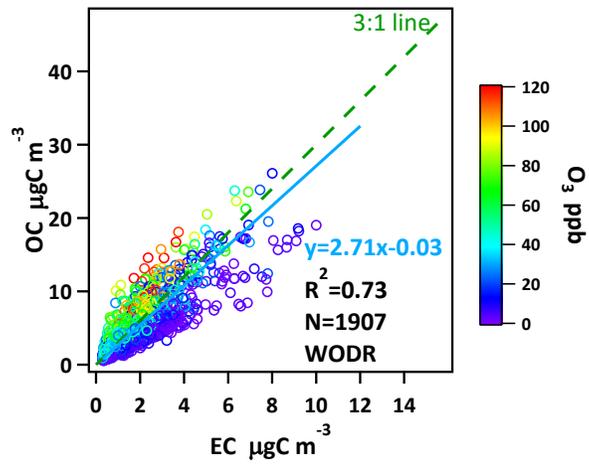


Scatter Plot



Scatter Plot使用说明书 Manual for Scatter Plot

吴晟

wucheng.vip@foxmail.com

2020-08-18

前言

Scatter Plot是一个可以最大限度地提高大气科学中数据可视化效率的便捷Igor工具包。虽然有许多现有的通用数据可视化软件获得了广泛应用，但不能满足许多大气科学特定的研究目的，这促使了Scatter Plot 程序的开发。本程序包括WODR, Deming和York算法进行线性回归，这三种算法考虑了X和Y都包含不确定性（观测误差），对大气的应用而言更加客观地反映真实情况。本工具包提供大量用于数据分析和图形绘图的有用功能，包括批量绘图，通过图形界面实现数据掩蔽，Z轴的颜色编码，根据数据或字符串进行过滤和分组。

有关Scatter Plot的评估和应用的更多细节，请参阅（**如果你在发表的文章中用到了本软件，请引用以下文章**）

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.

关于程序的最新信息可以在我的网站上找到：

<https://www.x-mol.com/groups/wucheng>

<https://wucheng.weebly.com/>

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

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

也可以在QQ群(258975942)的群共享文件中下载（**1号QQ群已满，建议加2号群**）：



群名称:Igor大气数据分析 1号群
群 号:258975942



群名称:Igor大气数据分析 2号群
群 号:780325682

吴晟

2020-08-18

目录

0 Igor Pro 运行环境的安装	1
1 关于数据结构的建议.....	3
2 跟其他程序的总体比较	5
3 导入数据.....	6
3.1 在 MS excel 中的时间线示例	6
3.2 从Excel复制	8
3.3 将数据粘贴到Igor中.....	9
3.4 更新列表.....	10
3.5 指定时间轴	11
4 设置介绍.....	12
4.1 General settings.....	12
4.2 Trace appearance settings.....	14
4.3 Axes settings.....	15
5 分页 “Input” 简介.....	16
6 分页 “Linear regression ” 简介:.....	18
6.1 数据按时间筛选	18
6.2 用数据进行筛选.....	19
6.3 使用图形界面进行数据遮掩(数据剔除).....	22
6.4 选择多个变量用作X&Y	27
6.5 时间变量作为Z轴	28
6.6 批量绘图-每张单独输出	29
6.7 批量绘图-合并成1张图输出.....	31
7 分页 “Multiple Y time series” 简介.....	32
8 分页 “Percentile” 简介	33
9 使用了该工具包的已发表文章:	34

0 Igor Pro 运行环境的安装

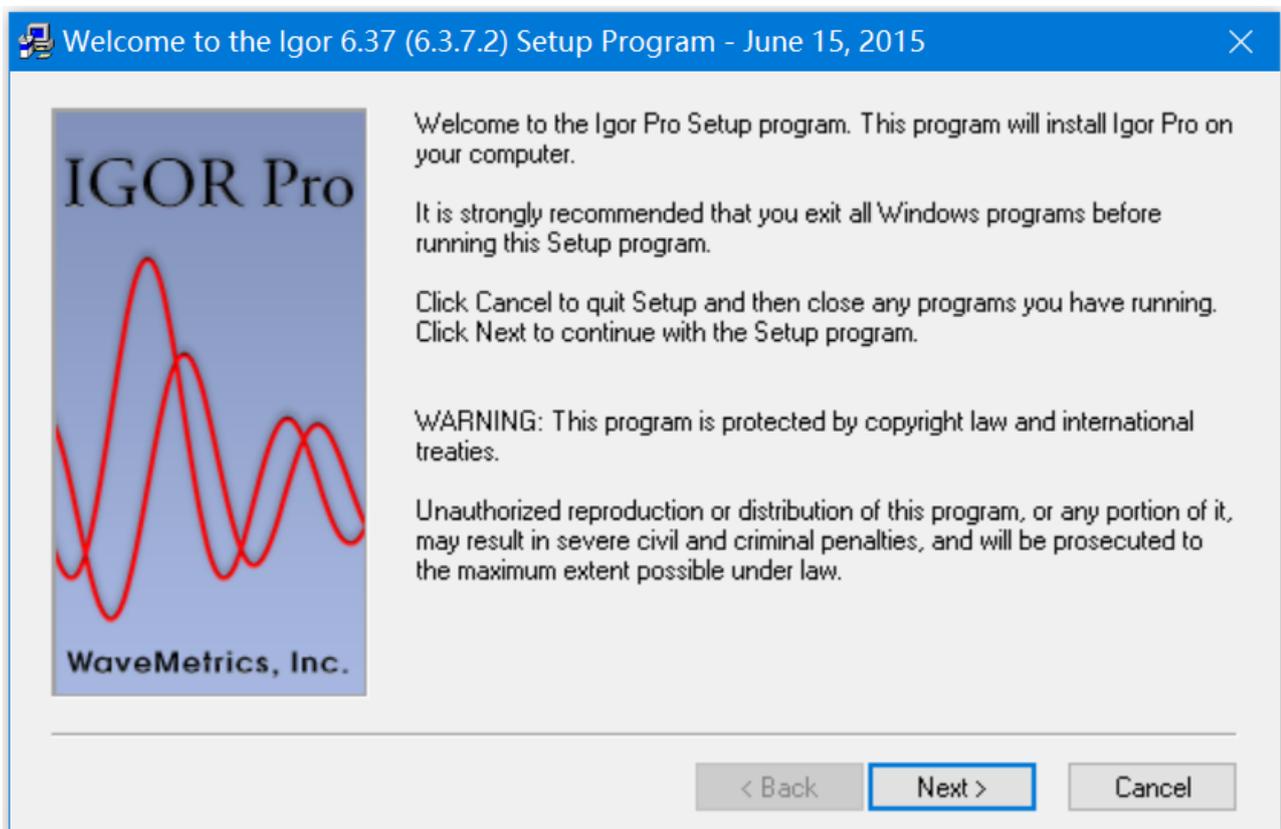
用户需要Igor Pro平台来运行Scatter Plot程序（pxp文件）。这类似于你需要安装微软 word来打开docx文件的场景。因此，您需要在计算机上安装Igor Pro。Igor Pro拥有Windows PC和Mac版本。建议使用Igor Pro 6.x版。Scatter Plot也可以在Igor Pro 7.x上运行，但在用户界面显示中存在一些问题，某些元素（按钮，下拉菜单）的比例失调，尤其是在计算机使用屏幕高DPI设置时。

- 1) 从 <https://www.wavemetrics.com/support/versions.htm> 下载Igor Pro
- 2) 双击Igor Pro安装文件



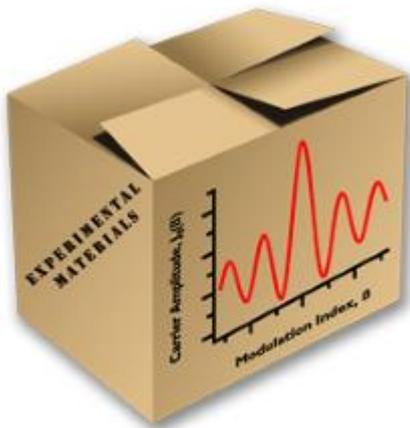
setupIgor6.37.exe

- 3) 按照步骤安装Igor Pro



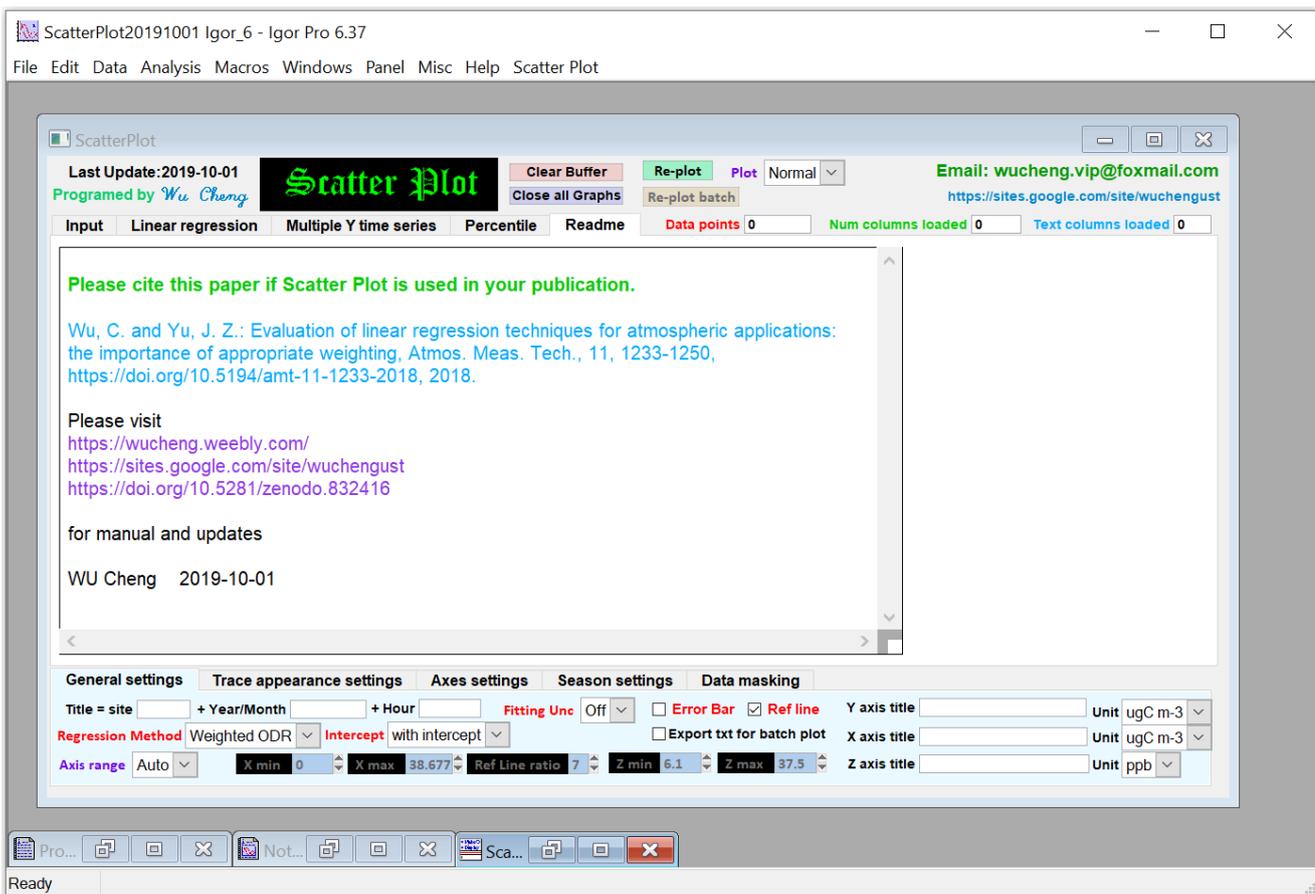
- 4) 成功安装后，系统可以识别pxp文件(如下图图标那样)。如果您没有Igor Pro的激活码，演示版本

Igor Pro会持续30天。 30天后，用户不能使用以下功能：a) 导出图和文件; b) 保存文件。



ScatterPlot20200818.pxp

5) 要运行Scatter Plot, 只需双击pxp文件即可。如果要同时打开多个pxp文件, 请使用光标选择pxp文件, 然后按“Ctrl + Enter”。



1 关于数据结构的建议

如果数据的大小小于 100 万行, Excel 被建议用于存储数据。否则, 建议使用.csv 文件。如果可能, 将所有数据与同一时间线都放在一张表格上以实现最大化的效率, 避免把它们都放在分离的表格, 因为子集可以通过筛选取。建议的数据的结构如下图 1.1 所示。第一行是表头 (文本格式)。导入 Igor 后表头将成为 wave (Igor 中关于数列的概念, 等同于 Excel 中的列) 的名称。表头中空格和其他非法字符 Igor 作为 wave 名称是不允许, 将由"_"替换。**此外, wave name(即数据的表头)必须是字母开头(A~Z), 不允许使用数字开头作为 wave name。**数据分为三类:

- 1) 时间戳 (时间轴)
- 2) 数值数据 (如空气污染物的浓度)
- 3) 文本数据 (例如标签、 站点名称, 后向轨迹聚类)

Recommended data structure in a sheet

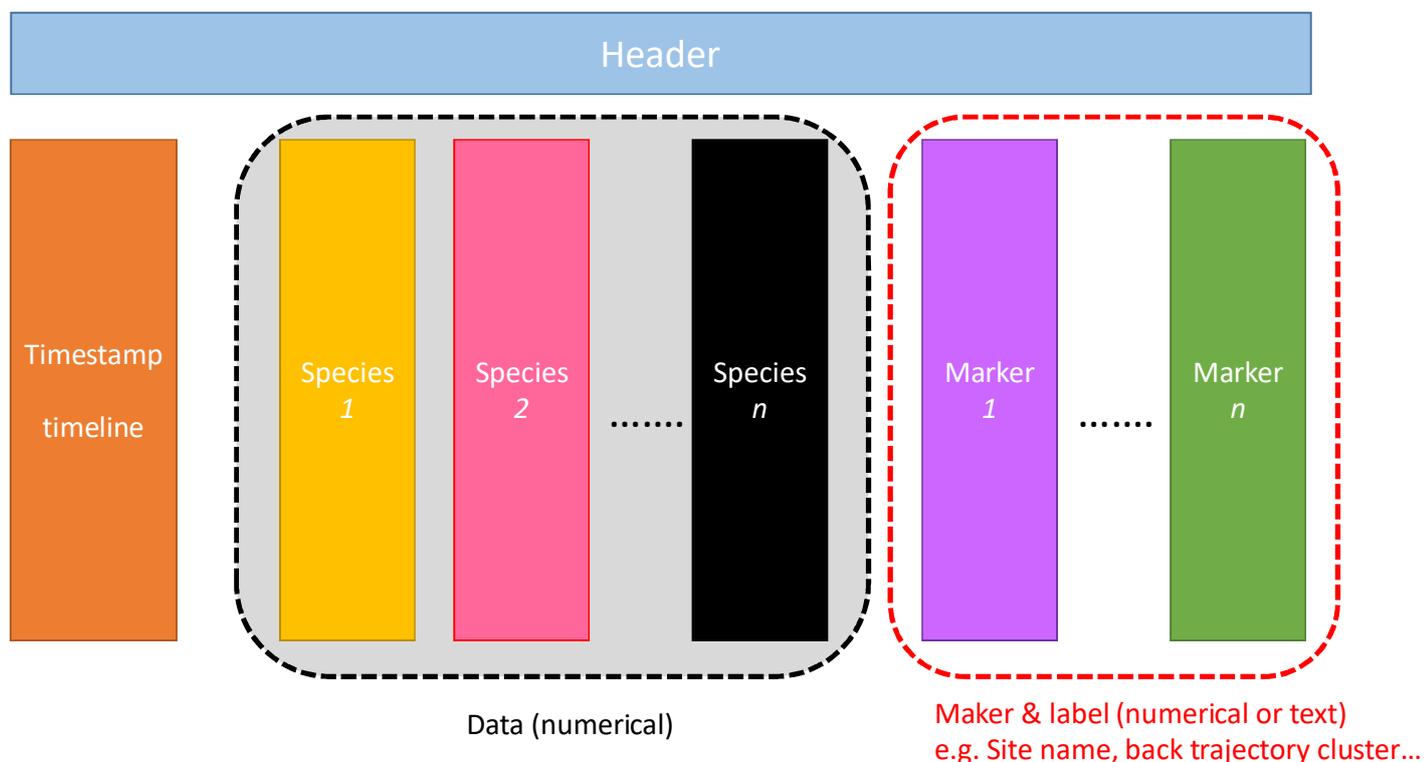


图 1.1 推荐在工作表中的数据结构

Excel 数据 (或.csv 文件) 的一个实际例子如图 1.2 所示。应该指出的是, 不同的数据列 (wave) 的顺序不一定是图 1.1 相同, 可以混合使用三个类别, 数据列的顺序没有限制。下例中, DateIndex 属于时间轴, SampleID和Site属于Marker (文本数据), 其余列属于数值列 (污染物浓度值)

	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

图 1.2 Excel数据的一个实际例子 (或.csv 文件)。

2 跟其他程序的总体比较

下表将本程序(Scatter Plot)与其他程序进行比较

软件	优势	不足之处
Excel	数据筛选	只能OLS线性回归, 不支持Deming 回归 行数有限制(一百万行数据) 不能做数据遮掩 不支持Z轴颜色
SPSS	数据筛选	只能OLS线性回归, 不支持Deming 回归 数据遮掩不能通过图形界面实现
Sigma Plot	支持Deming 回归	不支持数据筛选和数据遮掩
Origin	支持York回归 数据遮掩可以通过图形界面实现 支持Z轴颜色	不支持数据筛选
Scatter Plot Igor program	支持 OLS, Deming, Weighted orthogonal distance and York 回归。 支持数据筛选 数据遮掩可以通过图形界面实现 支持Z轴颜色 批量绘图	Igor普及率不及前四者

3 导入数据

3.1 在 MS excel 中的时间线示例

导入之前，数据在 excel 中，限制可以保存时间轴的格式如下所述。数据列中的时间轴必须遵循此格式"**MM/DD/YY hh: mm**"（在Igor7及以上版本，建议设置为"**MM/DD/YYYY hh: mm**"），位置必须是"英语（美国）"，如图 3.1 所示。

