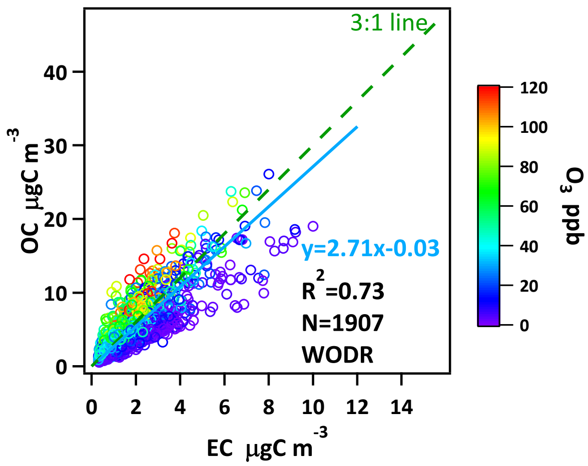
Scatter Plot



Manual for Scatter Plot

吴晟 Wu Cheng

wucheng.vip@foxmail.com

2020-08-18

Preface

Scatter Plot is a handy tool to maximize the efficiency of data visualization in atmospheric science. Many existing generalized data visualization software had been extensively used, but they remain unable to fulfill a number of specify research purposes in atmospheric science. That becomes the motivation of Scatter Plot development. The program includes Deming and York algorithm for linear regression, which considers uncertainties in both X and Y, and is more realistic for atmospheric applications. Scatter Plot is Igor based, and packed with a variety of useful features for data analysis and graph plotting, including batch plotting, data masking via GUI, color coding in Z-axis, data filtering and grouping on different time scales (year, season, month, hour, day of week, etc).

For more details regarding the evaluation and application of Scatter Plot, please refer to

Wu, C. and Yu, J. Z.: Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting, ***Atmos. Meas. Tech.***, 11, 1233-1250, doi: https://doi.org/10.5194/amt-11-1233-2018, 2018.

Please cite this paper if Scatter Plot is used in your publication.

The latest version of the program can be found on my website:

https://wucheng.weebly.com/

https://sites.google.com/site/wuchengust/

https://doi.org/10.5281/zenodo.832416

WU, Cheng

2020-08-18

Contents

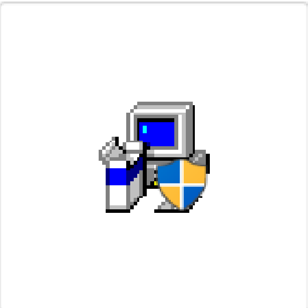
0 Installation of Igor Pro

The user needs Igor Pro platform to run the Scatter Plot program (pxp file). It is similar to the scenario that you need to install MS word to open the docx file. As a result, you need to install Igor Pro on your computer. Igor Pro has both windows PC and Mac version.

Igor Pro version 6.x is recommended. The Scatter Plot can also run on Igor Pro 7.x or 8.x, but there is some issue in the user interface, the scale of certain elements (buttons, drop down menu) is oversized, especially if the computer uses high DPI settings for the screen.

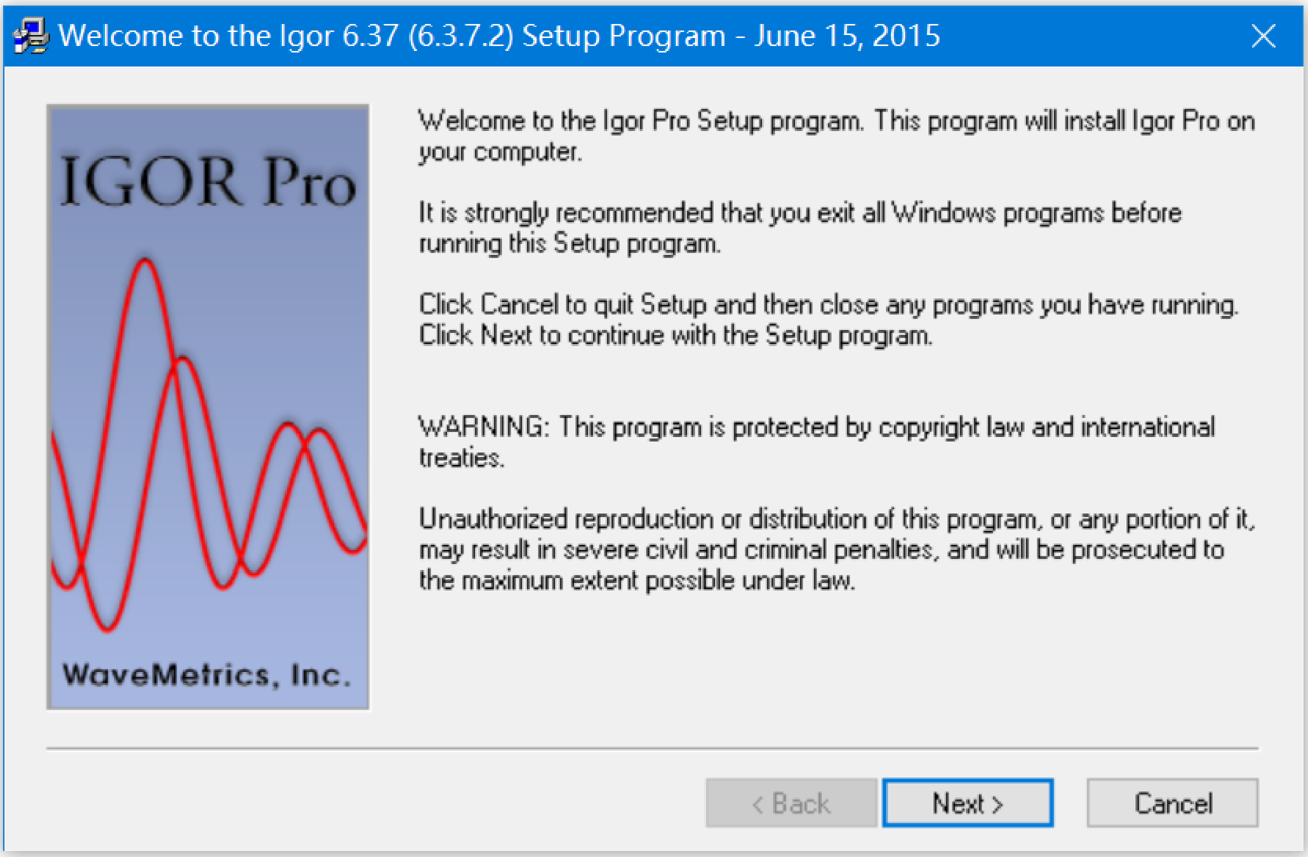
1) Download Igor Pro from https://www.wavemetrics.com/support/versions.htm

2) Double click the Igor Pro installation file

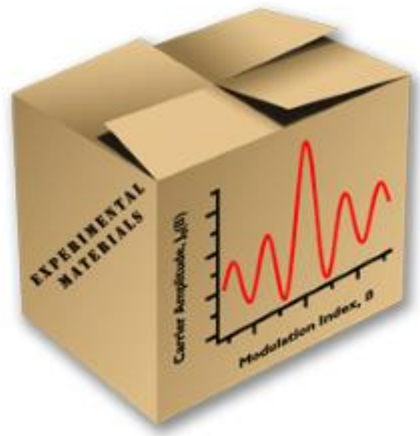


setupIgor6.37.exe

3) Follow the instructions to install Igor Pro.

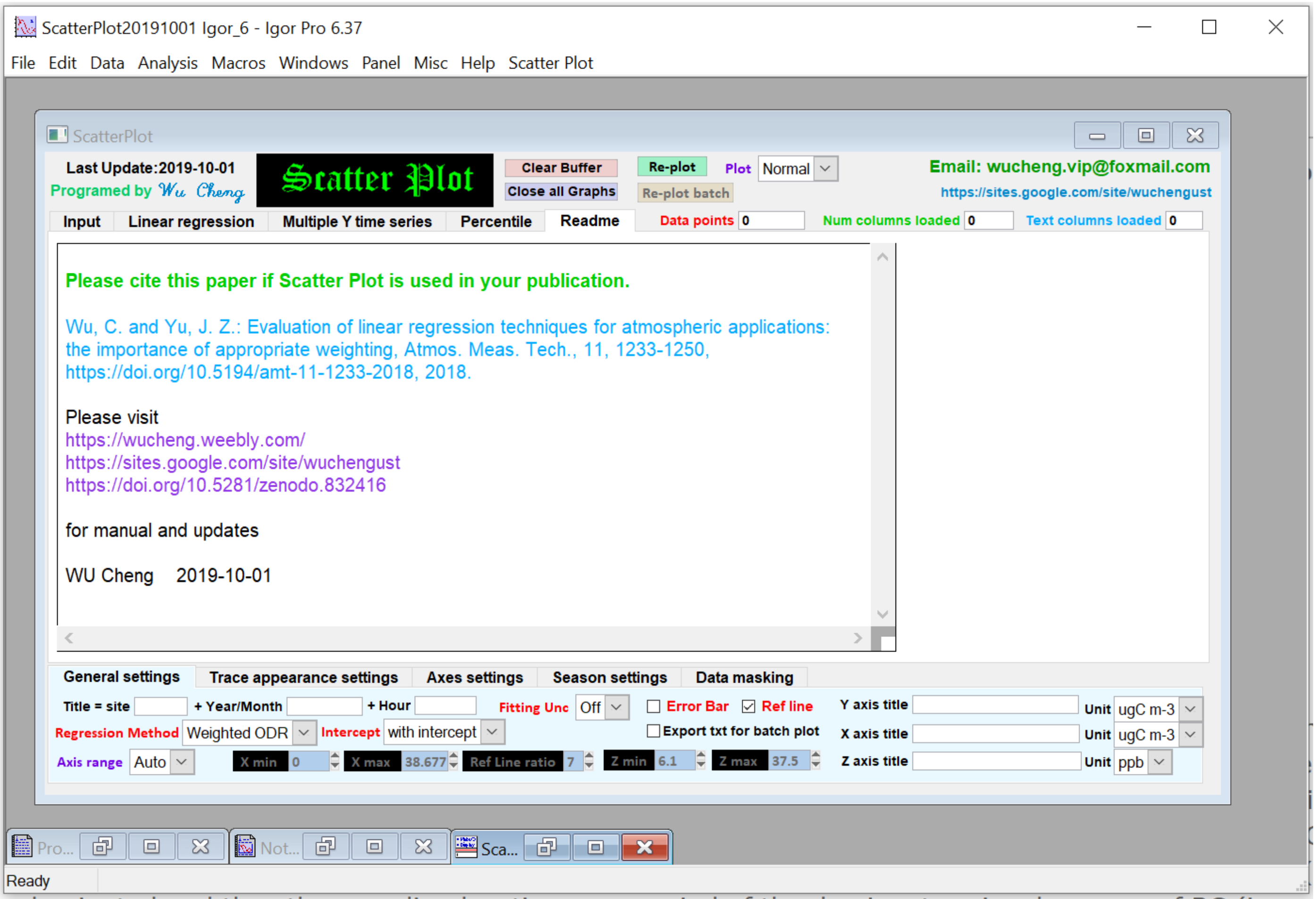


4) Once successfully installed, the system can recognize the pxp file (as the icon shown below). If you do not have an activation key of Igor Pro, the demo version Igor Pro would last for 30 days. After 30 days, the user will not be able to: a) export figure and file; b) save file.



ScatterPlot20200818.pxp

5) To run Scatter Plot, just simply double click the pxp file. If you want to open multiple pxp files at the same time, use the cursor to select the pxp file, then press “Ctrl+Enter”.



1 Recommendation on data structure

Excel is recommended for storing data if the size of the data is less than 1 million rows. Otherwise, .csv file is recommended. If possible, put all data with the same timeline in a single sheet to maximize the efficiency, subset can be extracted by filtering rather than manually put them in separated sheets. The structure of the data is shown below in Figure 1.1. The first row is header (text). After import in Igor, header will become the name of waves (columns in Excel). Space and other illegal characters are not allowed in Igor as wave name and will be replaced by "_". **It should be noted that the wave name must start with an alphabetic character (A - Z or a-z) rather than numbers.** Data falls into three categories:

1) Timestamp (timeline)

2) Numerical data (e.g. concentrations of air pollutants)

3) Text data (Makers, e.g. site name, back trajectories clusters)

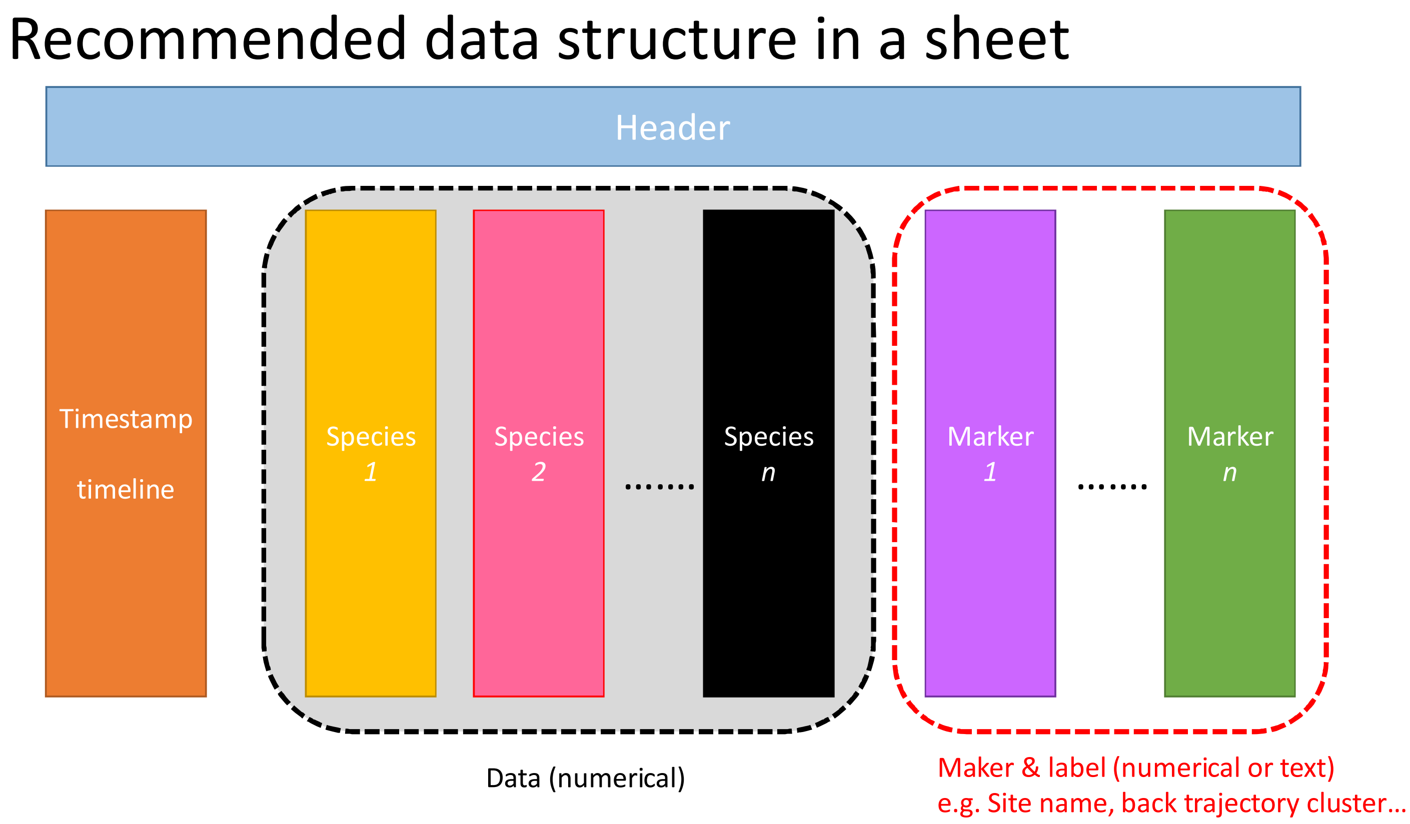


Figure 1.1 Recommended data structure in a sheet

An actual example of excel data (or .csv file) is shown in Figure 1.2. It should be noted that the order of different data columns (waves) is not necessarily the same as figure 1.1, the three categories can be mixed, there is no restrictions on data column order. As shown below, “DateIndex” is timeline, “Sample ID” and “Site” is text data (marker), rest columns are numerical data.

	A	B	E	F	G	H	I	J	K	L	CL
1	**DateIndex**	**Sample ID**	**TGC**	**QGC**	**NaIC_C**	**NH4_C**	**KIC_C**	**CLIC_C**	**NO3_C**	**SO4_C**	**Site**
2	1/13/11	MK110113	61.9167	69.2917	1.9802	6.4493	0.5765	0.8873	11.6865	10.2778	MK
3	1/25/11	MK110125	89.8333	101.3750	2.3110	10.9636	0.9455	0.9994	12.1388	22.2418	MK
4	1/27/11	MK110127	59.0417	66.6250	2.5072	5.8765	0.4537	0.8605	9.2997	10.7022	MK
5	1/31/11	MK110131	66.6667	73.9167	0.2254	7.7103	0.8675	0.4770	5.0206	16.8996	MK
6	2/5/11	MK110205	64.7500	73.0000	0.2102	8.3566	1.4050	0.1443	7.8915	17.7907	MK
7	2/9/11	MK110209	65.3333	72.7500	0.4780	9.0343	1.1588	0.4825	7.6093	19.9455	MK
8	2/11/11	MK110211	59.3750	65.8750	0.4283	6.6911	1.1546	0.1361	7.0313	13.4828	MK
9	2/15/11	MK110215	49.9583	52.3750	0.1801	6.4300	0.6436	0.6742	5.4804	13.6615	MK
10	2/23/11	MK110223	45.7083	48.7083	0.4691	4.6793	0.3861	0.2296	3.7037	10.7737	MK
11	2/25/11	MK110225	53.6667	63.6667	0.4837	6.7622	0.3832	0.3557	5.1990	15.1039	MK
12	3/1/11	MK110301	45.9167	53.5417	0.3452	5.0612	0.1890	0.2956	4.2997	10.6106	MK
13	3/10/11	MK110310	48.1667	53.3750	0.1575	3.2226	0.2605	0.2542	4.2285	11.9927	MK
14	3/13/11	MK110313	76.2500	79.7917	0.2476	10.6859	0.4086	0.1520	11.0401	20.7006	MK
15	3/25/11	MK110325	63.8750	70.8750	0.3131	7.1179	0.6857	0.2667	3.8859	17.7513	MK
16	3/29/11	MK110329	66.7500	74.0417	0.4628	6.7928	0.8272	0.2829	4.7515	15.8878	MK
17	3/31/11	MK110331	44.7917	50.7083	0.4477	4.2772	0.3880	0.2136	2.7727	10.1044	MK
18	4/9/11	MK110409	49.8750	56.9583	0.5887	6.7063	0.3916	0.3034	3.8906	14.8415	MK
19	4/12/11	MK110412	64.3333	74.2500	1.3417	7.3976	0.6255	0.1737	1.8103	21.5748	MK
20	4/18/11	MK110418	33.5417	44.2500	0.1214	3.0270	0.2772	0.0202	0.7076	8.1754	MK
21	4/24/11	MK110424	43.0417	54.2500	0.2250	4.8682	0.3361	0.0564	1.7277	13.0045	MK
22	4/30/11	MK110430	54.5833	61.6667	0.7725	6.1938	0.4845	0.0502	1.1593	20.6579	MK
23	5/6/11	MK110506	31.2917	41.9167	0.4799	3.4637	0.1624	0.0148	0.2584	10.8408	MK
24	5/18/11	MK110518	31.5000	38.9583	0.2331	3.0178	0.1924	0.0224	0.4353	8.6534	MK
25	5/20/11	MK110520	28.7083	34.8333	0.2930	2.7701	0.1236	0.0201	0.3417	8.0928	MK

Figure 1.2 An actual example of excel data (or .csv file).

2 Overall comparison with other programs

The following table compares scatter plot with other programs

Software	Advantage	Disadvantage
Excel	Powerful data filter	OLS only, no Deming regression row size limited to 1 million Can't mask data No color coding
SPSS	data filter	OLS only, no Deming regression Data masking can't by done by GUI
Sigma Plot	Deming regression	No data filter, can't mask data
Origin	York regression Data masking via GUI color coding	No data filter
Scatter Plot **Igor program**	OLS, Deming, Weighted orthogonal distance and York Regression Data filter Data masking via GUI color coding Batch plotting	Require Igor to run the program

3 Import Data

3.1 Timeline example in MS Excel

Before import, data could be kept in Excel, restrictions on timeline format are described as follows.

The timeline in data column **must** follow this format "MM/DD/YY hh:mm"(In Igor 7 or above, use "MM/DD/YYYY hh:mm"), Location **must** be set to "English (United States)"as shown in Figure 3.1.

