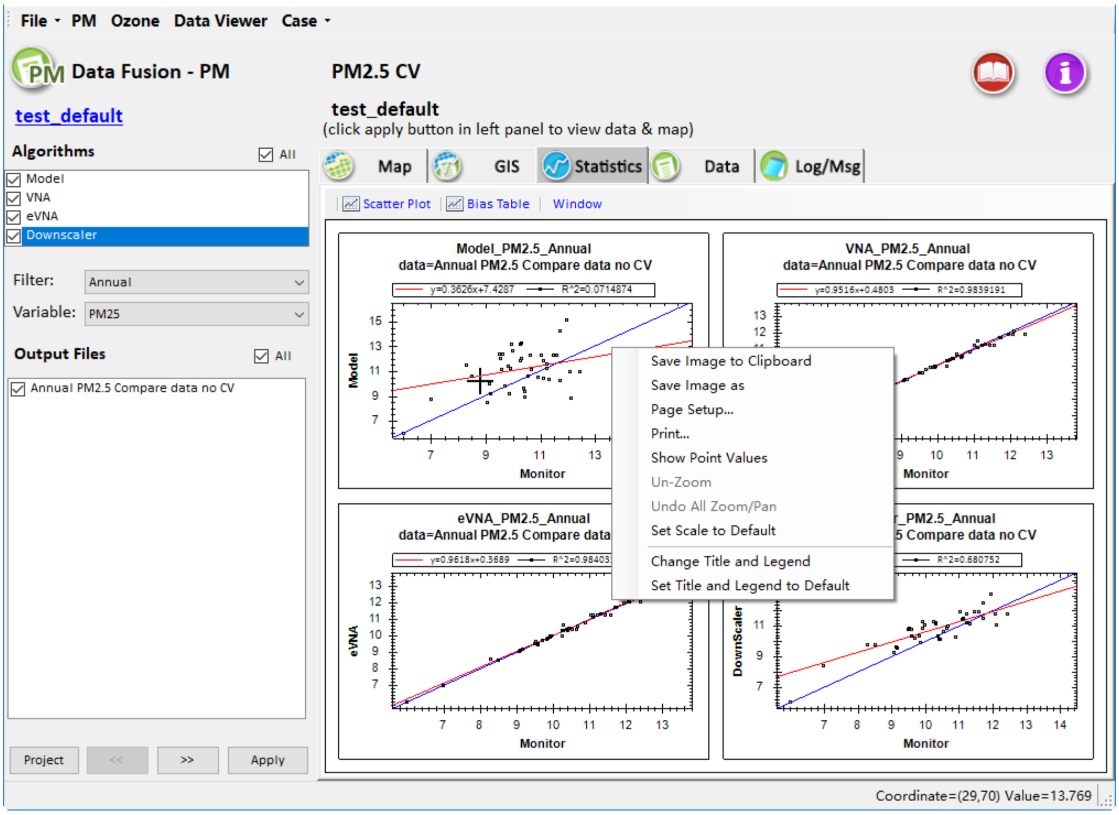
Output Files ☑ All

☑ Annual PM2.5 Compare data no CV

Coordinate=(29,70) Value=13.769

Fig. 53 Show chart of different output files

➢ As shown in Fig. 54, right-clicking on the chart will display a selection box with options to Copy the chart to the clipboard, export (Save Image As) an image of the chart, Print the chart, and zoom out (Un-Zoom) from the chart. The Show Point Values right-click option enables the values of a monitor to be displayed when the mouse hovers over the monitor. The Change Title and Legend right-click option enables customization of the variable names, title, and axes, as shown in Fig. 55.

