



Chinese FLUX Observation and Research Network
ChinaFLUX



ChinaFLUX第十八次通量观测理论与技术培训

数据质量评价和控制

李彦磊 (yanlei.li@Campbell.com.cn)
Campbell Scientific Inc.

数据质量控制和评价

由于环境因素或电子器件故障等导致的不合理高值或低值

- 由于雨、雪、尘粒等对传感器声/光程干扰，瞬间断电等产生的异常值，标记为Hard spikes
- 电子电路（如A/D转换器）、长电缆、电源不稳定等产生的异常值，标记为Soft spikes

观测环境并不十分满足涡动相关法的测量要求

仪器观测为空间测量而不是点测量

空气密度效应影响

.....



观测设备被恶劣天气影响



10Hz原始数据的质量控制



1.1 De-Spike

根据Vickers and Mahrt (1996) 等统计方法进行数据的预处理:

- spikes
- amplitude resolution
- absolute limits
- Higher-moment statistics
- drop out
- Non-stationarity of the horizontal wind
- discontinuities
- Lag correlation
- Vertical structure

Vickers and Mahrt, 1996, *Quality Control and Flux Sampling Problems for Tower and Aircraft Data*



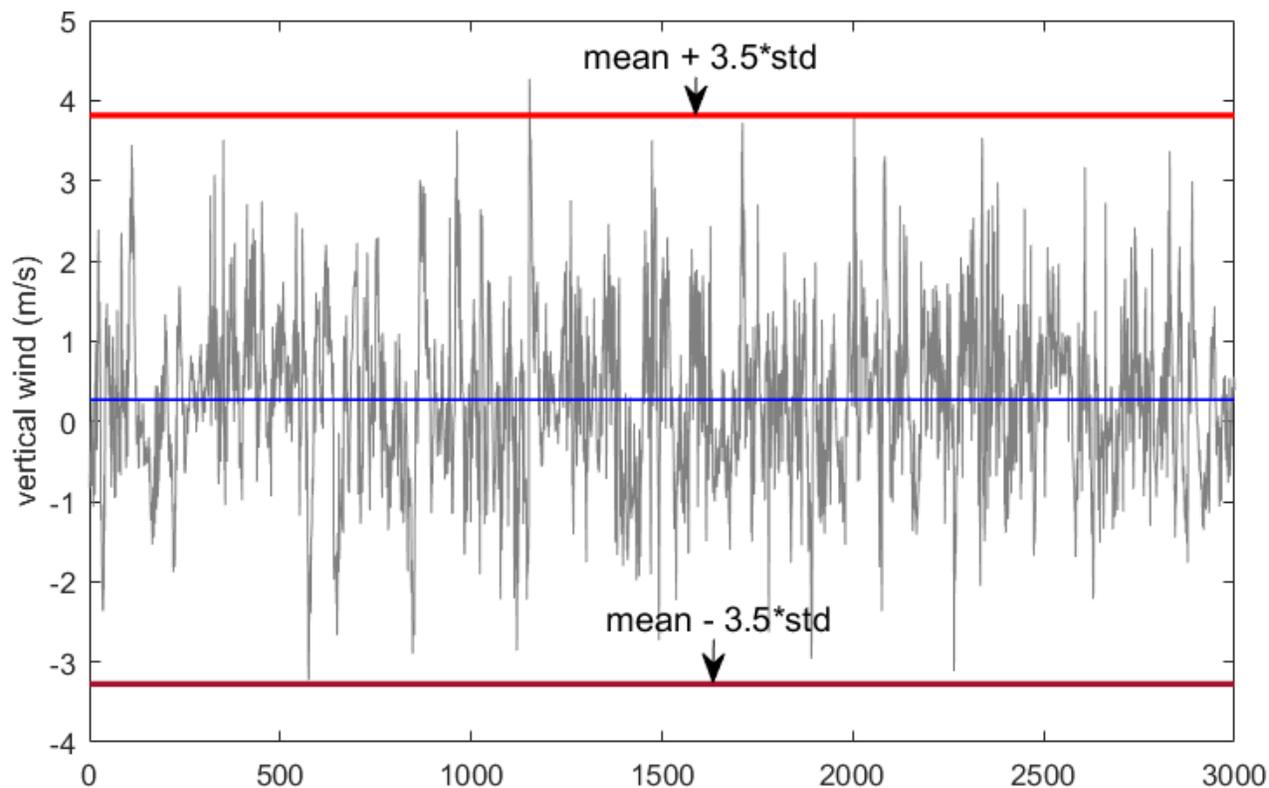
1.1 De-Spike

De-spike

对一移动窗口（通常为5 min），计算平均值 \bar{x} 和标准差 σ ，若 $|x_i - \bar{x}| \geq n\sigma$ ，则剔除 x_i

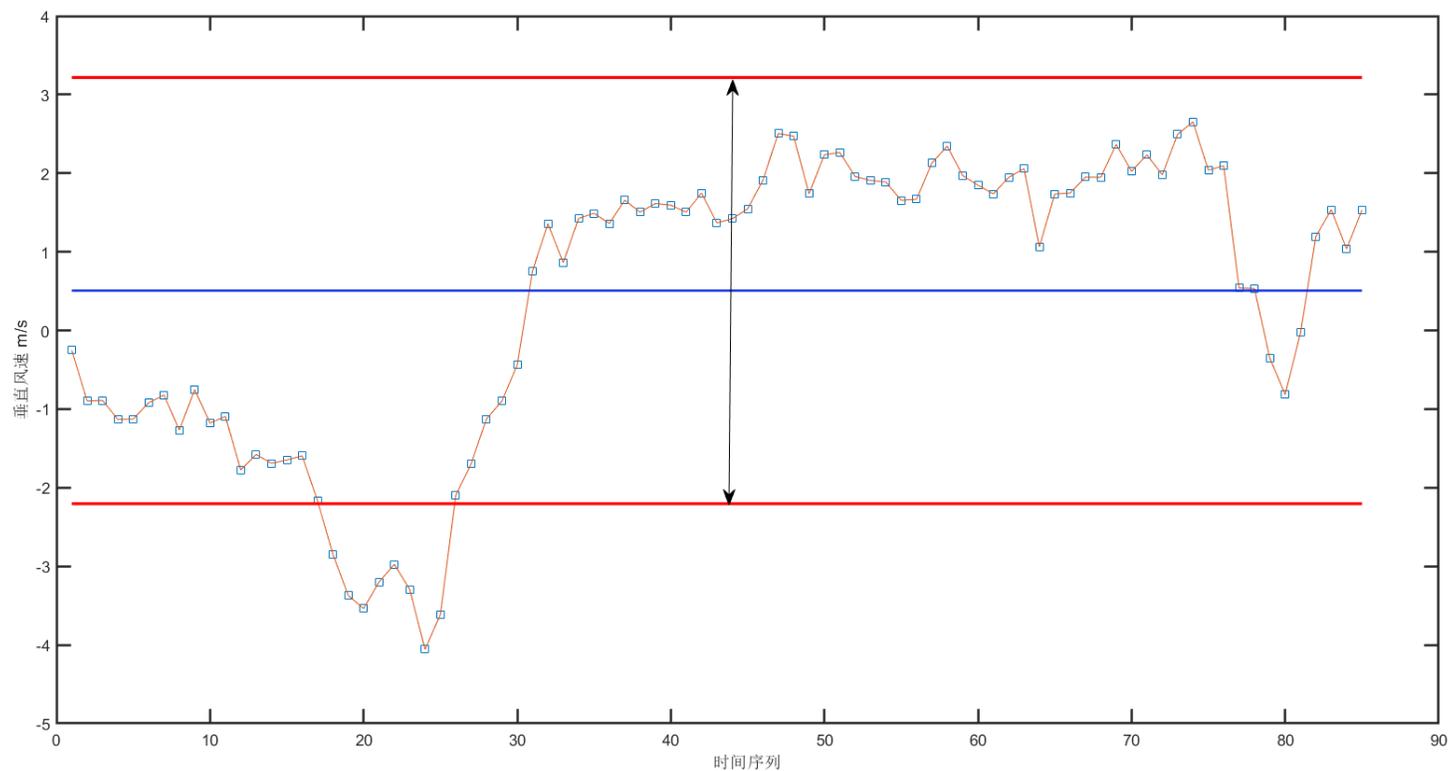
步骤如下：

1. 移动窗口通常为5min
2. n 通常为3.5，不同要素的 n 值可能不同
3. 剔除后的数据采用线性插补并计数
4. 完成一次处理后，需要重新迭代过滤，直至无野点存在为止
5. 野点数量不应超过1%



1.1 De-Spike

De-spike



当超出阈值的连续个数超过3时，通常认为是正常响应，不属于野点



1.2 系统诊断值

硬件诊断值

Sonic diagnostic flag 0

Gas diagnostic flag 0

信号强度

CO2 signal strength 0.7

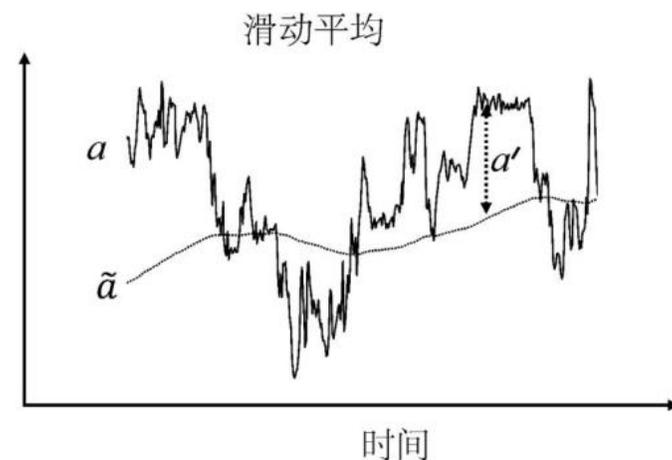
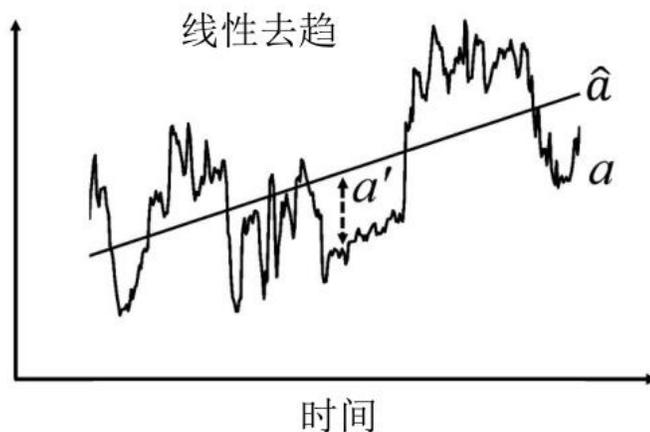
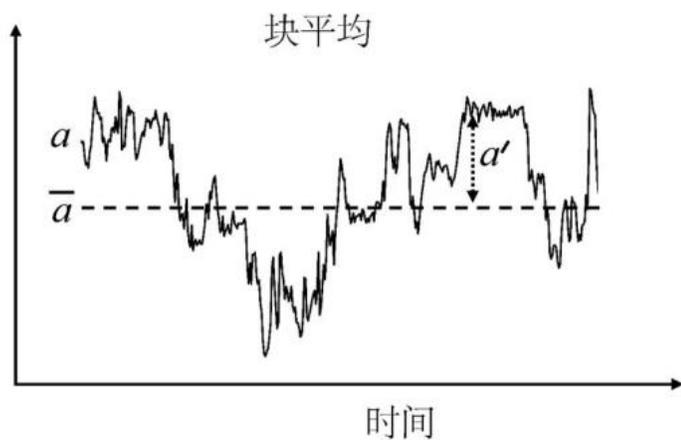
H2O signal strength 0.7