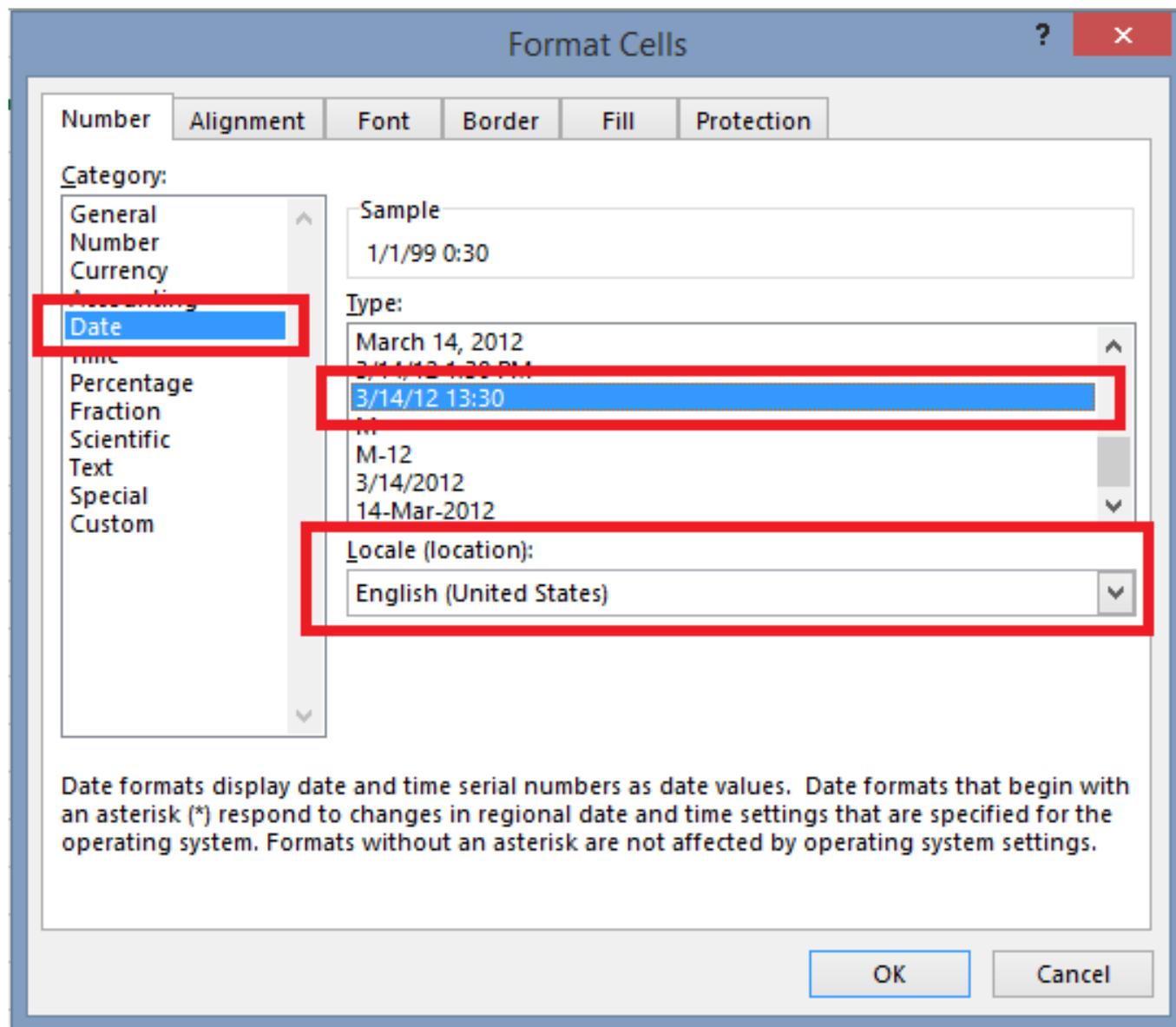


图3.1 时间轴列的MS Excel中的单元格格式配置

确保时间轴列的单元格格式与图2.1所示的完全相同，否则Igor无法识别它

为确保兼容性，应检查操作系统的日期时间格式。推荐美国格式（如下图所示，在win10系统中，转到控制面板设置日期时间格式）。

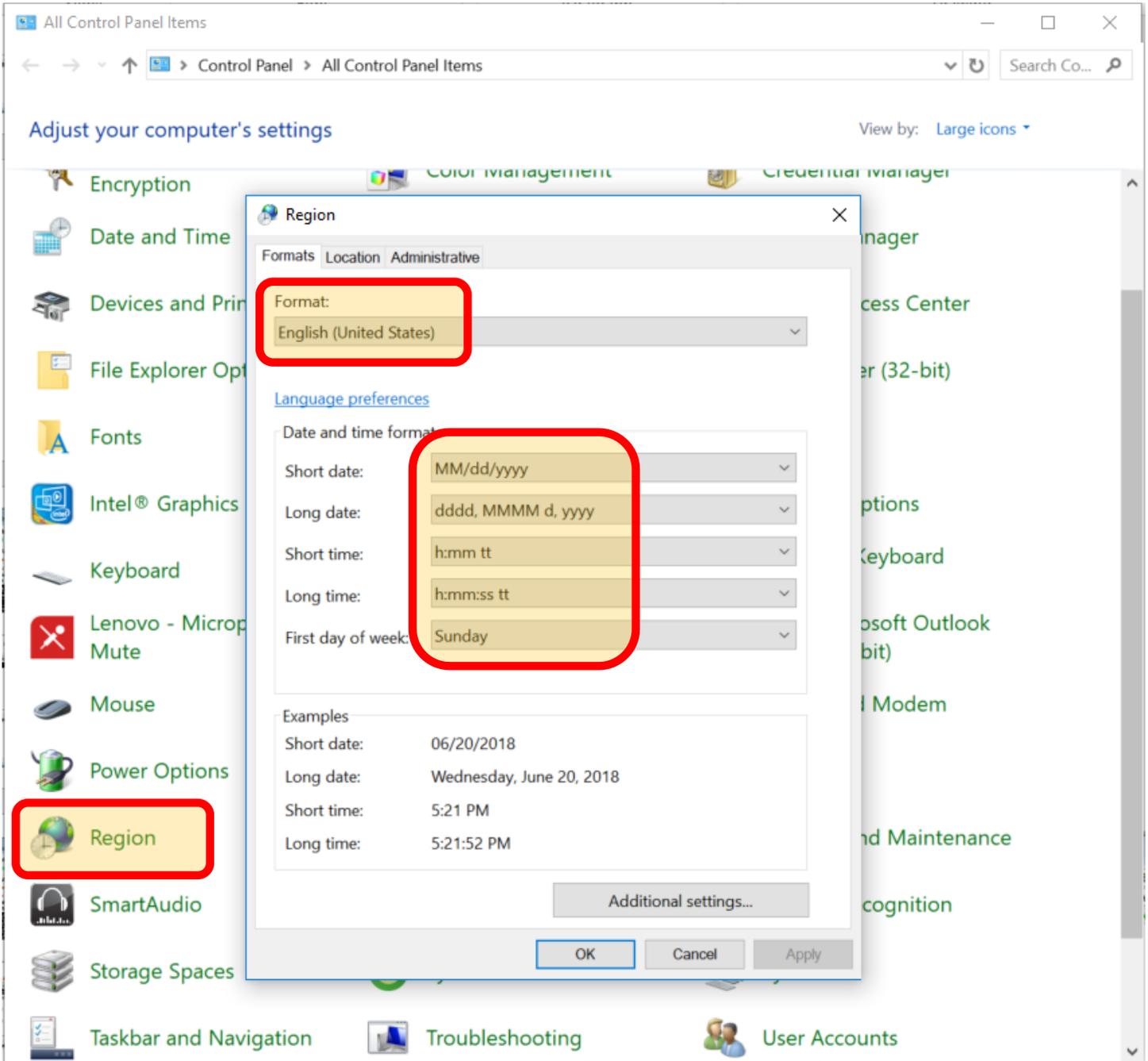
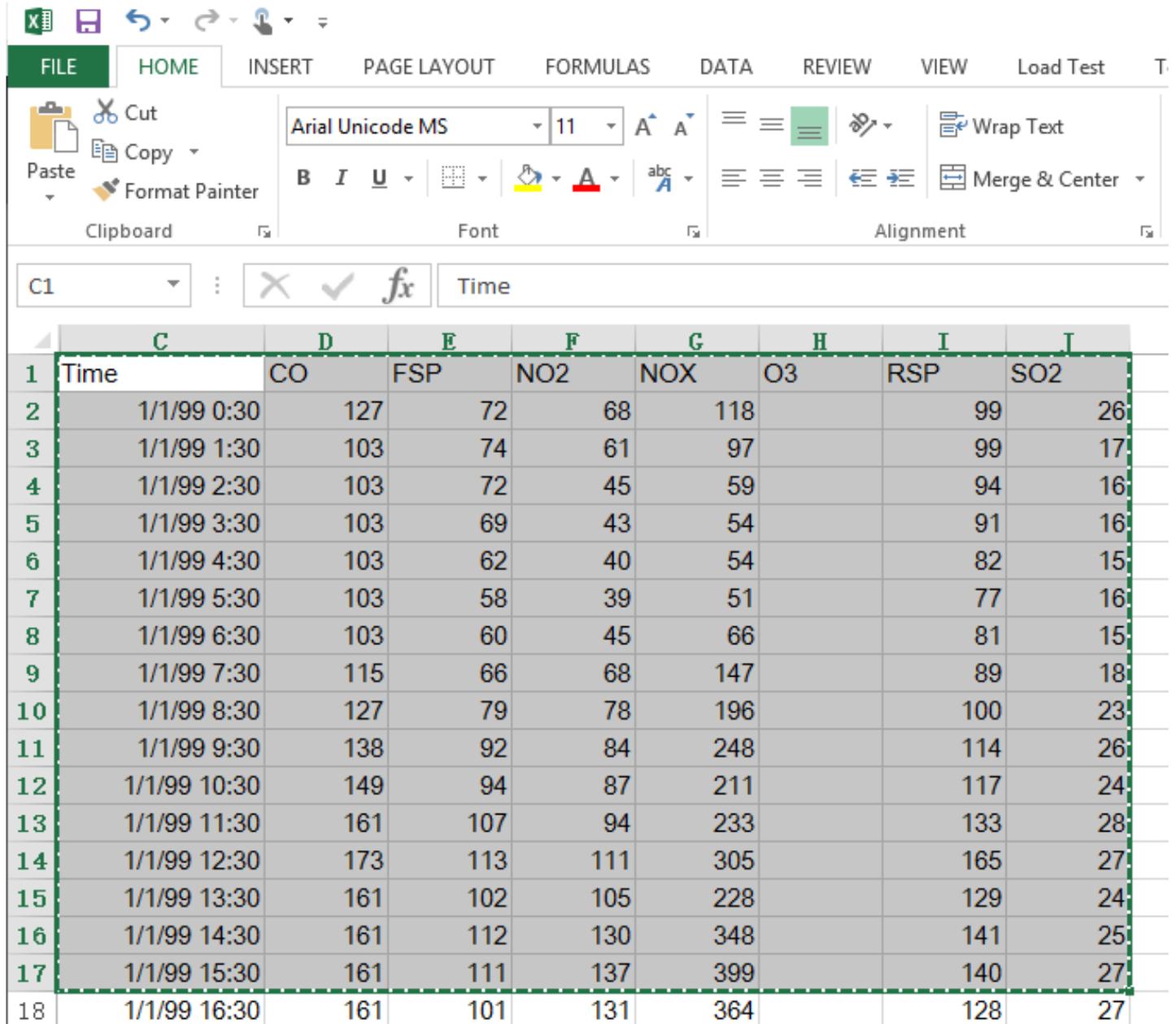


图 3.1.2 win10系统中，在控制面板设置日期时间格式。

3.2 从Excel复制

数据可以通过从Excel复制和粘贴导入，如图3.2所示。建议将时间轴放在第一列。



	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

图3.2 MS Excel中数据选择和复制 (Ctrl + C) 的示例。每列的表头将用作Igor中的wave名称。

3.3 将数据粘贴到Igor中

将光标选中表格左上角（第一行第一列）的单元格，将数据粘贴到Igor程序界面中的表格（高亮橙色区域），如图3.3.1所示

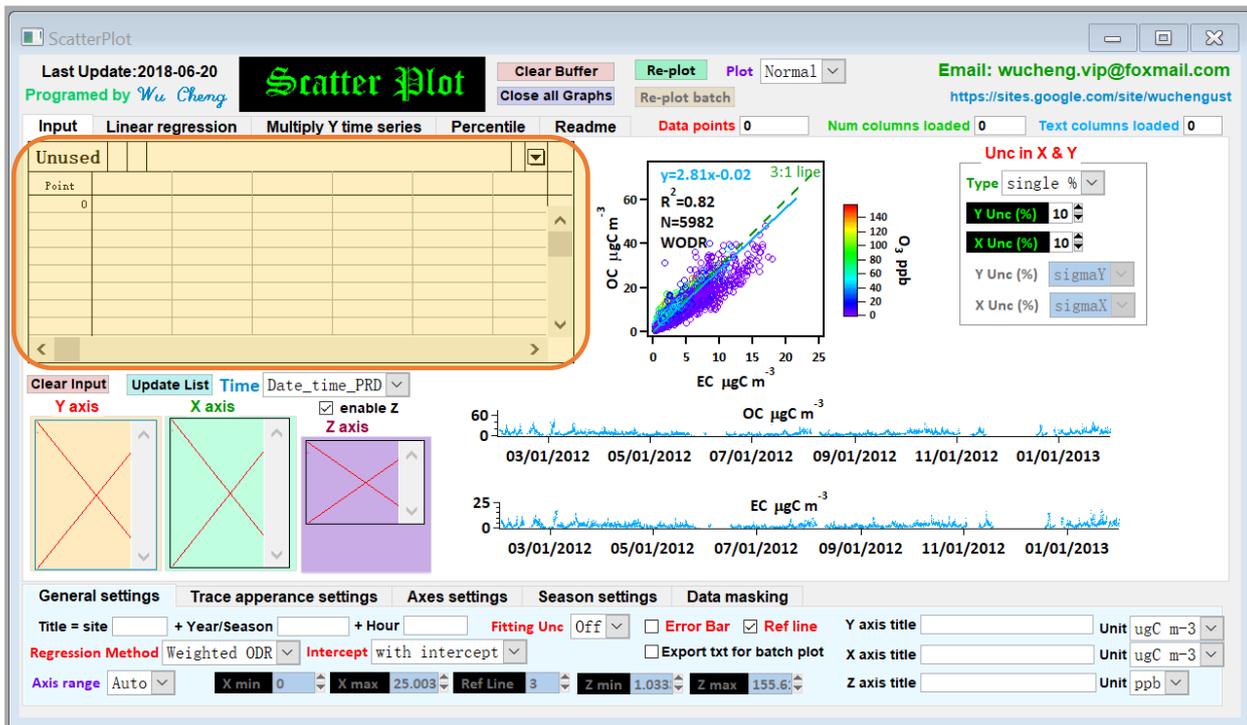


图3.3.1 粘贴数据之前Igor Pro中用户界面的示例。

应用粘贴 (ctrl + V) 后，数据将显示在表格区域中，确保时间线被Igor Pro正确识别。应当注意，数据点的索引从Igor中的0开始。

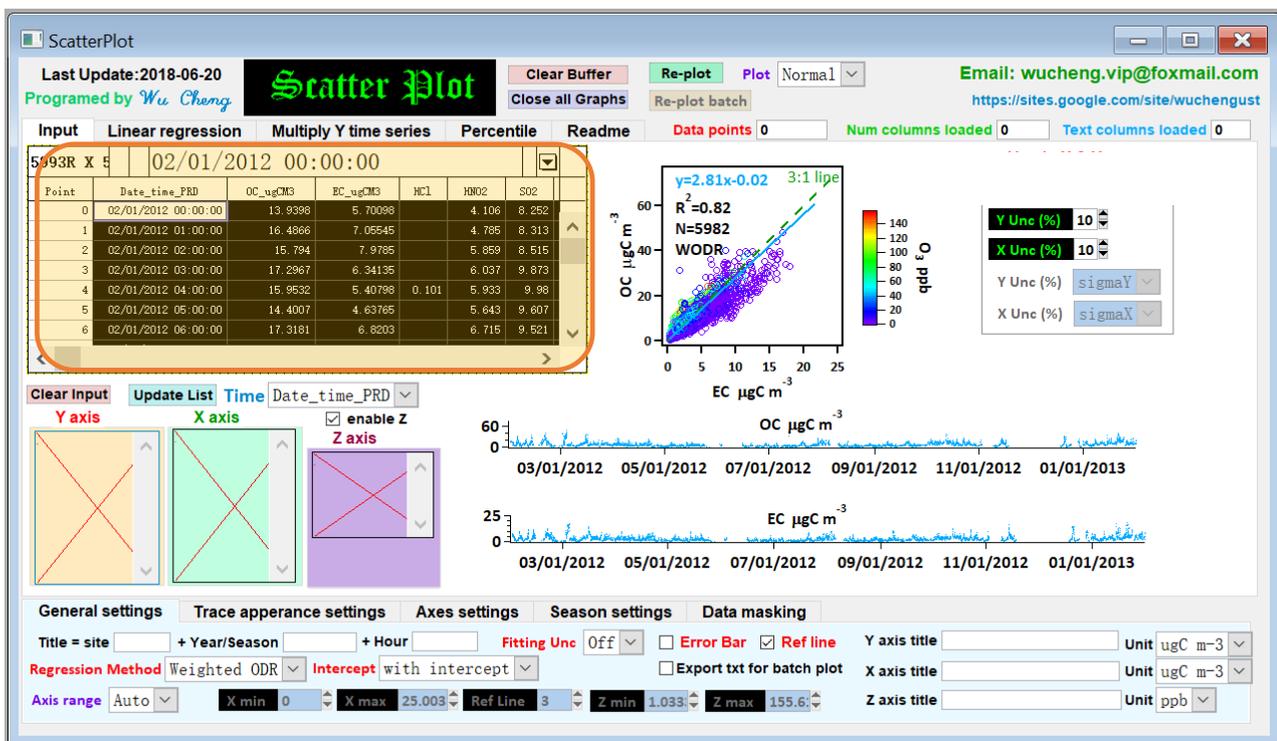


图3.3.2 粘贴数据后Igor Pro中的用户界面示例

3.4 更新列表

- (a) 点击“Update List” 按钮(图3.4, 高亮显示的区域a)
- (b) 然后列表数值数据系列 (在Excel中称为列和Igor Pro中的波) 将被更新 (图2.4.1, 高亮显示区域b)。
- (c) 加载数据的统计信息如图2.4.1高亮显示的区域c所示, 包括数字列, 文本列和数据点 (行) 的数量。