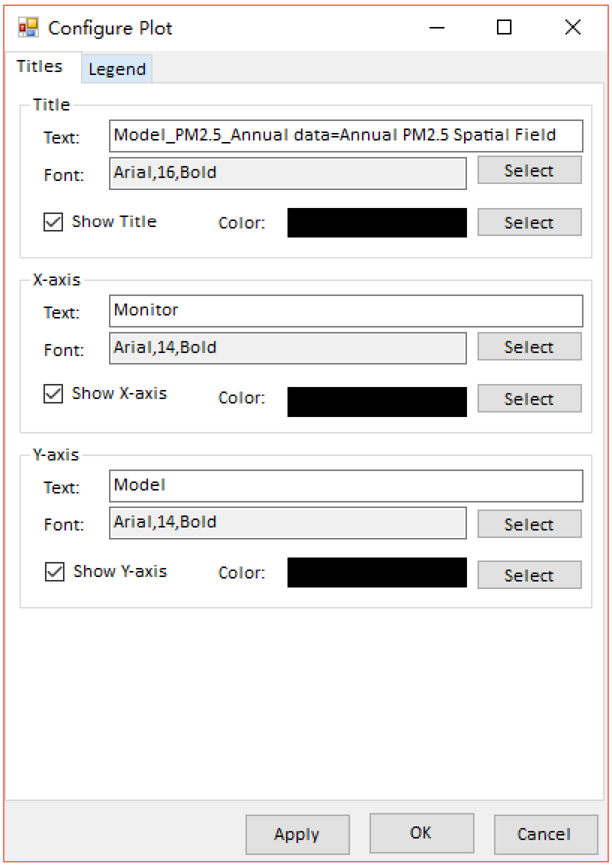


Filter: Annual

Variable: PM25

Coordinate=(29,70) Value=13.769

Fig. 54 Right-click the chart



Configure Plot

Titles | Legend

Title

Text: Model_PM2.5_Annual data=Annual PM2.5 Spatial Field

Font: Arial,16,Bold

☑ Show Title Color: ■

X-axis

Text: Monitor

Font: Arial,14,Bold

☑ Show X-axis Color: ■

Y-axis

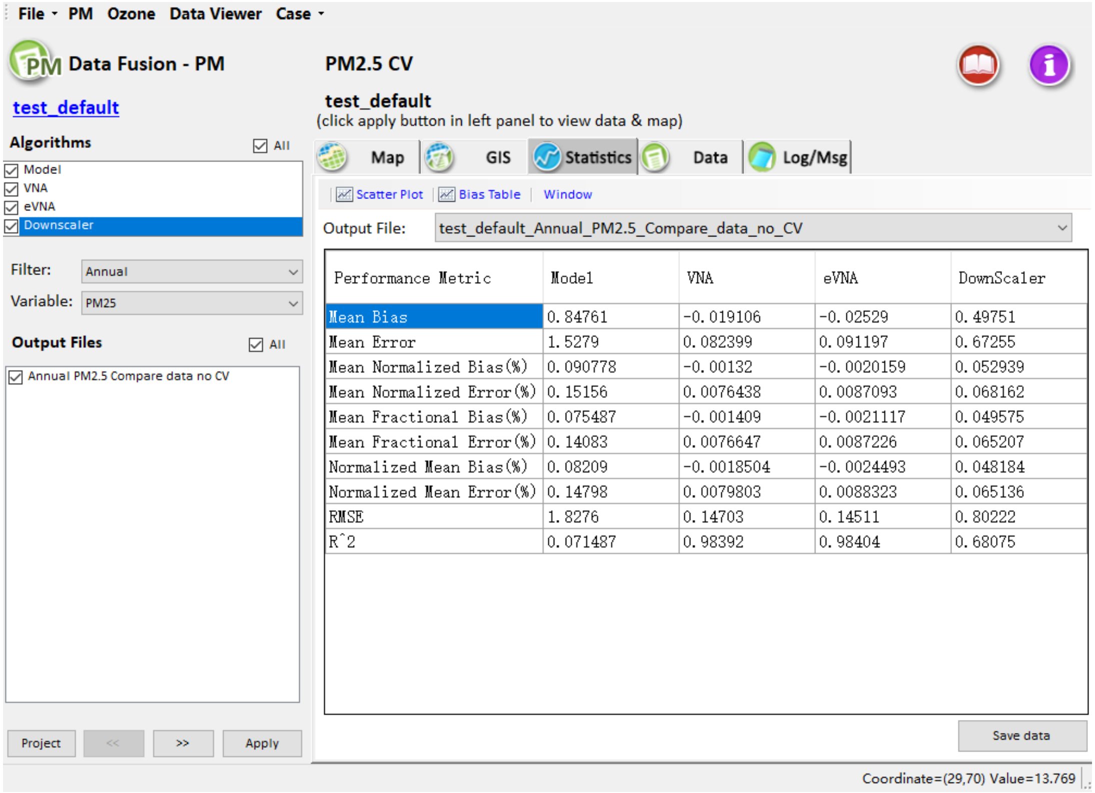
Text: Model

Font: Arial,14,Bold

☑ Show Y-axis Color: ■

Fig. 55 Change Title and Legend

➢ As shown in Fig. 56, click the **Bias Table** menu item with the blue font above the DF Chart viewer interface can display the estimated performance of data fusion results interpolated by the selected algorithms. Click the **Save data** button at the bottom right can export the data to a CSV file.



PM2.5 CV

test_default

(click apply button in left panel to view data & map)

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

☑ Annual PM2.5 Compare data no CV

Output File: test_default_Annual_PM2.5_Compare_data_no_CV

Performance Metric	Model	VNA	eVNA	DownScaler
Mean Bias	0.84761	-0.019106	-0.02529	0.49751
Mean Error	1.5279	0.082399	0.091197	0.67255
Mean Normalized Bias(%)	0.090778	-0.00132	-0.0020159	0.052939
Mean Normalized Error(%)	0.15156	0.0076438	0.0087093	0.068162
Mean Fractional Bias(%)	0.075487	-0.001409	-0.0021117	0.049575
Mean Fractional Error(%)	0.14083	0.0076647	0.0087226	0.065207
Normalized Mean Bias(%)	0.08209	-0.0018504	-0.0024493	0.048184
Normalized Mean Error(%)	0.14798	0.0079803	0.0088323	0.065136
RMSE	1.8276	0.14703	0.14511	0.80222
R^2	0.071487	0.98392	0.98404	0.68075