TOA5_11180_Time_Series_367_2020_11_12_0000.dat - Excel

1	TOA5	11180	CR6	11180	CR6.Std.05	CPU:Easyf	7276	Time_Series												
2	TIMESTAN	RECORD	Ux	Uy	Uz	T_SONIC	diag_sonic	CO2_dens	CO2_dens	H2O_dens	diag_irga	T_SONIC	TA_1.1.1	PA	CO2_sig_s	H2O_sig_strgth				
3	TS	RN	m s-1	m s-1	m s-1	deg C	adimensic	mg m-3	g m-3	adimensic	deg C	deg C	kPa	fraction	fraction					
4			Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp				
5	00:00.1	70337521	0.451387	-1.3885	0.029965	12.77218	0	268.7003	271.1707	8.937676	0	11.44922	9.744843	80.85374	0.684096	0.530777				
6	00:00.2	70337522	0.468052	-1.40105	0.023577	12.84877	0	268.3875	270.8808	8.958587	0	11.52649	9.744843	80.85541	0.684098	0.530774				
7	00:00.3	70337523	0.495217	-1.393	0.025156	12.84931	0	268.351	270.8442	8.956136	0	11.51785	9.744843	80.8531	0.684095	0.530784				
8	00:00.4	70337524	0.4724	-1.39738	0.021664	12.85648	0	268.4471	270.9426	8.977779	0	11.52219	9.744843	80.85766	0.684077	0.530767				
9	00:00.5	70337525	0.472141	-1.40394	0.012059	12.84489	0	268.5677	271.0613	8.954818	0	11.52267	9.744843	80.85484	0.684109	0.530794				
10	00:00.6	70337526	0.450438	-1.4075	0.011022	12.83741	0	268.3877	270.8793	8.915041	0	11.51361	9.744843	80.85095	0.684105	0.530803				
11	00:00.7	70337527	0.46345	-1.44968	0.039384	12.84092	0	268.4651	270.9572	8.940692	0	11.52664	9.744843	80.85542	0.684094	0.53079				
12	00:00.8	70337528	0.491126	-1.42594	0.031909	12.87994	0	268.3302	270.8339	8.959976	0	11.55649	9.744843	80.85466	0.684086	0.530774				
13	00:00.9	70337529	0.516179	-1.49178	0.044502	12.92411	0	268.0396	270.5562	8.962878	0	11.58783	9.744843	80.85588	0.684085	0.530784				
14	#####	70337530	0.516499	-1.46283	0.034446	12.94001	0	268.3079	270.8314	8.983394	0	11.61545	9.744843	80.85805	0.684104	0.530772				
15	00:01.1	70337531	0.509392	-1.44183	0.048322	12.91887	0	268.3876	270.9046	8.976799	0	11.5903	9.747406	80.85856	0.6841	0.53077				
16	00:01.2	70337532	0.494532	-1.42694	0.013421	12.92326	0	268.2302	270.7485	8.946343	0	11.60666	9.747406	80.8533	0.68408	0.530785				
17	00:01.3	70337533	0.522239	-1.44437	0.023022	12.93833	0	268.369	270.8933	8.958384	0	11.61948	9.747406	80.85466	0.6841	0.530779				
18	00:01.4	70337534	0.528063	-1.44598	0.023089	12.93063	0	267.9595	270.4769	8.967772	0	11.59296	9.747406	80.8561	0.684082	0.530778				
19	00:01.5	70337535	0.502047	-1.39704	0.003816	12.93996	0	268.1196	270.6425	8.95253	0	11.60776	9.747406	80.8549	0.684075	0.530785				
20	00:01.6	70337536	0.455285	-1.39399	0.025687	12.92064	0	268.2467	270.7631	8.968275	0	11.58932	9.747406	80.8588	0.684082	0.530763				
21	00:01.7	70337537	0.508342	-1.42821	0.016226	12.92154	0	268.2197	270.7368	8.955704	0	11.60577	9.747406	80.8573	0.684083	0.530776				
22	00:01.8	70337538	0.493317	-1.42187	0.072098	12.91921	0	268.3824	270.8991	8.975021	0	11.58502	9.747406	80.8562	0.684082	0.530761				
23	00:01.9	70337539	0.485488	-1.44076	0.02741	12.93196	0	268.2154	270.7361	8.955953	0	11.61023	9.747406	80.8574	0.684076	0.530755				
24	#####	70337540	0.458419	-1.41938	0.042274	12.8897	0	268.2188	270.7242	8.964753	0	11.56534	9.747406	80.8562	0.684083	0.530765				
25	00:02.1	70337541	0.475734	-1.42477	0.076665	12.88478	0	268.1348	270.6377	8.966874	0	11.54736	9.749084	80.8593	0.684069	0.530755				
26	00:02.2	70337542	0.489232	-1.44494	0.074258	12.92152	0	268.1098	270.6241	8.984787	0	11.58362	9.749084	80.8585	0.684059	0.530753				
27	00:02.3	70337543	0.4785	-1.41464	0.087049	12.92658	0	268.2477	270.7667	8.952171	0	11.60681	9.749084	80.8565	0.684061	0.530771				

1.3 趋势去除

自相关函数和互相关函数在数据中出现趋势的情况下不能反映真实情况，因此，在计算不确定性之前，需要对变量序列进行线性去趋势



野外观测时，块平均（时间平均）是单点测量唯一可行的平均方法
线性去趋和滑动平均不符合雷诺法则；
飞机上测湍流时需使用线性去趋

李旭辉，王伟等，2018



1.4 时间延迟校正

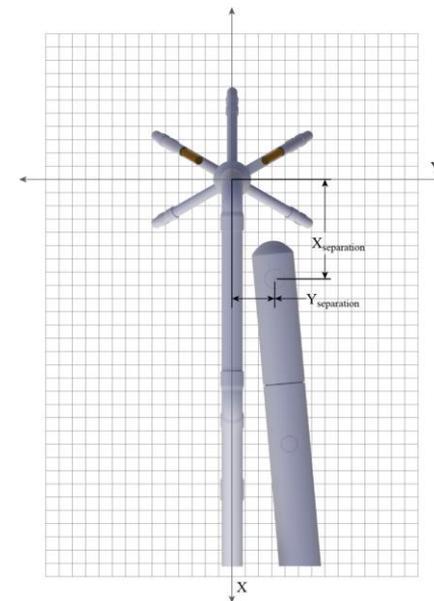
时间延迟的原因:

- OPEC: 传感器的空间分离, 一个气团从一个传感器到达另一个传感器需要一定时间

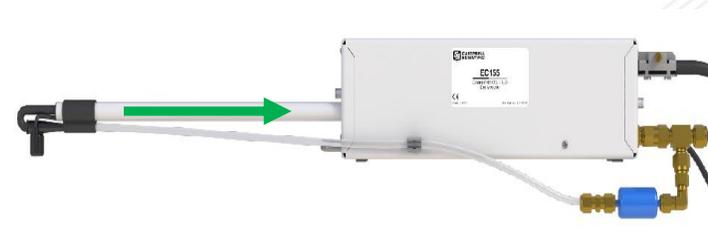
主风向上的距离: $d = x * \cos(\theta) + y * \sin(\theta)$

主风向上的时间延迟: $t = d/v$

- CPEC: 通过抽气的方式将气体送到分析室腔室, 导致气体分析仪与超声风速仪测量同一气团出现延迟。此外, 由于水汽分子的极性特点, 当温度, 相对湿度不同时, 其在管路内的停留时间 (residence time) 不同