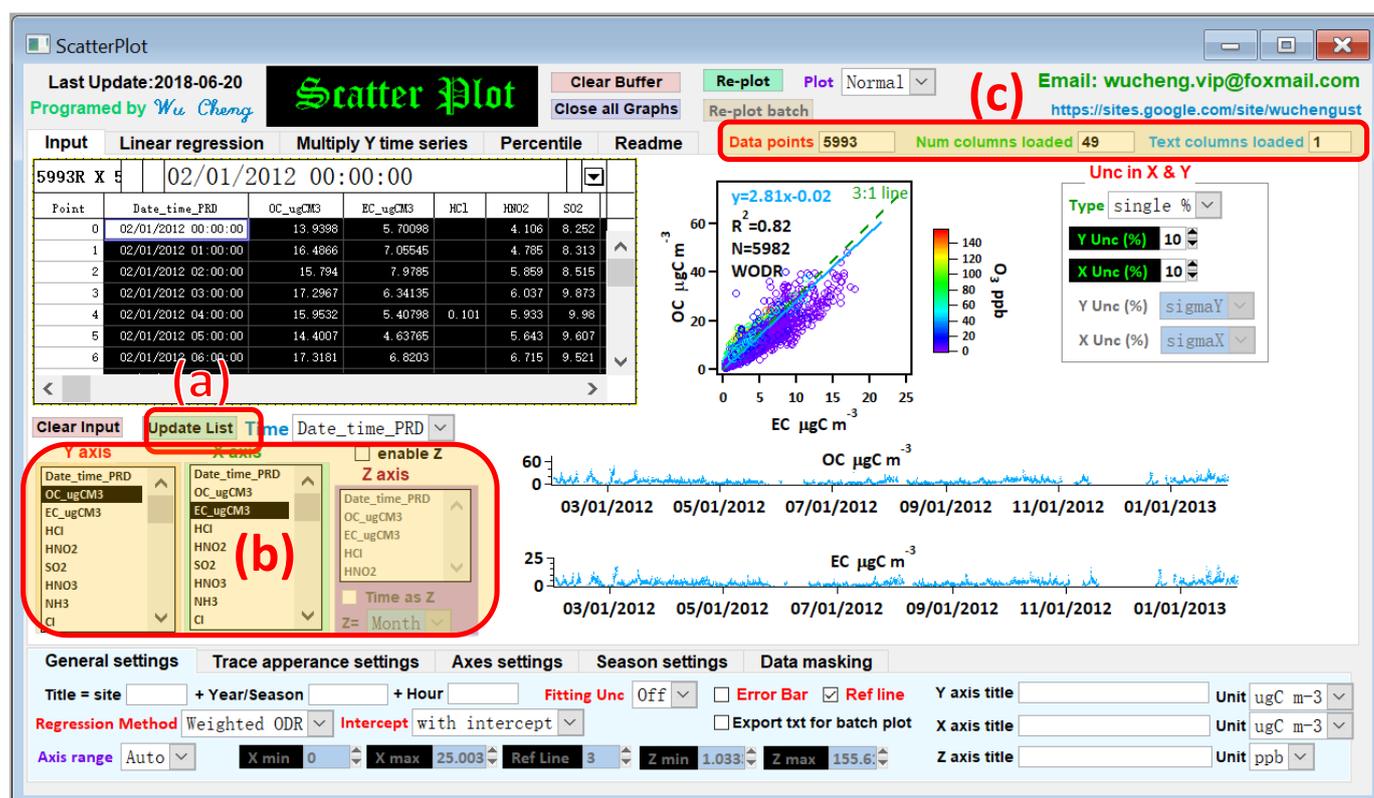


图3.4 Igor中的更新列表示例。

3.5 指定时间轴

下一步是告诉程序哪个列是时间戳。它可以通过使用弹出菜单来完成，如图3.5所示。

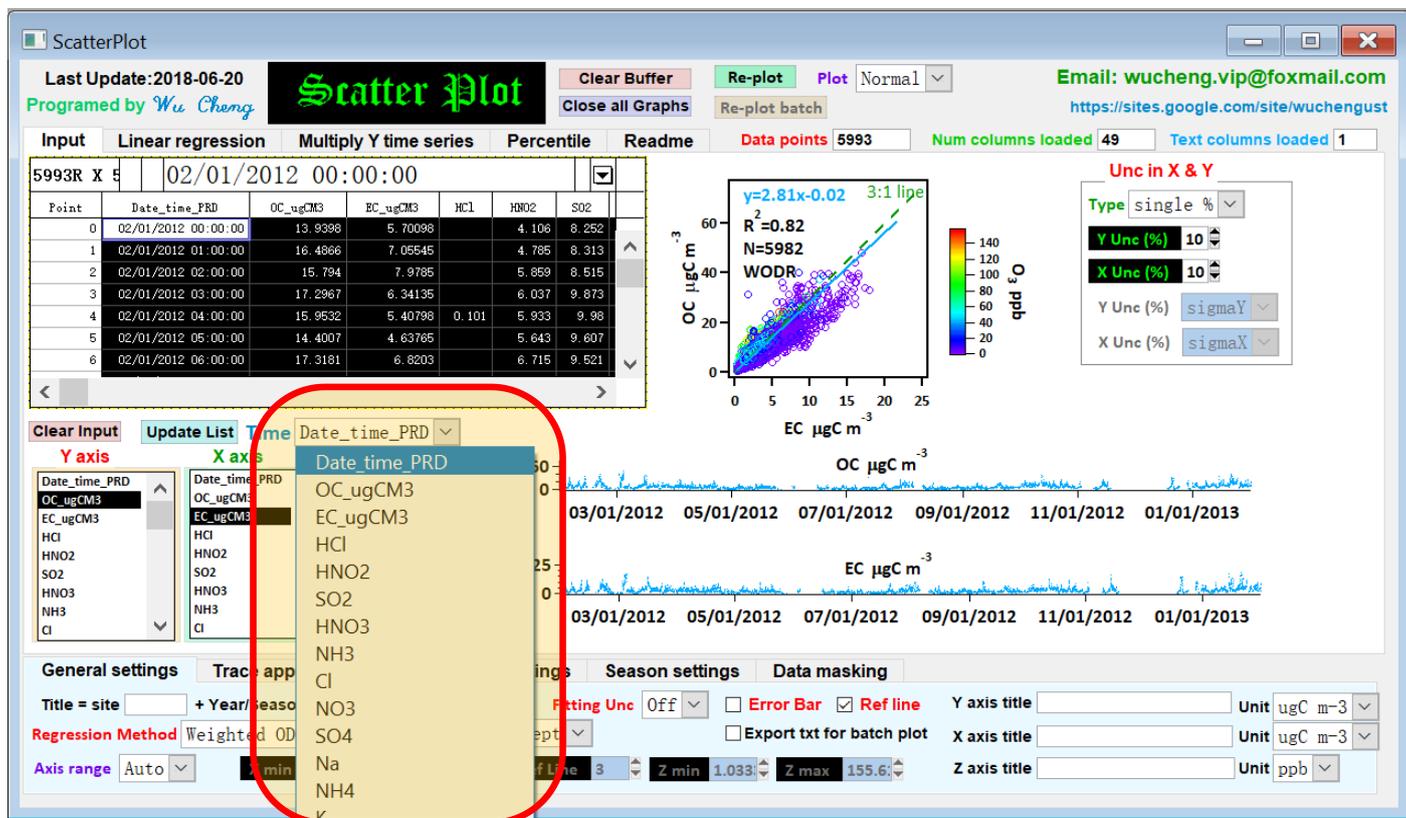
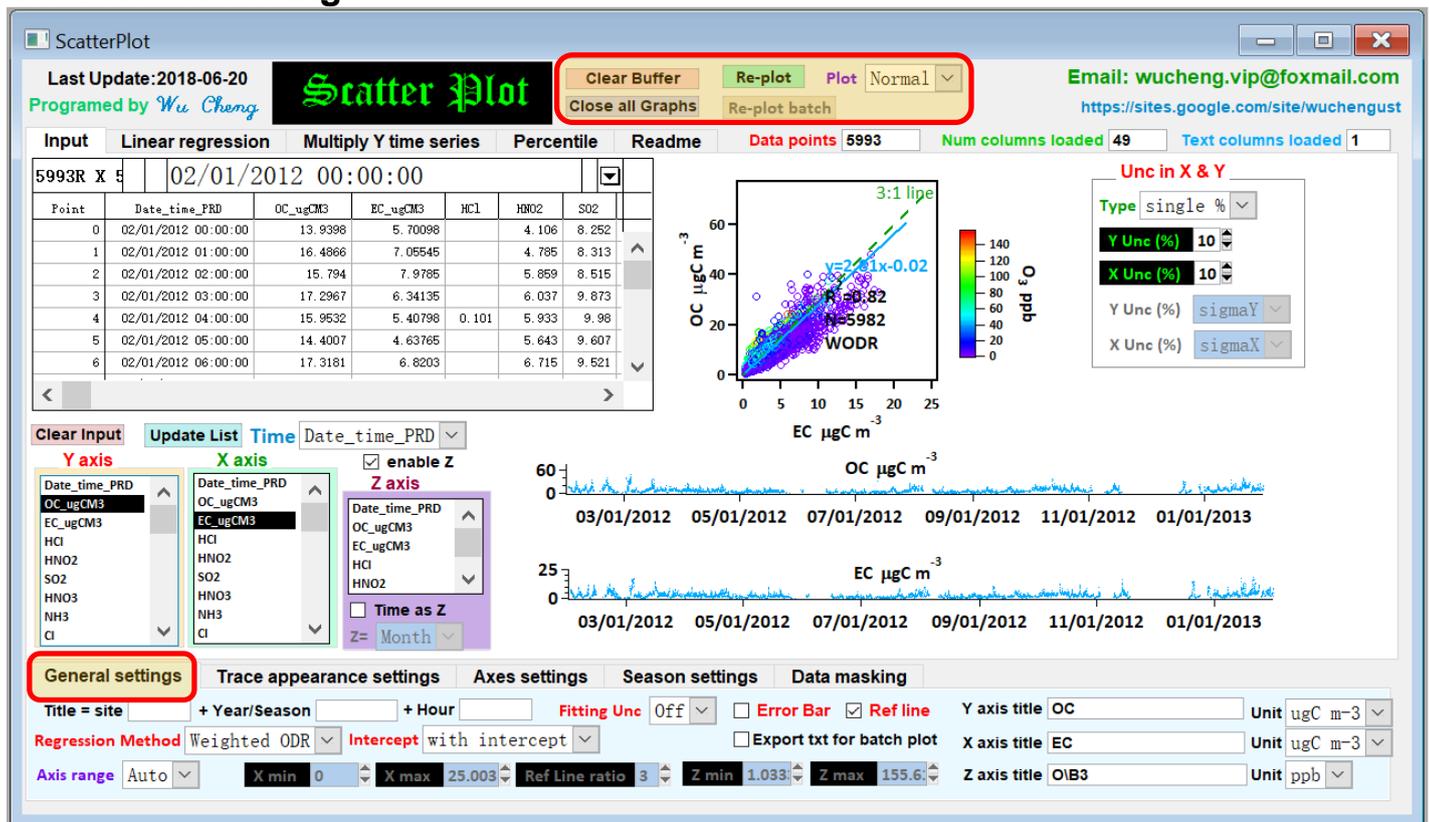


图3.5 在Scatter plot Igor程序中指定时间轴的示例

4 设置介绍

4.1 General settings



Close all Graphs: 关闭所有新窗口中的图形。

Clear Buffer: 删除批处理图的缓存数据，避免程序文件过大。

Replot: 在当前选项卡中重新进行计算并重绘图。注意：非批量模式画图下，每次重新绘图的时候，上次绘图的数据都会被覆盖。即使之前画的图已经生成在新窗口（比如说已经画了图在Graph1窗口），其对应数据也会在重新绘图的时候被最新的数据覆盖（生成了Graph2窗口后，Graph1的数据已经被覆盖，其数据点跟Graph2是一样的，只有标注的文字还是原来的）。因此如果想手动绘制多张图，每画完一张图都要及时保存（粘贴到ppt中，或者存为图片文件），然后再画下一张。否则画了多张图，在Igor中其实都只是最后一张图的数据。

Plot option: Normal, 只重绘当前选项卡; New window, 也可以在一个新窗口中生成绘图，可以复制并粘贴到MS办公室; Export PNG, 不仅在新窗口中生成绘图，还生成PNG文件; Export EMF, 而不是PNG文件, EMF是一个矢量文件，可以无限放大。生成的PNG和EMF文件跟本Igor程序(.pxp文件)存放在同一个文件夹。

Title: 绘图的标题，包含三个字段（采样站，年月，小时和物种），如果它们留空,这些字段将在扫描时使用自动命名，否则会使用用户输入的字符。

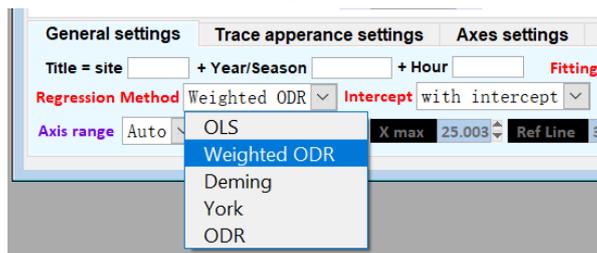
X/Y/Z title: X/Y/Z轴标题，如果留空，会默认使用对应wave的wave name作为标题。

Fitting Unc: 选择是否显示拟合的斜率和截距的不确定性(正负偏差的形式，仅限于WODR)。

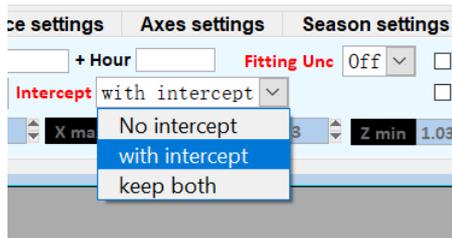
Error Bar: 是否显示误差棒。

Ref line: 是否显示参考线。

Regression method: 选择拟合算法, 可选项有Ordinary Least Squared (OLS), Weighted orthogonal distance regression (WODR), Deming regression, York regression. 其中OLS只考虑Y存在误差, 其他算法考虑XY都存在误差.



Intercept: 如选择 "No intercept", 拟合线会强制通过原点 (不适用于 Deming 和 York Regression). 如选择 "with intercept", 则结果会报道截距. 如选择"keep both" 两种结果都会同时给出 (分别用红蓝色表示. 不适用于 Deming 和 York Regression).



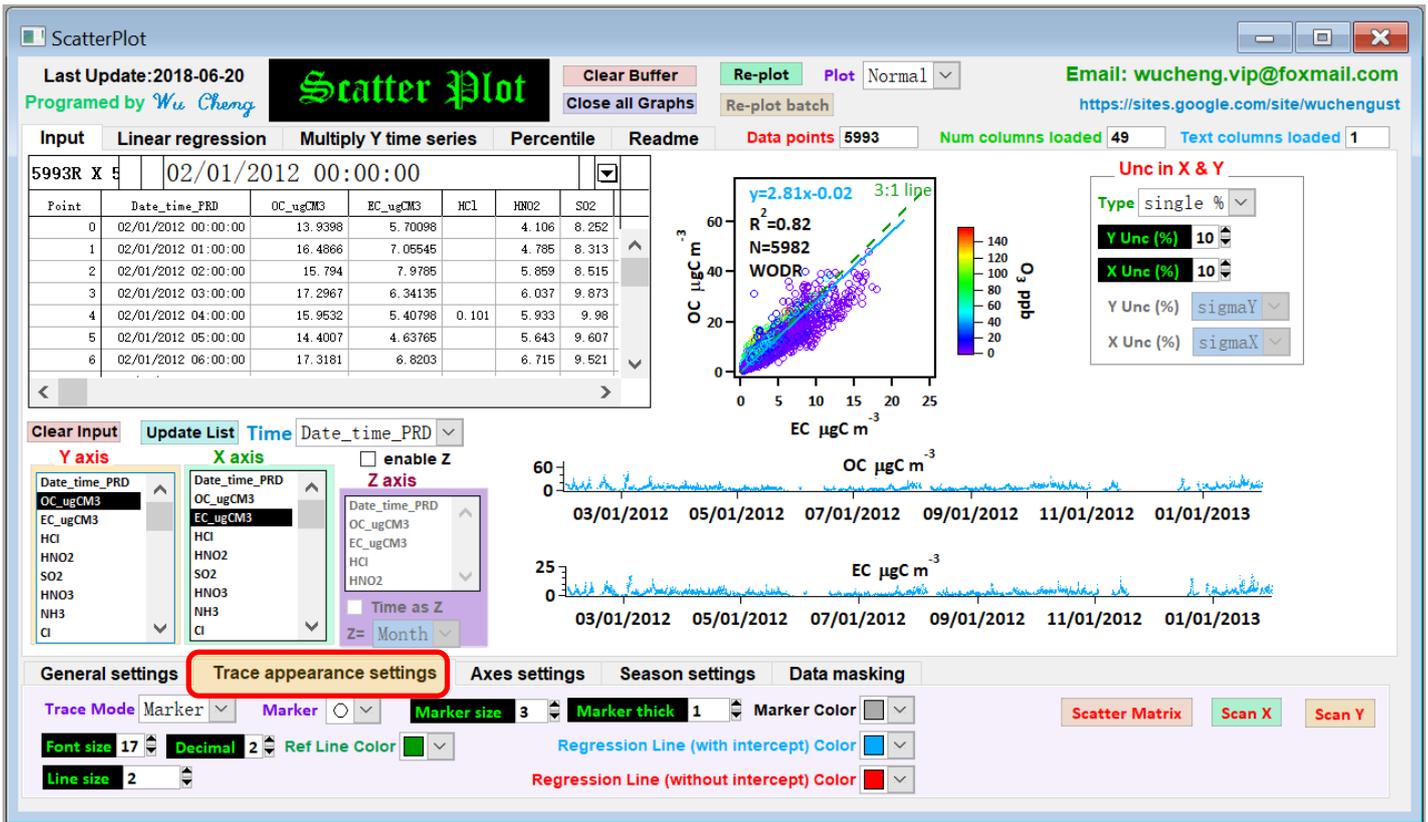
Axis range: 坐标轴模式开关. 在manual mode下, x, z 以及 ref line ratio 可以由用户指定. 自动模式下会根据数据范围自行取值.

Ref Line ratio: 设置Y:X 参考线的比值 (仅限Axis range 在manual mode下).

Unit: 用户可以选择预设的一系列大气科学常用单位, 节省输入所需的时间.

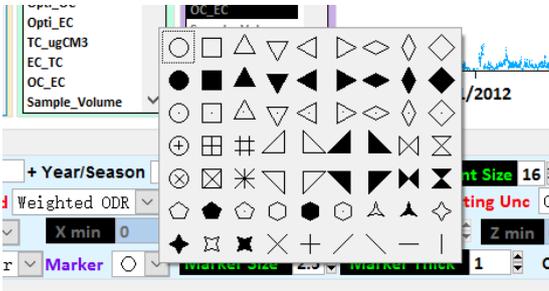
Export txt for batch plot: 在批量画图模式下, 是否输出拟合结果到txt文件, 如果输出, 会保存在pxp 同一个目录下, 其内容包含, 斜率, 截距, 不确定性, 数据点个数, R^2 等等.