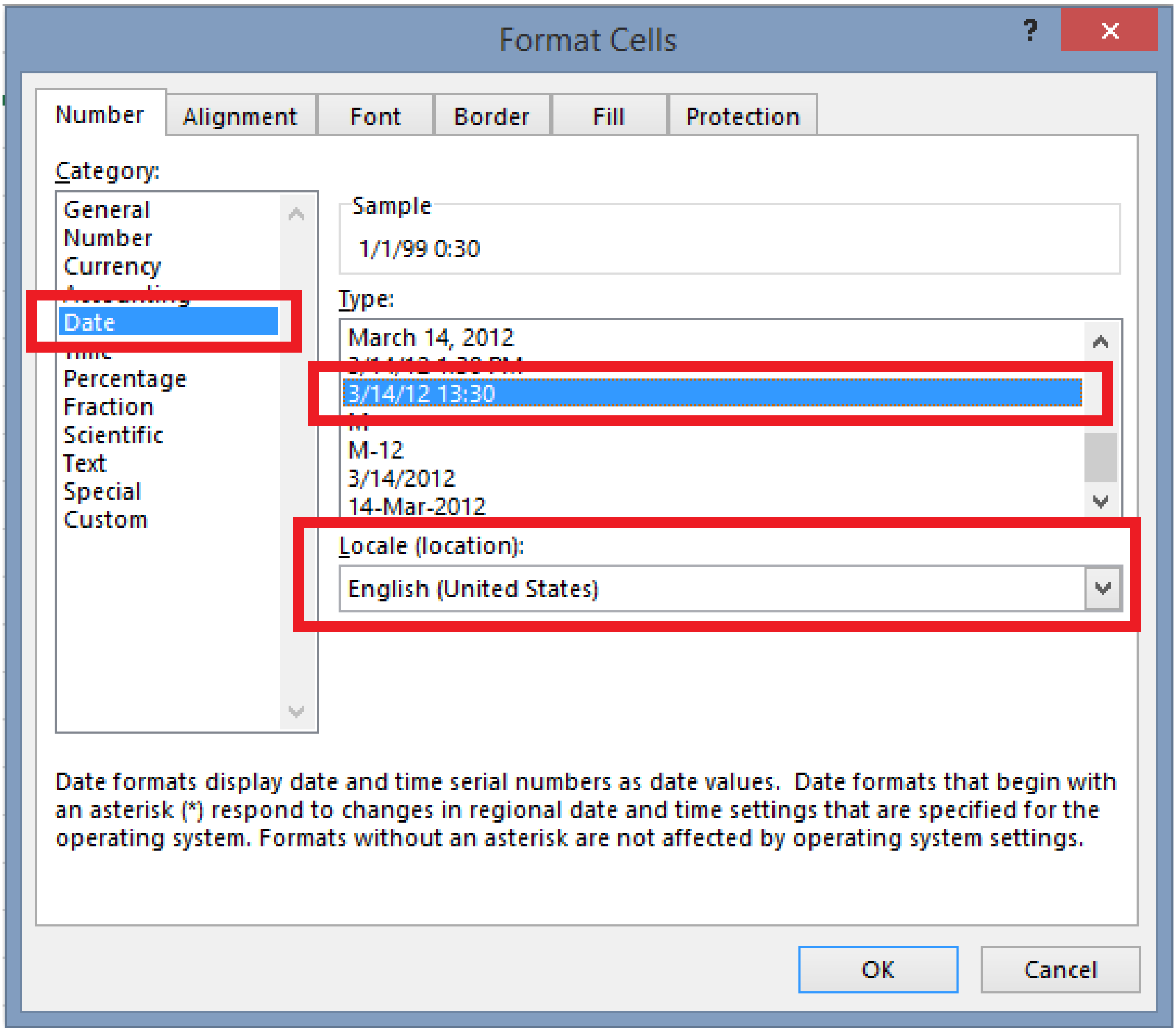


Figure 3.1 Cell format configuration in MS Excel for the timeline column

Make sure the cell format of the timeline column is exactly the same as shown in Figure 3.1, otherwise Iogr cannot recognize it.

To ensure the compatibility, the date time format of the operating system should be checked. The US format is recommended (In window 10 system, go to Control Panel to set the date time format).

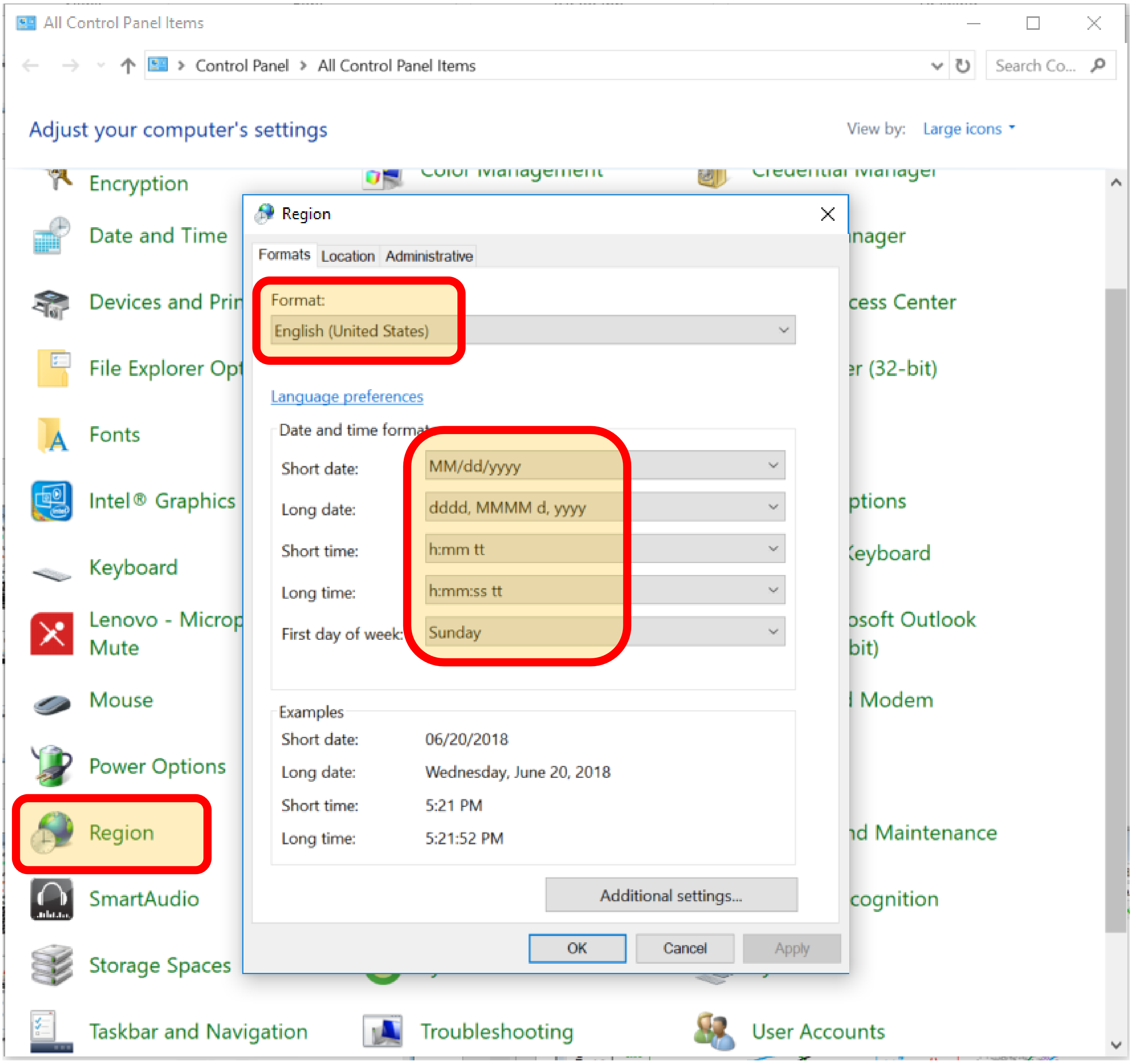
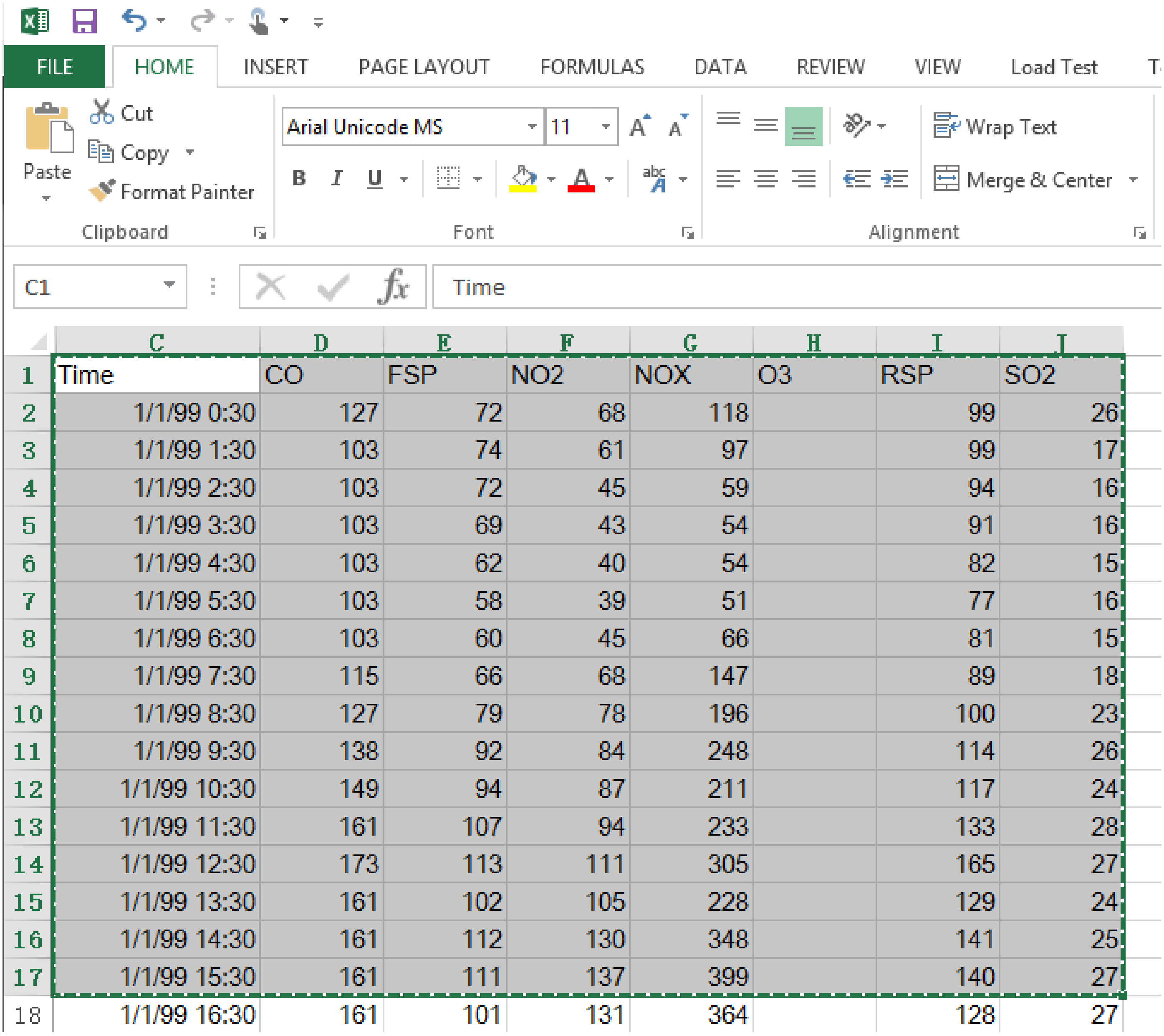


Figure 3.1.2 Date time format settings in window 10.

3.2 Copy from Excel

Data can be imported via copy & paste from MS Excel, as shown in Figure 3.2. It is recommended to put timeline in the first column.



	C	D	E	F	G	H	I	J
1	Time	CO	FSP	NO2	NOX	O3	RSP	SO2
2	1/1/99 0:30	127	72	68	118		99	26
3	1/1/99 1:30	103	74	61	97		99	17
4	1/1/99 2:30	103	72	45	59		94	16
5	1/1/99 3:30	103	69	43	54		91	16
6	1/1/99 4:30	103	62	40	54		82	15
7	1/1/99 5:30	103	58	39	51		77	16
8	1/1/99 6:30	103	60	45	66		81	15
9	1/1/99 7:30	115	66	68	147		89	18
10	1/1/99 8:30	127	79	78	196		100	23
11	1/1/99 9:30	138	92	84	248		114	26
12	1/1/99 10:30	149	94	87	211		117	24
13	1/1/99 11:30	161	107	94	233		133	28
14	1/1/99 12:30	173	113	111	305		165	27
15	1/1/99 13:30	161	102	105	228		129	24
16	1/1/99 14:30	161	112	130	348		141	25
17	1/1/99 15:30	161	111	137	399		140	27
18	1/1/99 16:30	161	101	131	364		128	27

Figure 3.2 Example of data selection and copy (Ctrl + C) in MS Excel. The header of each column will be used as wave name in Igor.

3.3 Paste data into Igor

Put the cursor on up left corner and paste the data into the table in the Igor program interface (the highlighted area in orange) as shown in Figure 3.3.1

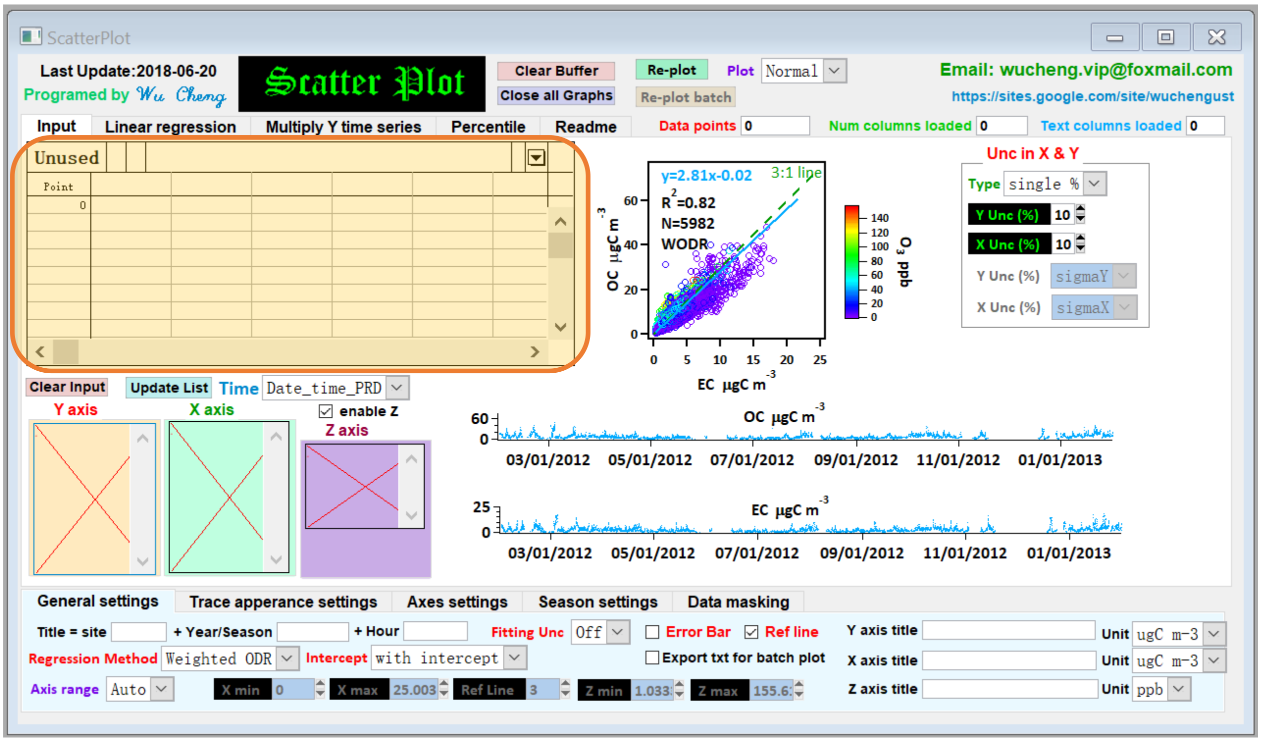


Figure 3.3.1 Example of user interface in scatter plot Igor Pro program before pasting data.

After applying paste (ctrl + V), the data will show up in the table area, make sure the timeline is recognized properly by Igor Pro. It should be noted that the index of data points starts from 0 in Igor.

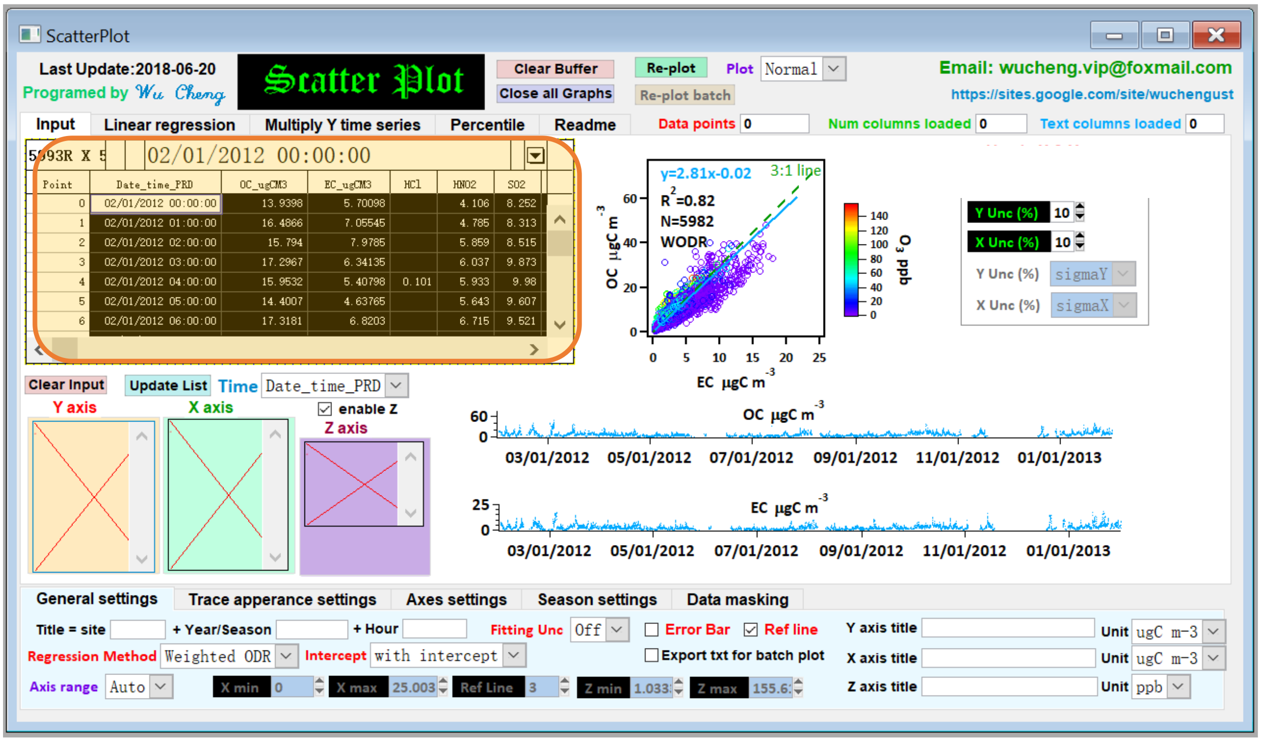


Figure 3.3.2 Example of user interface in scatter plot Igor Pro program after pasting data

3.4 Update list

Click “Update List” button (Figure 3.4, highlighted area a), then the list numerical data series (known as column in Excel and wave in Igor Pro) will be updated (Figure 3.4, highlighted area b). The statistics of loaded data are shown in Figure 3.4 highlighted area c, including number of numerical columns, text columns and data points (rows).