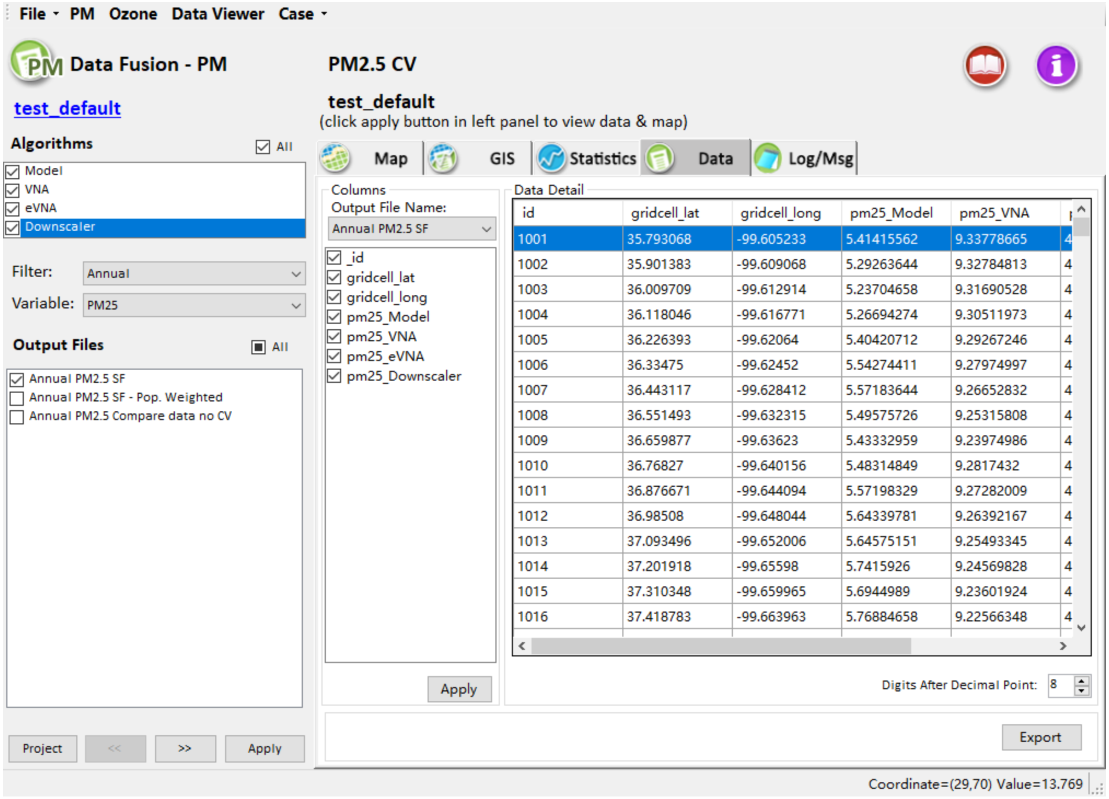
Coordinate=(29,70) Value=13.769

Fig. 56 Show Bias Table

➢ Click the **Window** menu item with the blue font above the DF Chart Viewer window, then choose **Close** to close the window.

6.4 Result Data Viewer

The Result Data window displays the DF input and output data in a tabulated format. The Data view will display the analysis results as a table on the right-hand panel of the Data Viewer window (Fig. 57).



test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ▣ All

☑ Annual PM2.5 SF
☐ Annual PM2.5 SF - Pop. Weighted
☐ Annual PM2.5 Compare data no CV

Columns

Output File Name: Annual PM2.5 SF

☑ _id
☑ gridcell_lat
☑ gridcell_long
☑ pm25_Model
☑ pm25_VNA
☑ pm25_eVNA
☑ pm25_Downscaler

Data Detail

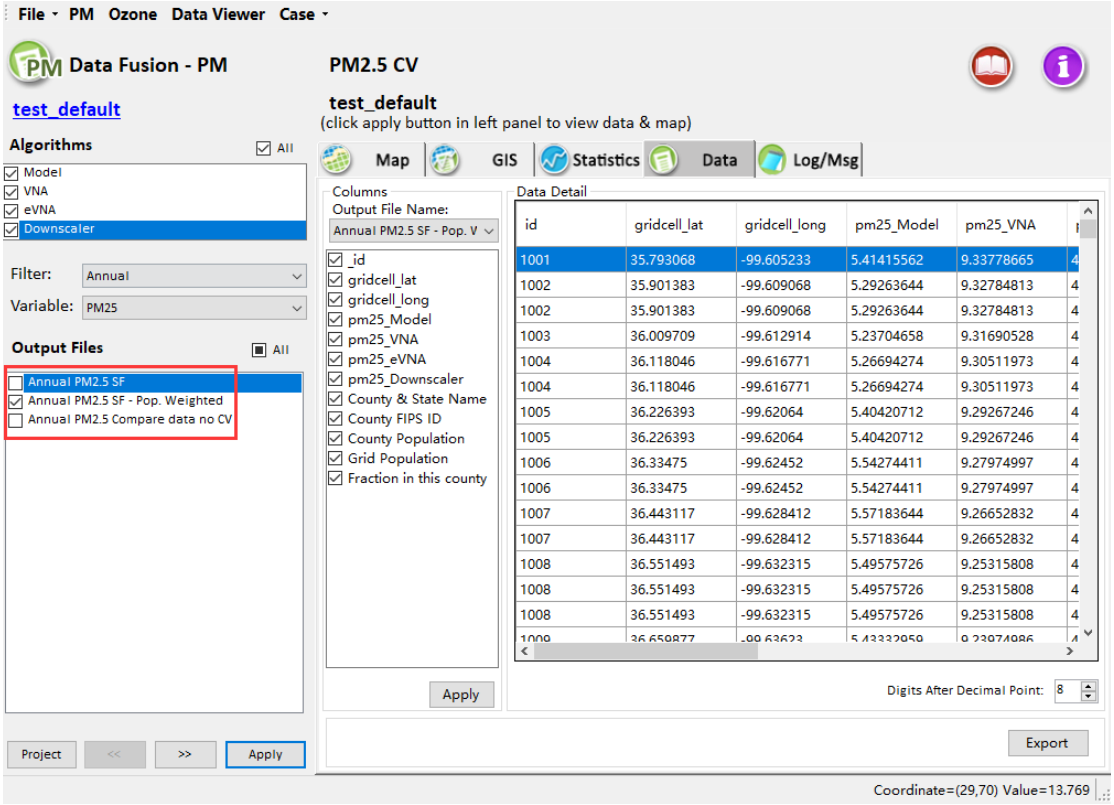
id	gridcell_lat	gridcell_long	pm25_Model	pm25_VNA	p
1001	35.793068	-99.605233	5.41415562	9.33778665	4
1002	35.901383	-99.609068	5.29263644	9.32784813	4
1003	36.009709	-99.612914	5.23704658	9.31690528	4
1004	36.118046	-99.616771	5.26694274	9.30511973	4
1005	36.226393	-99.62064	5.40420712	9.29267246	4
1006	36.33475	-99.62452	5.54274411	9.27974997	4
1007	36.443117	-99.628412	5.57183644	9.26652832	4
1008	36.551493	-99.632315	5.49575726	9.25315808	4
1009	36.659877	-99.63623	5.43332959	9.23974986	4
1010	36.76827	-99.640156	5.48314849	9.2817432	4
1011	36.876671	-99.644094	5.57198329	9.27282009	4
1012	36.98508	-99.648044	5.64339781	9.26392167	4
1013	37.093496	-99.652006	5.64575151	9.25493345	4
1014	37.201918	-99.65598	5.7415926	9.24569828	4
1015	37.310348	-99.659965	5.6944989	9.23601924	4
1016	37.418783	-99.663963	5.76884658	9.22566348	4