4.2 Trace appearance settings



Trace Mode: 数据点使用符号(marker)抑或者是点(dot)来表示。

Maker: 菜单提供marker(符号)式样的选择。



Marker Size: marker的大小。

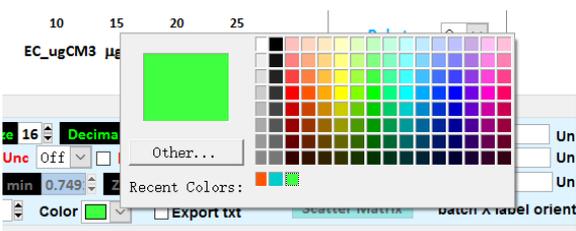
Marker thick: Maker外轮廓线的粗细。

Decimal: 标注中, 显示多少位小数点(适用于斜率, 截距和 R^2)。

Font Size: 图中字体大小。

Line size: 图中线条的宽度。

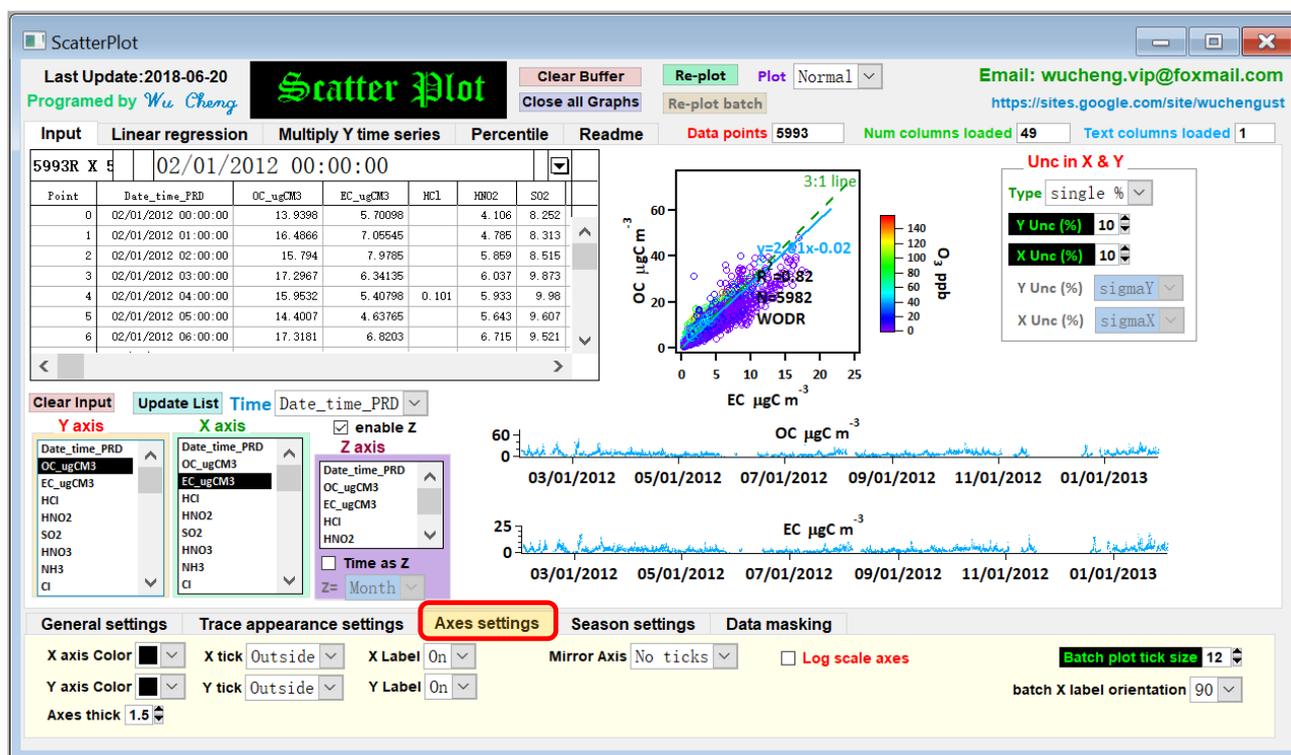
Maker Color: 设置Maker 的颜色(只适用于Z 轴关闭的情况下)



Regression line Color: 设置拟合线的颜色。

Ref Line color: 设置参考线的颜色。

4.3 Axes settings



X/Y axis Color: 设置轴的颜色.

X/Y tick: 设置刻度位置.

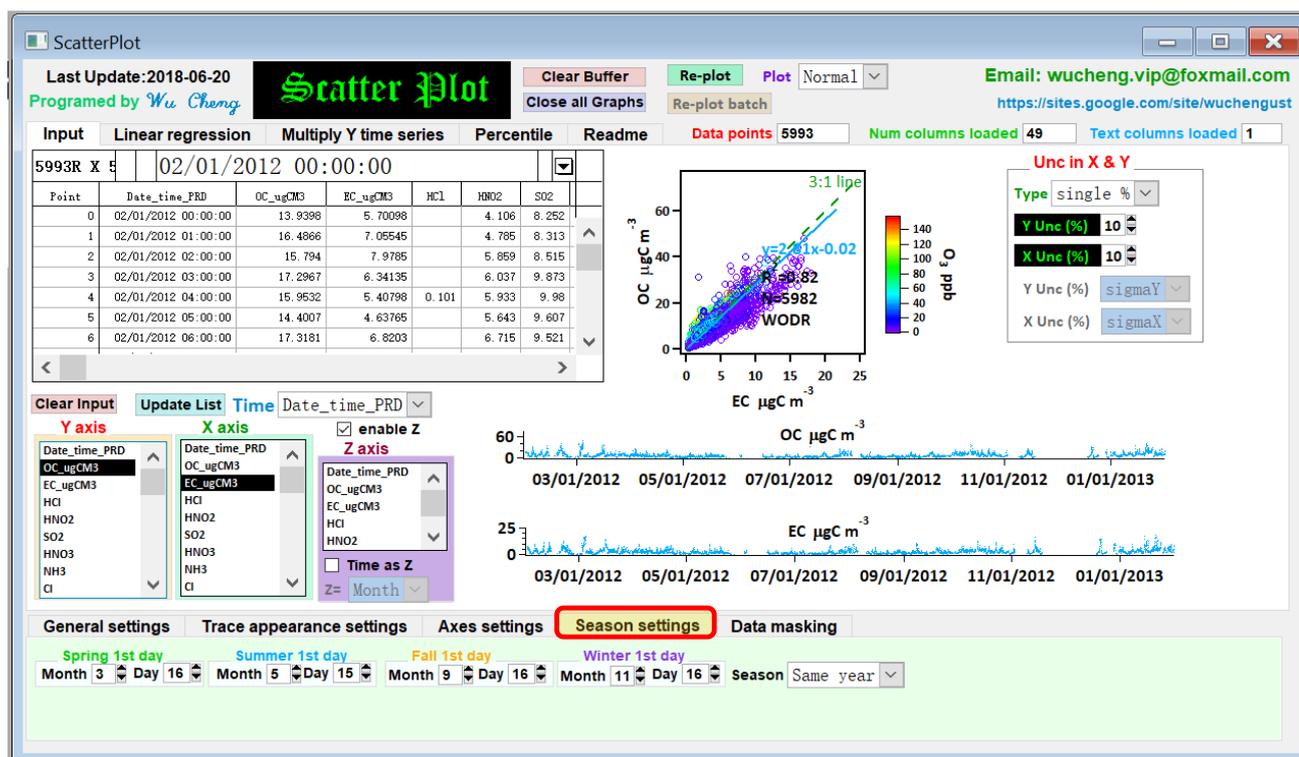
X/Y Label: 是否显示坐标轴标题.

Axes thick: 轴的线宽设置.

Mirror Axis: 显示或隐藏镜像轴.

4.4 Season settings

输入每个季节的第一天, 从而定义四季.



5 分页 “Input” 简介

(a) 用户可以选择哪些变量为X, Y和Z (Z可以打开或关闭)。 通过选择不同的X Y Z组合, 散点图和两个时间序列图将立即更新, 故而用户可以快速查看数据

(b) X&Y中的不确定性(误差)

WODR, Deming和York回归的不确定性设置。 有两种类型的输入可用, “Single%” 意味着用户只需要提供一个数字来指示X和Y中的相对不确定性。 Input Data” 意味着用户需要为单个数据点提供误差权重(单独一列数据的形式, 内容为标准差)。 用户需要使用弹出菜单来指定Y和X的相应的权重wave。

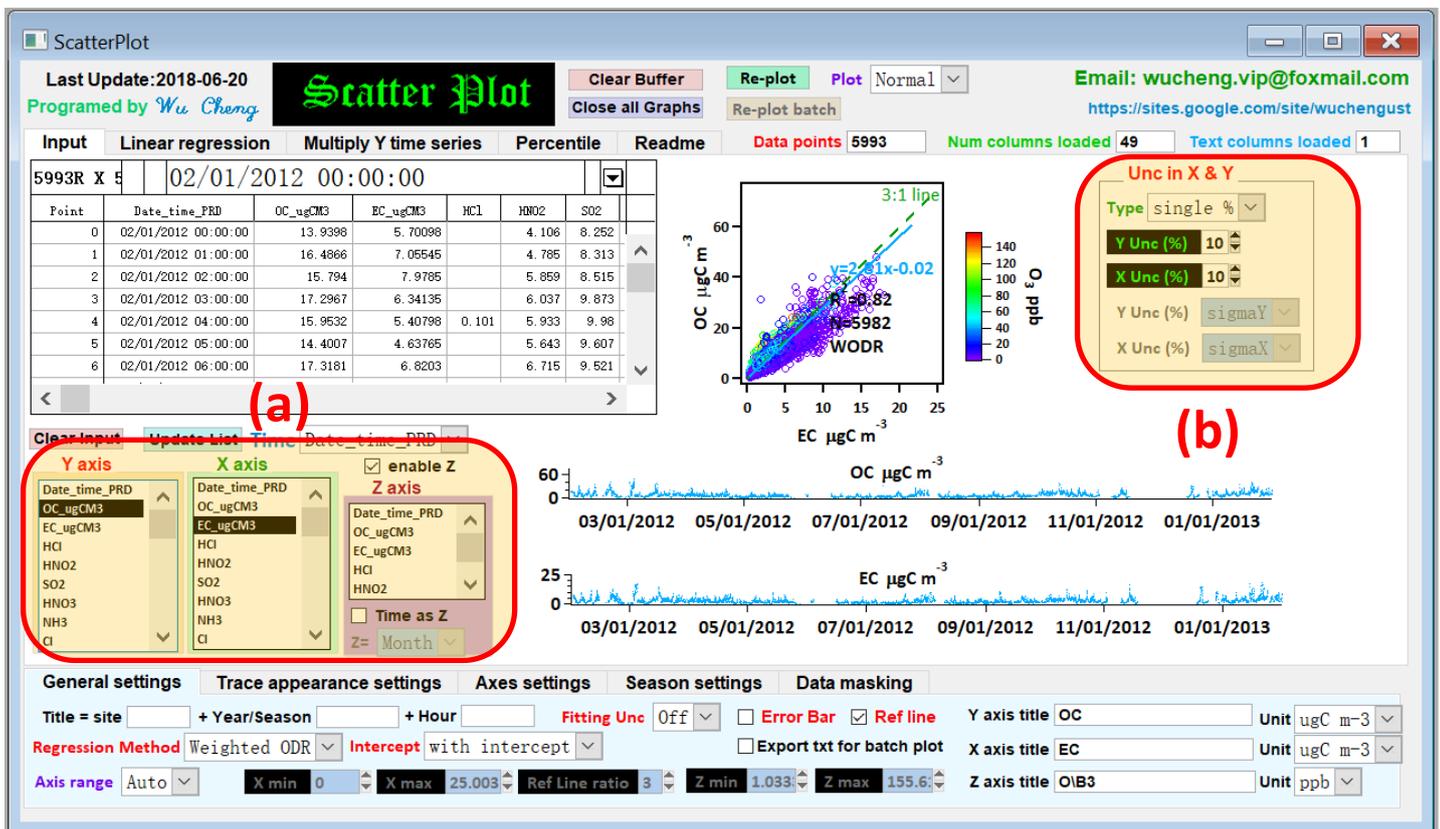


图5.1 Scatter plot 设置

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

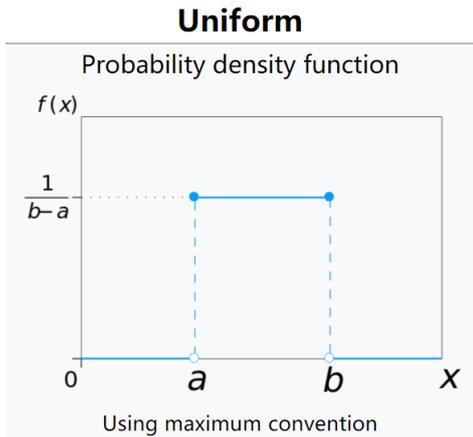
$$\lambda = \frac{\omega(X_i)}{\omega(Y_i)}$$

and the weights are:

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

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 $\pm m\%$ uncertainty, which follow a uniform distribution (in the range of $[a, b]$). the variance of the uniform distribution becomes

$$\begin{aligned} \sigma_{X_i}^2 &= \frac{1}{12} (b - a)^2 \\ &= \frac{1}{12} (X_i + m \times X_i - (X_i - m \times X_i))^2 \\ &= \frac{1}{12} (2 \times m \times X_i)^2 = \frac{(mX_i)^2}{3} \end{aligned}$$



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 分页 “Linear regression” 简介:

6.1 数据按时间筛选

三种类型的时间标度用于数据过滤：**YSM**（年季月），**Dow**（星期几）和**Hour**（0: 00~23: 00）

(a) **YSM**可以进一步分为五种情况：**年**；**年/月**；**月**；**年/季**；**季节**；**年/月/日**。季节由突出显示的区域 (d) 定义，使用每个季节的第一天作为分野。有两个选择是可用的，跨年和同年。例如，如果选择跨年，1999年12月和2000年1月分组在一起作为1999年冬季。在选择期间可以使用shift键进行多项选择。

(b) **Dow**（星期几）。勾选Multi select后，既可以使用shift键进行自定义多项选择，也可以使用**weekday, weekend**进行快速选择。

(c) **Hour**（0: 00~23: 00）。勾选Multi select后，既可以使用shift键进行多项自定义选择，也可以使用**daytime nighttime**进行快速选择。用户可以在(d)处自定义daytime nighttime的时间段。

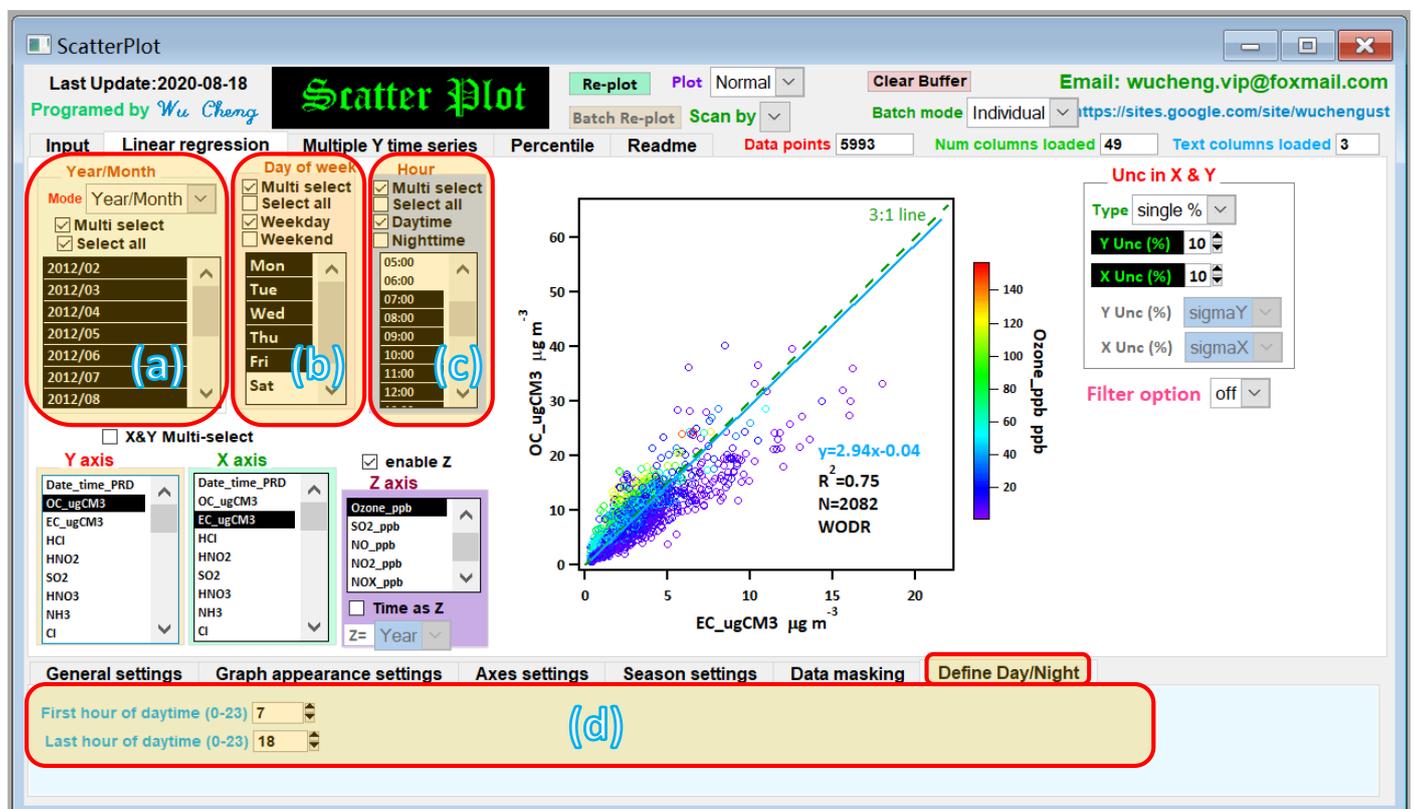


图 6.1.1 时间筛选的列表框

6.2 用数据进行筛选

可以有三种类型的数据过滤：按文本（字符串）数据列表，按数值数据列表，按各列的数值范围

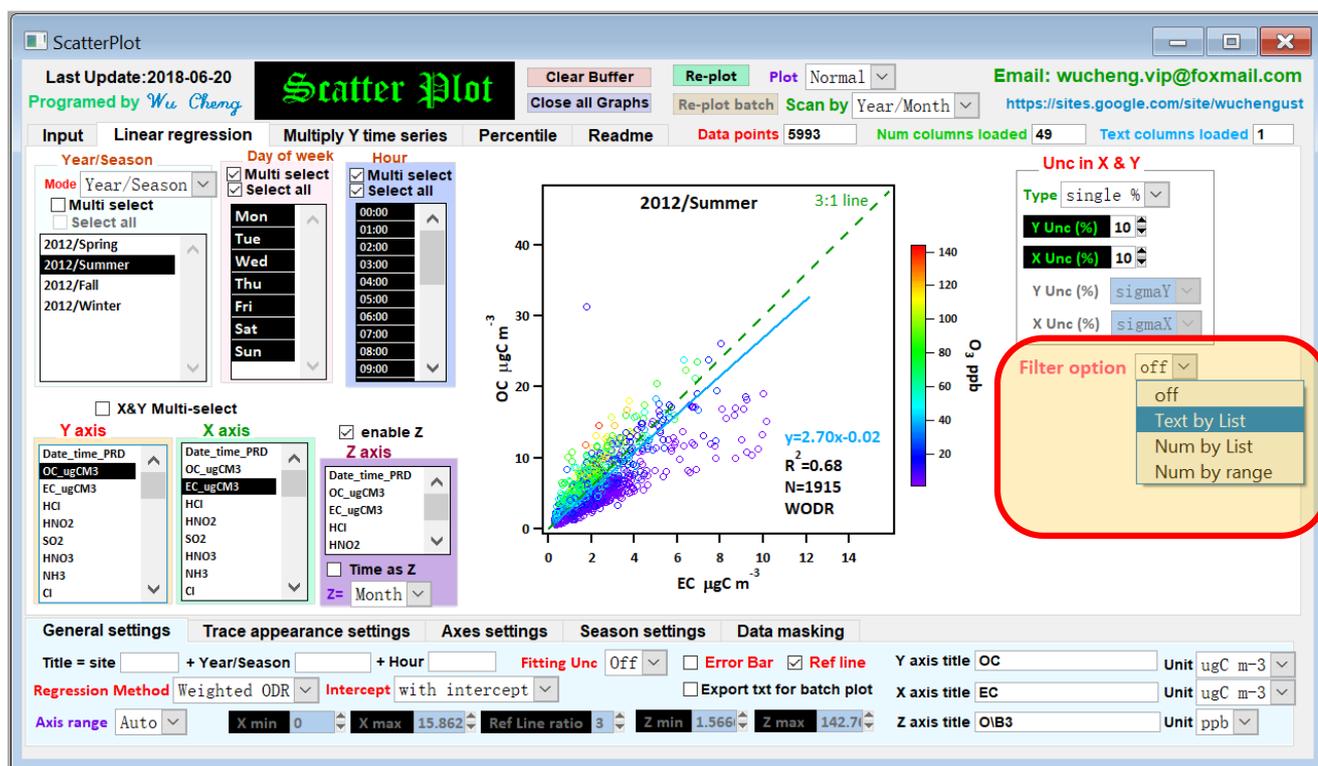


图 6.2.1 按数据过滤数据

(a) 使用字符串筛数据Text by list: 例如，提供包括C1~C4的后面轨迹分组信息的列，使用该函数，绘制子集（例如，如下所示，仅C2）。系统也会自动生成weekend和day/night字符串的列来进行数据筛选。可以进行多项选择。