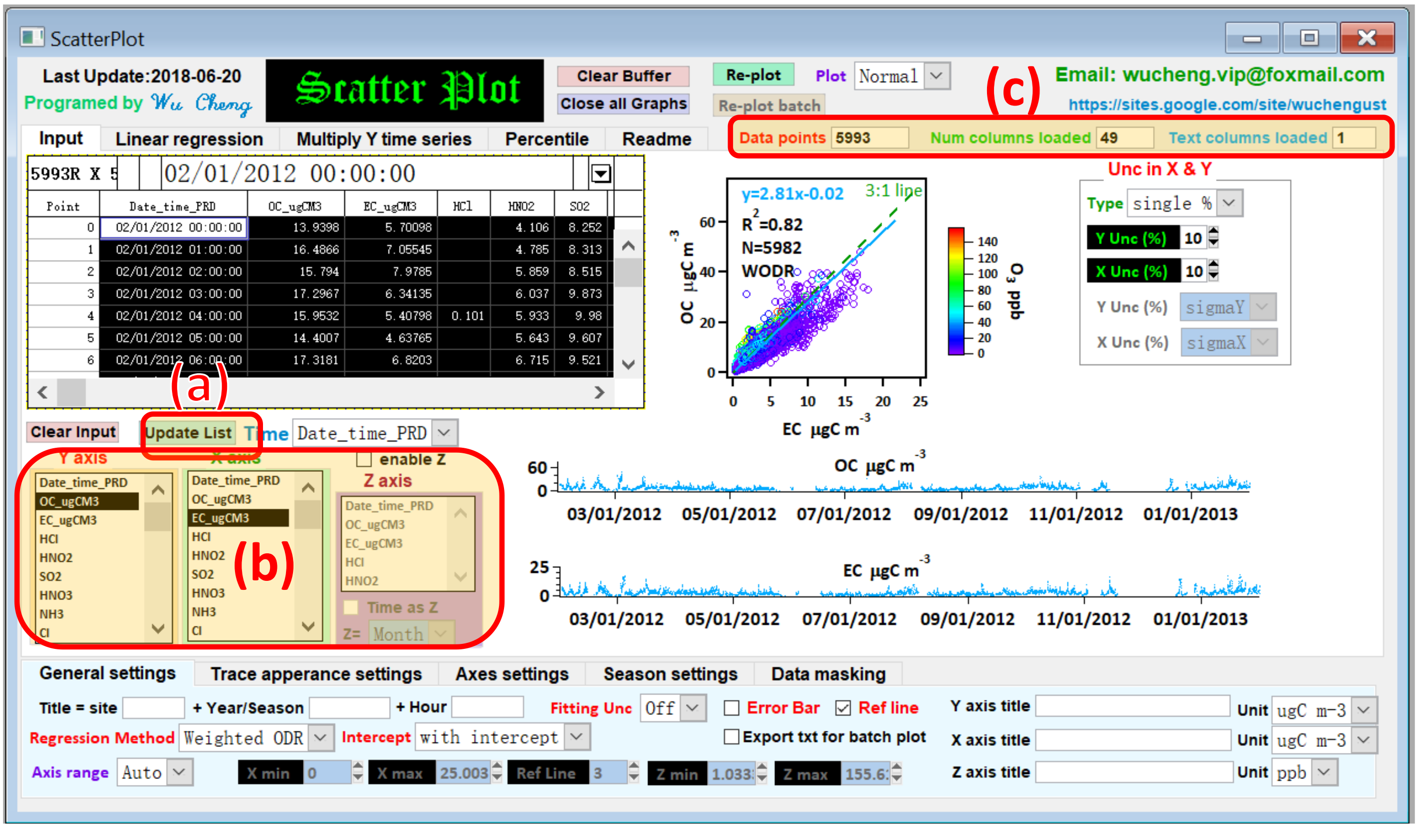


Figure 3.4 Example of Update list in Igor.

3.5 Specify timeline

The next step is to tell the program which column is the timestamp. It can be done by using the pop-up menu as shown in Figure 3.5.

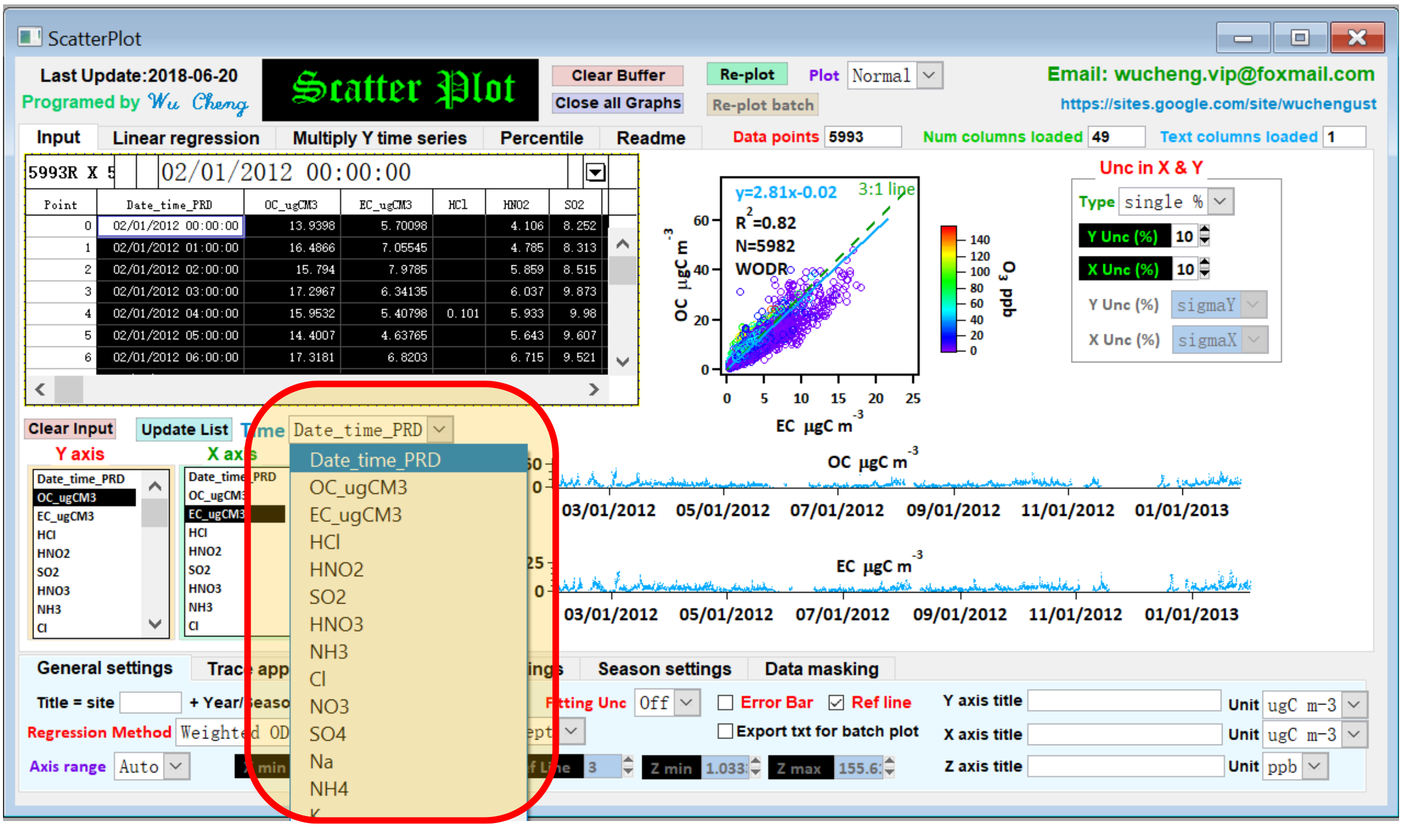
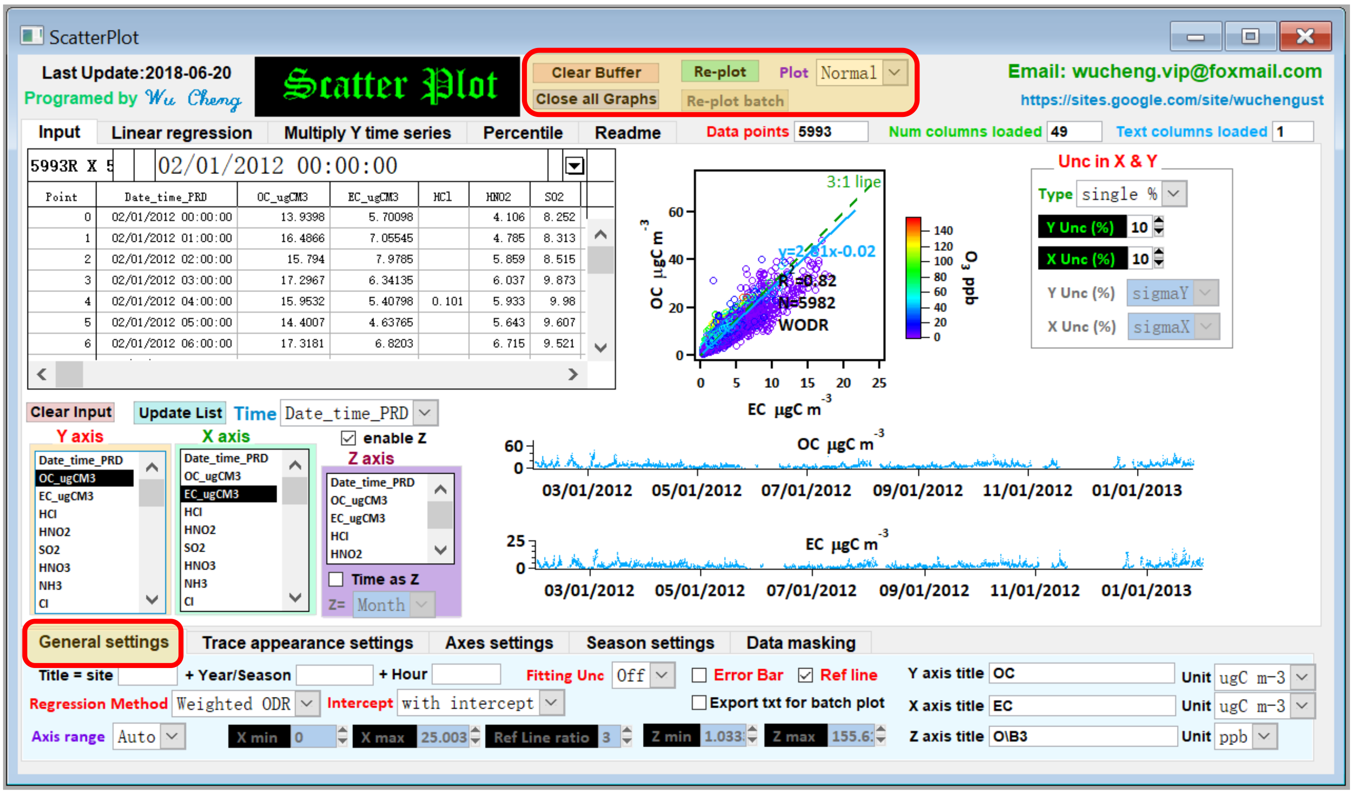


Figure 3.5 Example of specifying timeline in Scatter plot Igor program.

4 Introduction to the setting tabs

4.1 General settings



Close all Graphs: Close all Graphs in the new windows.

Clear Buffer: erase data for batch plots and keep file size of the program.

Replot: replot the figure in the current tab. **Notes: for non-batch-Replot mode, the data will be overwritten once a new graph is generated. For example, Graph1 is generated in a new window, then parameters are changed (e.g. subset of data) and Graph2 is generated in another window. In this case the data of Graph1 has been overwritten and replaced by Graph2. If the user would like to create multiple plots, the plot must be saved (e.g. copy&paste to ppt or save to a file) before the generation of the next plot.**

Plot option: Normal, only replot the current tab; New window, also generate the plot in a new window that can be copy&paste to MS office; Export PNG, not only generate plots in new windows, but also a PNG file; Export EMF, instead of PNG file, EMF is a vector file that can zoom in infinitely. The exported PNG and EMF files are placed in the same folder that igor .exp file (this Scatter plot program) is placed.

Title: Title of plot, contain three fields (Site, Year/Month, Hour and species), these fields will use auto naming when doing a scan if they left as blanks.

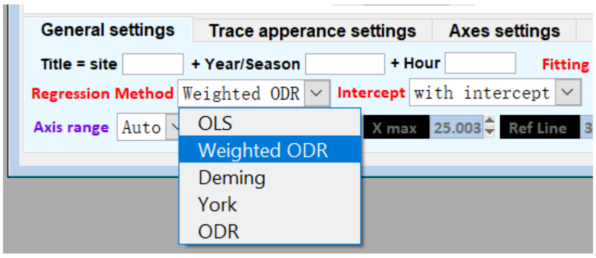
X/Y/Z title: If left as blank, wave name will be adopted as titles, otherwise will be overridden by user input.

Fitting Unc: report the uncertainties of the regressed slope and intercept (WODR only).

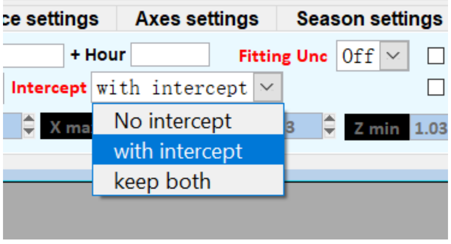
Error Bar: show error bar on data points.

Ref line: show or hide the reference line.

Regression method: Ordinary Least Squared (OLS), Weighted orthogonal distance regression (WODR), Deming regression, York regression. The OLS only consider errors in Y, while the later three consider uncertainties in both Y and X.



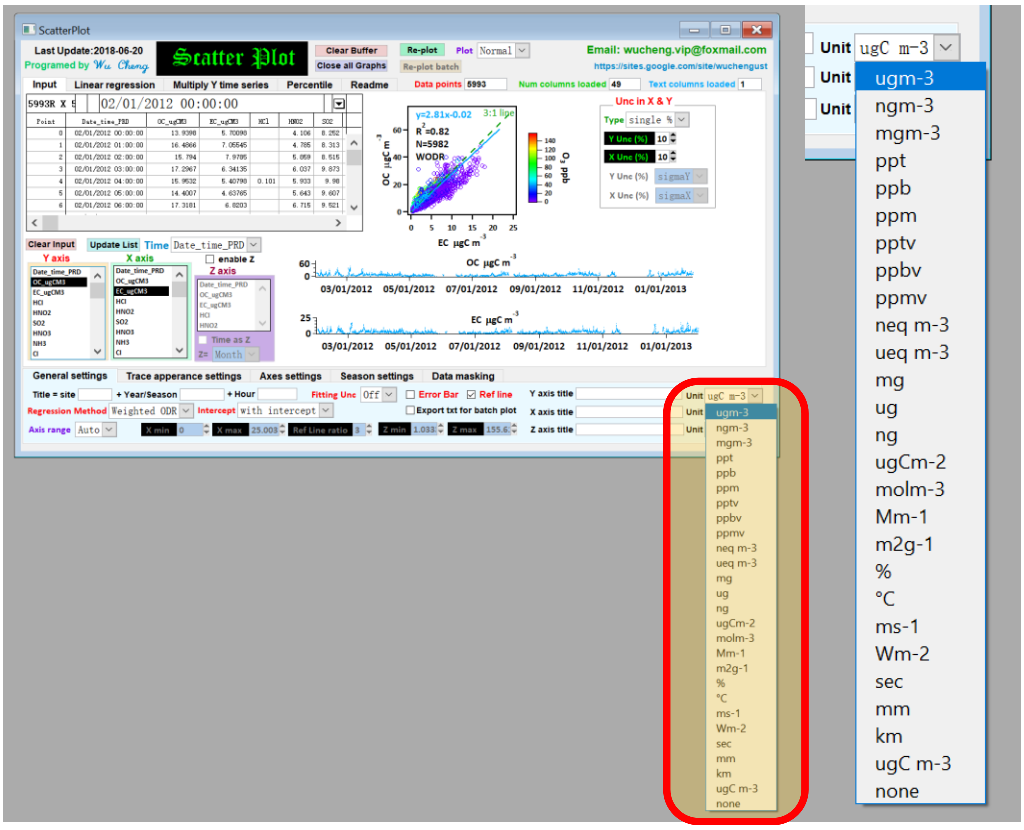
Intercept: If "No intercept" is selected, regression will be performed through origin (Not available for Deming and York Regression). If "with intercept" is selected, all regression methods are available. If "keep both" is selected, both with and without intercept regression will be performed (Not available for Deming and York Regression).



Axis range: turn on or off auto scale. In manual mode, the range of x and z, the ref line ratio can be set manually.

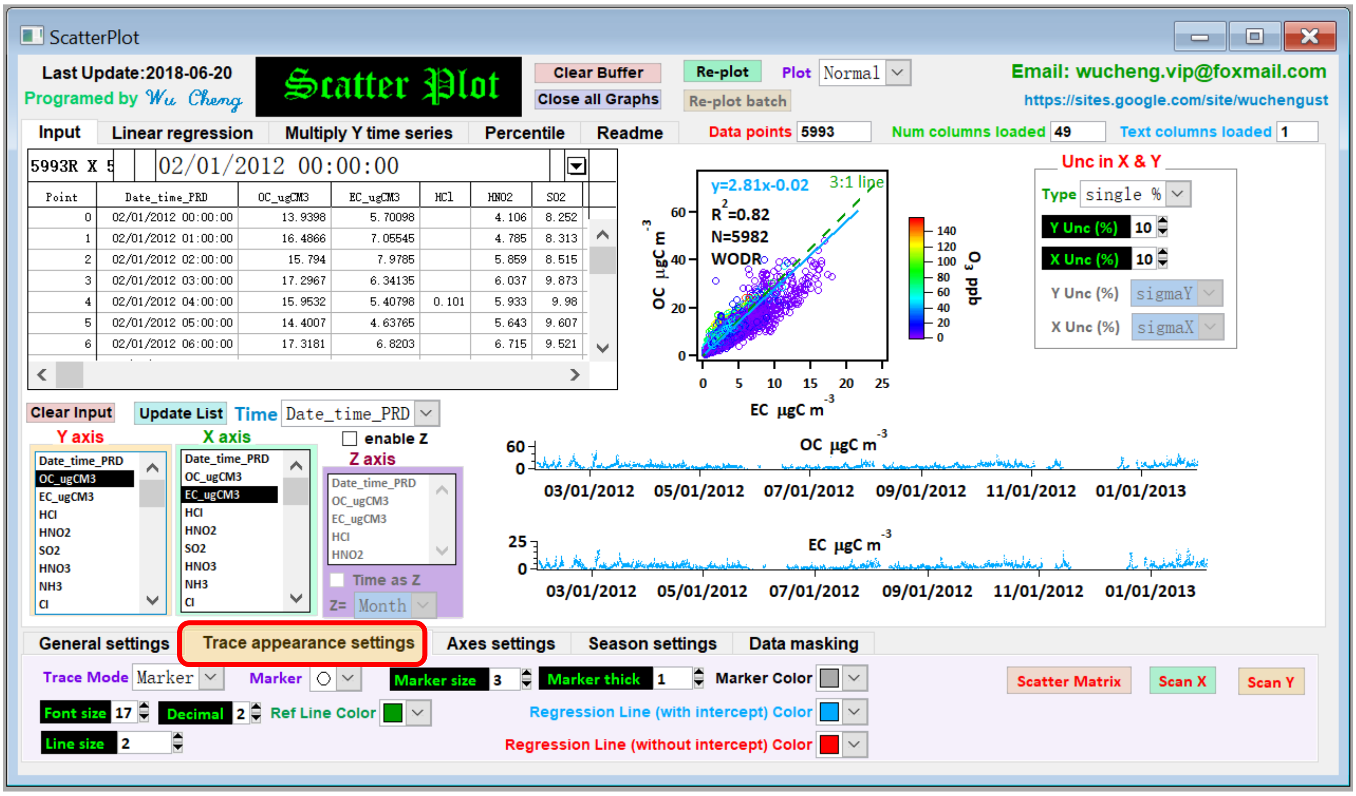
Ref Line ratio: set the ratio of Y:X for the reference dash line (Axis range manual mode).

Unit: use pre-set unit on conc. axes.



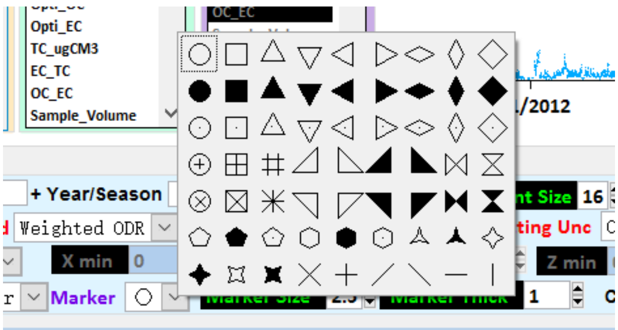
Export txt for batch plot: Export a txt file (placed in the same folder of this .pxp file) the includes the slopes, intercepts and R^2 etc. info for the batch plot.

4.2 Trace appearance settings



Trace Mode: Choose between dot and Marker for display the data points on scatter plot.

Maker: Choose the symbol for the marker



Marker Size: Control the size of the marker.

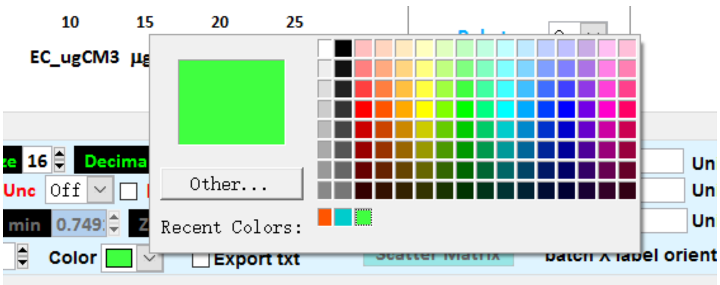
Marker thick: Control the thickness of the marker outline.

Decimal: set how many decimal points to show in the annotations (slope, intercept R^2).

Font Size: control font size in figure.

Line size: control the thickness of lines in the plot.