OPEC

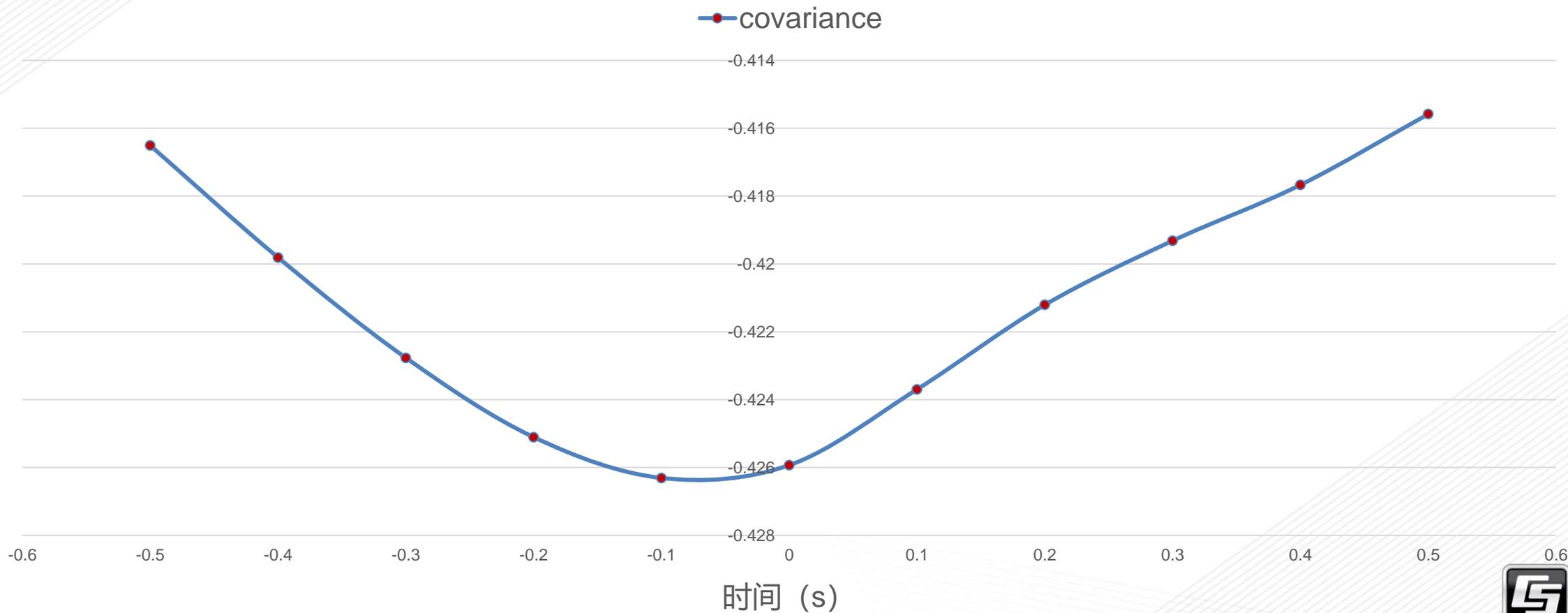


CPEC



1.4 时间延迟校正

计算通量时常用的方法是**协方差最大化**



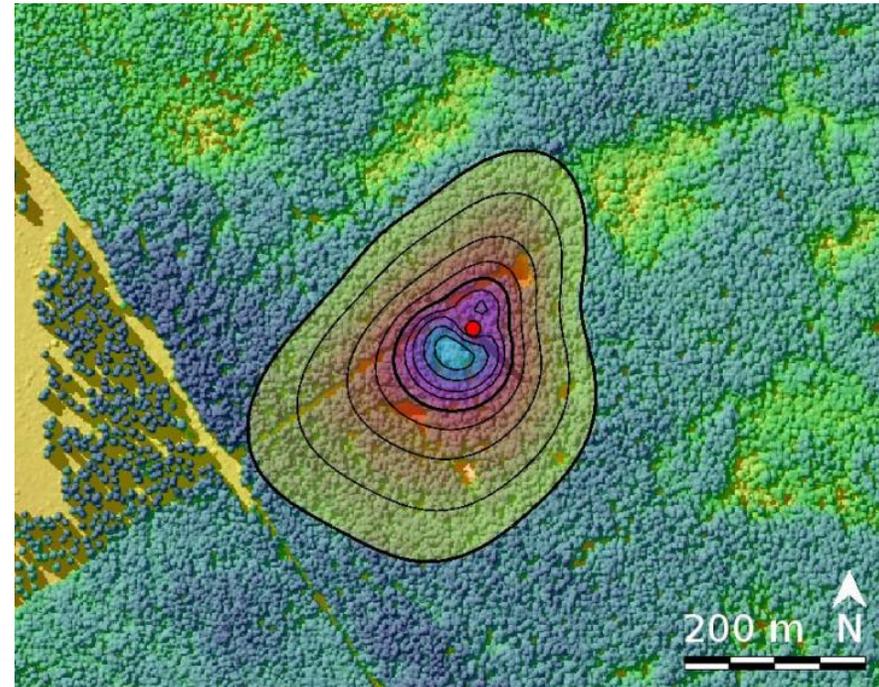
通量数据质量评价



2.1 源区分析

Göckede(2008)通过定义目标面积通量贡献的阈值来将源区面积内的通量源的异质性进行分类:

- **均一测量**: $\geq 95\%$ 的通量来自目标源区
- **代表性测量**: 80%~95%
- **可接受测量**: 50%~80%
- **受干扰测量**: $< 50\%$



2.2 总体质量标记

湍流平稳性检验 (Steady State Tests, SST)

指一个观测时次内主要统计量保持稳定。非平稳情况下，在30 min的观测时段内，一些主要的统计量会有变化趋势，或出现结构性变化。

检验方法参考Foken, 2004, 2012:

$$RN = \left| \left(\sum_{i=1}^6 CV_i / 6 - CV_{30} \right) / CV_{30} \right| * 100\%$$

根据长期实践得到的经验，**如果非平稳指数小于30%**，通常认为这个观测的时间序列是稳态的

RANGE(%)	0-15	16-30	31-50	51-75	76-100	101-250	251-500	501-1000	>1000
CLASS	1	2	3	4	5	6	7	8	9



2.2 总体质量标记

湍流充分发展的情况下，Monin-Obukhov相似理论成立。近地面层大气的许多归一化无量纲参数比如方差，能谱协谱等，只是稳定度 z/L 的函数，参考Foken, 2012，对“标准”的归一化方差与实际的计算值做比较，计算得到**总体湍流特征指数** (Integrated Turbulence Characteristics, ITC)

$$ITC = \left| \frac{(\sigma_{\chi}/\chi^*)_{model} - (\sigma_{\chi}/\chi^*)_{measurement}}{(\sigma_{\chi}/\chi^*)_{model}} \right| \times 100\%$$

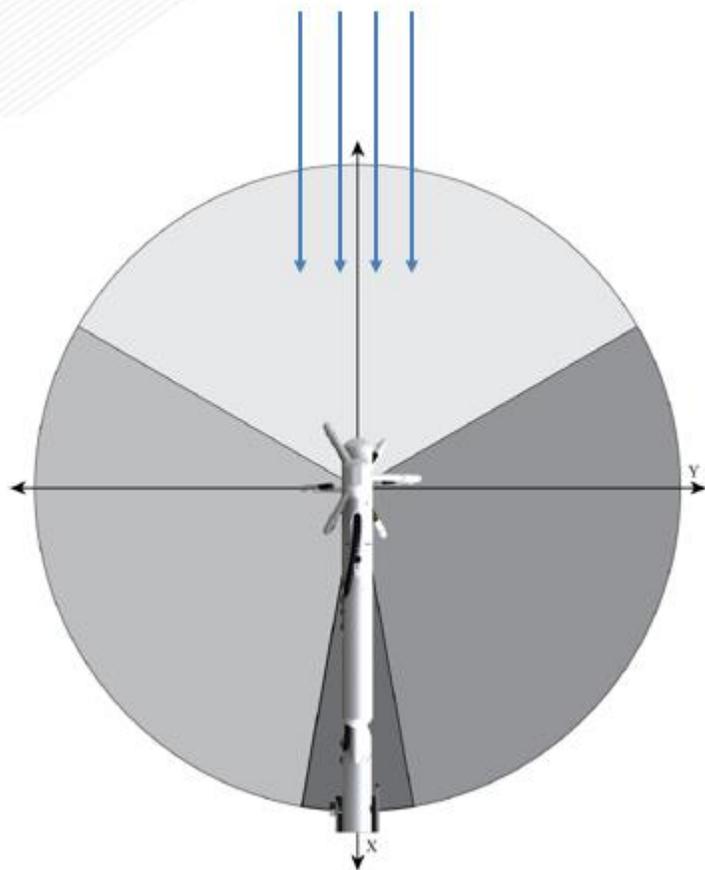
根据ITC的计算结果，可以划分为9个等级

RANGE(%)	0~15	16~30	31~50	51~75	76~100	101~250	251~500	501~1000	>1000
CLASS	1	2	3	4	5	6	7	8	9

如果总体湍流特征指数ITC**小于30%**，通常认为这个观测时段内的湍流充分混合



2.2 总体质量标记



Horizontal orientation
of the sonic anemometer

Class	Range
1	$\pm 0-30^\circ$
2	$\pm 31-60^\circ$
3	$\pm 60-100^\circ$
4	$\pm 101-150^\circ$
5	$\pm 101-150^\circ$
6	$\pm 151-170^\circ$
7	$\pm 151-170^\circ$
8	$\pm 151-170^\circ$
9	$> \pm 171^\circ$