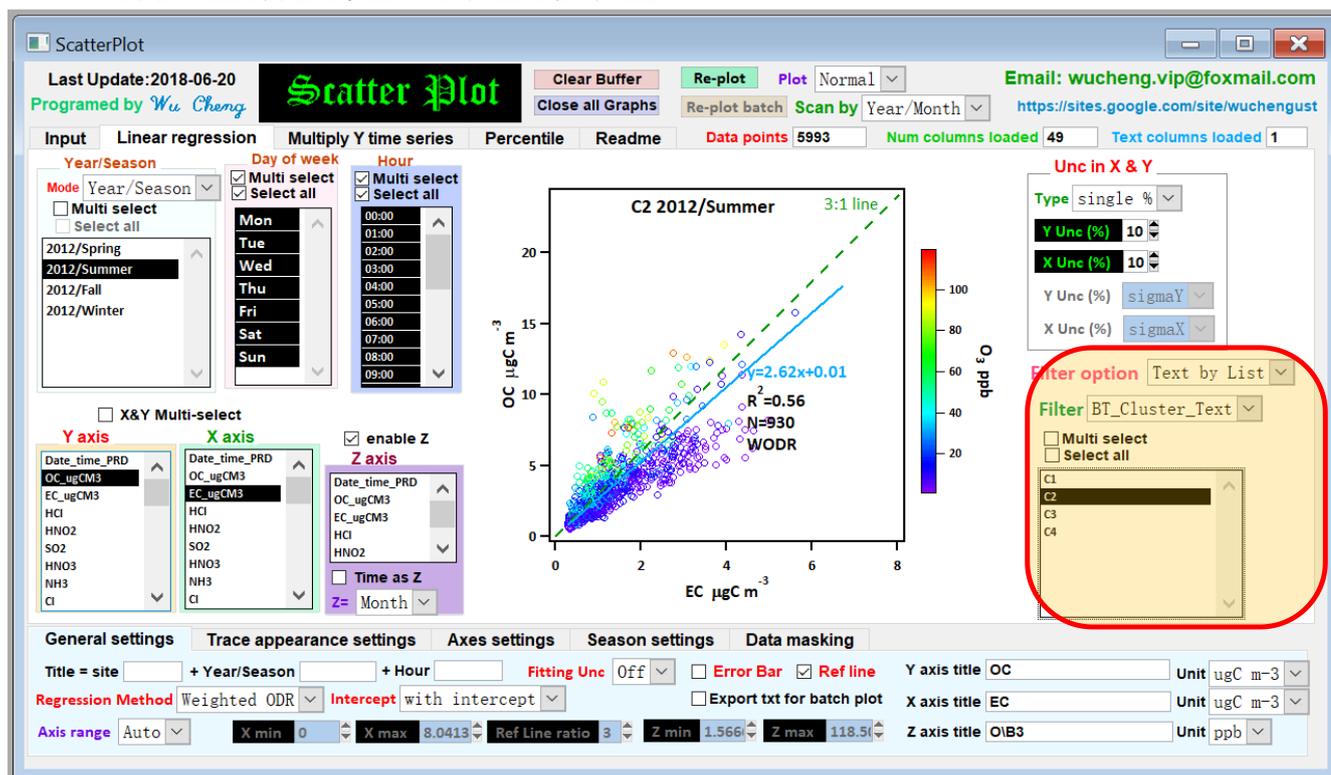


图 6.2.2 按文本数据列表过滤数据 - Text by List

下图是一个例子，使用 Day/night 字符串批量做图输出到同一张图。

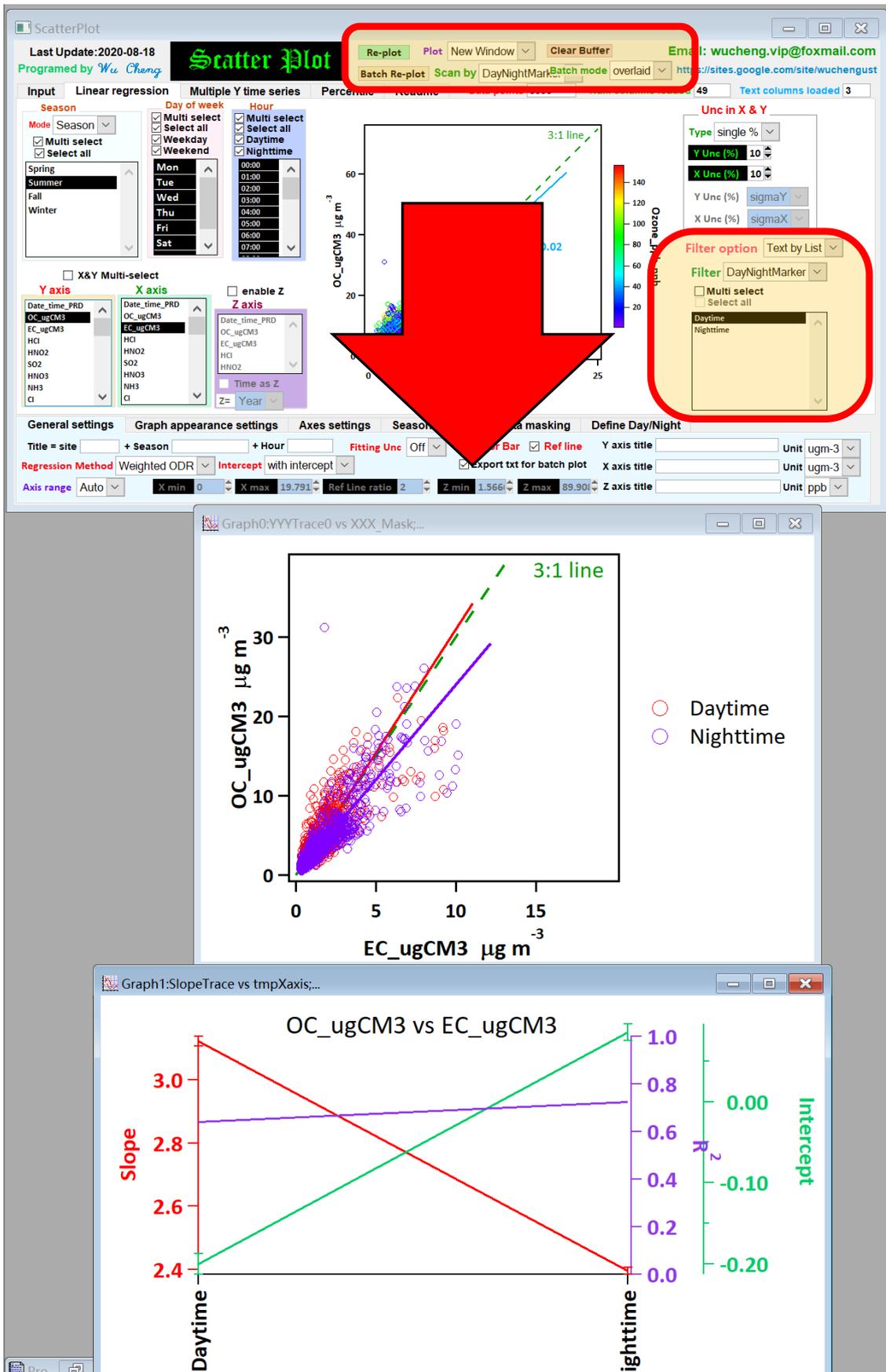


图 6.2.3 使用 Day/night 字符串批量做图输出到同一张图示例

(b) 使用数值型数据来筛数据 Num by List: 具有数值的列可以用作数据分组的过滤器。当唯一的数量远小于总计数时, 它很有用。例如, 数据按照如下所示的RH分组。

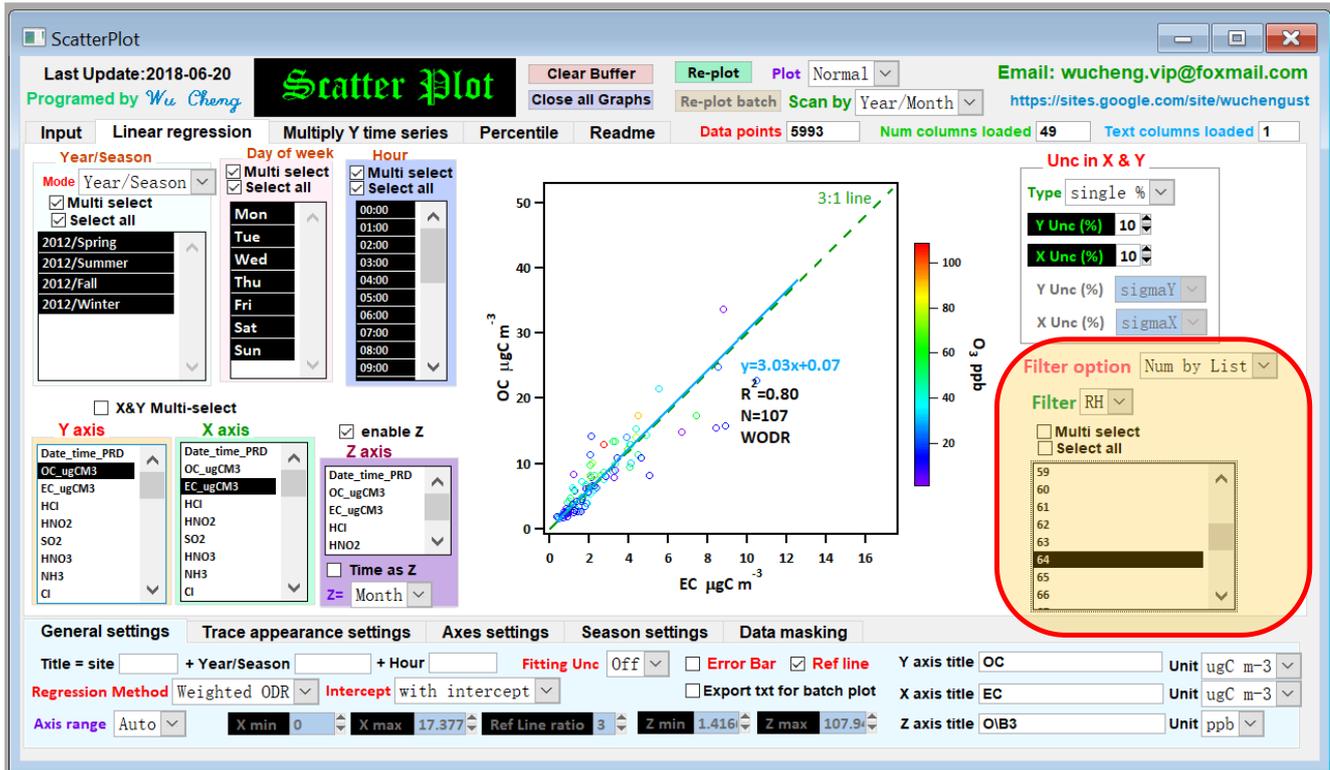


图 6.2.4 按数据过滤数据- Num by List

(c) 使用数值型数据的范围来筛数据 Num by range: 范围 (由min和max定义) 可用于单个列以筛选子集。当使用多个列时, 会取这些条件的交集。以下是使用 $30 < \text{RH} < 50$ 的实例。

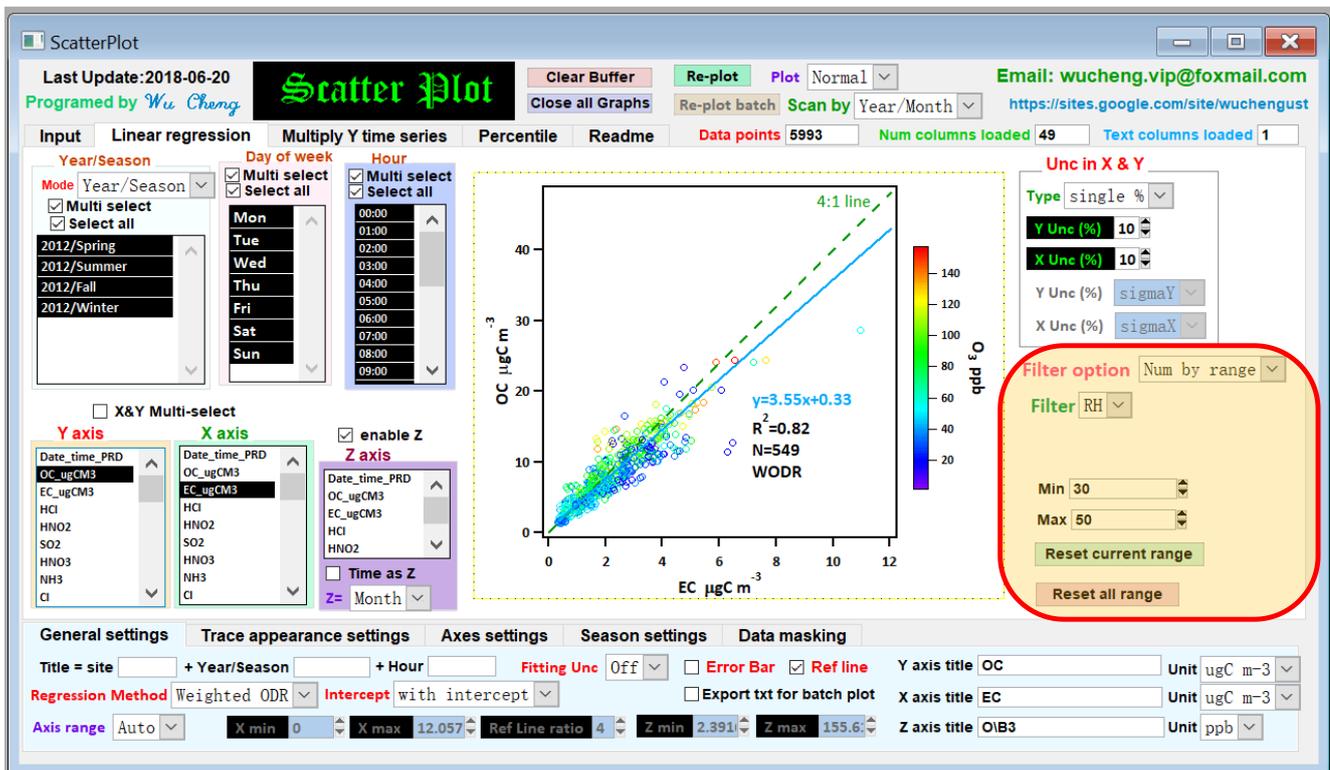


图 6.2.5 按数据范围过滤- Num by range

6.3 使用图形界面进行数据遮掩(数据剔除)