Digits After Decimal Point: 8

Coordinate=(29,70) Value=13.769

Fig. 57 The interface of the Result Data Viewer

➢ To plot the results from a completed DF analysis, select one of the algorithms in the left-hand pane of the Data window to display the result data table on the DF Data for analysis; Select one of the periods in the **Filter** drop-down boxes and one of the pollutants in the **Variable** drop-down boxes for analysis (the default case calculates the average value for $PM_{2.5}$ only, the cases for analysis of other variables within different periods can implement this operation); select one of the project **output files** in the Output Files pane to show the file data on the Data for analysis (Fig. 58). Click the **Apply** button to display the selected analysis choices on the Data table of the DF.



test_default

test_default
(click apply button in left panel to view data & map)

Algorithms ☑ All
- ☑ Model
- ☑ VNA
- ☑ eVNA
- ☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ■ All
- □ Annual PM2.5 SF
- ☑ Annual PM2.5 SF - Pop. Weighted
- □ Annual PM2.5 Compare data no CV

Columns

Output File Name: Annual PM2.5 SF - Pop. V

- ☑ _id
- ☑ gridcell_lat
- ☑ gridcell_long
- ☑ pm25_Model
- ☑ pm25_VNA
- ☑ pm25_eVNA
- ☑ pm25_Downscaler
- ☑ County & State Name
- ☑ County FIPS ID
- ☑ County Population
- ☑ Grid Population
- ☑ Fraction in this county

Data Detail

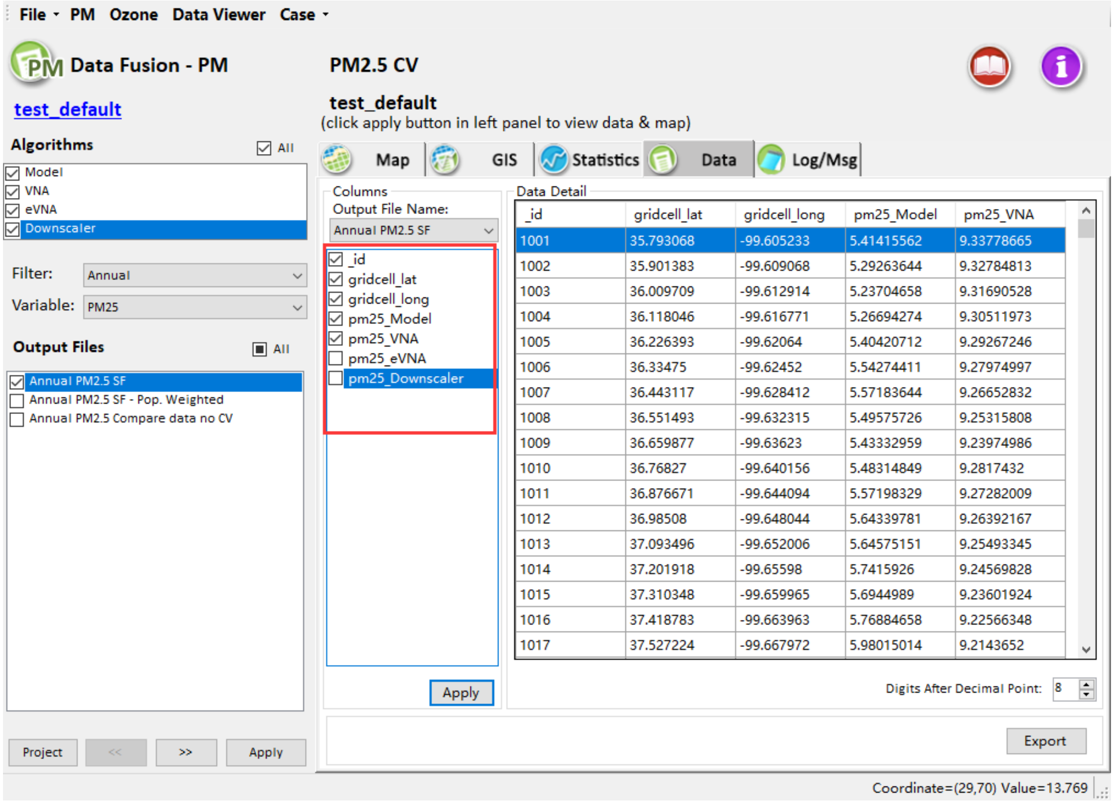
id	gridcell_lat	gridcell_long	pm25_Model	pm25_VNA	
1001	35.793068	-99.605233	5.41415562	9.33778665	4
1002	35.901383	-99.609068	5.29263644	9.32784813	4
1002	35.901383	-99.609068	5.29263644	9.32784813	4
1003	36.009709	-99.612914	5.23704658	9.31690528	4
1004	36.118046	-99.616771	5.26694274	9.30511973	4
1004	36.118046	-99.616771	5.26694274	9.30511973	4
1005	36.226393	-99.62064	5.40420712	9.29267246	4
1005	36.226393	-99.62064	5.40420712	9.29267246	4
1006	36.33475	-99.62452	5.54274411	9.27974997	4
1006	36.33475	-99.62452	5.54274411	9.27974997	4
1007	36.443117	-99.628412	5.57183644	9.26652832	4
1007	36.443117	-99.628412	5.57183644	9.26652832	4
1008	36.551493	-99.632315	5.49575726	9.25315808	4
1008	36.551493	-99.632315	5.49575726	9.25315808	4
1008	36.551493	-99.632315	5.49575726	9.25315808	4
1009	36.659877	-99.63623	5.43332959	9.23974986	4