在安装设备之前，需要确定当地的主风向，并将设备布置在目标区域的下风向



2.2 总体质量标记

综合湍流平稳性检验，湍流发展充分性检验以及风向方向，可以得到总体质量标记,可划分为1~9
九个等级

Table 4.5 Proposal for the combination of the single quality flags into a flag of the general data quality (Foken et al. 2004)

Flag of the general data quality	Steady state test according to Eq. 4.38	Integral turbulence characteristics according to Eq. 4.41	Horizontal orientation of the sonic anemometer
1	1	1-2	1-5
2	2	1-2	1-5
3	1-2	3-4	1-5
4	3-4	1-2	1-5
5	1-4	3-5	1-5
6	5	≤ 5	1-5
7	≤ 6	≤ 6	≤ 8
8	≤ 8	≤ 8	≤ 8
9	≤ 8	6-8	≤ 8
9	one flag equal to 9		



2.2 总体质量标记

- 或者采用简化的等级划分，可划分为0, 1, 2三个等级

湍流平稳性	湍流发展性	总体质量级
(1-2) 或 (<30%)	(1-2) 或 (<30%)	0
(≤5) 或 (<100%)	(≤5) 或 (<100%)	1
(≤6) 或 (>100%)	(≤6) 或 (>100%)	2

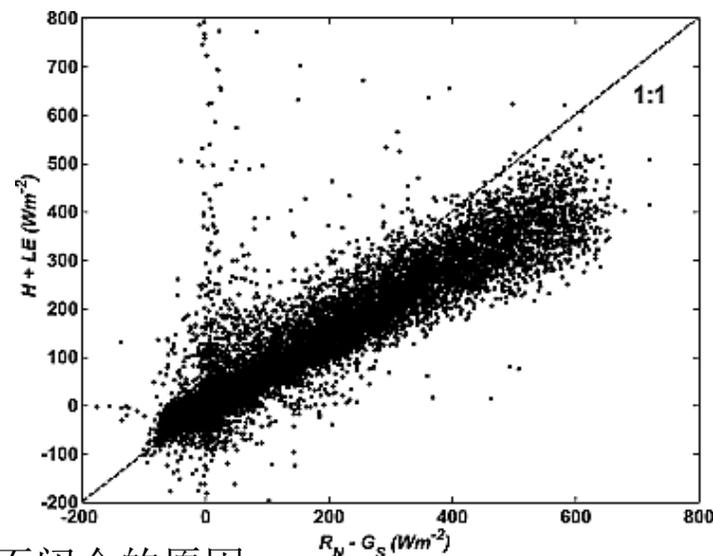
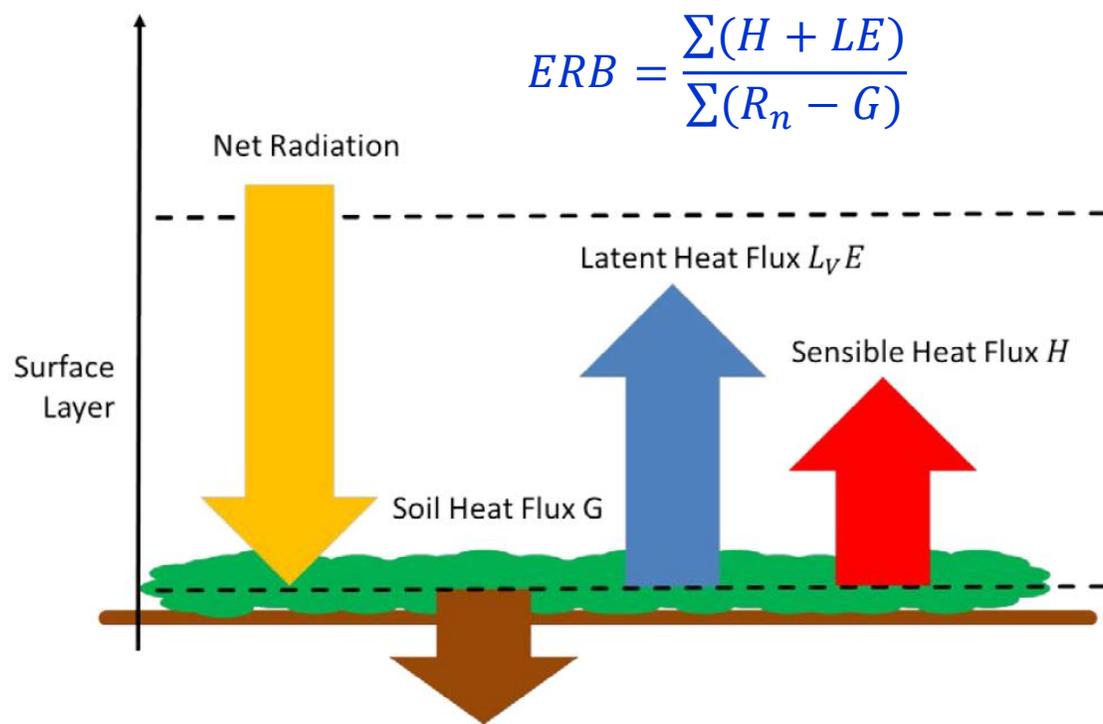
无论是哪种等级划分方法，质量等级数字越小，表示数据质量越高，反之则越差。

- ◆ 质量较好的数据，通常用于科学研究分析或论文发表 (1~3)
- ◆ 质量一般的数据，可用于连续通量观测实验 (4~6)
- ◆ 质量较差的数据，只能作为定性分析参考，但通常会好于插补数据 (7~8)
- ◆ 质量较差，建议参考或舍弃