数据遮掩功能可剔除不需要的数据点再进行线性回归。本程序可以使用图形用户界面直接便捷实现，不需要用户返回原始数据搜寻需要删除的点，大大提高了易用性。

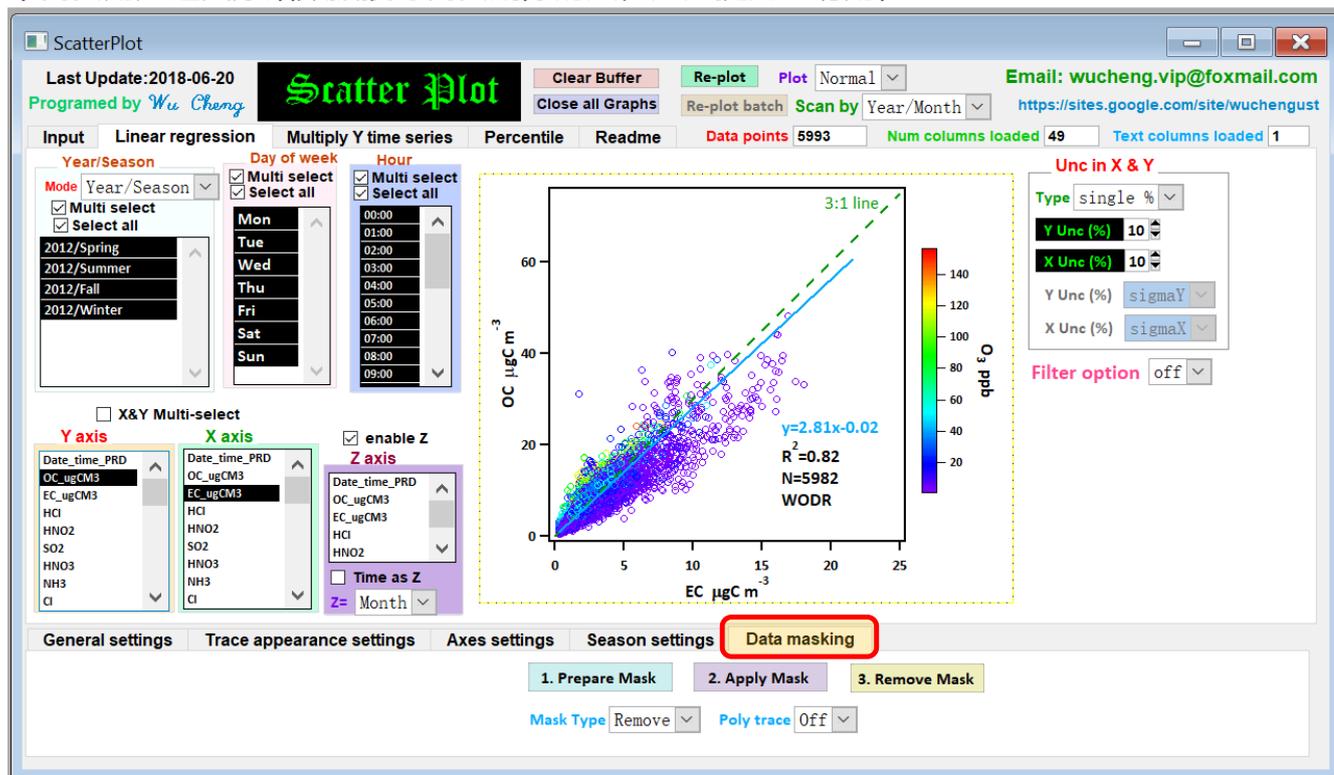


图 6.3.1 数据遮掩功能区一览

首先，点击"Prepare Mask" 按钮。然后可以使用光标绘制多边形。多边形由路径点定义。

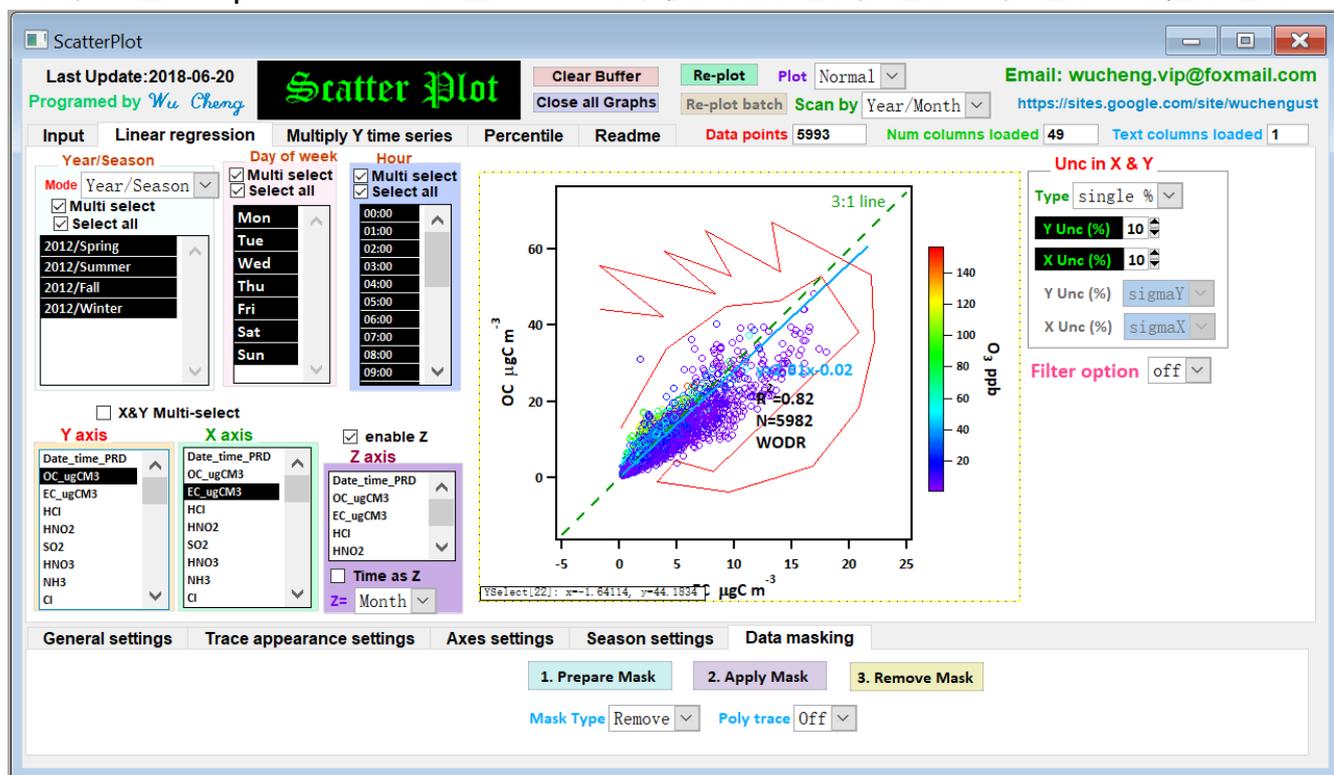


图 6.3.2 用光标开始进行多边形的绘制。

确保多边形闭合，如图6.3.3所示，每个点将以方形标记的形式显示。

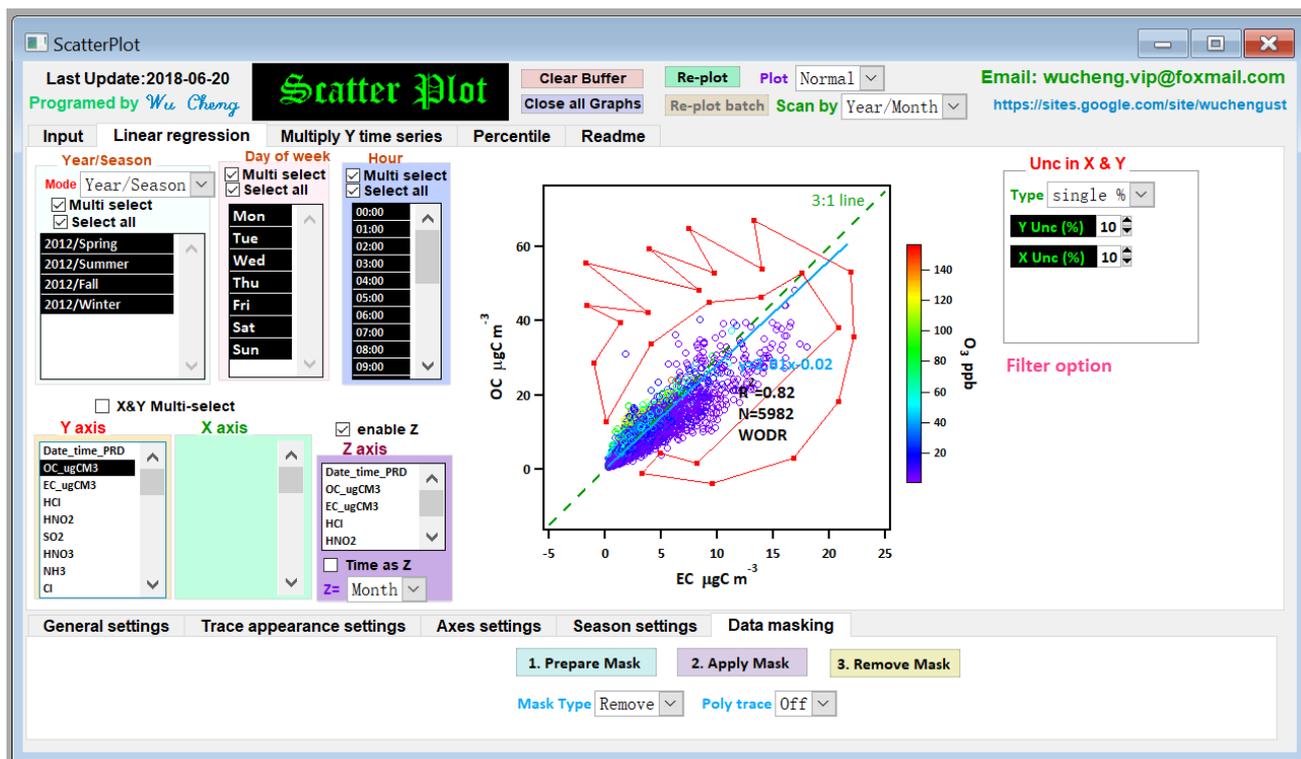


图6.3.3 闭合的多边形的示例

一旦多边形完成，选择“Mask type”。以下是选择“Highlighted”的示例，然后单击“Apply Mask”按钮。不需要的数据点会被标记为粉红色三角形。这样就实现了排除多边形内的数据点进行回归（注意，数据点个数N从图6.3.3中的5982变为图6.3.4中的5980）。对于这个特定的例子，去除不需要的数据点没有改变斜率和截距。

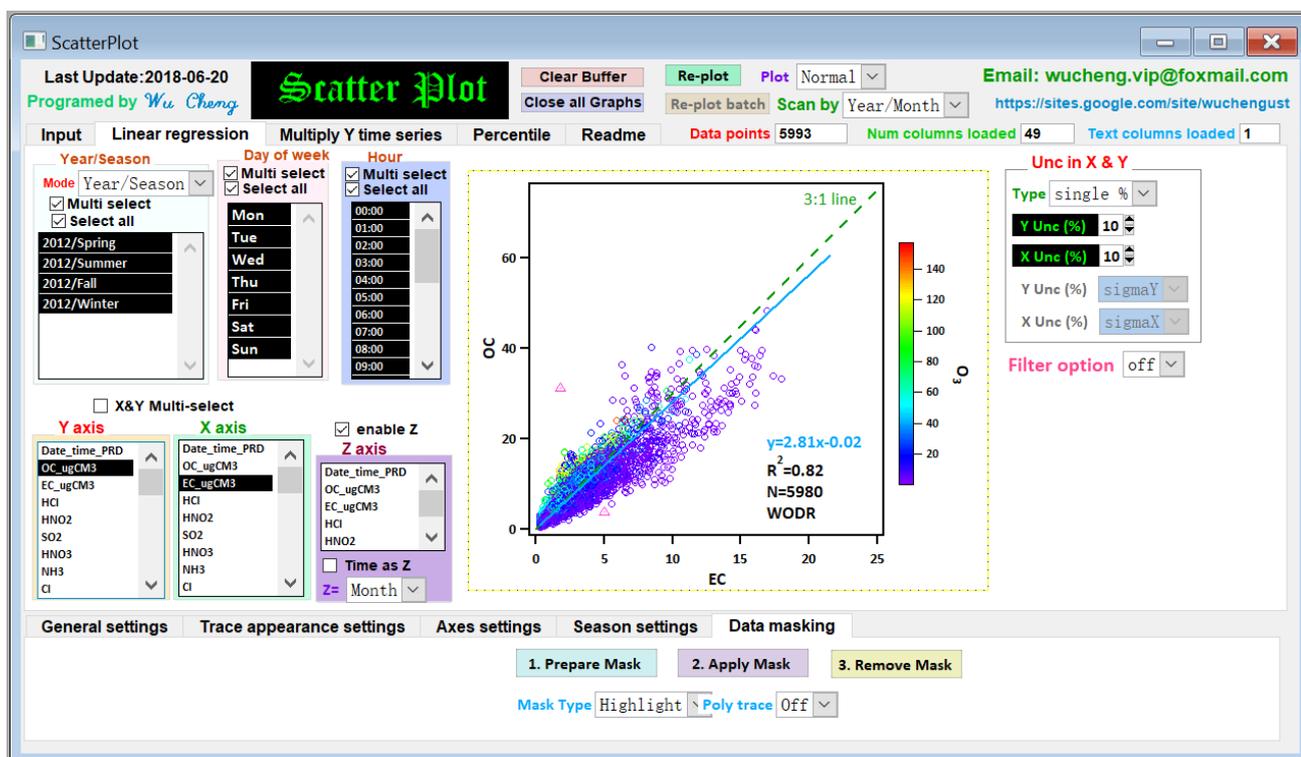


图6.3.4 在“Mask type”中选择“Highlighted”的示例。

以下是在“Mask type”中选择“Remove”的示例，然后单击“Apply Mask”按钮。删除不需要的数据点。这样就实现了排除多边形内的数据点进行回归（注意，数据点个数N从图6.3.3中的5982变为图6.3.5中的5980）。对于这个特定的例子，去除不需要的数据点不影响斜率和截距。

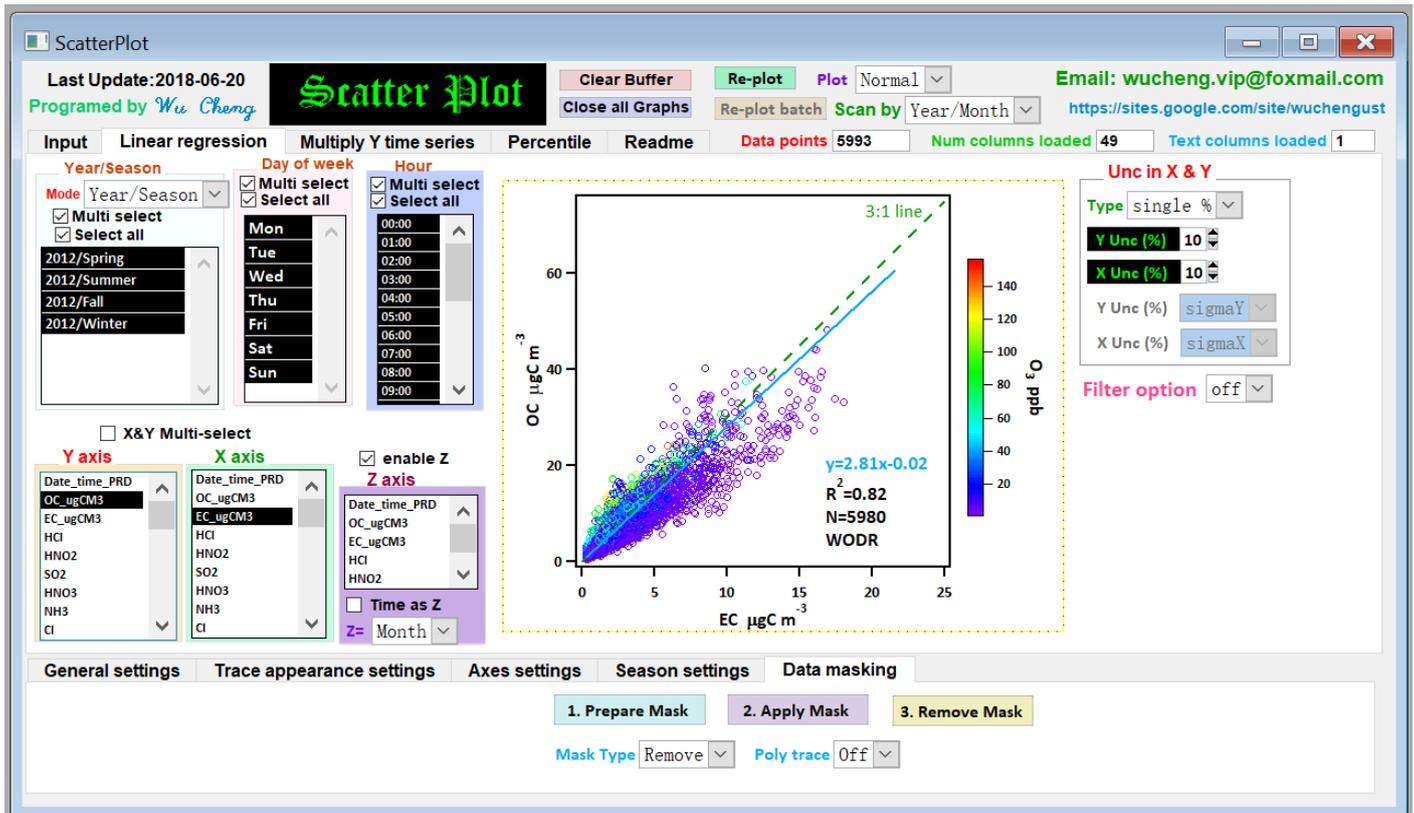


图 6.3.5 在“Mask type”中选择“Remove”的示例。

以下是在“Trace type”中选择“On”的示例，然后单击“Apply Mask”。除去不需要的数据点，并以虚线显示多边形。多边形内的数据点被排除回归（注意，数据点个数N从图6.3.3中的5682变为图6.3.6中的5876。对于这个特定的例子，去除不需要的数据点改变了斜率（2.81->2.85）和截距（-0.02->-0.05）。

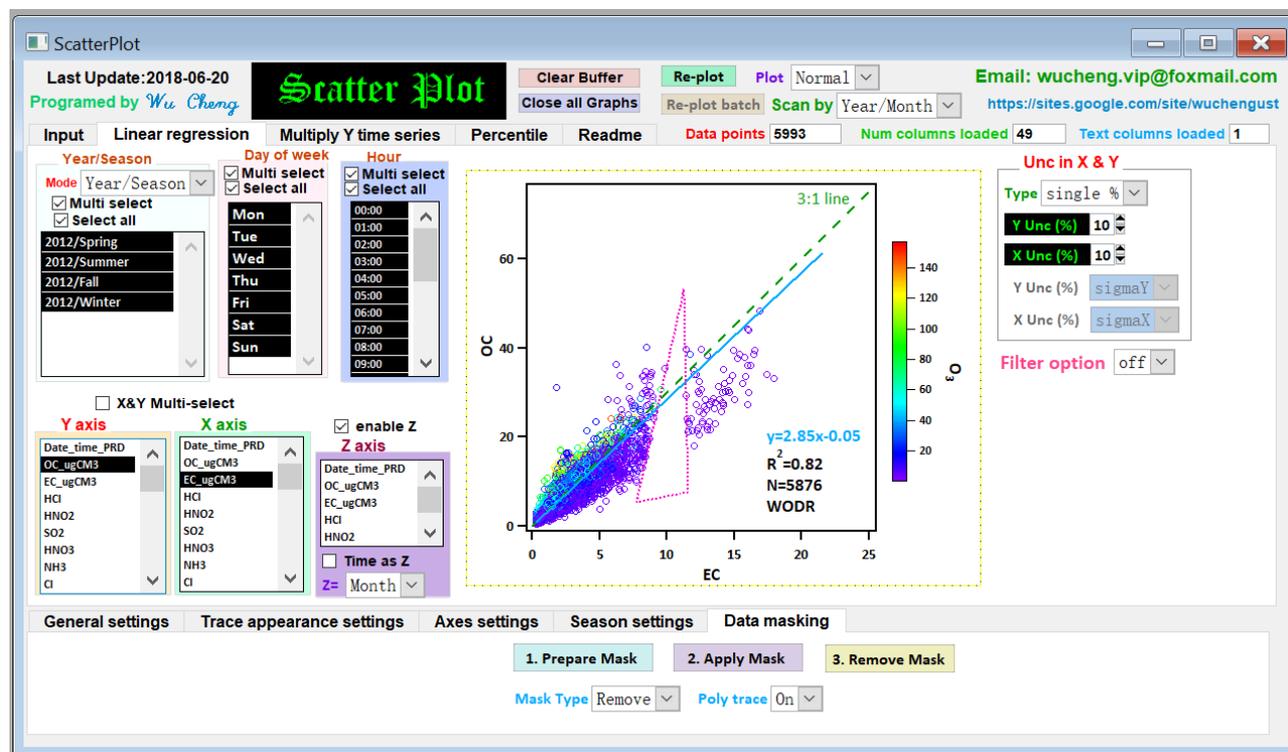


图 6.3.6 在“Trace type”中选择“On”的示例。

如果需要输出数据遮掩的标记数组用于后续数据处理，可以点击“Export mask as txt”

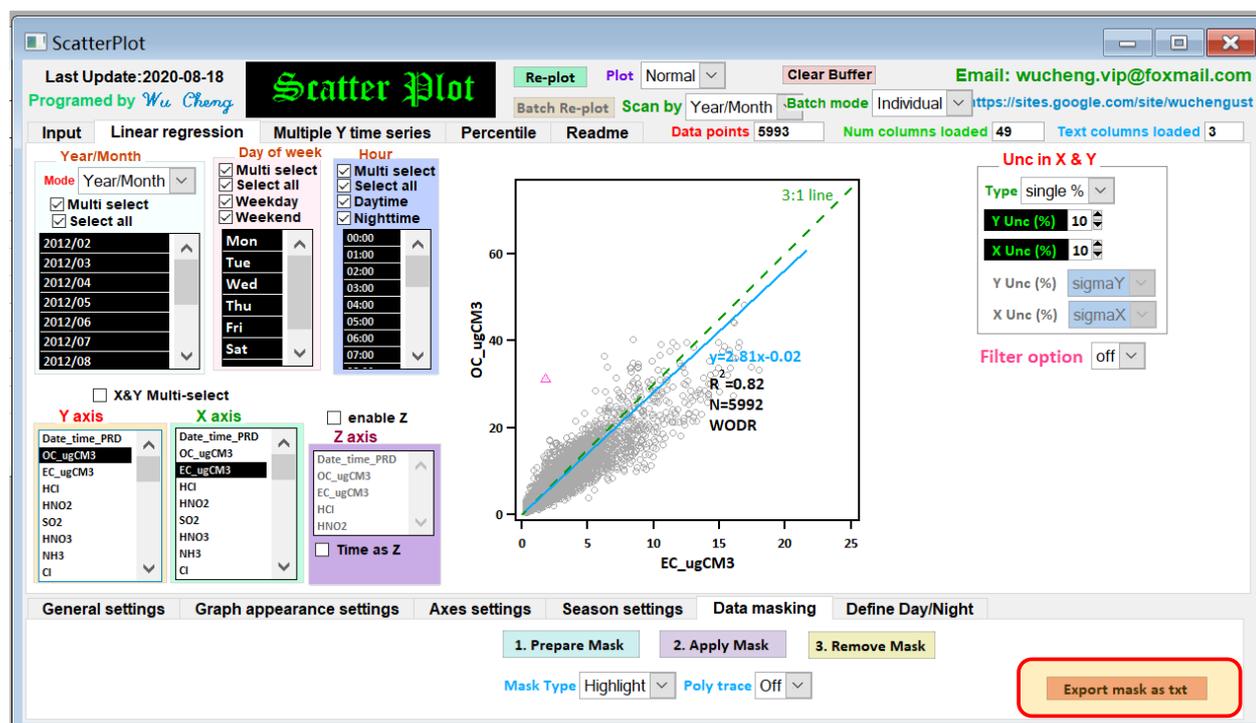


图 6.3.7 Export mask as txt

输出的数据遮掩的标记数组（CSV文件）用Excel打开如下图所示，被剔除的那行数据会标记为1，正常数据会标记为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			