Digits After Decimal Point: 8

Coordinate=(29,70) Value=13.769

Fig. 58 Show result data of different output files

➢ As shown in Fig. 59, the list shows the output files generated in the calculations of data fusion and the relative columns in the data table. Checking the boxes next to the columns and clicking "Apply" will filter out rows in the table where the values of the selected column(s) are null (-9). The data table shows the tabulated results from the selected DF data file. The table can be sorted by any of the columns by left-clicking on the column header.



test_default

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ▣ All

☑ Annual PM2.5 SF
□ Annual PM2.5 SF - Pop. Weighted
□ Annual PM2.5 Compare data no CV

PM2.5 CV

test_default
(click apply button in left panel to view data & map)

Columns

Output File Name: Annual PM2.5 SF

☑ _id
☑ gridcell_lat
☑ gridcell_long
☑ pm25_Model
☑ pm25_VNA
□ pm25_eVNA
□ pm25_Downscaler

Data Detail

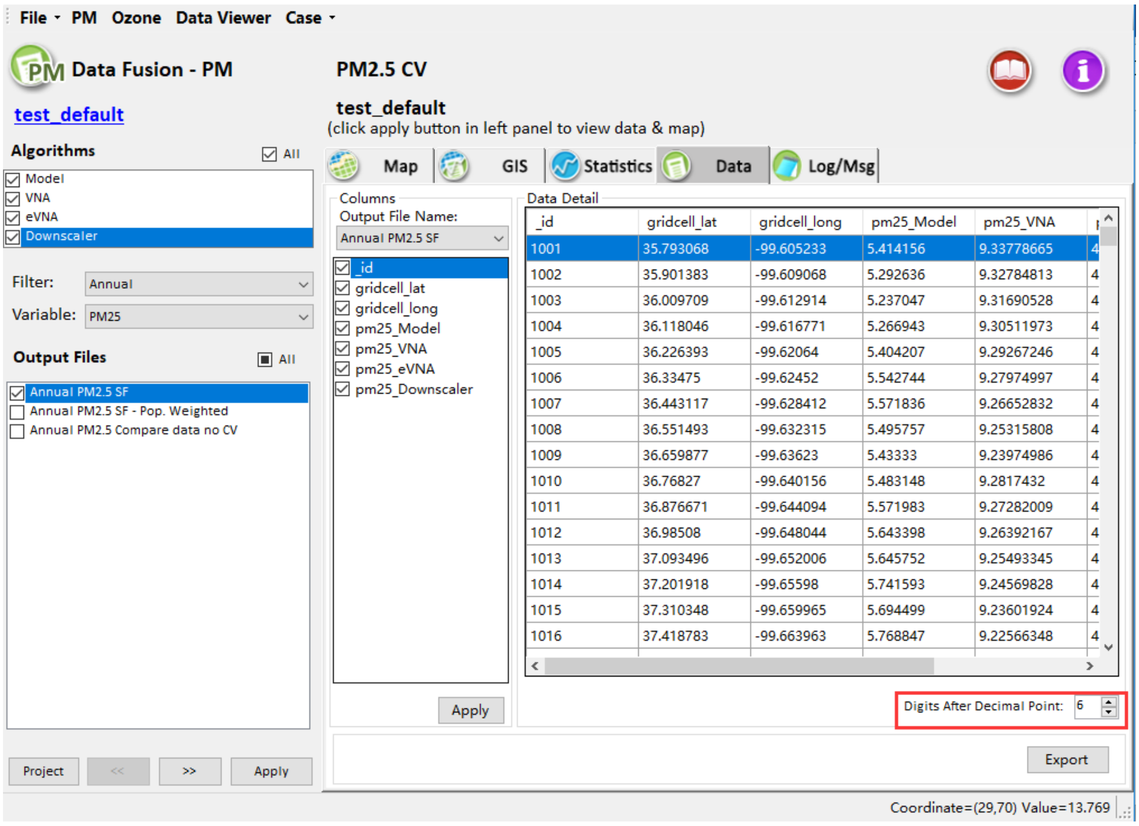
_id	gridcell_lat	gridcell_long	pm25_Model	pm25_VNA
1001	35.793068	-99.605233	5.41415562	9.33778665
1002	35.901383	-99.609068	5.29263644	9.32784813
1003	36.009709	-99.612914	5.23704658	9.31690528
1004	36.118046	-99.616771	5.26694274	9.30511973
1005	36.226393	-99.62064	5.40420712	9.29267246
1006	36.33475	-99.62452	5.54274411	9.27974997
1007	36.443117	-99.628412	5.57183644	9.26652832
1008	36.551493	-99.632315	5.49575726	9.25315808
1009	36.659877	-99.63623	5.43332959	9.23974986
1010	36.76827	-99.640156	5.48314849	9.2817432
1011	36.876671	-99.644094	5.57198329	9.27282009
1012	36.98508	-99.648044	5.64339781	9.26392167
1013	37.093496	-99.652006	5.64575151	9.25493345
1014	37.201918	-99.65598	5.7415926	9.24569828
1015	37.310348	-99.659965	5.6944989	9.23601924
1016	37.418783	-99.663963	5.76884658	9.22566348
1017	37.527224	-99.667972	5.98015014	9.2143652