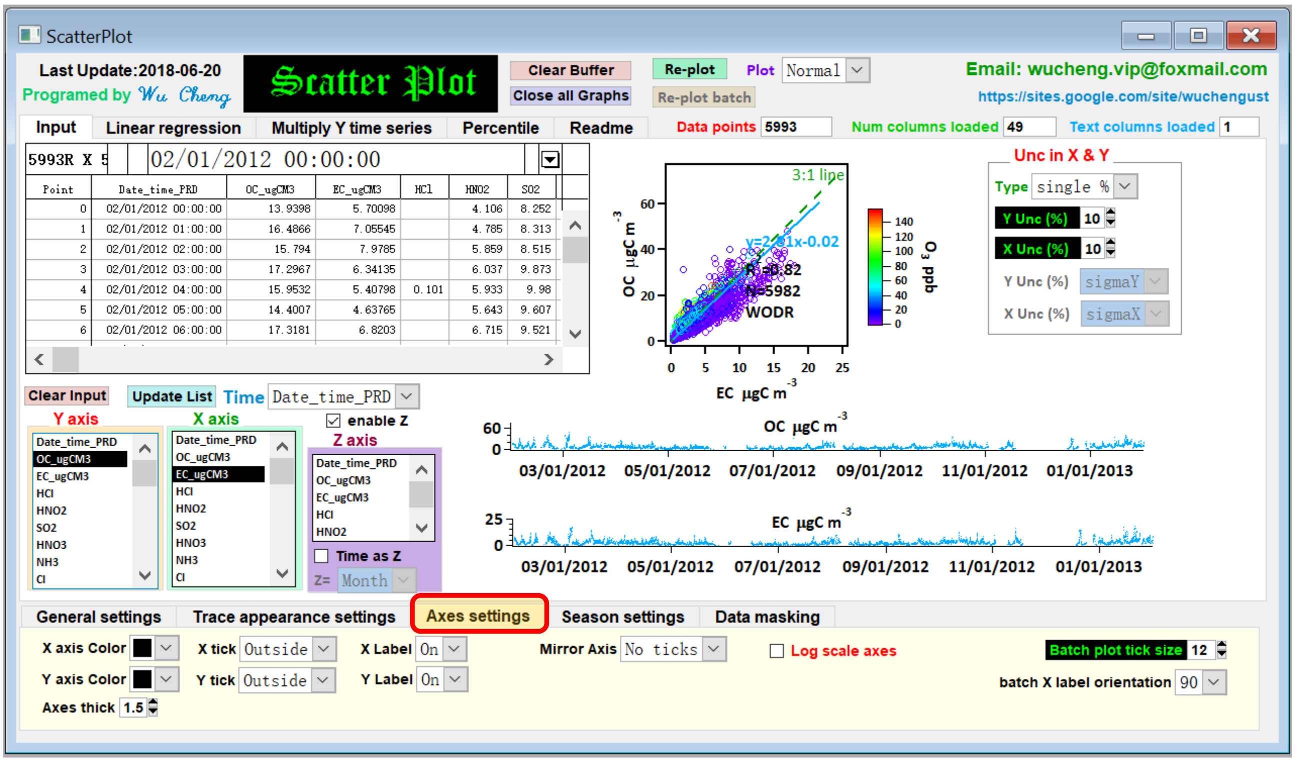
Maker Color: Set marker color (only works when Z axis if **off**)



Regression line Color: Set the color of the regression line.

Ref Line color: Set the color of the reference line.

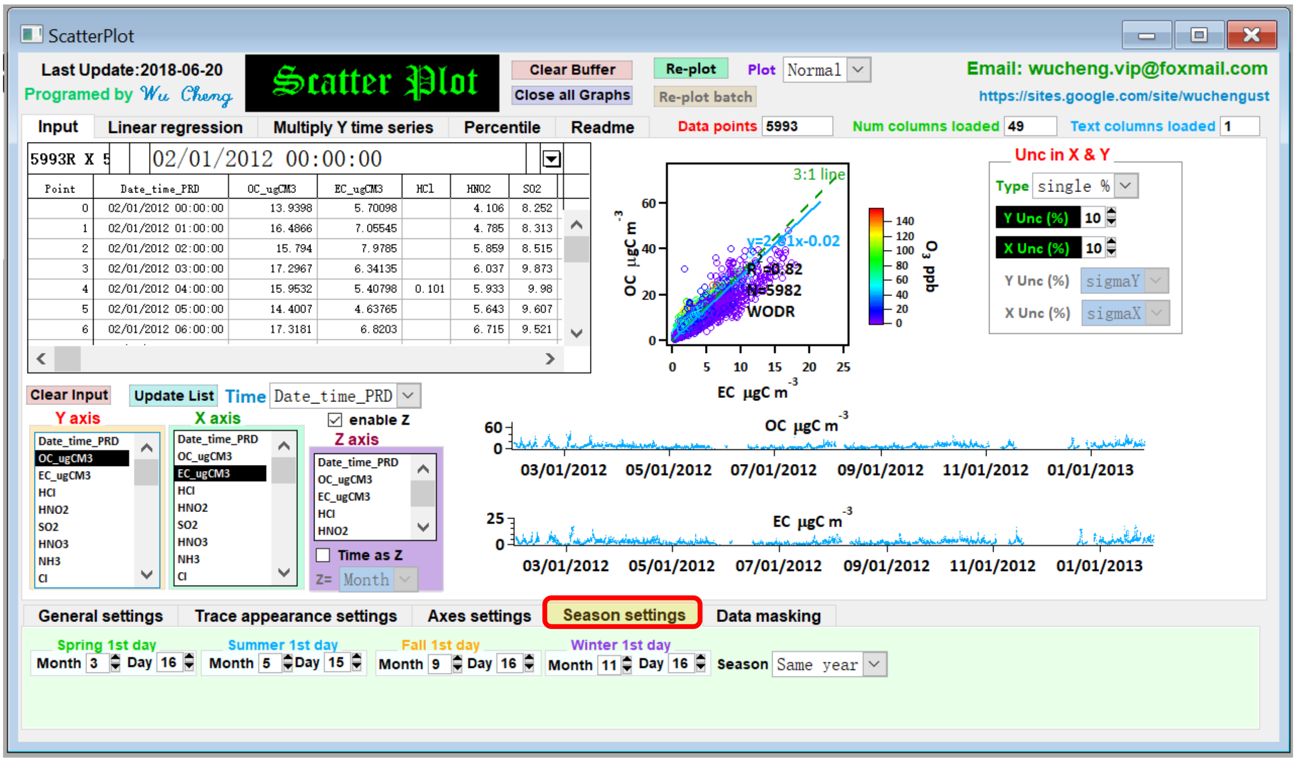
4.3 Axes settings



X/Y axis Color: Set axes color.
X/Y tick: Set the location of the axes ticks.
X/Y Label: Show or hide the axes labels.
Axes thick: Control the thickness of the axes.
Mirror Axis: hide or show mirror axis.

4.4 Season settings

Define the four seasons by inputting the 1st day of each season.



5 Tab "Input" Introduction

(a) User can choose which species to be X, Y and Z (Z can be switch on or off). By choosing different X Y Z combinations, the scatter plot and the two time series plots will be updated instantly so user can take a quick look at the dataset

(b) Unc in X&Y

Uncertainty settings for WODR, Deming and York regression. Two types of input are available, "Single %" means user just need to provide a single number to indicate the relative uncertainty in X and Y. "Input Data" means user need to provide uncertainty for individual data points (a separate wave in the form of standard deviation). User need to use the pop-up menu to indicate the corresponding weighting wave for Y and X.

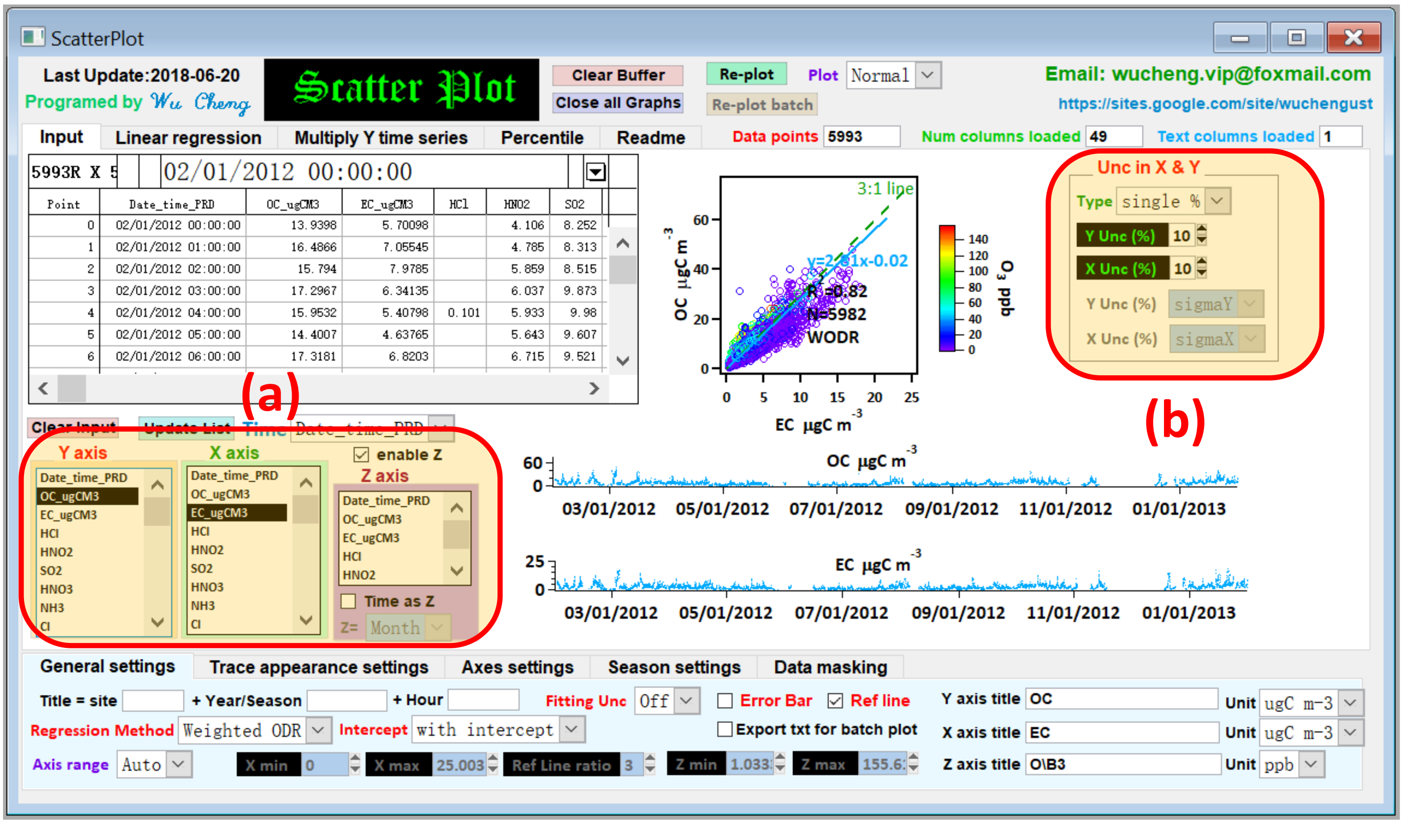


Figure 5.1 Scatter plot settings.

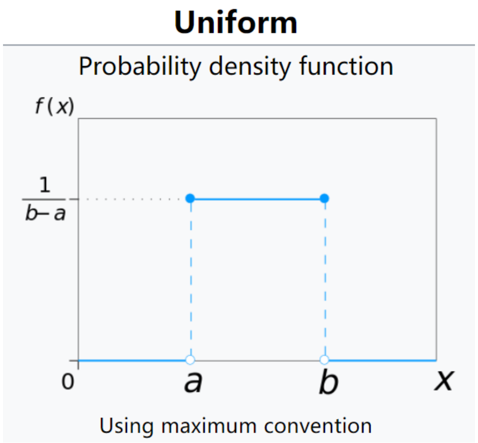
In Deming regression, the key parameter λ is the ratio of the weights:

$$\lambda = \frac{\omega(X_i)}{\omega(Y_i)}$$

and the weights are:

$$\omega(X_i) = \frac{1}{\sigma_{X_i}{}^2}\ , \omega(Y_i) = \frac{1}{\sigma_{Y_i}{}^2}$$

σ_{X_i} and σ_{Y_i} are the standard deviations of the error in measurement of X_i and Y_i respectively. For example, say data point X_i has a ±m% uncertainty, which follow a uniform distribution (in the range of [a,b]). the variance of the uniform distribution becomes



$$\sigma_{X_i}{}^2 = \frac{1}{12}(b-a)^2$$
$$= \frac{1}{12}\left(X_i + m \times X_i - (XX_i - m \times X_i)\right)^2$$
$$= \frac{1}{12}(2 \times m \times X_i)^2 = \frac{(mX_i)^2}{3}$$

As a result, the standard deviation of the error can be written as

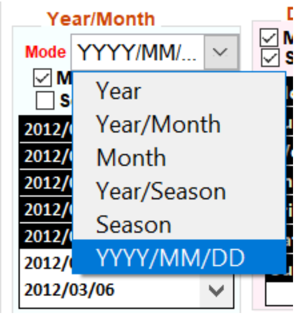
$$\sigma_{X_i} = \frac{mX_i}{\sqrt{3}}$$

In "Input Data" mode, σ_{X_i} and σ_{Y_i} are required as measurement error input for WODR and YR. σ_{X_i} and σ_{Y_i} are also used to calculate λ for Deming regression.

6 Tab “Linear regression” Introduction:

6.1 data filter by time

Three type of time scales are used for data filtering: **YSM** (year season month), **Dow** (day of week) and **Hour** (0:00~23:00)



(a) **YSM** can be further divided into six scenarios: **Year; Year/Month; Month; Year/Season; Season; YYYY/MM/DD**. Season is defined by the highlighted area (d) using the first day of each season as cut-off. Two options are available, across year and same year. For example, if select across year, 1999 Dec and 2000 Jan are grouped together as 1999 winter. Multiple selection is possible using the shift key during selection.

(b) **Dow** (day of week). User defined multiply selection is possible using the shift key during selection. User can also use the **weekday/weekend** check box for quick selection.

(c) **Hour** (0:00~23:00). Multiply selection is possible using the shift key during selection. User can also use the **daytime/nighttime** check box for quick selection. User can define daytime duration in (d).

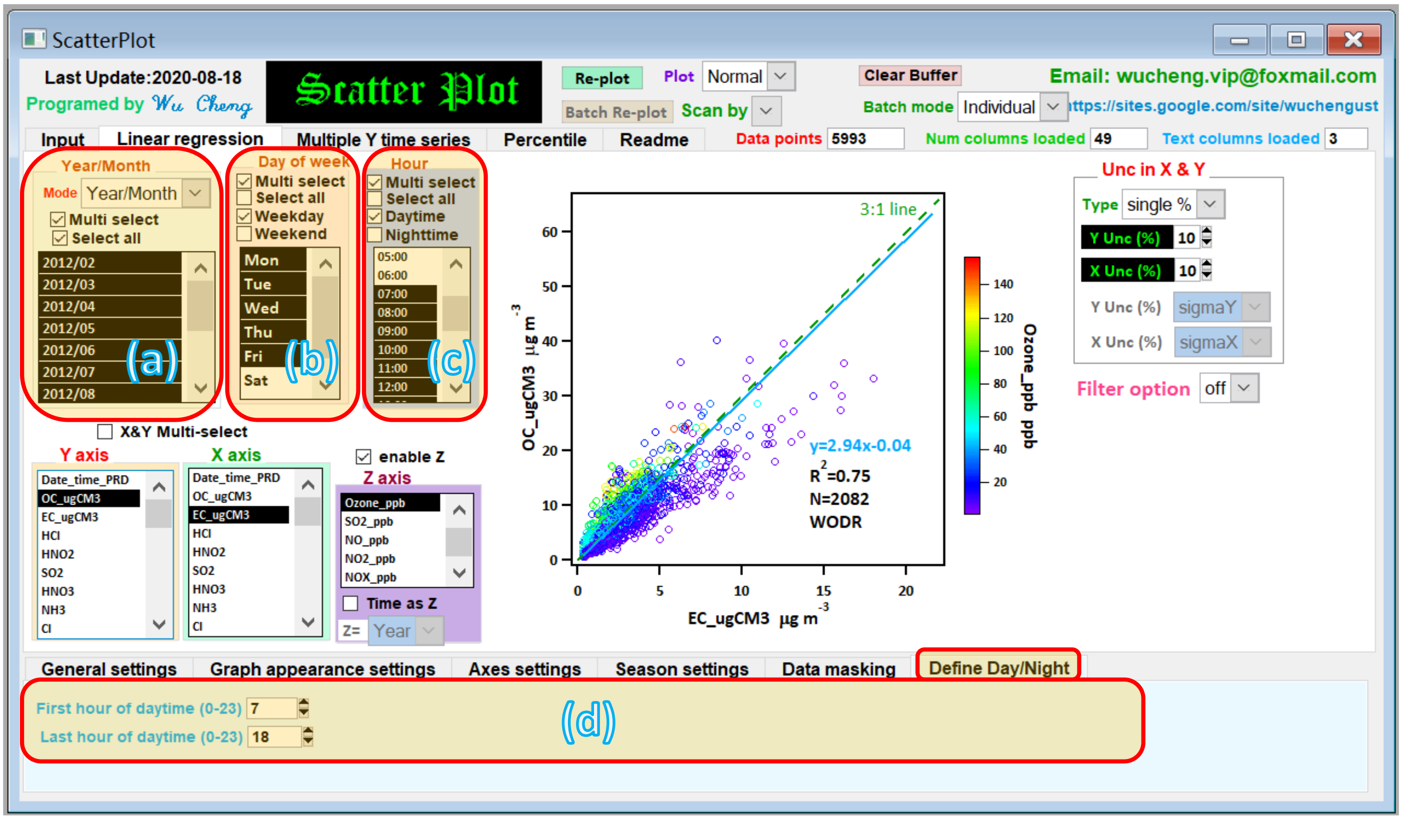


Figure 6.1.1 List boxes for time filtering

6.2 Data filter by data

Three types of data filtering are possible: Text data by list, Numerical data by List and Numerical data by range.

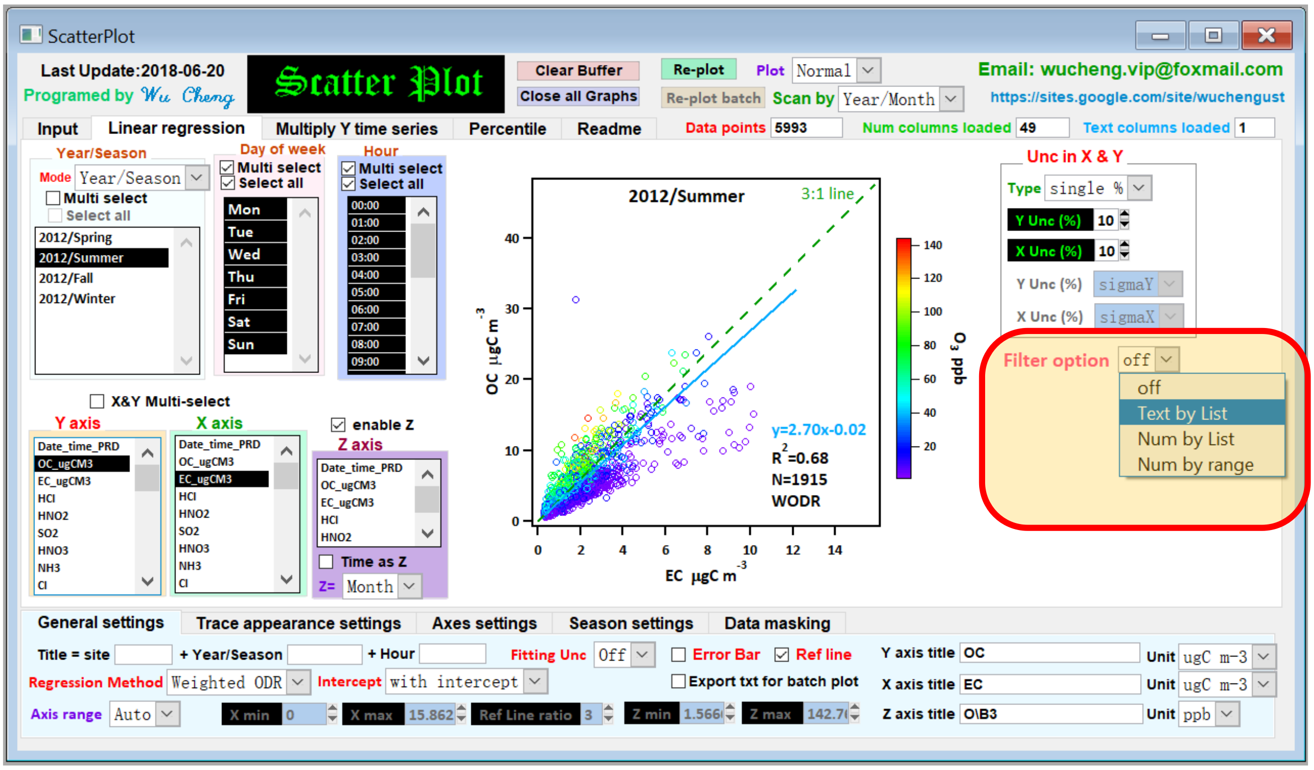


Figure 6.2.1 Data filter by data

(a) **Text by list**: say a column provide the back trajectories grouping info that includes C1~C4, using this function, a subset (e.g. as shown below, only C2 in summer) is plotted. Multiple selection is possible. Daytime/nighttime and weekend/weekday options are also provided.