2.3 能量平衡

能量平衡闭合被认为是检验通量数据的重要标准之一，FLUXNET许多站点都把能量平衡闭合状况分析作为一种标准的程序用于通量数据的质量评价。



能量不闭合的原因：

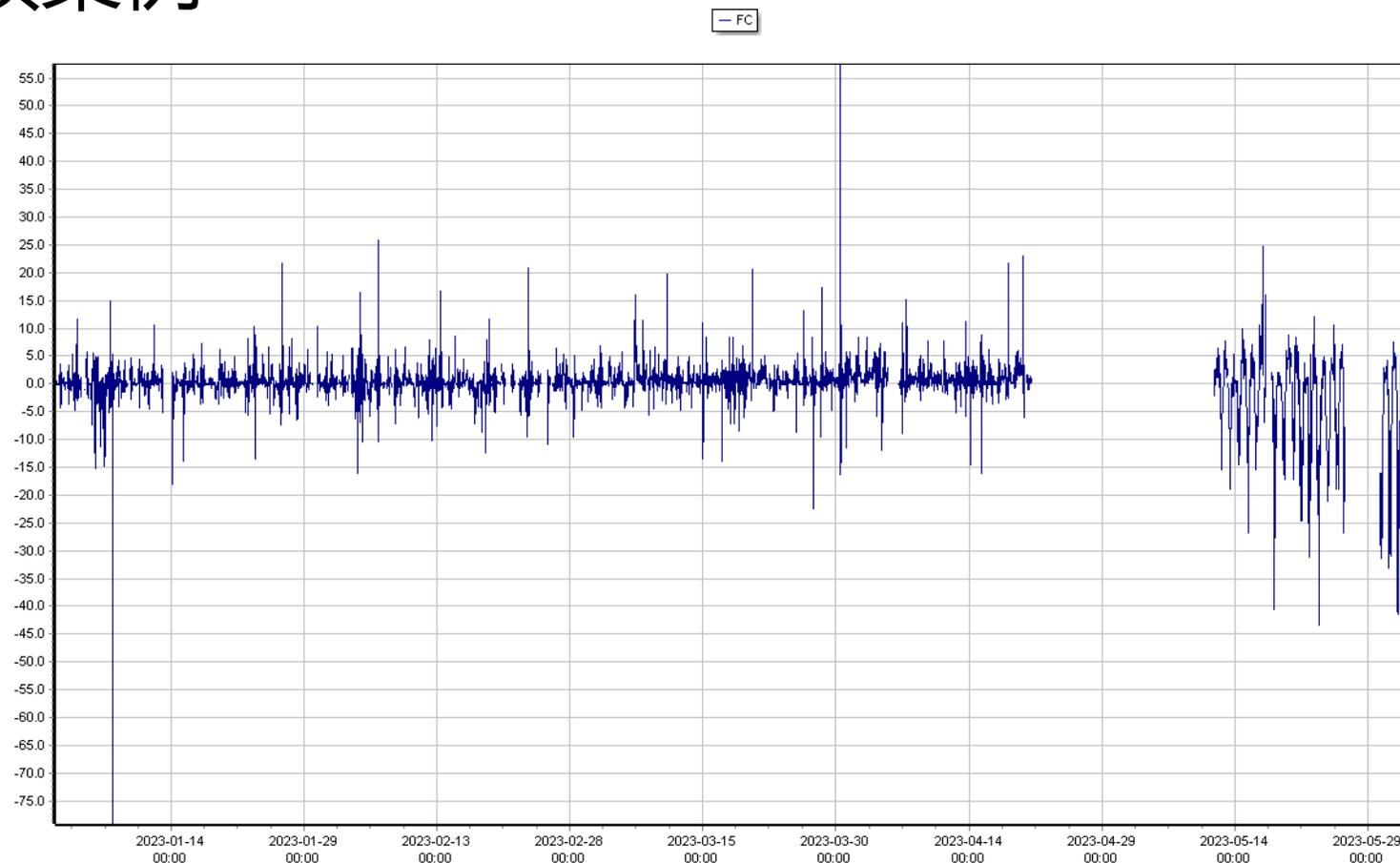
- 仪器的准确性
- 边界层弱湍流或者大气层结稳定
- 储藏项的遗漏
- 平流项的遗漏（复杂地形，非均一植被）
- 测量源区或者目标区的不匹配