图 6.3.7 数据遮掩标记数组示例

最后，要重置数据遮掩，请单击“Remove Mask”，然后将擦除所有数据屏蔽，如下所示。

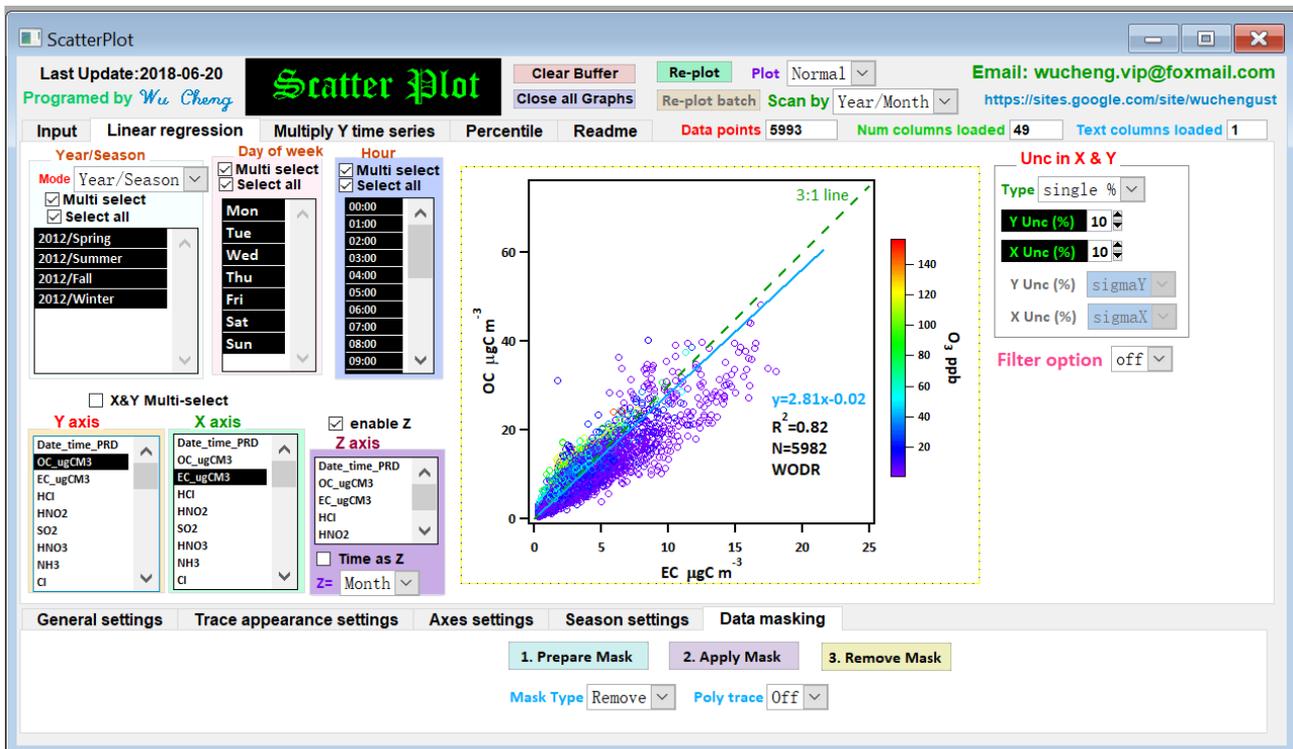


图 6.3.8 应用“Remove Mask”后的示例。

6.4 选择多个变量用作X&Y

有时X和Y不是一对一的变量，有可能是多对一或者多对多。在X和Y变量中多项选择允许用户通过使用它们的加和用作X和Y。可以使用“shift”键和光标选择X和Y中的多项变量（wave）。X和Y中各自所选的总和将被用于线性回归。以下是气溶胶的离子色谱数据的QA / QC示例。使用硫酸盐和硝酸盐的总和作为Y，将铵根离子作为X，以检查离子的电荷平衡。

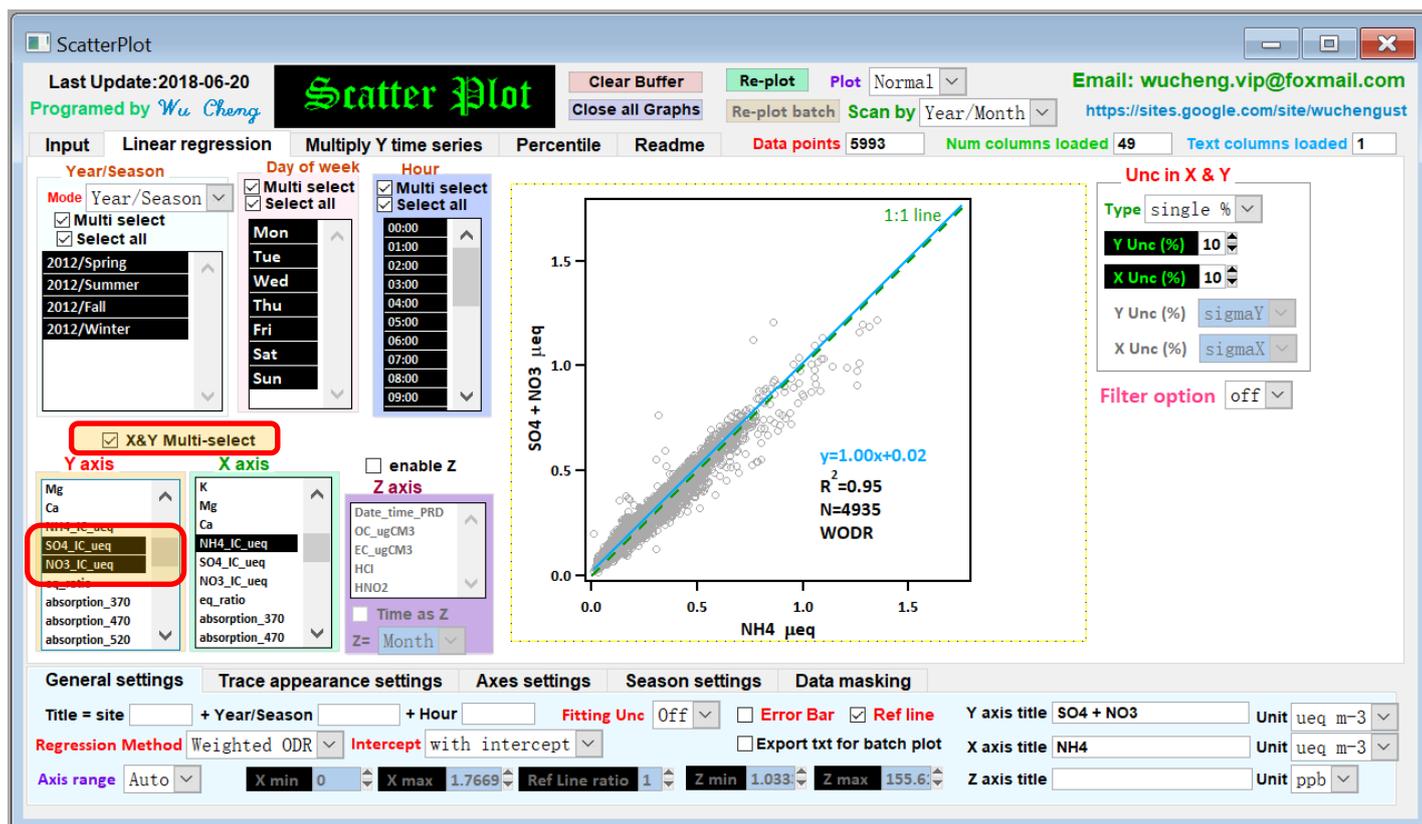


图 6.4.1 选择多个变量用作X&Y的示例.

6.5 时间变量作为Z轴

除了使用用户直接输入变量作为Z轴, 包括YSM (年季节月), Dow (星期几) 和小时 (0: 00~23: 00) 的派生变量可以用作Z。

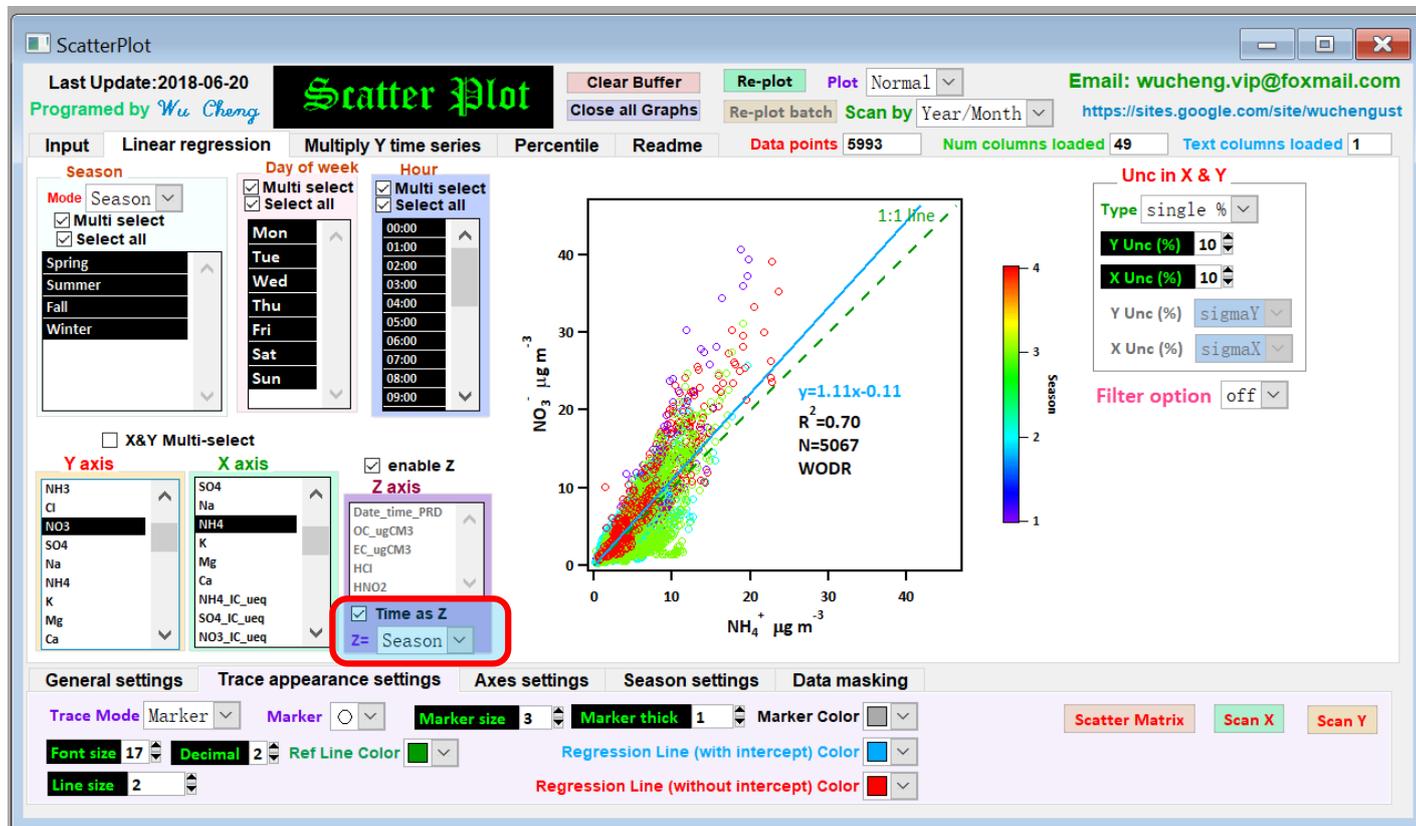


图 6.5.1 使用季节作为Z轴颜色编码的示例。

6.6 批量绘图-每张单独输出

当绘图选项不是“normal”时，则批处理绘图被激活。批量绘图可以在三个时间维度（Scan by）进行：年/季/月，星期几，小时，这对应于按时间进行数据分组的三个维度的列表框。