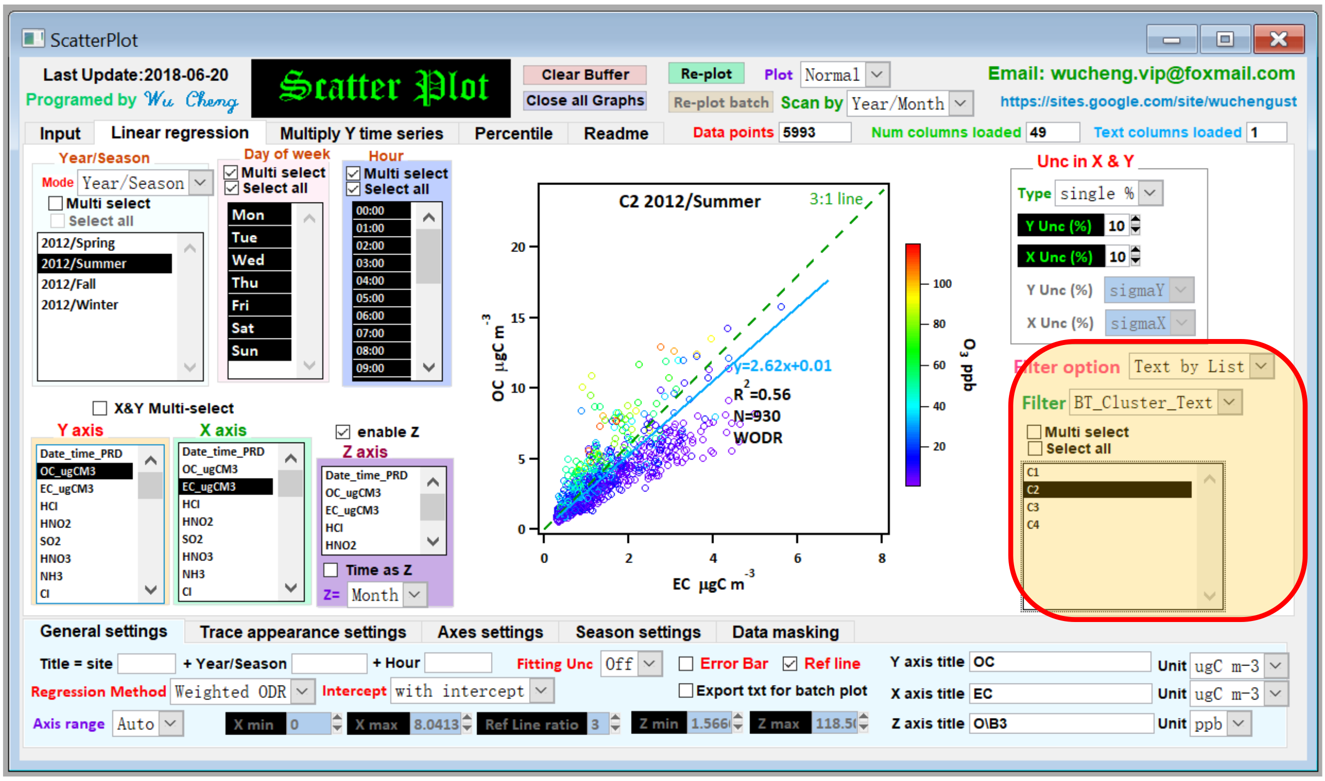


Figure 6.2.2 Data filter by data – Text by List

An example is shown below, using Text by list (daytime/nighttime) + overlaid style batch plotting.

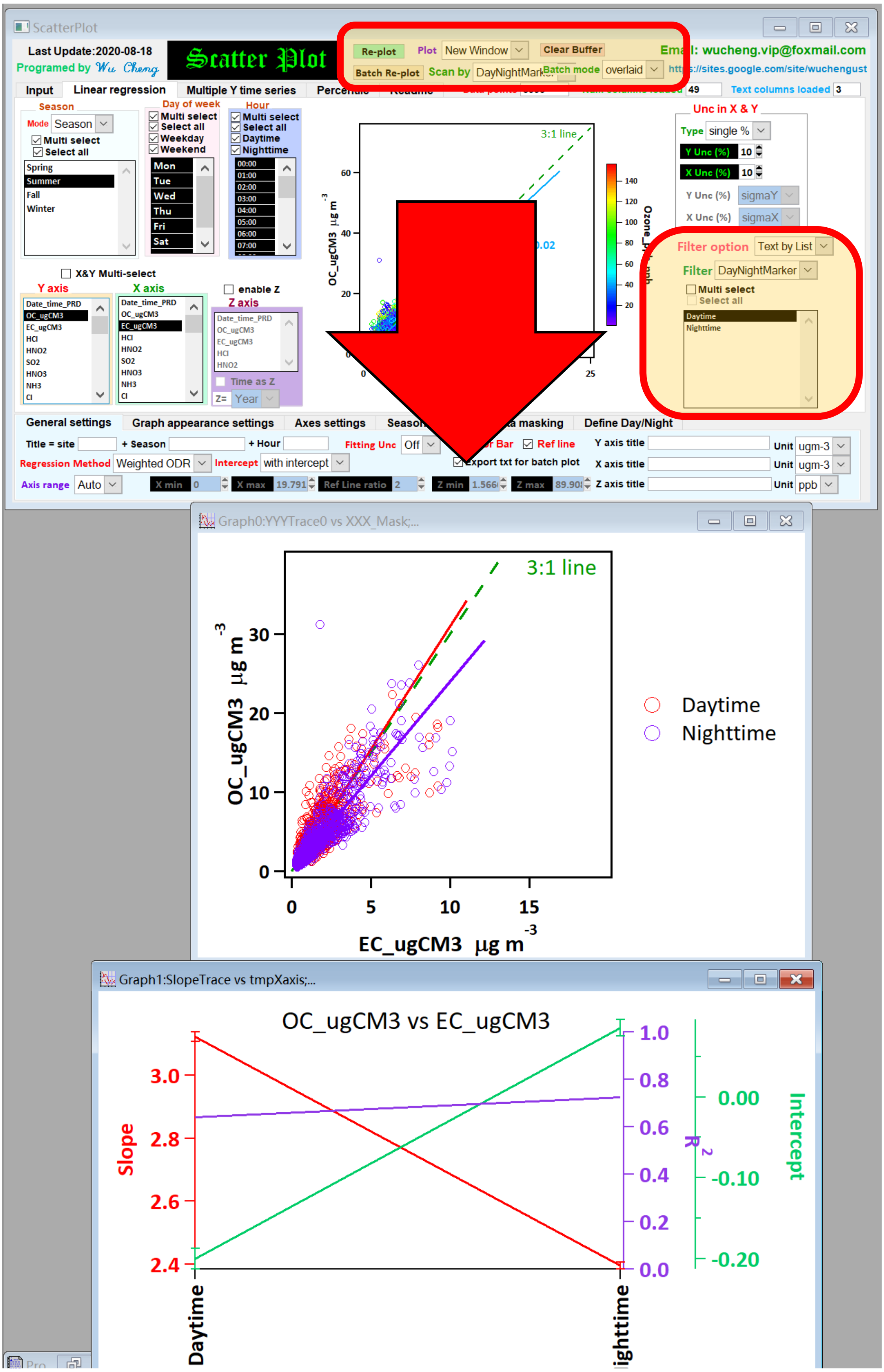


Figure 6.2.3 Example of Text by list (daytime/nighttime) + overlaid style batch plotting

(b) **Num by List**: a column with numerical values can be used a filter for data grouping. It's useful when the number of unique values are much smaller than the total counts. For example, data is grouped by RH as shown below.

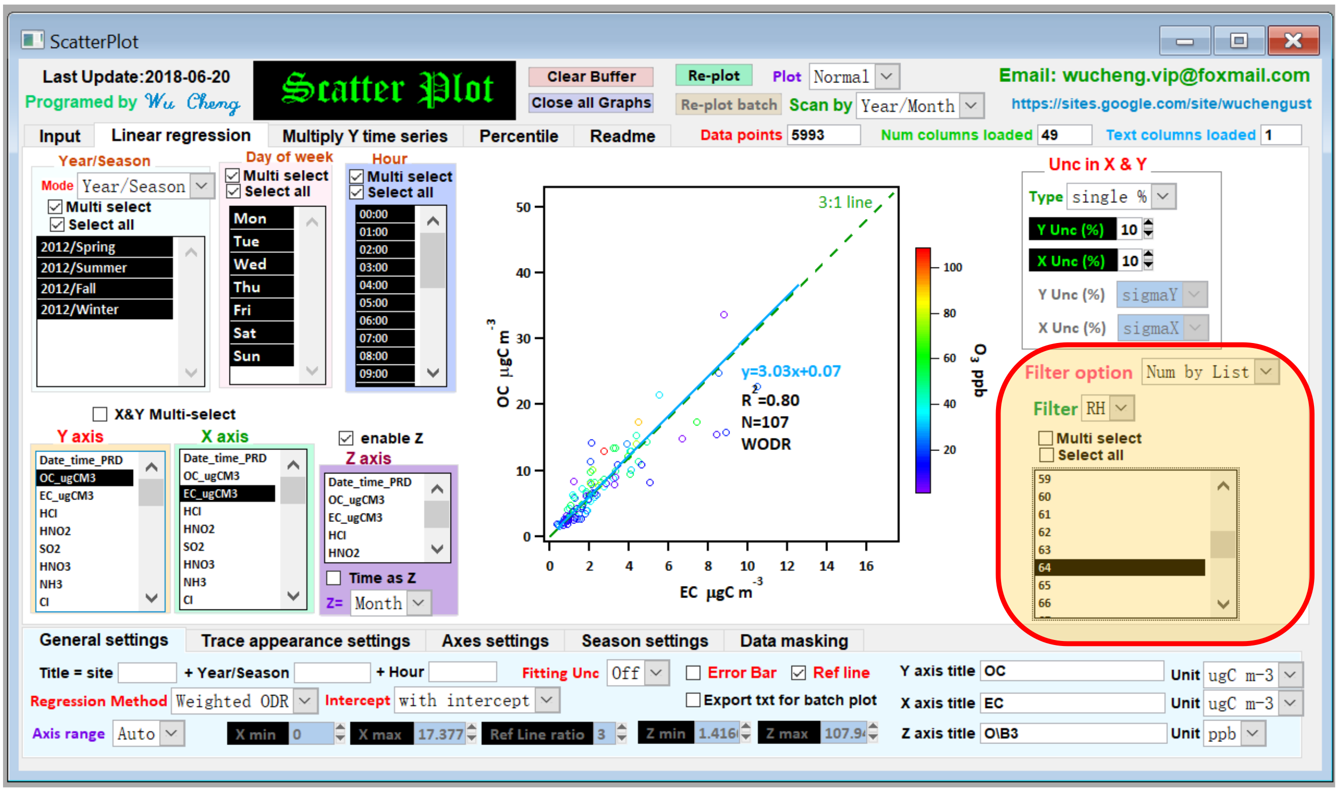


Figure 6.2.4 Data filter by data – Num by List

(c) **Num by range**: A range (defined by min and max) can be used for individual columns to extract a subset. When multiple columns are used, intersection of these conditions are applied. Following is an example of using 30<RH<50.

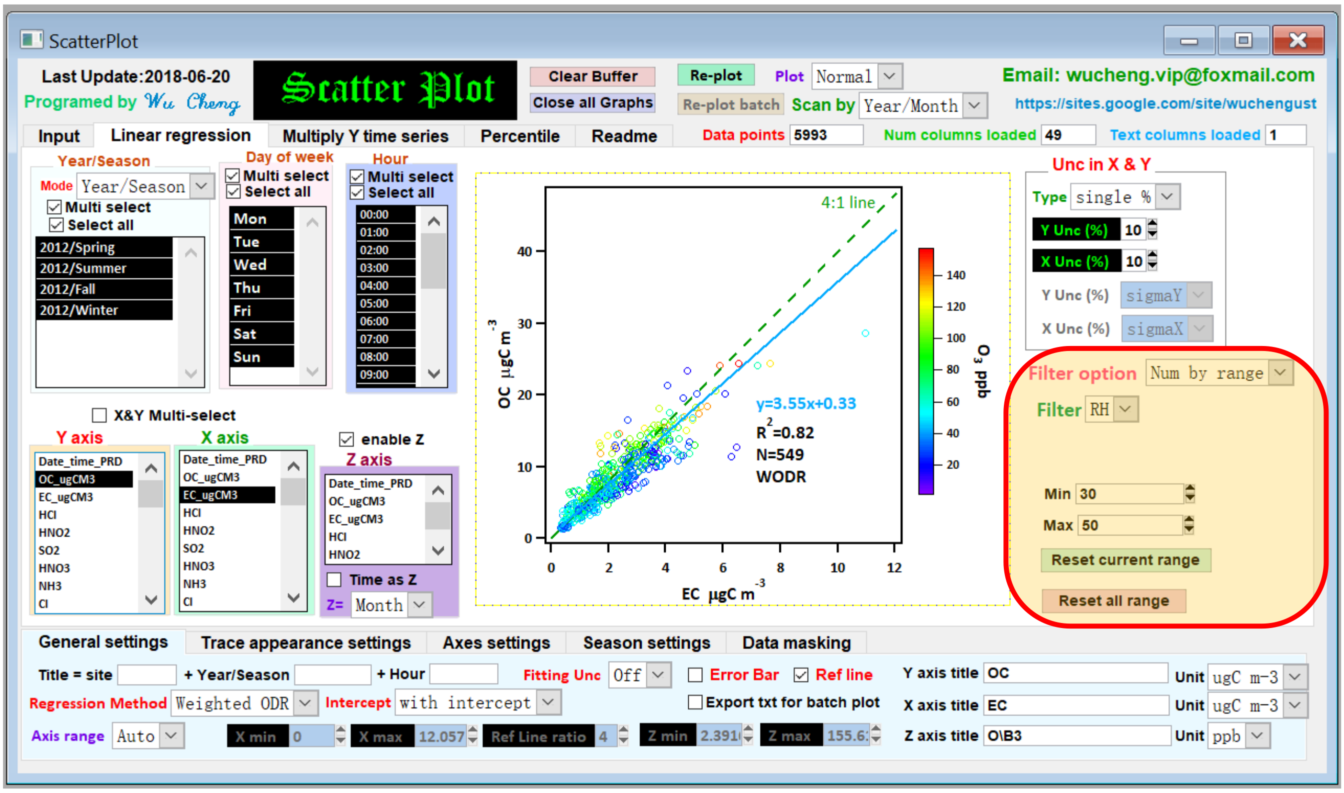


Figure 6.2.5 Data filter by data – Num by range

6.3 Data masking via GUI

Data Masking feature to exclude unwanted data points for regression. Data Masking can be applied using graphic user interface.

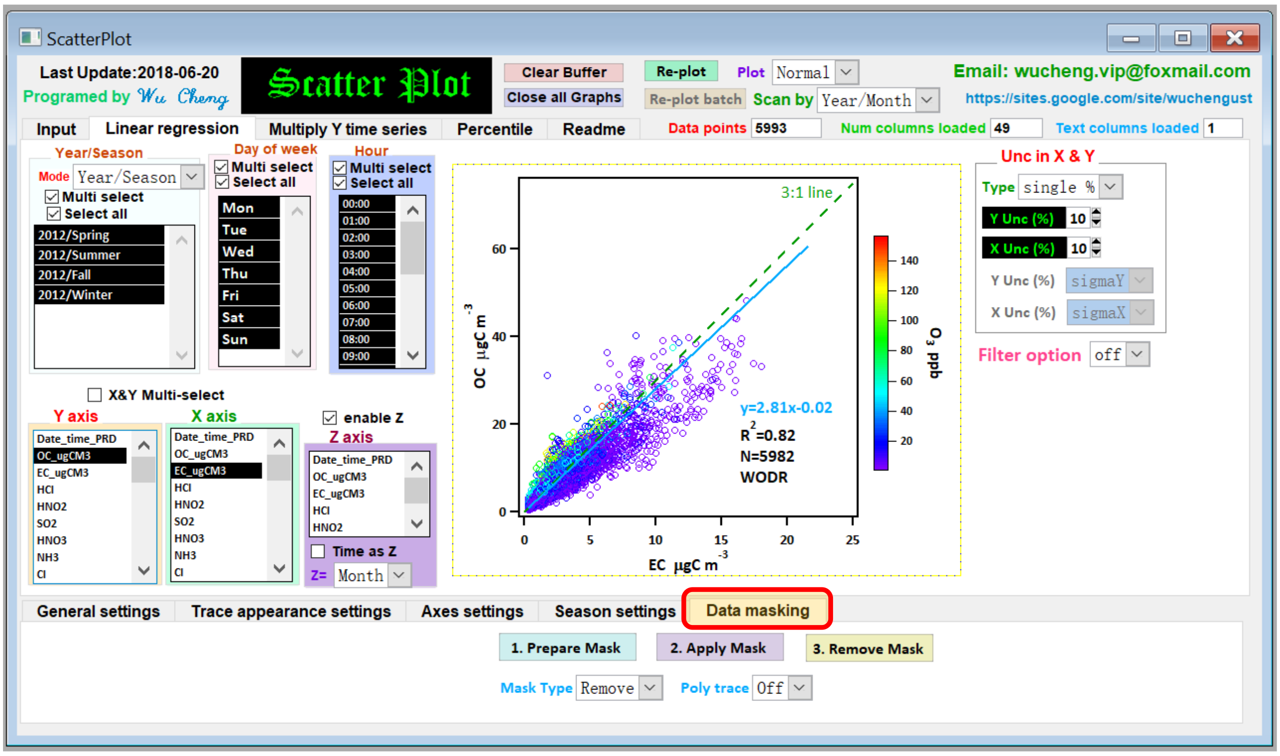


Figure 6.3.1 Data Masking overview

First, click the "Prepare Mask" button. Then a polygon can be drawn using the cursor. The polygon is defined by way points.

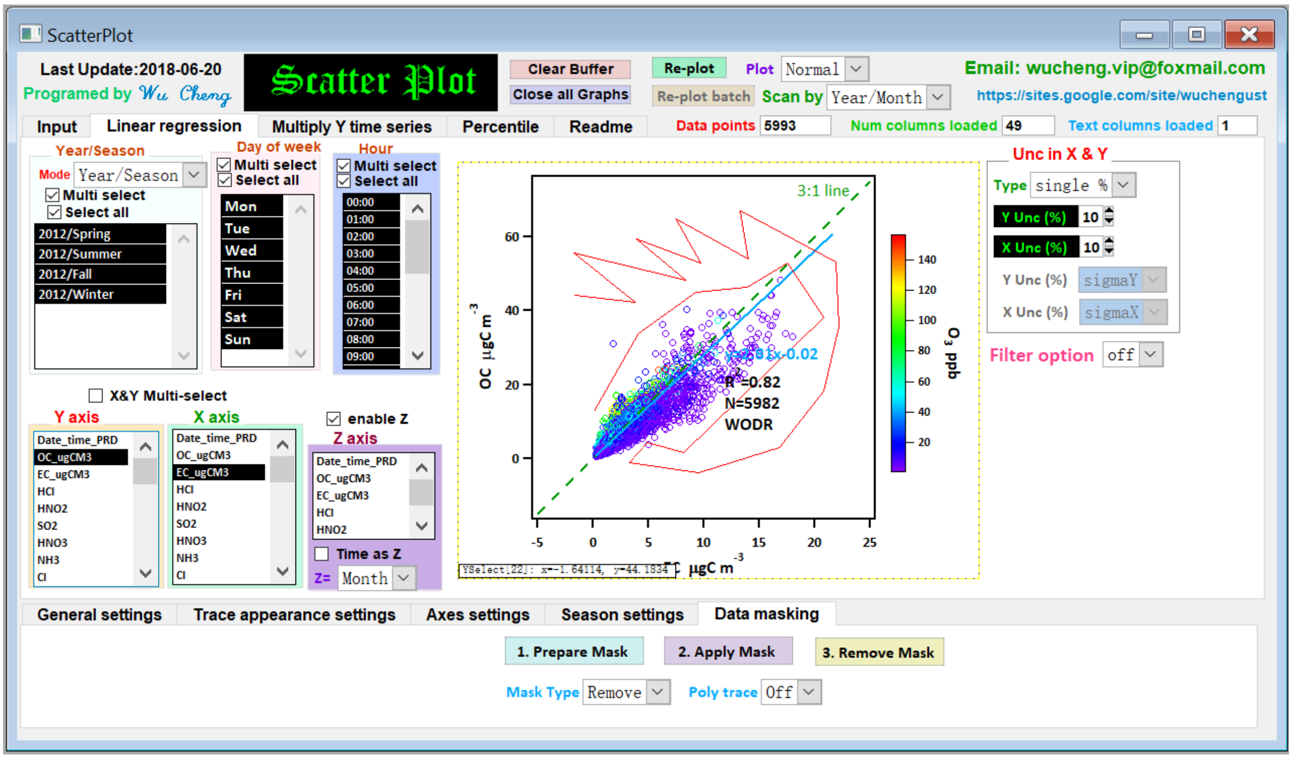


Figure 6.3.2 Start a polygon drawing

Make sure the polygon is closed as shown in Figure 6.3.3, each way points will be shown in the form of square markers.