资料质量问题实例

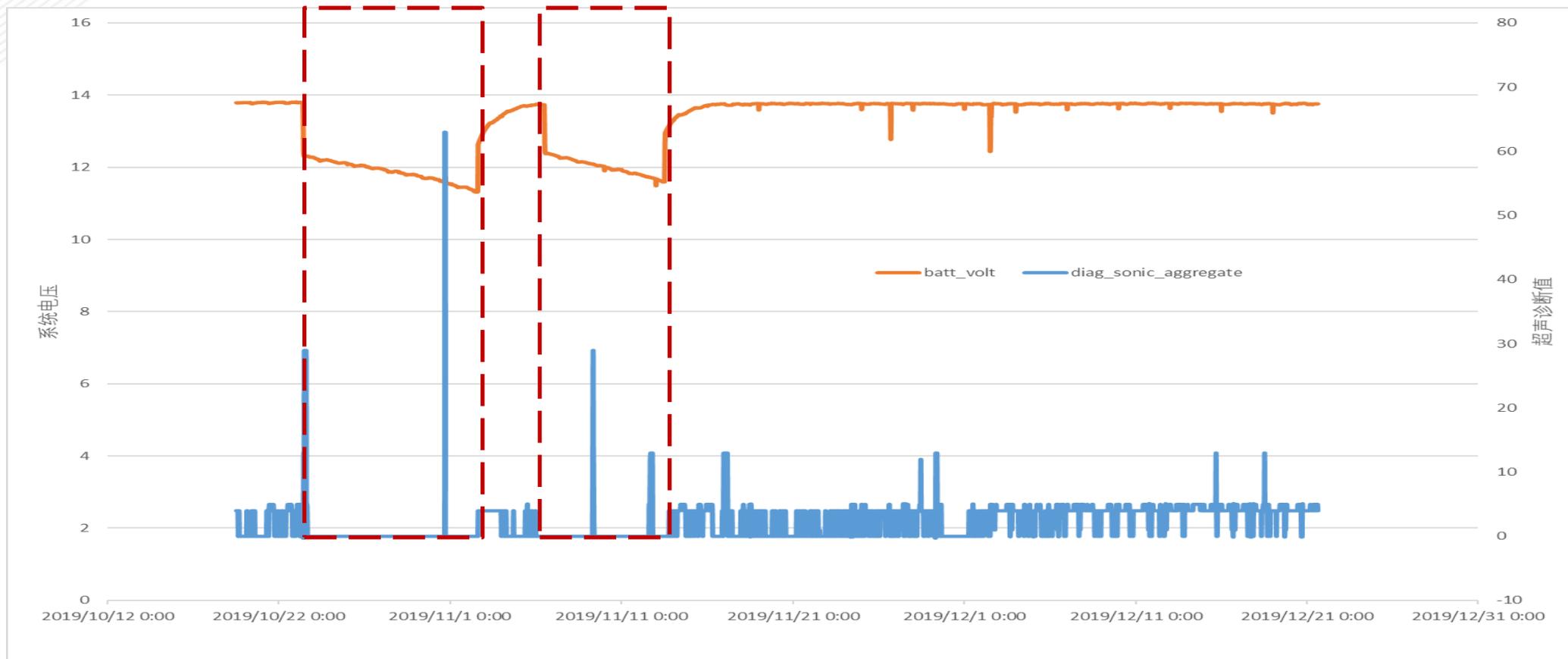
仪器硬件故障

数据不连续案例



系统接地故障

三维超声风速仪诊断值报错



涡动系统供电方式：交流电+蓄电池

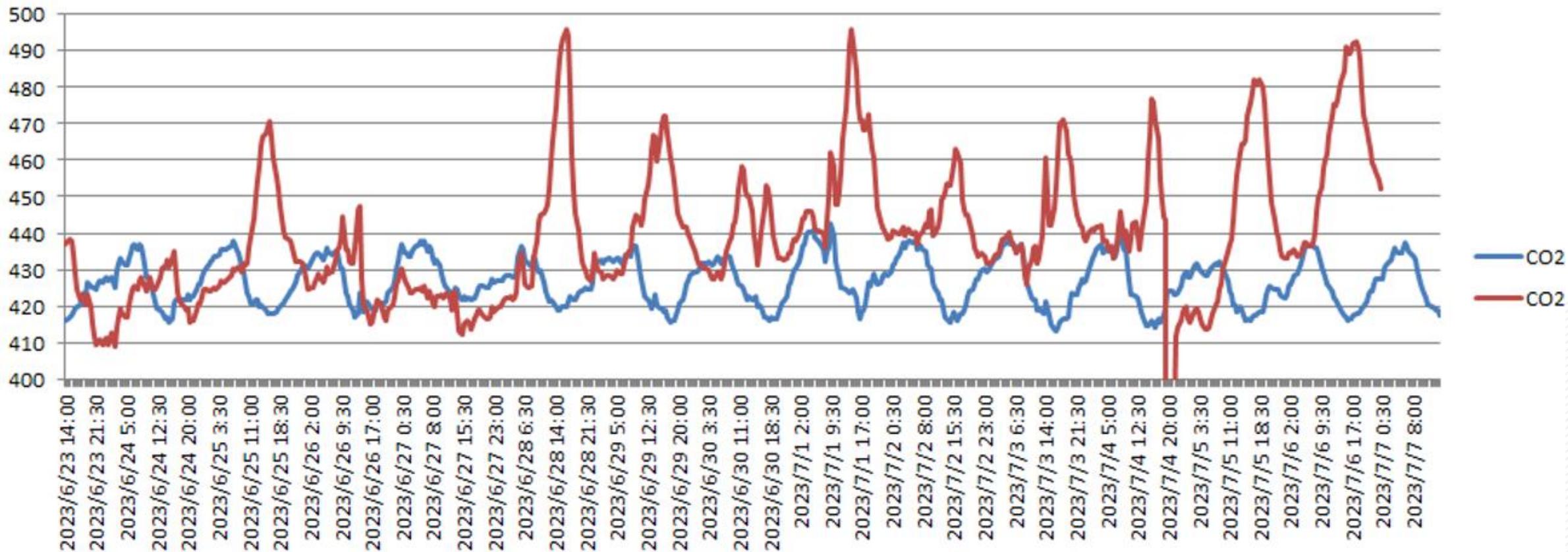


仪器老化或者标定漂移

CO2浓度平均值偏低



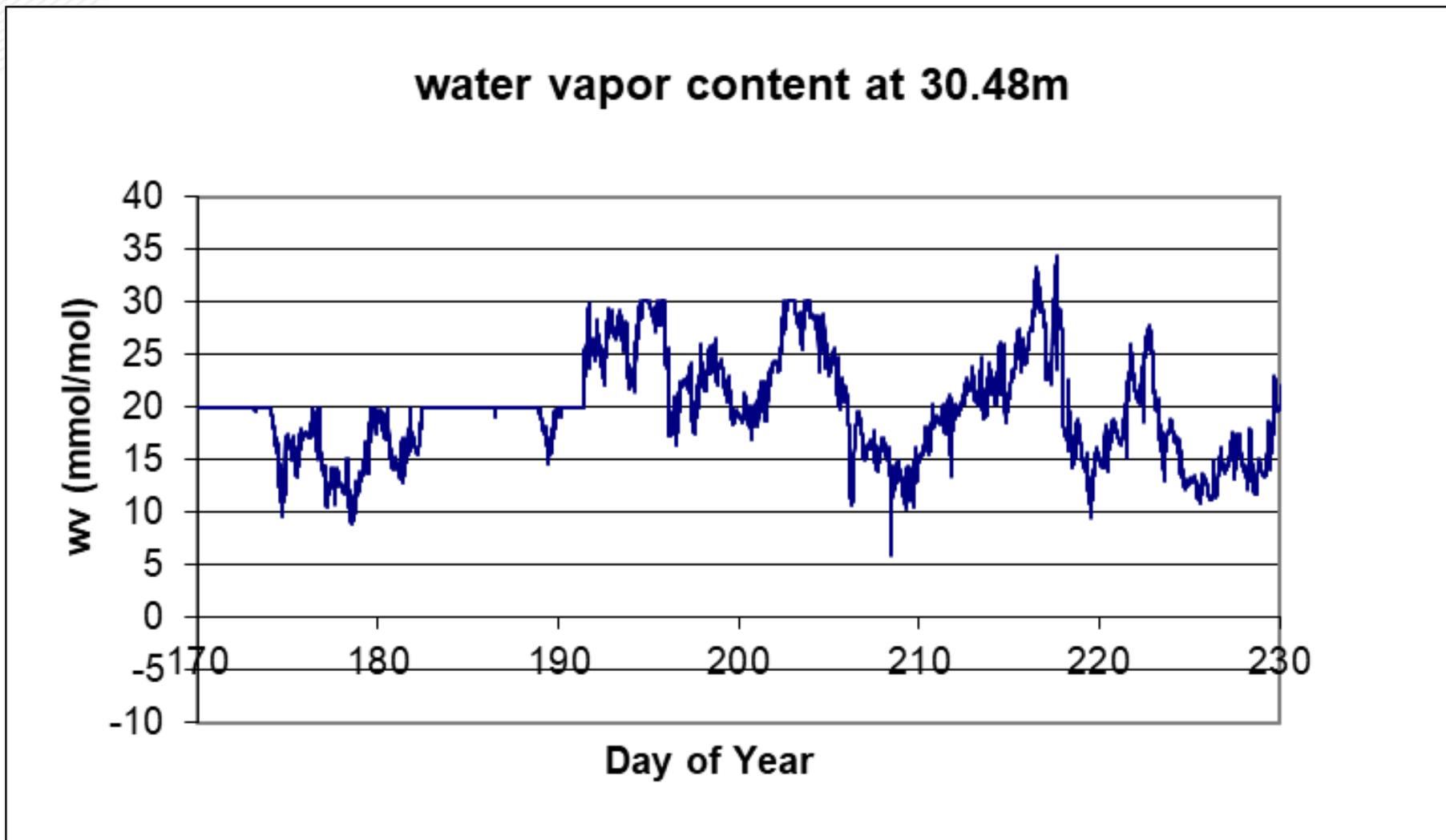
错误的日变化



腔室分子筛维护案例



传感器参数设置错误



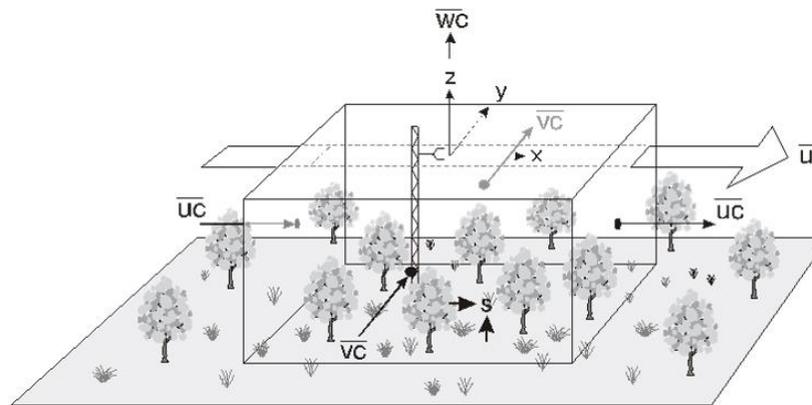
提高数据质量的方法

提高数据质量的方法 (1)