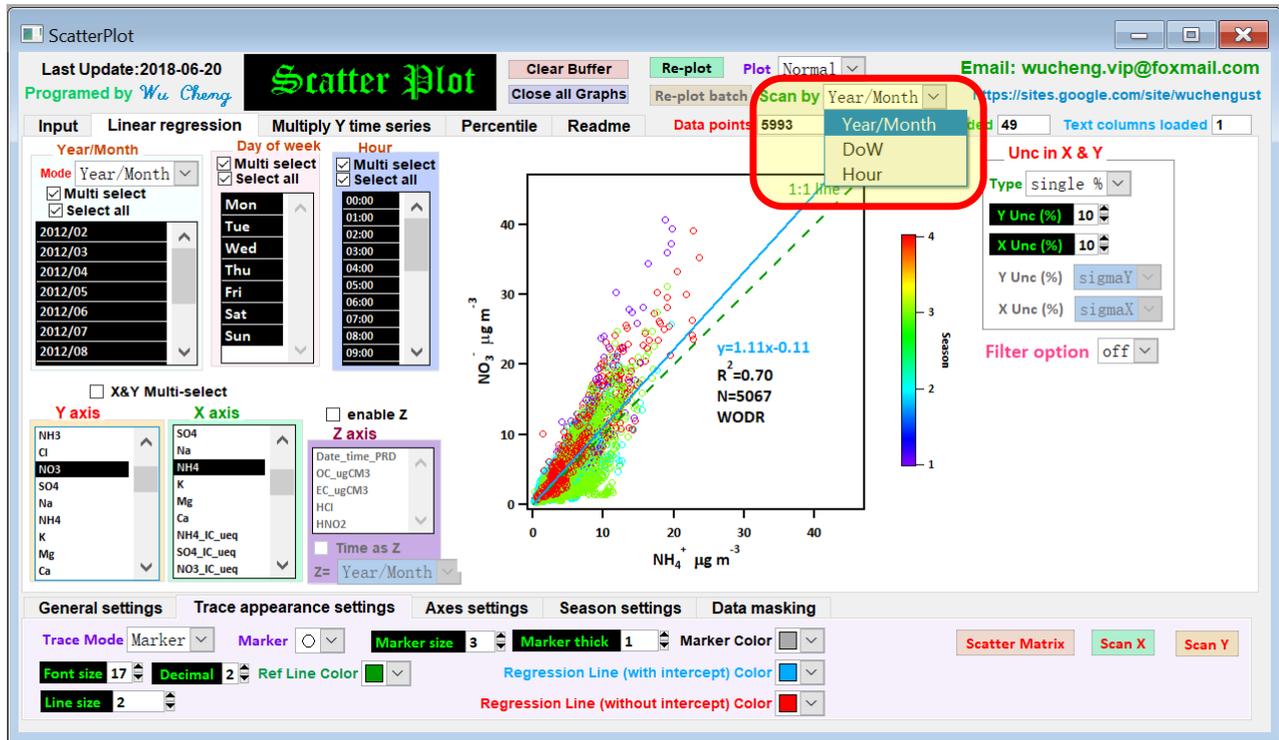


图 6.6.1 按时间维度进行批量绘图的设置

第四种方式是通过文本标记进行扫描。当使用“Text by list”激活数据筛选器时，第4个选项将显示在“Scan by”弹出菜单中。

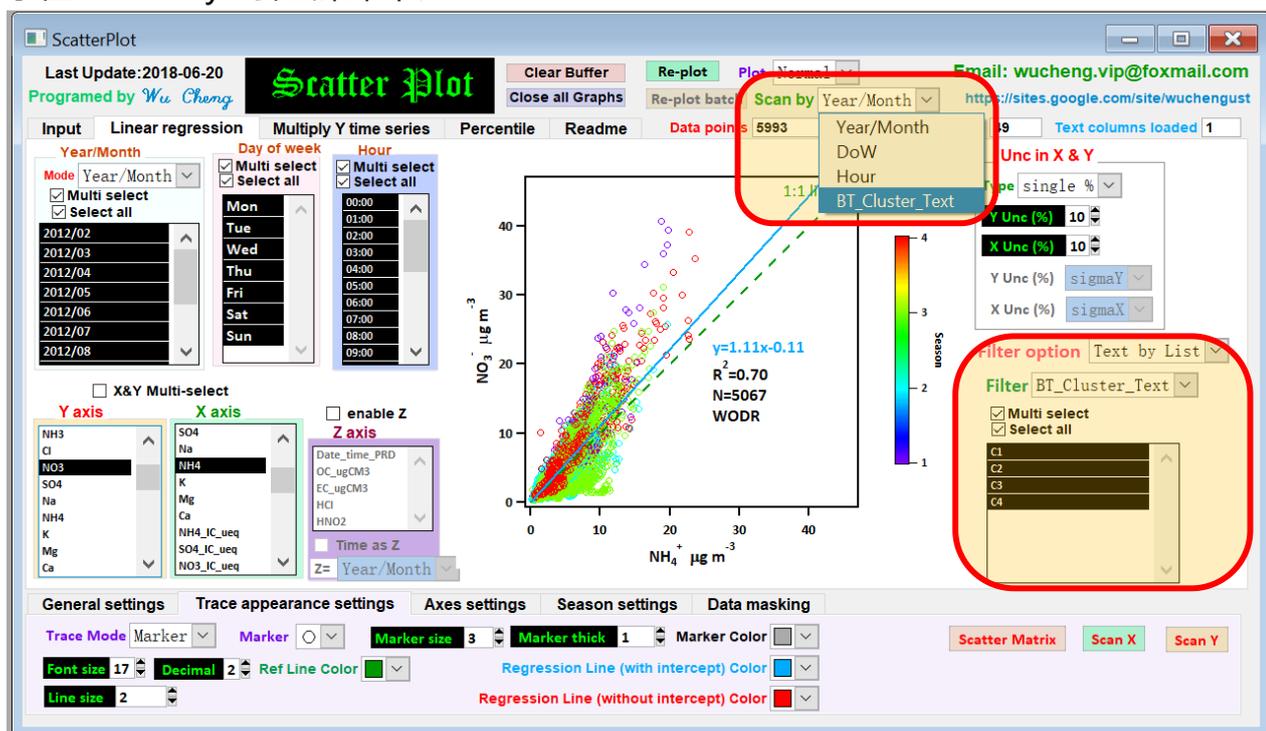


图6.6.2 通过文本标记进行批量绘图

以下是实施批量绘图的示例，按年/月（12个月）扫描。除了单个散点图，还将给出总结斜率，截

距和 R^2 随年/月的变化的图。如下图所示，硝酸盐对温度（挥发）敏感，因此夏季（6月 - 9月）的斜率远低于冬季（12月 - 2月）。

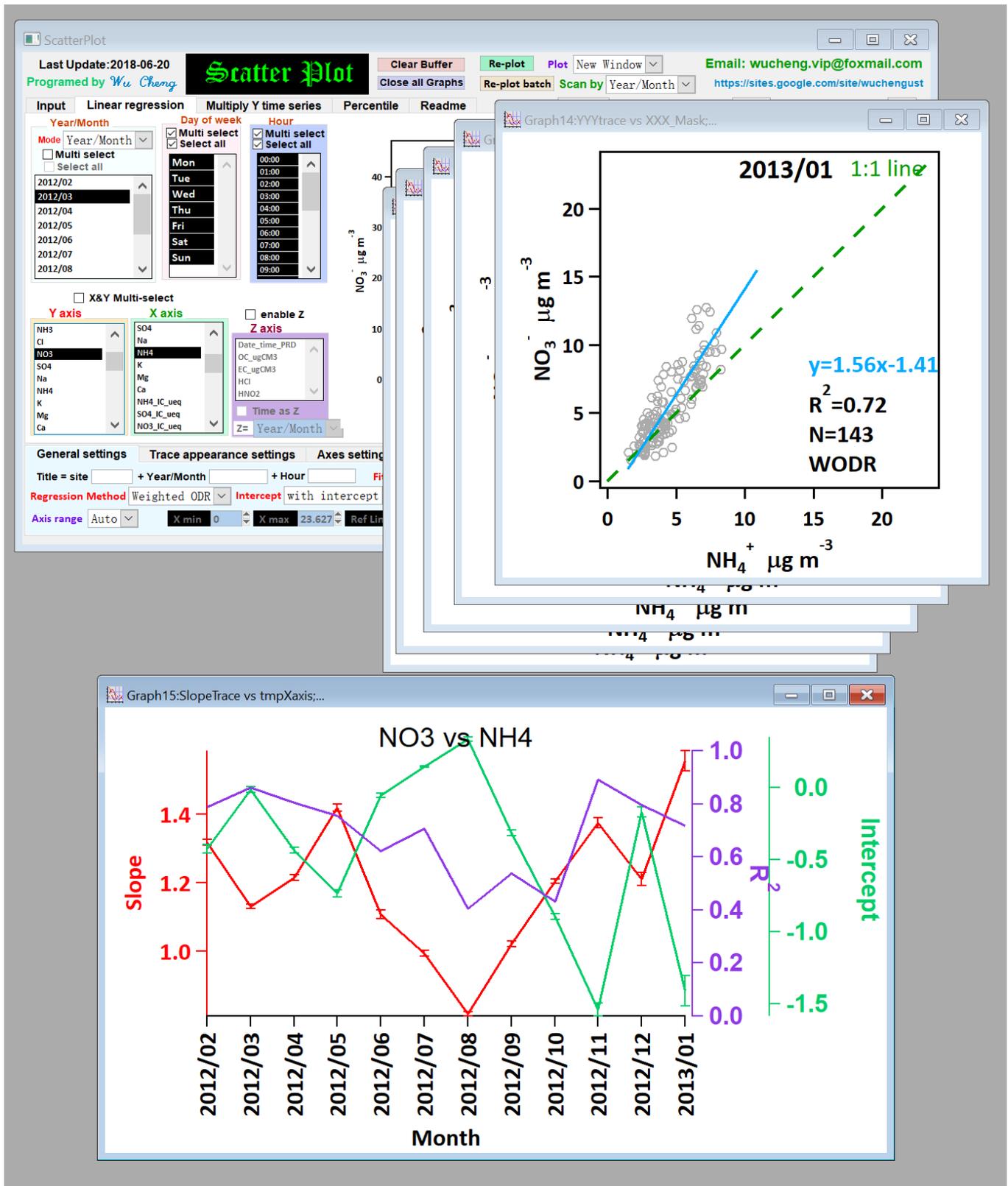
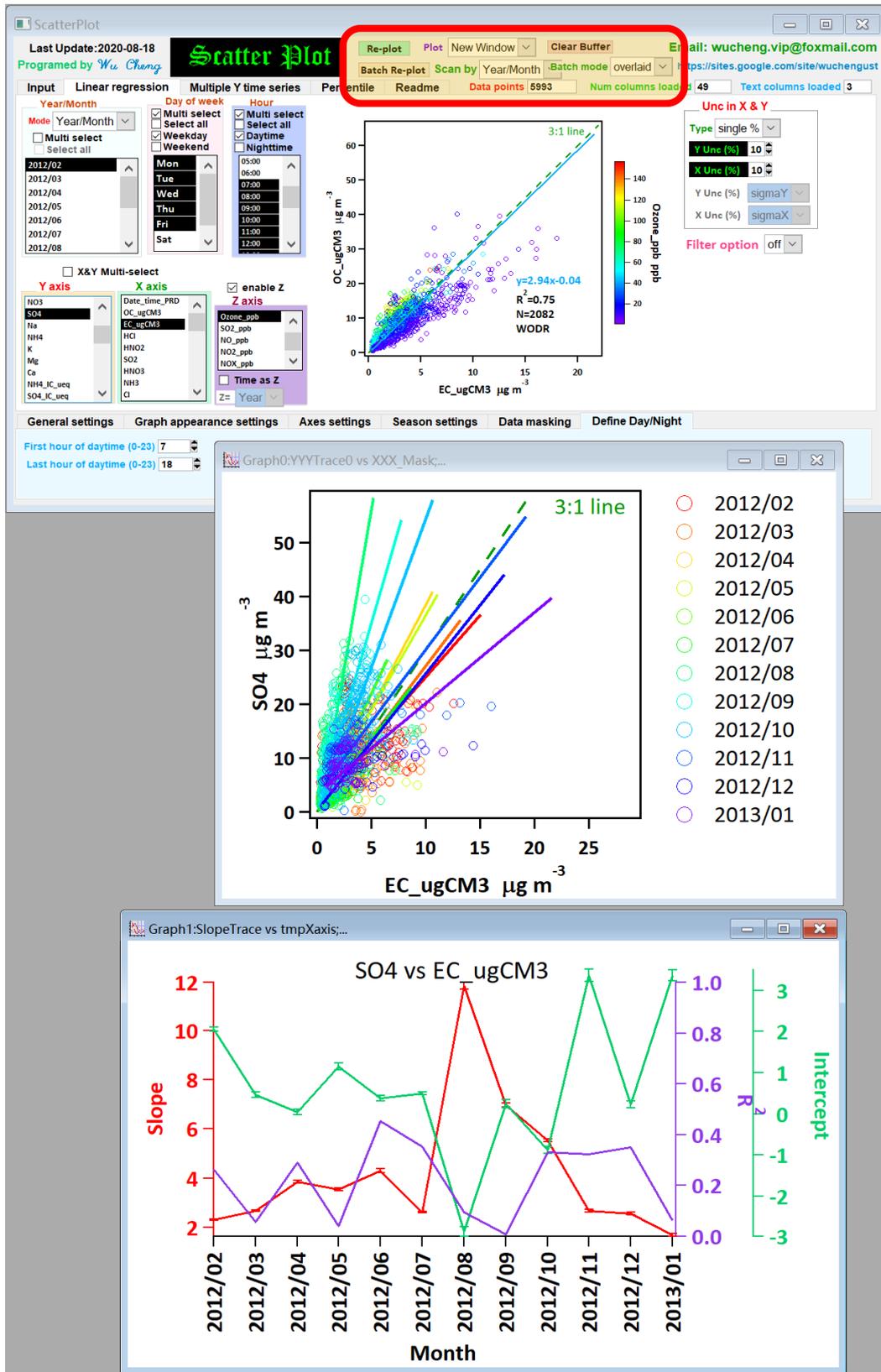


图 6.6.3 根据年/月执行批量绘图的范例。

6.7 批量绘图-合并成1张图输出

合并模式(**Batch mode: overlaid**)下, 批量绘图的结果都呈现在同一张图上, 如下图所示。每个线条会按照彩虹色表自动分配颜色。“Scan by” 批量绘图的可以操作的维度包括: YSM 年/季/月(具体可分为**年; 年/月; 月; 年/季; 季节; 年/月/日**), **星期, 小时**, 以及**字符串数组(用户自定义字符串, 工作日/周末, 昼/夜)**。下图例子中展示的是按照“**年/月**”筛选了**工作日**并且是**白天**的数据进行批量画图, 批量拟合的斜率, 截距, R^2 等信息在勾选了 txt 后会以 csv 文件输出到 pxp 文件所在的文件夹。



7 分页 “Multiple Y time series” 简介

多变量Y时间序列图通常用于呈现各种污染物的时间变化。如下所示，可以使用“添加”按钮选择所需的Y。

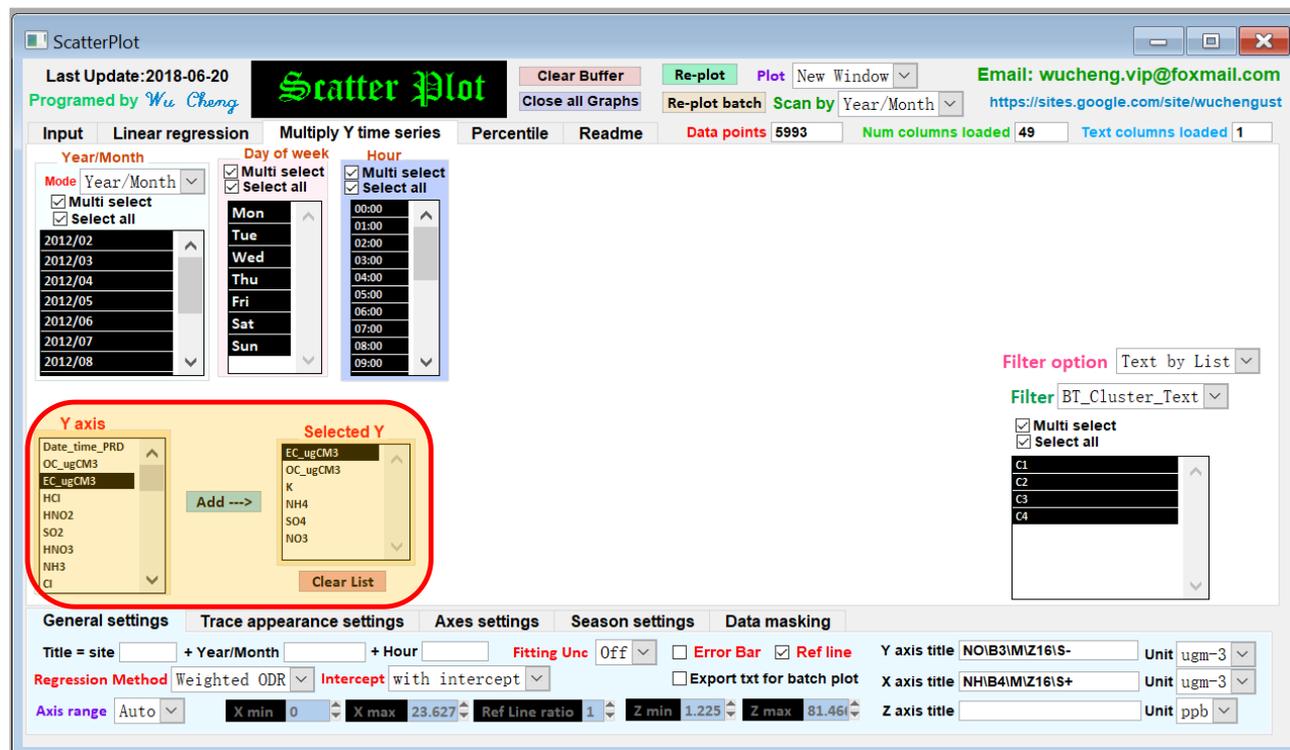


图7.1通过“Add”按钮选择所需Y的示例

“Plot option”应设置为“new”。然后，点击“Re-plot”将在新窗口中生成图形。而后用户可以在新窗口中设置颜色和线形，轴标题等外观。本功能主要目的是节省为单个轴设置Y方向的份额需要耗费的时间。

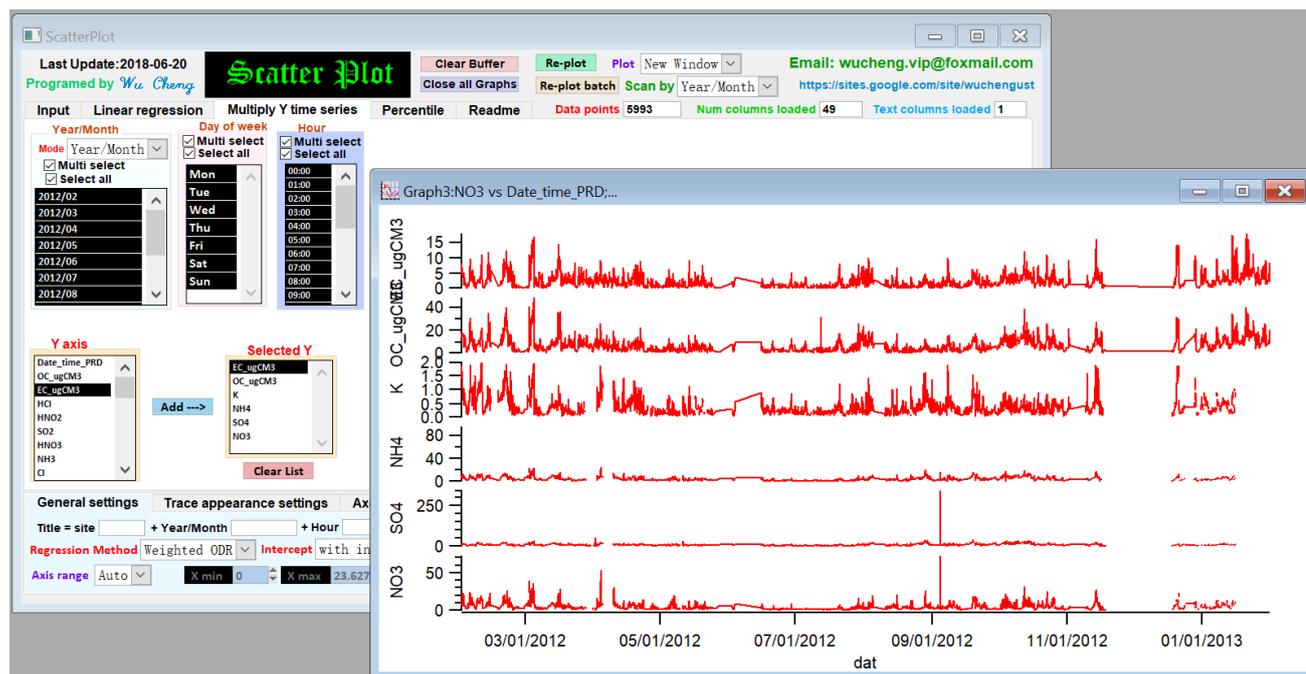


图7.2 点击“Re-plot”后在新窗口中生成多变量Y时间序列图的示例。

8 分页 “Percentile” 简介

在EC示踪法中，具有特定OC / EC百分位数的子集通常用于回归来确定 $(OC/EC)_{pri}$ 。

Step: 百分位数子集的步长。例如，0.005代表0.5%的间隔。

Slope Min & Max: 设置左图的斜率范围 (y轴)

Intercept Min & Max: 设置左图的截距范围 (y轴)

Replot: 计算子集上的逐步回归。例如，以0.5%的步长，从OC / EC 0.5% ~ 100%的所有子集上计算线性回归。



这个滑动条是为了选择百分位数。该选择仅在应用“Replot”后才可用。

用。

Redraw: 在新窗口中创建绘图(plot option: new window).

使用Tab “Percentile” 的步骤

1) 设置步长，斜率和截距范围

2) 点击 “Replot”

3) 使用滑杆检查结果。例如，要显示10%子集的回归结果，请将滑动条拖动到10%，右侧的散点图将相应更新。紫色的数据点代表回归的选定子集，灰色的数据点代表未使用的数据。

4) 导出绘图 (在新窗口/导出到文件)，点击 “Redraw” 。

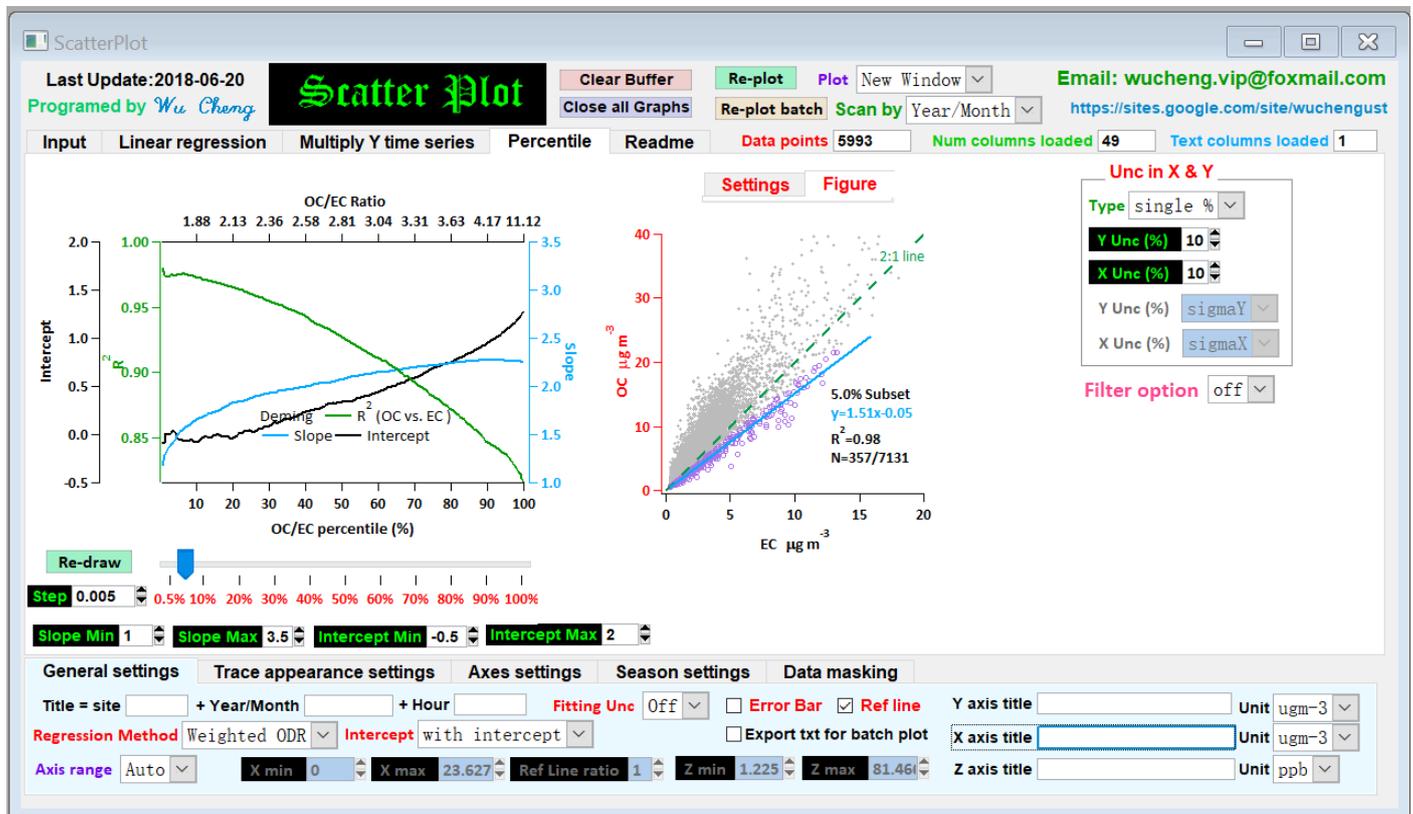


图 8.1 分页 “Percentile” .

9 使用了该工具包的已发表文章:

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 (PM_{2.5}) 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](https://doi.org/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 WSIs) 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。



No. 01760115