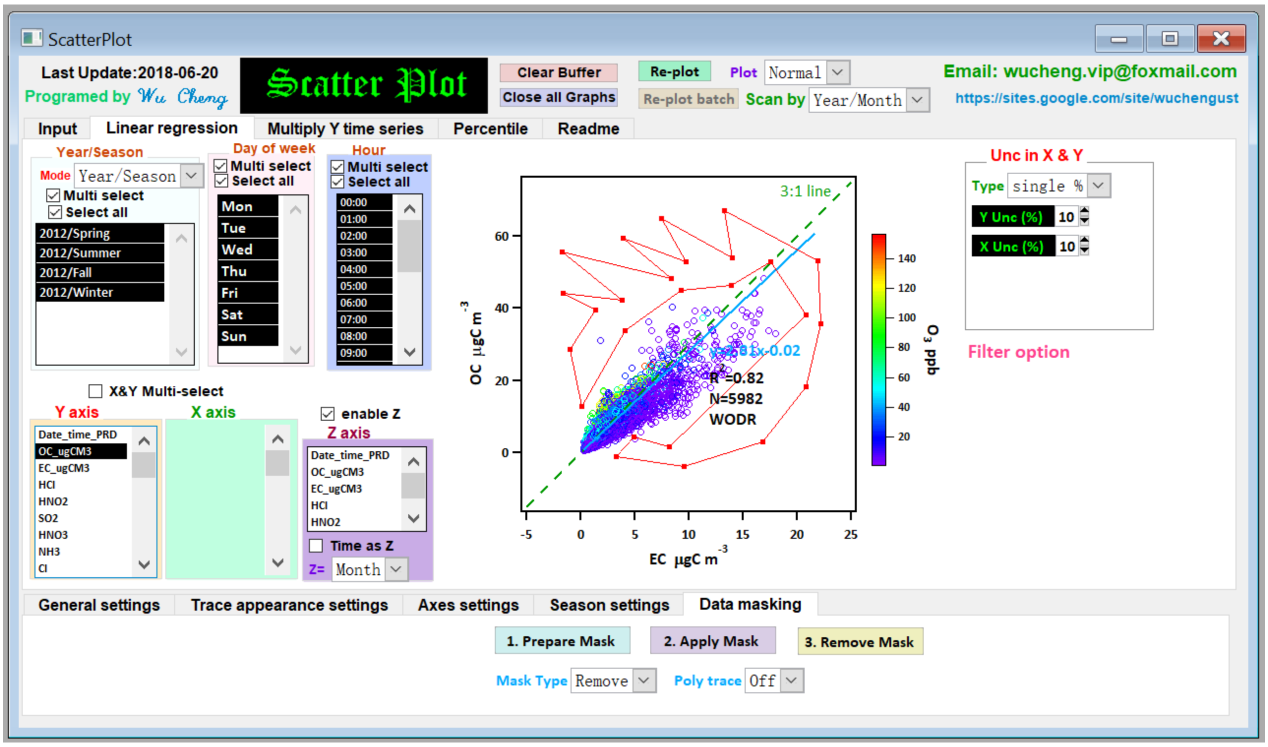


Figure 6.3.3 An example of closed polygon.

Once polygon is finished, choose "Mask type". Following is an example of choosing “Highlighted”, then click "Apply Mask" button. Unwanted data points are marked as pink triangles. Data points inside the polygon are excluded for regression (N changed from 5982 in Figure 6.3.3 to 5980 in Figure 6.3.4). For this particular example, the removal of unwanted data points didn’t affect the slope and intercept.

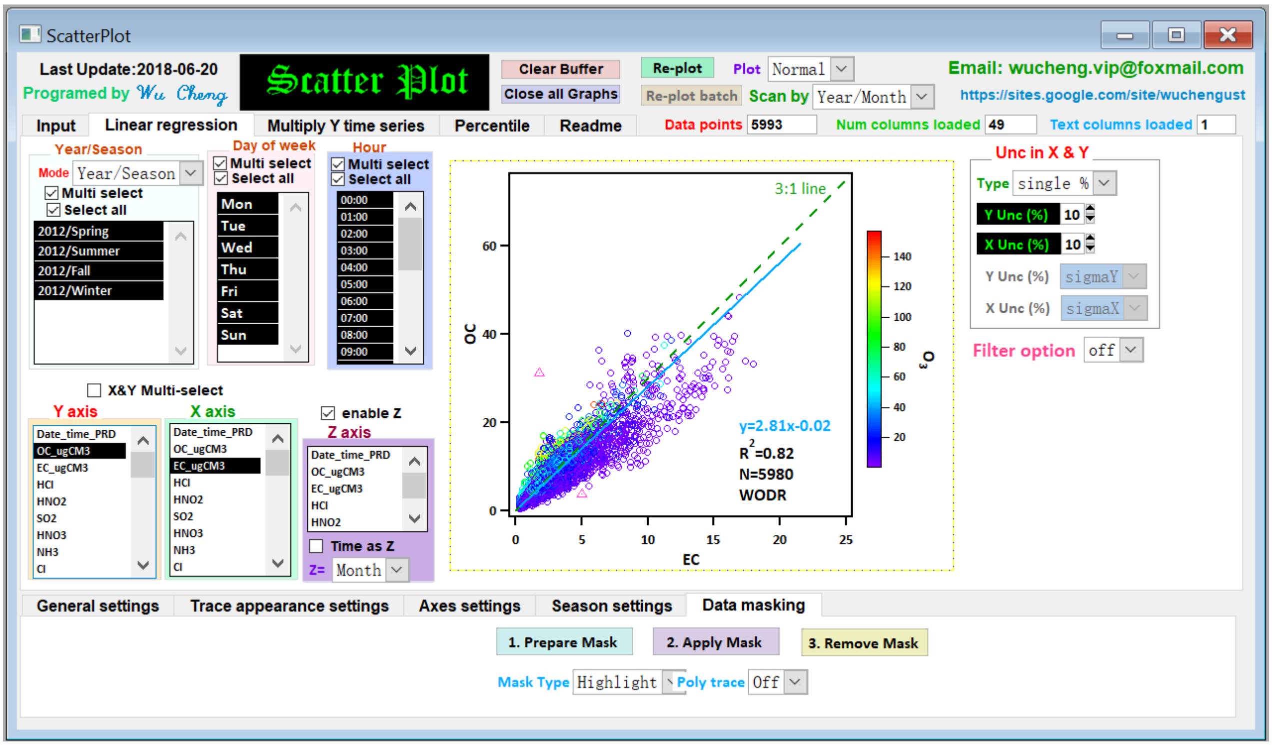


Figure 6.3.4 An example of choosing “Highlighted” in "Mask type".

Following is an example of choosing "Remove" in "Mask type", then click "Apply Mask" button. Unwanted data points are removed. Data points inside the polygon are excluded for regression (N changed from 5982 in Figure 6.3.3 to 5980 in Figure 6.3.5). For this particular example, the removal of unwanted data points didn't affect the slope and intercept.

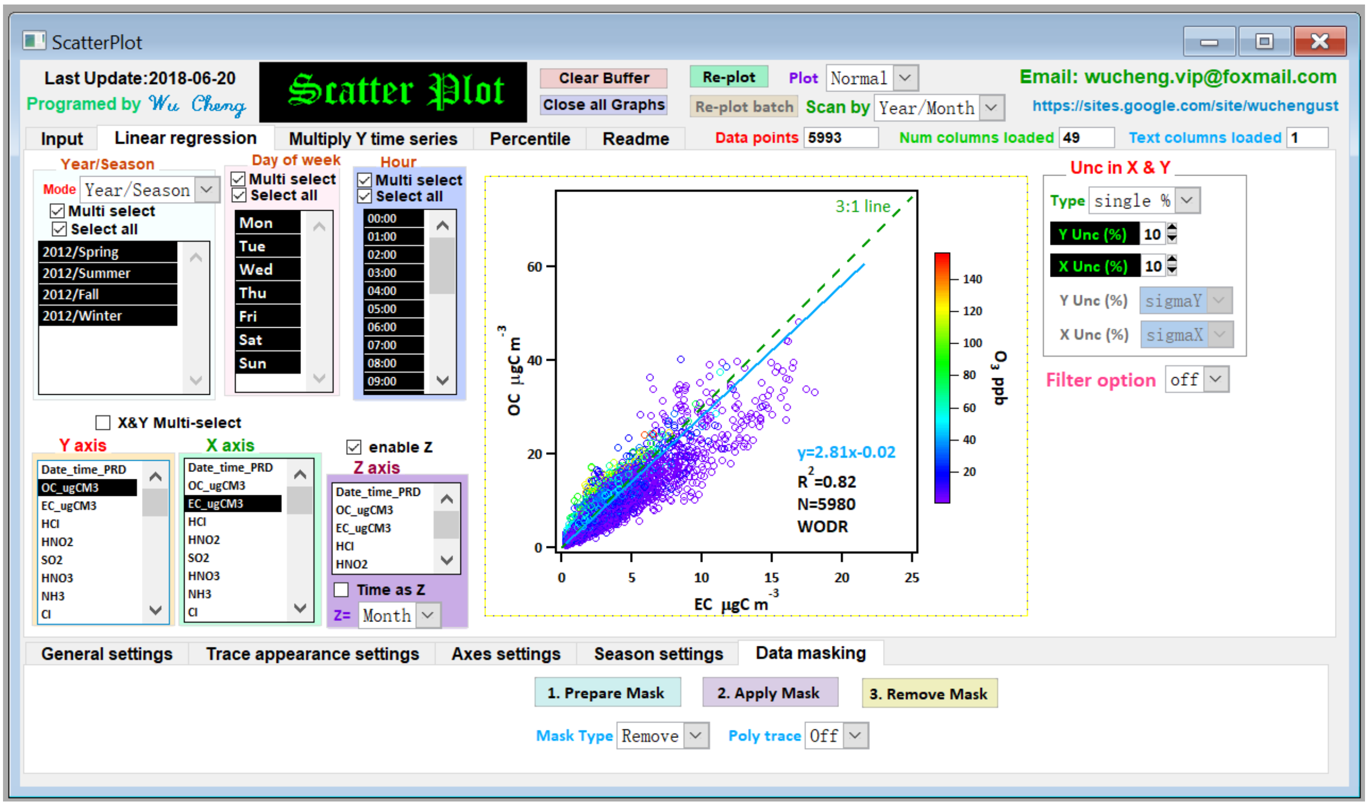


Figure 6.3.5 An example of choosing "Remove" in "Mask type".

Following is an example of choosing “On” in "Trace type”, then click "Apply Mask". Unwanted data points are removed and polygon is shown in dash line. Data points inside the polygon are excluded for regression (N changed from 5982 in Figure 6.3.3 to 5876 in Figure 6.3.6). For this particular example, the removal of unwanted data points results in changed slope (2.81->2.85) and intercept (-0.02->-0.05).

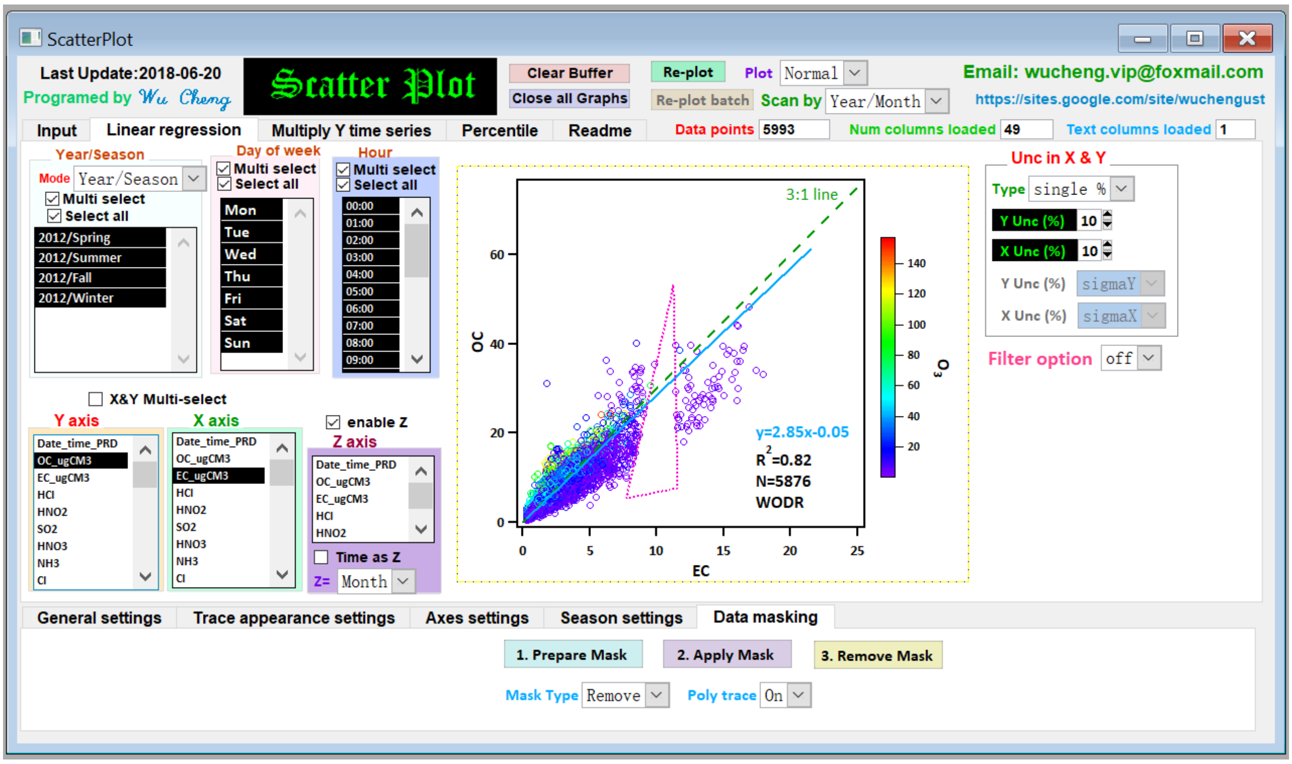


Figure 6.3.6 An example of choosing “On” in "Trace type".

The mask wave can be exported to a csv file for further use. Click “Export mask as txt” to export.

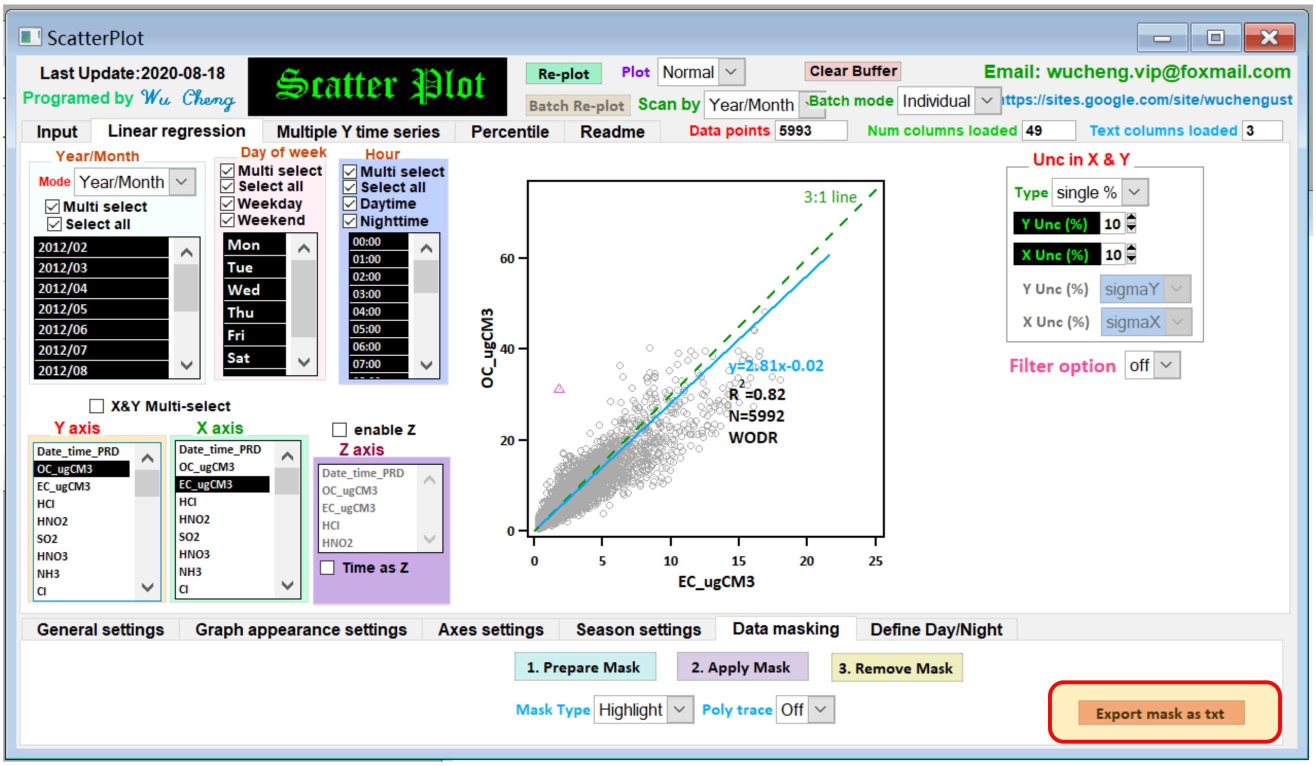


Figure 6.3.7 Export mask as txt

The exported mask wave in csv format can be open by Excel as shown below. The masked data points are marked by 1 while the normal data points are marked as 0.

	A	B	C	D	E
1	TimeText	W_inPoly			
2706	7/12/2012 07:00	0			
2707	7/12/2012 08:00	0			
2708	7/12/2012 09:00	0			
2709	7/12/2012 10:00	0			
2710	7/12/2012 11:00	0			
2711	7/12/2012 12:00	0			
2712	7/12/2012 13:00	0			
2713	7/12/2012 14:00	0			
2714	7/12/2012 15:00	0			
2715	7/12/2012 16:00	0			
2716	7/12/2012 17:00	0			
2717	7/12/2012 18:00	0			
2718	7/12/2012 19:00	0			
2719	7/12/2012 20:00	0			
2720	7/12/2012 21:00	1			
2721	7/12/2012 22:00	0			
2722	7/12/2012 23:00	0			
2723	7/13/2012 00:00	0			

Figure 6.3.8 Example csv file of exported mask wave.

To reset data masking, click “Remove Mask”, then all data masking will be wiped as shown below.

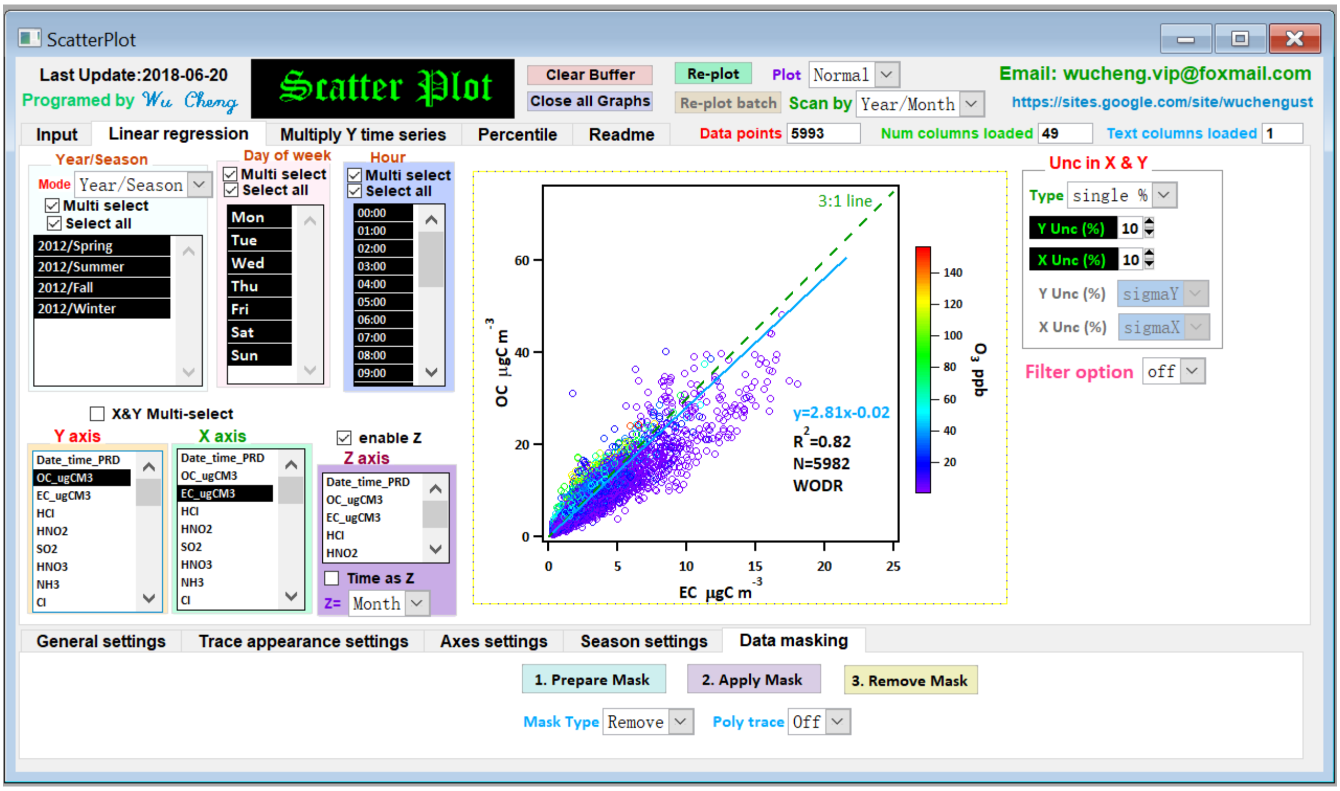


Figure 6.3.9 An example after “Remove Mask” applied.