Digits After Decimal Point: 8

Coordinate=(29,70) Value=13.769

Fig. 59 Show data detail of checked column

➢ As shown in Fig. 60, this option of the data window sets the precision of the values in the table (we set it to 6 here).



test_default

Algorithms ☑ All

☑ Model
☑ VNA
☑ eVNA
☑ Downscaler

Filter: Annual

Variable: PM25

Output Files ▣ All

☑ Annual PM2.5 SF
□ Annual PM2.5 SF - Pop. Weighted
□ Annual PM2.5 Compare data no CV

PM2.5 CV

test_default
(click apply button in left panel to view data & map)

Columns

Output File Name: Annual PM2.5 SF

☑ _id
☑ gridcell_lat
☑ gridcell_long
☑ pm25_Model
☑ pm25_VNA
☑ pm25_eVNA
☑ pm25_Downscaler

Data Detail

_id	gridcell_lat	gridcell_long	pm25_Model	pm25_VNA	
1001	35.793068	-99.605233	5.414156	9.33778665	4
1002	35.901383	-99.609068	5.292636	9.32784813	4
1003	36.009709	-99.612914	5.237047	9.31690528	4
1004	36.118046	-99.616771	5.266943	9.30511973	4
1005	36.226393	-99.62064	5.404207	9.29267246	4
1006	36.33475	-99.62452	5.542744	9.27974997	4
1007	36.443117	-99.628412	5.571836	9.26652832	4
1008	36.551493	-99.632315	5.495757	9.25315808	4
1009	36.659877	-99.63623	5.43333	9.23974986	4
1010	36.76827	-99.640156	5.483148	9.2817432	4
1011	36.876671	-99.644094	5.571983	9.27282009	4
1012	36.98508	-99.648044	5.643398	9.26392167	4
1013	37.093496	-99.652006	5.645752	9.25493345	4
1014	37.201918	-99.65598	5.741593	9.24569828	4
1015	37.310348	-99.659965	5.694499	9.23601924	4
1016	37.418783	-99.663963	5.768847	9.22566348	4