6.4 Multiple selection in X&Y

Sometimes X and Y are not one to one variables, but multiple to one or multiple to multiple. Multiple selection in X&Y function allows user to use multiple variables in both X and Y via their sum. Multiple waves in X and Y can be selected using "shift" key with cursor. The sum of selected waves in X and Y are used for regression. Following is an example of QA/QC for ion chromatograph data of aerosols. Sum of sulfate and nitrate is used as Y against ammonium as X to check the charge balance of ions.

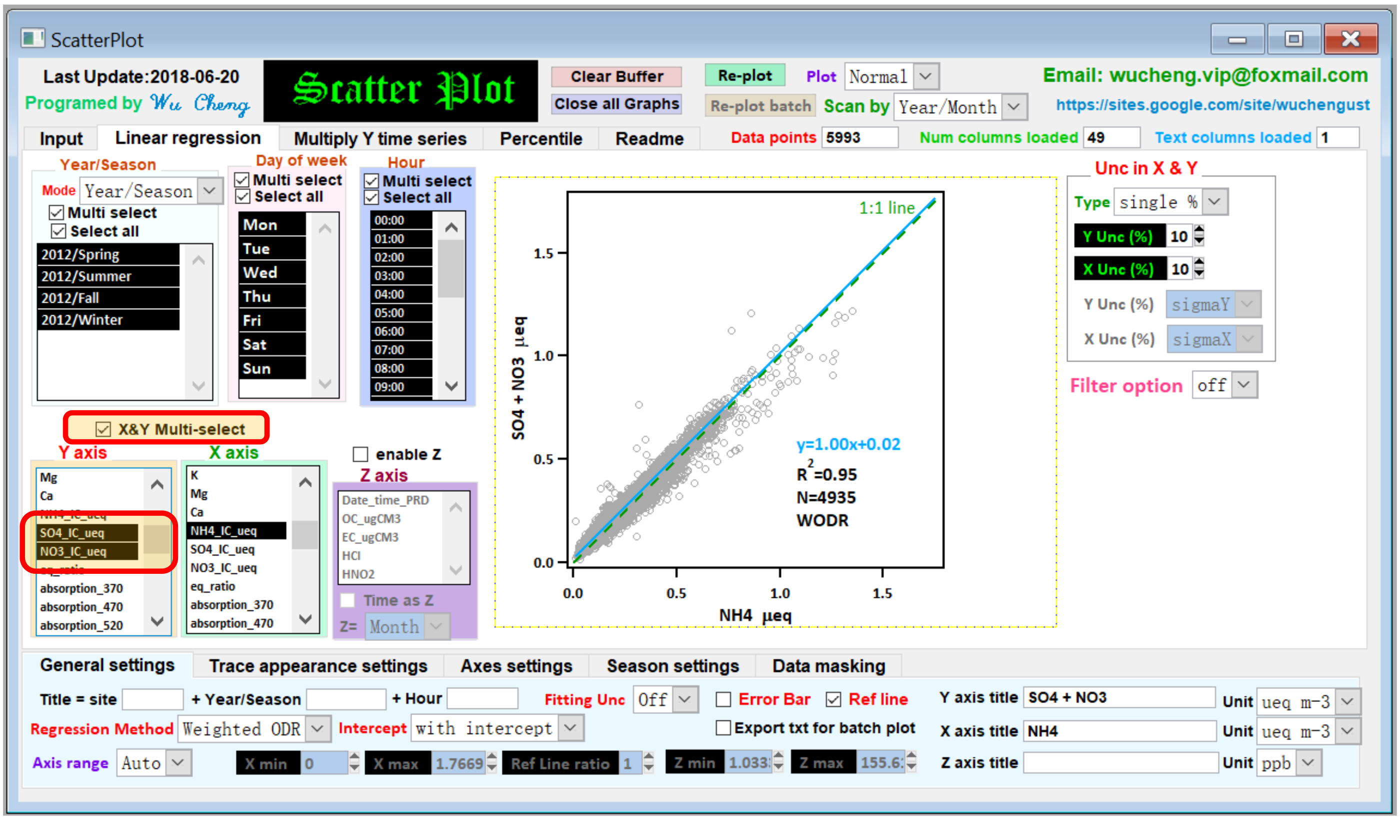


Figure 6.4.1 An example of multiple selection in X&Y.

6.5 Time as Z

Besides using user direct input variables as Z axis, derived variables including YSM (year season month), Dow (day of week) and Hour (0:00~23:00) can be used as Z.

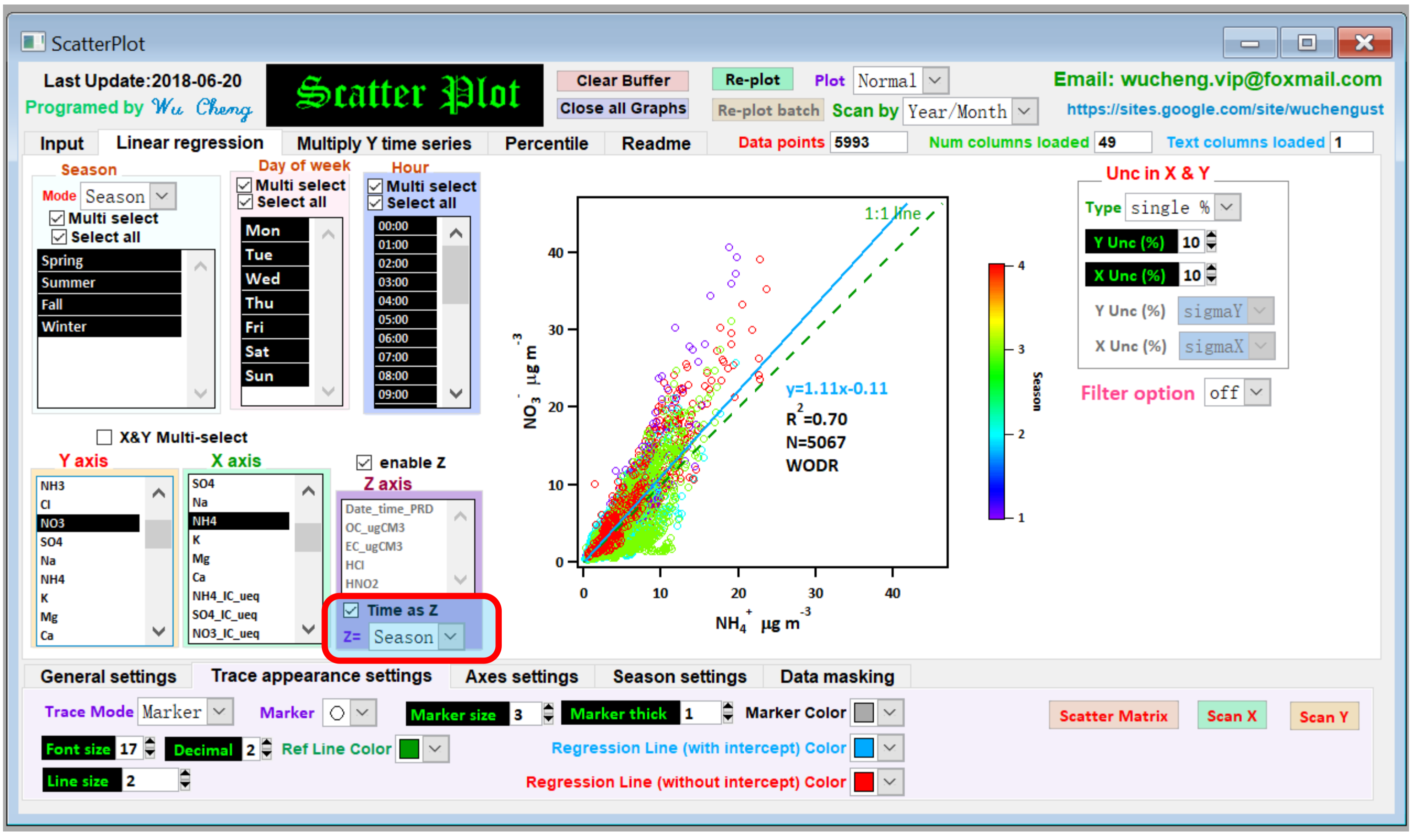


Figure 6.5.1 An example of using season as Z axis color coding.

6.6 Batch plotting- individual plots

When plot option is other than "normal", then batch plotting is activated. Batch plotting can be done on three temporal aspects (Scan by): Year/season/month, day of week, hour, which is corresponding to the three list boxes for data grouping by time.

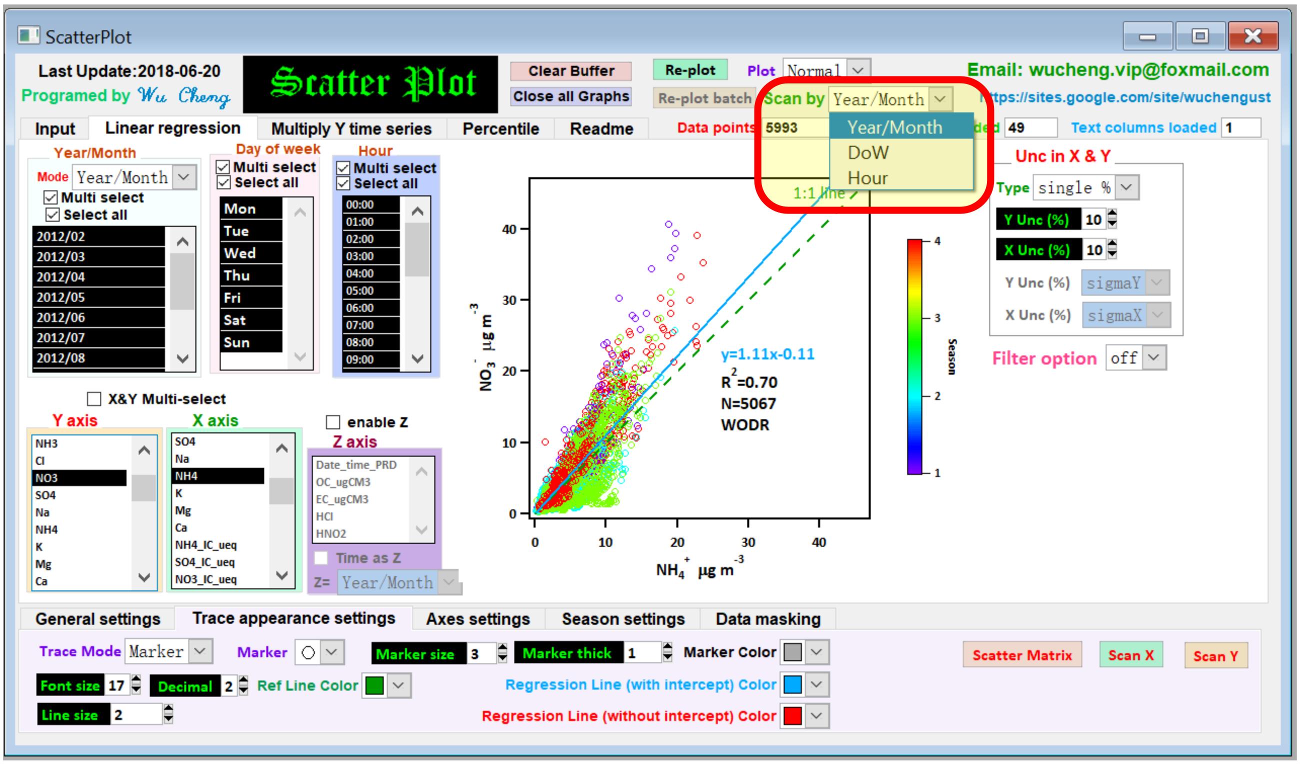


Figure 6.6.1 Settings of batch plotting by time

The fourth way is scan by text markers. When data filter is activated using "Text by list", a 4th option will show up in the "Scan by" pop-up menu.

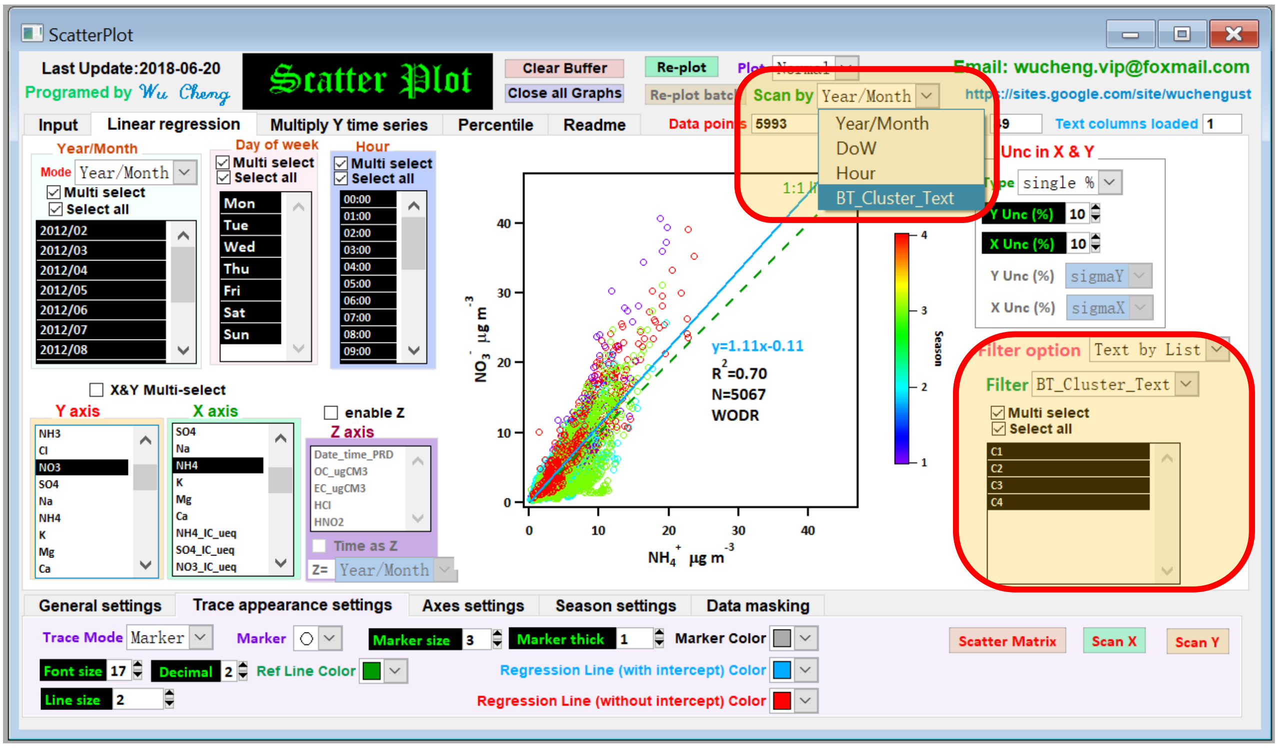


Figure 6.6.2 Batch plotting by text filtering

Following is an example of implementation of batch plotting, scan by year/month (12 months). Besides individual scatter plots, a plot summarizing the variations of slope, intercept and R^2 vs. year/month will also be given. As shown below, nitrate is sensitive to temperature (vaporization), as a result the slope during summer time (Jun-Sep) is much lower than the winter time (Dec-Feb).

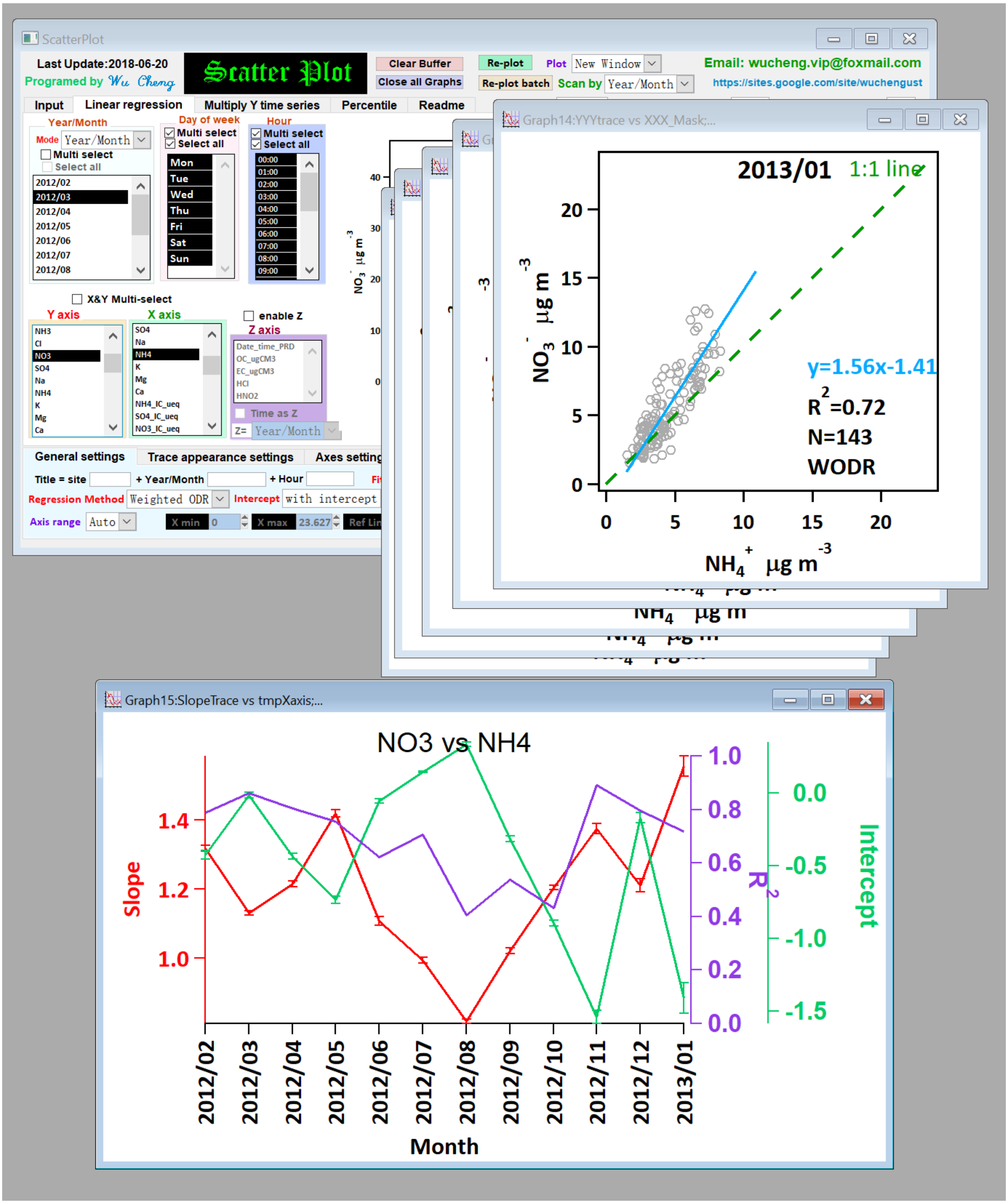
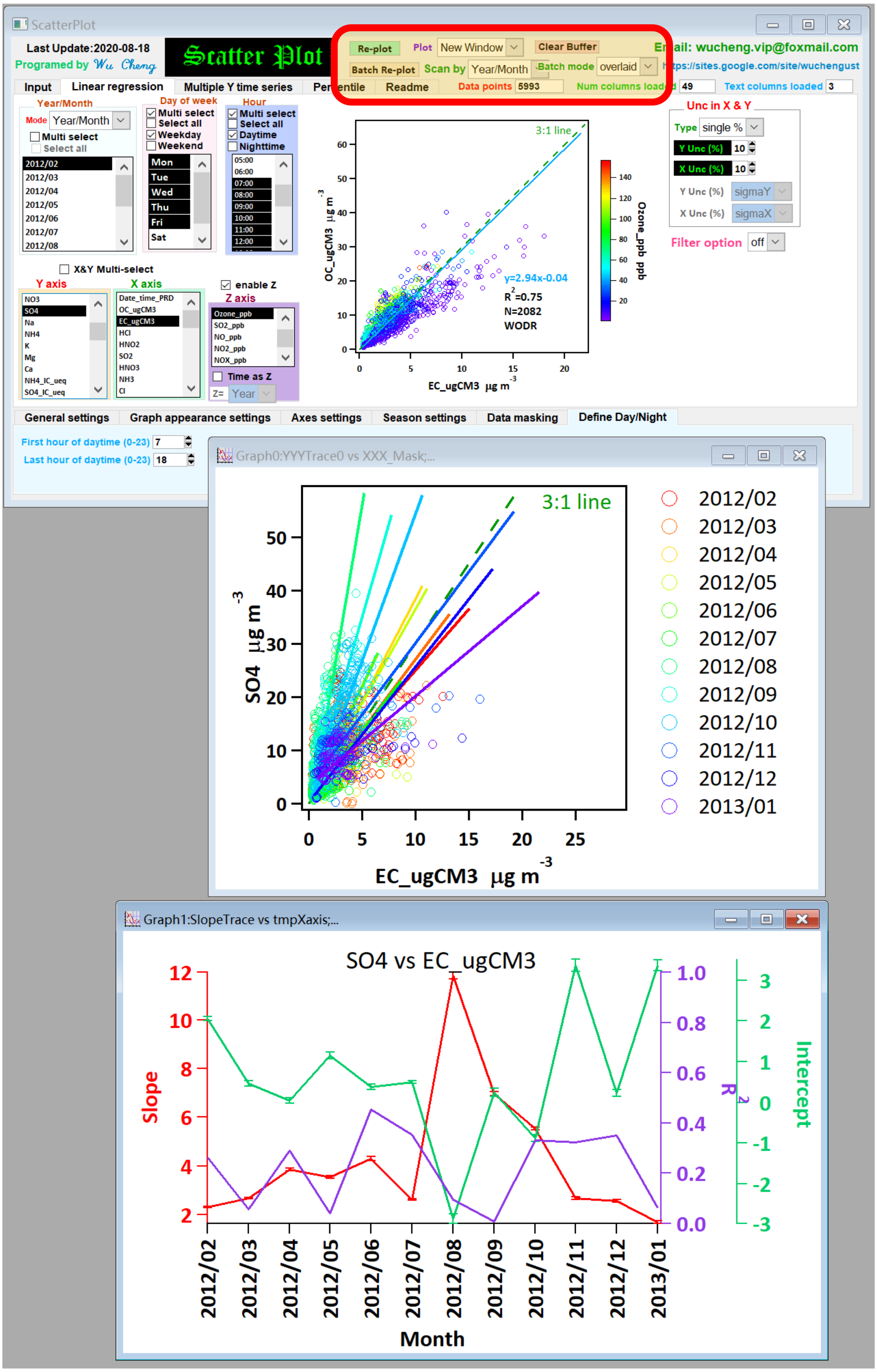


Figure 6.6.3 An example of batch plotting by year/month.

6.7 Batch plotting- overlaid plots

In overlaid mode, all traces are combined in a single plot, as shown below. Each trace will be assigned to a specified color according to the rainbow color table. Batch plotting can be applied on following scales: **Year**; **Year/Month**; **Month**; **Year/Season**; **Season**; **YYYY/MM/DD; Day of week; Hour of day; text data (User defined, Weekday/weekend, Daytime/nighttime).** Regressed slopes, intercepts, R2 and other metrics can be found in the csv file if "Export txt for batch plot" was selected.



7 Tab “Multiple Y time series” Introduction

Multiple Y time series plot is commonly used for presenting temporal variations of various pollutants. As shown below, desired Y can be selected using the “Add” button.

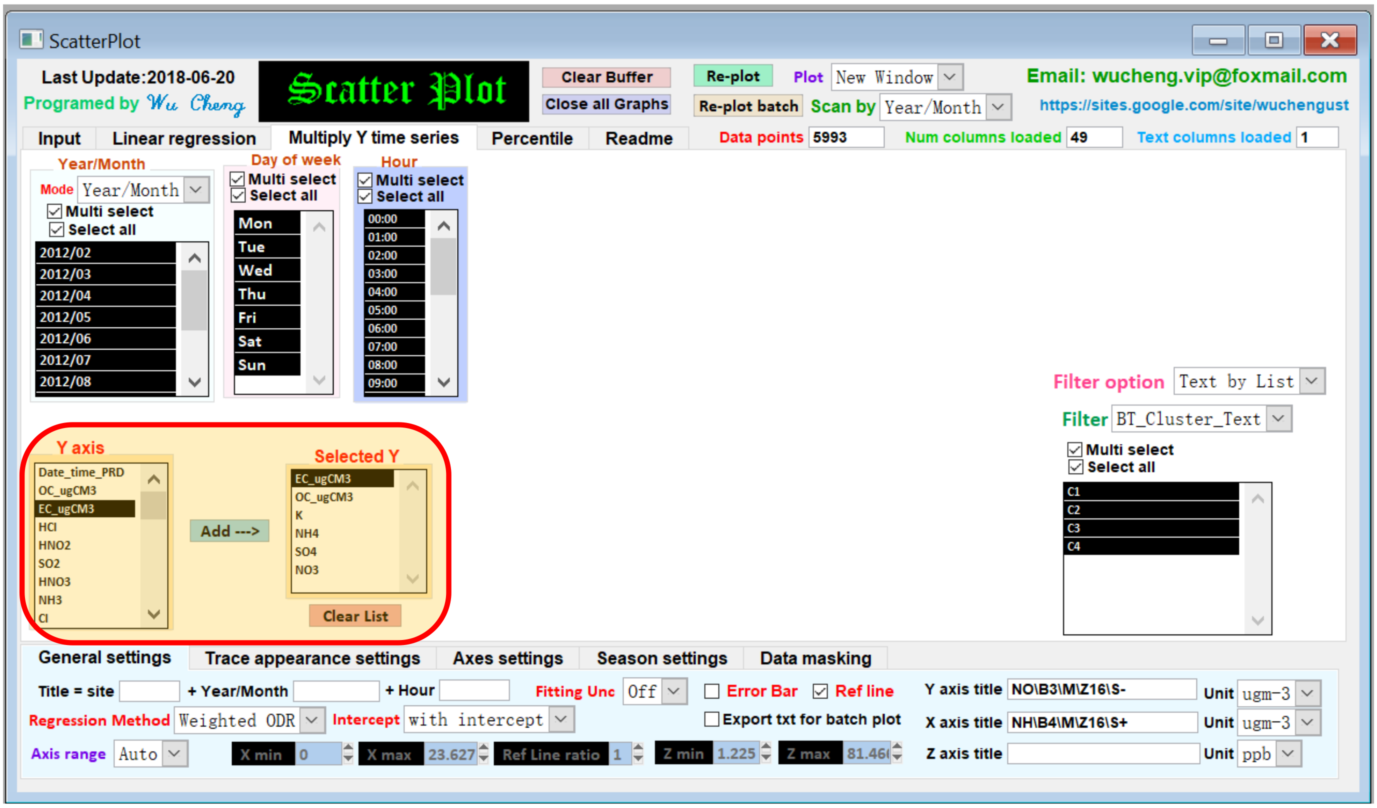


Figure 7.1 Example of selecting desired Y by the “Add” button.

“Plot option” should be set to new. Then click “Re-plot”, the graph will be generated in a new window. User can set the color and line shape, axis title in the new window. The main purpose is to save the time in setting the % portion in Y direction for individual axis.

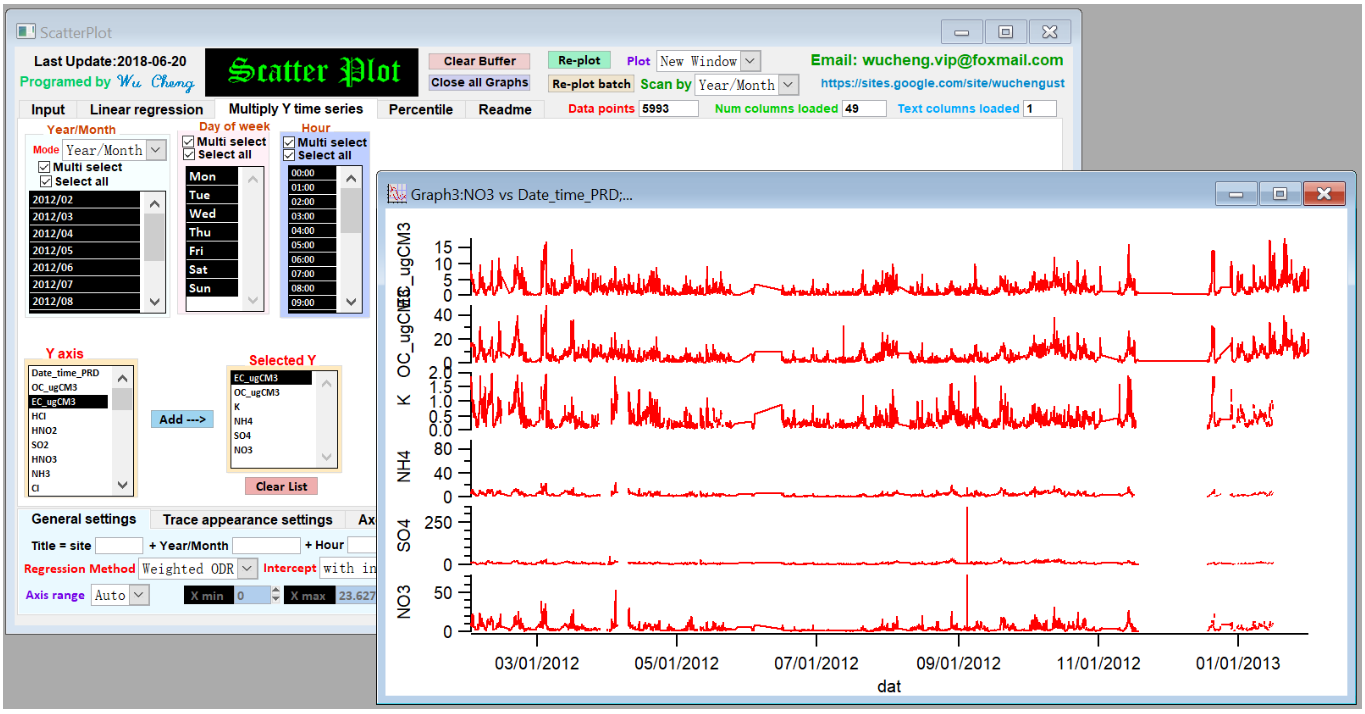


Figure 7.2 Example of multiple Y time series plot in a new window.

8 Tab “Percentile” Introduction

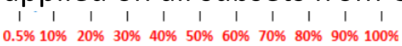
In EC tracer method, a subset with certain OC/EC percentile is often used for regression to determine $(OC/EC)_{pri}$.

Step: interval for stepwise percentile subset regression. For example, 0.005 represents a interval of 0.5%.

Slope Min & Max: Set the slope range (y axis) of the left figure

Intercept Min & Max: Set the intercept range (y axis) of the left figure

Replot: calculate stepwise regression on subsets. For example, with a step of 0.5%, regressions are applied on all subsets from OC/EC 0.5%~100%.



: this slide bar is to select percentile. The selection is only available once “Replot” is applied.

Redraw: create figure in new window (plot option: new window).

Steps to use Tab “Percentile”

1) Set Step, slope and Intercept range
2) Click “Replot”. The figure on the left shows the slops and intercept as a function of
3) Check the results using the slide bar. For example, to show the regression result of 10% subset, drag the slide bar to 10%, the scatter figure on the right will update accordingly. The data points in purple represent selected subset for regression, the data points in grey represent unused data.
4) To export the plots (in new window of export to files), click “Redraw”.

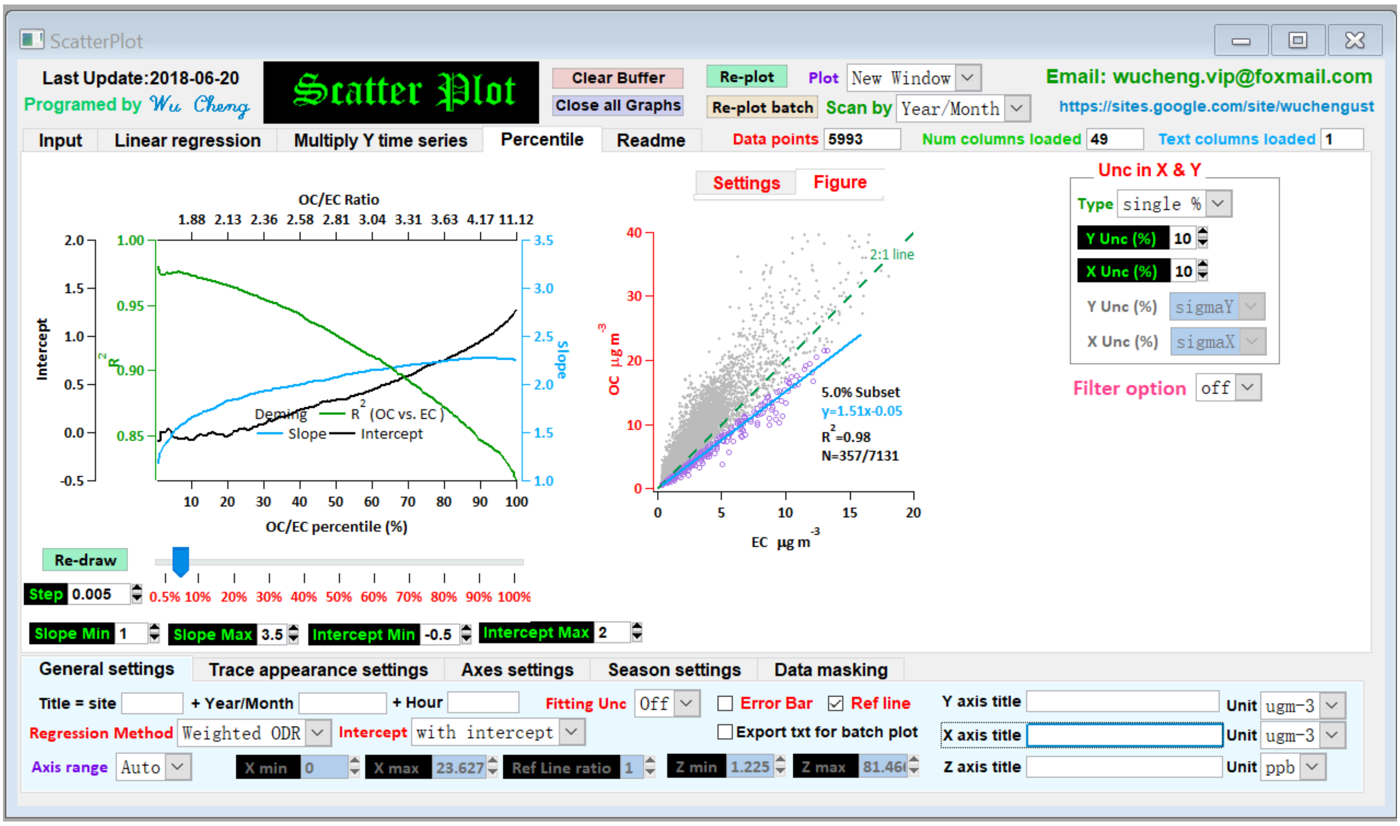


Figure 8.1 Tab “Percentile”.

9 Adoption in research publications:

Sun, J. Y., Wu, C.*, Wu, D.*, Cheng, C., Li, M., Li, L., Deng, T., Yu, J. Z., Li, Y. J., Zhou, Q., Liang, Y., Sun, T., Song, L., Cheng, P., Yang, W., Pei, C., Chen, Y., Cen, Y., Nian, H., and Zhou, Z.*: Amplification of black carbon light absorption induced by atmospheric aging: temporal variation at seasonal and diel scales in urban Guangzhou, ***Atmos. Chem. Phys.***, 20, 2445-2470, doi: https://doi.org/10.5194/acp-20-2445-2020, 2020.

Sun, T., Wu, C.*, Wu, D.*, Liu, B., Sun, J. Y., Mao, X., Yang, H., Deng, T., Song, L., Li, M., Li, Y. J., and Zhou, Z.*: Time-resolved black carbon aerosol vertical distribution measurements using a 356-m meteorological tower in Shenzhen, ***Theor. Appl. Climatol.***, doi: https://doi.org/10.1007/s00704-020-03168-6 , 2020.

Liu, B., Wu, C., Ma, N., Chen, Q., Li, Y., Ye, J., Martin, S. T., and Li, Y. J.*: Vertical profiling of fine particulate matter and black carbon by using unmanned aerial vehicle in Macau, China, ***Sci. Total. Environ.***, 136109, doi: https://doi.org/10.1016/j.scitotenv.2019.136109, 2020.

Wu, C*., Wu, D., and Yu, J. Z*.: Estimation and Uncertainty Analysis of Secondary Organic Carbon Using One-Year of Hourly Organic and Elemental Carbon Data. ***J. Geophys. Res.-Atmos***, 124, 2774-2795 doi:https://doi.org/10.1029/2018JD029290, 2019

Ji, D., Gao, W., Maenhaut, W., He, J., Wang, Z., Li, J., Du, W., Wang, L., Sun, Y., Xin, J., Hu, B., and Wang, Y.: Impact of air pollution control measures and regional transport on carbonaceous aerosols in fine particulate matter in urban Beijing, China: insights gained from long-term measurement, ***Atmos. Chem. Phys.***, 19, 8569-8590, doi: https://doi.org/10.5194/acp-19-8569-2019, 2019.

Liu, B., He, M. M., Wu, C., Li, J., Li, Y., Lau, N. T., Yu, J. Z., Lau, A. K. H., Fung, J. C. H., Hoi, K. I., Mok, K. M., Chan, C. K., and Li, Y. J*.: Potential exposure to fine particulate matter (PM2.5) and black carbon on jogging trails in Macau, ***Atmos. Environ.***, 198, 23-33, doi:https://doi.org/10.1016/j.atmosenv.2018.10.024, 2019.

Wang, N. and Yu, J. Z.: Size distributions of hydrophilic and hydrophobic fractions of water-soluble organic carbon in an urban atmosphere in Hong Kong, ***Atmos. Environ.,*** 166, 110-119, doi: 10.1016/j.atmosenv.2017.07.009, 2017.

Cheng, C., Li, M.*, Chan, C. K., Tong, H., Chen, C., Chen, D., Wu, D., Li, L., Wu, C., Cheng, P., Gao, W., Huang, Z., Li, X., Zhang, Z., Fu, Z., Bi, Y., and Zhou, Z.*: Mixing state of oxalic acid containing particles in the rural area of Pearl River Delta, China: implications for the formation mechanism of oxalic acid, ***Atmos. Chem. Phys.,*** 17, 9519-9533, doi:https://doi.org/10.5194/acp-17-9519-2017, 2017.

Wu, C., Huang, X. H. H., Ng, W. M., Griffith, S. M., and Yu, J. Z.: Inter-comparison of NIOSH and IMPROVE protocols for OC and EC determination: Implications for inter-protocol data conversion, ***Atmos. Meas. Tech***. doi: https://doi.org/10.5194/amt-9-4547-2016, 2016.

Zhou, Y., Huang, X. H. H., Griffith, S. M., Li, M., Li, L., Zhou, Z., Wu, C., Meng, J., Chan, C. K., Louie, P. K. K., and Yu, J. Z.: A field measurement based scaling approach for quantification of major ions, organic carbon, and elemental carbon using a single particle aerosol mass spectrometer, ***Atmos. Environ.***, 143, 300-312, http://dx.doi.org/10.1016/j.atmosenv.2016.08.054 2016.

Adoption in research publications (without acknowledgment):

Qiao, T., Zhao, M., Xiu, G., and Yu, J.: Seasonal variations of water soluble composition (WSOC, Hulis and WSIIs) in PM1 and its implications on haze pollution in urban Shanghai, China, ***Atmos. Environ.***, 123, Part B, 306-314, 2015. http://dx.doi.org/10.1016/j.atmosenv.2015.03.010

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第1901293号

软 件 名 称：Scatter plot海量数据线性回归分析及绘图软件
[简称：Scatter plot]
20170206

著 作 权 人：吴晟

开发完成日期：2017年02月06日

首次发表日期：未发表

权利取得方式：原始取得

权 利 范 围：全部权利

登 记 号：2017SR316009

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权
登记专用章

No. 01760115

2017年06月27日