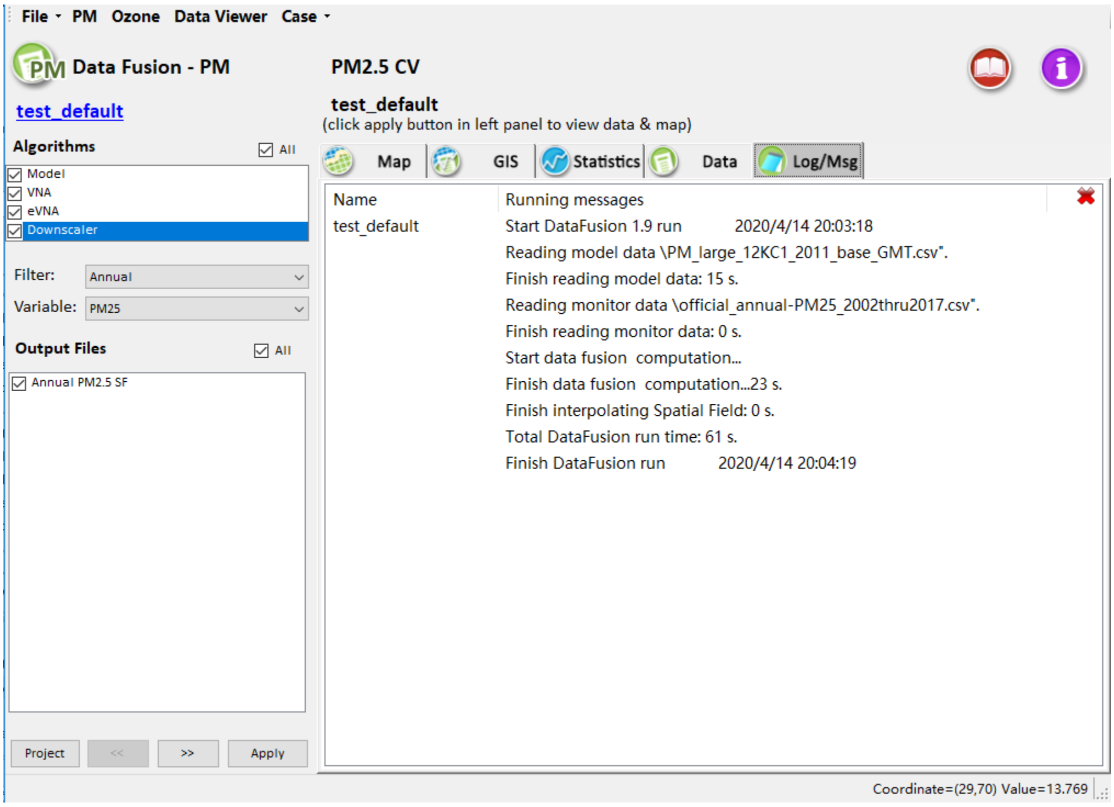
Digits After Decimal Point: 6

Coordinate=(29,70) Value=13.769

Fig. 60 Set the digits after decimal point

6.5 Log/Msg Viewer

The DF Log/Mag Viewer window displays the logging information generated during the currently loaded project. Clicking on the **Log & Msg** menu item displays the steps used to create the DF results and the time that the run completed (Fig. 61).



```
Name           Running messages
test_default   Start DataFusion 1.9 run        2020/4/14 20:03:18
               Reading model data \PM_large_12KC1_2011_base_GMT.csv".
               Finish reading model data: 15 s.
               Reading monitor data \official_annual-PM25_2002thru2017.csv".
               Finish reading monitor data: 0 s.
               Start data fusion  computation...
               Finish data fusion  computation...23 s.
               Finish interpolating Spatial Field: 0 s.
               Total DataFusion run time: 61 s.
               Finish DataFusion run        2020/4/14 20:04:19
```

Fig. 61 The interface of the Log/Msg Viewer

This Chapter describes the multiple display forms under the Data Viewer module in the Data Fusion tool. The interface design and operation methods of the Data Viewer module for $PM_{2.5}$ are similar to that for Ozone.

7 Appendix I

7.1 Evaluation principle of Data Fusion Techniques

1. Voronoi Neighbor Averaging (VNA): It is a deterministic spatial interpolation method that calculates the concentration at the center of each grid cell as the inverse-distance-squared weighted average of $PM_{2.5}$ concentrations at neighboring monitors, where the neighboring monitors are identified using Delaunay triangulation. The equation is shown below:

$$VNA(E) = \sum_{i=1}^{n} Weight_i \times Monitor_i \tag{1}$$

where: n is the number of neighboring sites, $Weight_i$ is the inverse distance squared weight for monitor *i*, *Monitor$_i$* is the observed concentration at monitor *i*, and *VNA(E)* is the estimation at grid cell *E*.

2. Enhanced Voronoi Neighbor Averaging (eVNA): In the eVNA approach, the monitored concentrations used in VNA interpolation are multiplied by the ratio of the modeled concentration in the target grid cell to that in the monitor-containing grid cell. Therefore, eVNA places strong weight on CMAQ gradients between the target cell and cells with nearby monitors. The equation of eVNA is defined as:

$$eVNA(E) = \sum_{i=1}^{n} Weight_i \times Monitor_i \times \frac{Model_E}{Model_i} \tag{2}$$

where: *Model$_E$* is the modeled concentration for cell E, and *Model$_i$* is the modeled concentration in the grid cell containing monitor site *i*.

3. Downscaler (DS): The DS interpolation method is a relatively complex statistical prediction method, but DS resembles a simple linear regression model with spatially varying coefficients at a high level. DS uses Markov chain Monte Carlo (MCMC) methods with Gibbs sampling to develop a relationship between observed and modeled concentrations, and then use the relationship to predict concentrations at points in the spatial domain. The DS method is frequently used to develop exposure fields for health studies.

The general structure of DS can be expressed in an equation similar to that of linear regression:

$$Y_t(\boldsymbol{s}) = \beta_{0,t} + \beta_{0,t}(\boldsymbol{s}) + \beta_{1,t} \times \tilde{X}_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}) \tag{3}$$

where:

$Y_t(\boldsymbol{s})$ is the observed monitoring station data at location $\boldsymbol{s}$ on day *t*. Furthermore, let $X_t(B_k)$ be the CMAQ output for grid block *B$_k$* on day *t*.

$\beta_{0,t}(\boldsymbol{s})$ is a mean zero Gaussian process with covariance function,

$$\mathbb{C}_Q\left(\boldsymbol{h}, u \middle| \phi_{\beta_0,t}\right) \equiv \mathbb{C}\mathrm{ov}\left(\beta_{0,t}(\boldsymbol{s}), \beta_{0,t+u}(\boldsymbol{s}+\boldsymbol{h})\right) = \sigma^2_{\beta_0,t} \exp\left\{-\phi_{\beta_0,t}\|\boldsymbol{h}\|\right\} \mathbb{I}_{\{u=0\}}, \tag{4}$$

the predictors $\tilde{X}_t(\boldsymbol{s})$ are defined as,

$$\tilde{X}_t(\boldsymbol{s}) = \sum_{k=1}^{K} \omega_{k,t}(\boldsymbol{s}) X_t(B_k) \tag{5}$$

with

$$\omega_{k,t}(\boldsymbol{s}) = \frac{\exp\{-\phi_\omega\|\boldsymbol{s}-\boldsymbol{c}_k\|\}\exp\{Q_t(\boldsymbol{c}_k)\}}{\sum_{k=1}^{K}\exp\{-\phi_\omega\|\boldsymbol{s}-\boldsymbol{c}_k\|\}\exp\{Q_t(\boldsymbol{c}_k)\}} \tag{6}$$

K is the total number of numerical model grid cells, ϕ_ω is a decay parameter, ***c$_k$*** is the centroid for the k^{th} grid cell, and $Q_t(\boldsymbol{s})$ is a mean zero Gaussian process with covariance function $\mathbb{C}_Q(\boldsymbol{h}, u|\phi_Q)$ which is defined similarly to (4). In this way $\exp\{Q_t(\boldsymbol{c}_k)\}$ defines the multiplicative increase to the spatial weight given by $\exp\{-\phi_\omega \|\boldsymbol{s} - \boldsymbol{c}_k\|\}$. $\epsilon_t(\boldsymbol{s})$ is a white noise Gaussian process with variance τ_t^2.

The various parameters of DS and how the estimation of these parameters are described below:

1. $(\tau_t^2, \beta_{0,t}, \beta_{1,t})$. Vague conjugate prior distributions are assumed for these parameters. Additionally, $\{\beta_{0,t}\}$ and $\{\beta_{1,t}\}$ are assumed to be independent in time. Thus, conditional on $\{\tilde{X}_t(\boldsymbol{s})\}$, $(\tau_t^2, \beta_{0,t}, \beta_{1,t})$ can be sampled via composition by first sampling τ_t^2 then sampling $(\beta_{0,t}, \beta_{1,t})$ conditional on the obtained value of τ_t^2.
2. $\{\beta_{0,t}(\boldsymbol{s})\}$. The prior distribution for $\beta_{0,t}(\boldsymbol{s})$, as stated above, is a mean zero Gaussian process with exponential covariance function and decay parameter $\phi_{\beta_0,t}$. Conditional on all the parameters, the entire vector of $\{\beta_{0,t}(\boldsymbol{s})\}$ can be sampled from its Gaussian complete conditional distribution.
3. $\sigma^2_{\beta_0,t}$. Using the conjugate inverse gamma distribution, $\sigma^2_{\beta_0,t}$ can be drawn directly from its complete conditional distribution.
4. $\phi_{\beta_0,t}$. This parameter represents the rate of decay of the spatial correlation in $\{\beta_{0,t}(\boldsymbol{s})\}$. Specifically, as $\phi_{\beta_0,t}$ increases, the spatial correlation of $\{\beta_{0,t}(\boldsymbol{s})\}$ decreases. Since decay parameters for the spatial process are particularly difficult to estimate1, a grid search for the spatial range with discrete prior which places mass at (10; 20; ; 90) percent of the maximum observed distance between $Y_t(\boldsymbol{s})$ and $Y_t(\boldsymbol{s}')$ is used. After performing the grid search, $\phi_{\beta_0,t}$ is fixed at the most likely value (from the grid search) for the remainder of the MCMC algorithm.
5. (ϕ_ω, ϕ_Q). These parameters represent the decay parameters for the weights $\omega_{k,t}(\boldsymbol{s})$ used to weight various neighboring numerical model grid cell observations. As very little information about these parameters is available2 , ϕ_ω and ϕ_Q are fixed such that the weights (correlation) are essentially zero beyond a distance of three grid cells.
6. $\{Q_t(\boldsymbol{c}_k)\}$. Each $Q_t(\boldsymbol{c}_k)$ defines a multiplicative increase in the weight assigned to the numerical model value $X_t(B_k)$. Intuitively, if $Q_t(\boldsymbol{c}_k)$ is large then $X_t(B_k)$ will have a large effect on $Y_t(\boldsymbol{s})$. Notice that there is one $Q_t(\boldsymbol{c}_k)$ for each CMAQ centroid. This represents, potentially, a very large number of $Q_t(\boldsymbol{c}_k)$. As such,

predictive process3 are used to reduce the dimensionality. Specifically, $Q_t(\boldsymbol{c}_k) \equiv \mathbb{E}(Q_t(\boldsymbol{c}_k)|Q_t(\ddot{\boldsymbol{c}}_j))$ where the $\{\boldsymbol{b}_j\}$ are a set of sparsely chosen locations with the number of $\boldsymbol{b}_j$ being substantially less than the number of $\boldsymbol{c}_k$. In this way, all $Q_t(\boldsymbol{c}_k)$ can be updated by updating relatively few $Q_t(\boldsymbol{b}_j)$. To update $\{Q_t(\boldsymbol{b}_j)\}$, a Metropolis-Hastings step is used with proposal distribution equal to the prior distribution.

4. Weighting Voronoi Neighbor Averaging (wVNA): The wVNA interpolation method combines the VNA and eVNA methods by setting the weights of VNA and eVNA.

$$\text{wVNA} = w \times eVNA + (1 - w) \times VNA \tag{7}$$

where:

w is the weight coefficient, when $w = 0$, $wVNA$ = VNA, when $w = 1$, $wVNA$ = eVNA.

5. Nearest Site (NS): The Nearest Site interpolation method is a deterministic spatial interpolation method that chooses the single closest monitor and applies that value as the grid cell value.

$$Model_E = Monitor_{nearest} \tag{8}$$

where:

$\text{Monitor}_{\text{nearest}}$ =the nearest monitor s species/$PM_{2.5}$ concentration to cell E, Model_E = modeled species/$PM_{2.5}$ concentration at cell E.

8 References

1. Zhang, H., Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics. *Journal of the American Statistical Association* **2004**, 99, 250-261.
2. Berrocal, V.J., Gelfand, A.E., Holland, D.M., Space-Time Data fusion Under Error in Computer Model Output: An Application to Modeling Air Quality. *Biometrics* **2012**, 68, 837-848.
3. Finley, A.O., Sang, H., Banerjee, S., Gelfand, A.E., Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis* **2009**, 53, 2873-2884.
4. Lv, B., et al., Daily estimation of ground-level $PM_{2.5}$ concentrations at 4km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. Science of The Total Environment, 2017. 580: p. 235-244.
5. Young, M.T., et al., Satellite-Based NO2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression, in Environmental Science & Technology. 2016, American Chemical Society. p. 3686-3694.