观测开始前，选择合适的观测地点，尽可能满足涡动相关法测量的基本要求（比如下垫面水平，植被种类较均一）

观测区域下垫面水平均匀：

- 水平梯度可以忽略，水平积分过程就不再需要
- 垂直风速平均值 $\bar{w} = 0$
- 储存项可以忽略

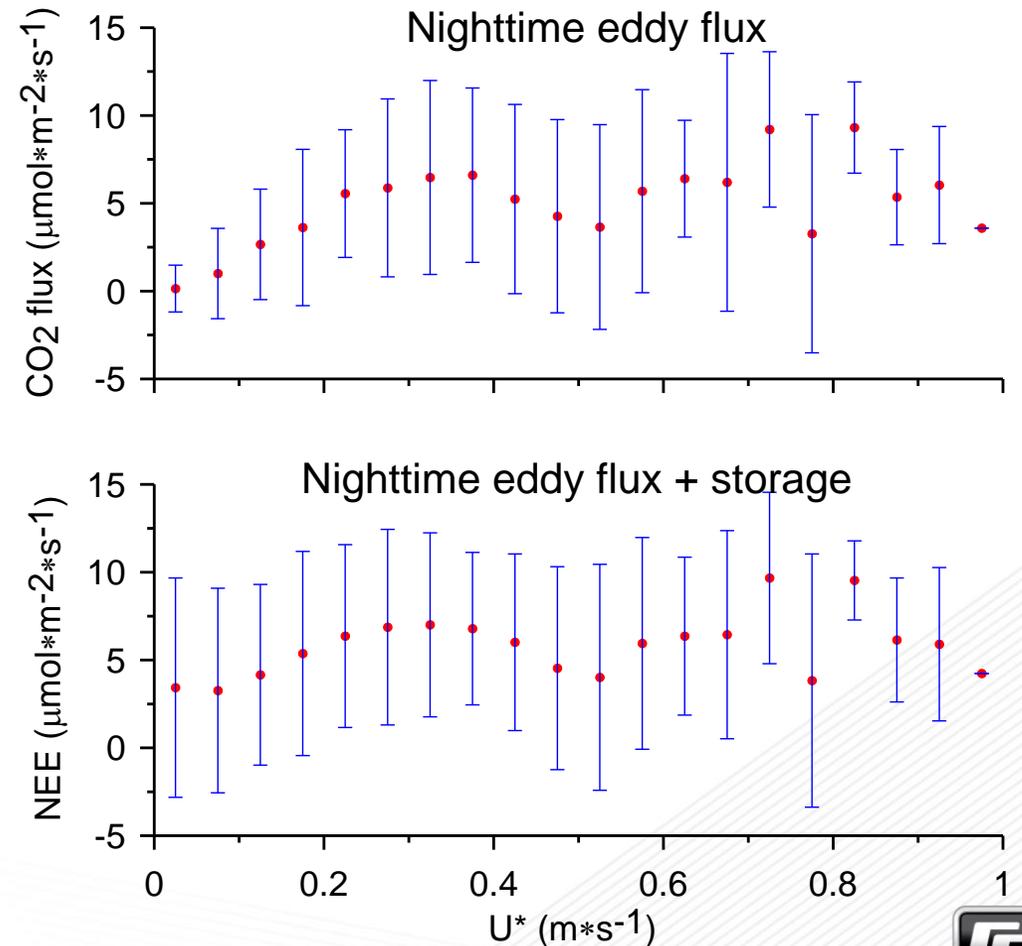


$$\int_{-L}^L \int_{-L}^L \int_0^{h_m} \overline{S_s} dz dx dy = \int_{-L}^L \int_{-L}^L \int_0^{h_m} \left[\cancel{\rho_d \frac{\partial \overline{\chi_s}}{\partial t}} + \cancel{\rho_d \bar{u} \frac{\partial \overline{\chi_s}}{\partial x}} + \cancel{\rho_d \bar{v} \frac{\partial \overline{\chi_s}}{\partial y}} + \cancel{\rho_d \bar{w} \frac{\partial \overline{\chi_s}}{\partial z}} + \cancel{\rho_d \frac{\partial \overline{u' \chi_s'}}{\partial x}} + \cancel{\rho_d \frac{\partial \overline{v' \chi_s'}}{\partial y}} + \rho_d \frac{\partial \overline{w' \chi_s'}}{\partial z} \right] dz dx dy$$



提高数据质量的方法 (2)

当涡动相关系统观测高度大于10m时，由于逆温层存在或植被冠层的影响，通常建议使用CO₂廓线系统，用于监测观测高度下方的碳储存项。如果没有CO₂廓线系统，至少应该计算观测高度上的碳储存项

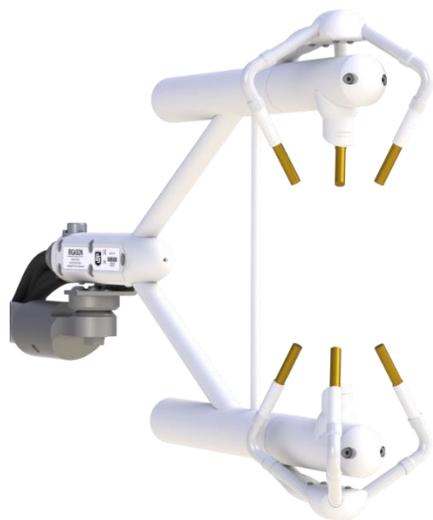


Growing season: 06/16/2004—10/31/2004;
bin size for U^* is 0.05 $\text{m}\cdot\text{s}^{-1}$.



提高数据质量的方法 (3)

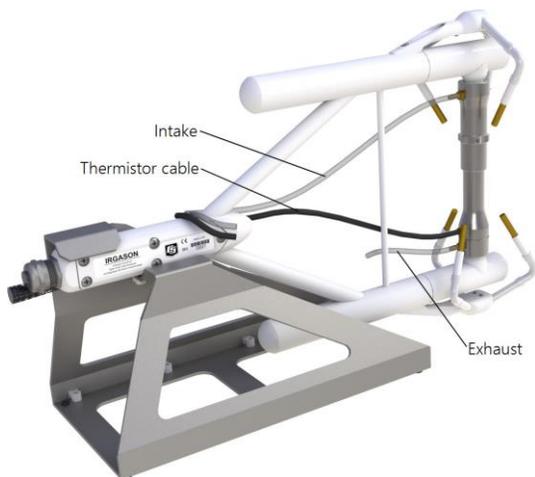
大多数环境条件及场景，开路和闭路涡动系统都可以获得准确的通量结果。实际应用时，根据应用环境特点和维护的难易程度**选择选择合适的测量设备**。一般情况下，恶劣天气条件（高湿或风沙），闭路涡动系统的数据连续性和可靠性优于开路



提高数据质量的方法 (4)

制定设备的维护计划:

- 定期查看数据, 及时发现问题 (使用辅助软件监视站点运行情况)
- 保证气体分析仪的信号强度 (推荐信号强度在0.9以上)
- 定期校准设备 (建议每月校准一次)



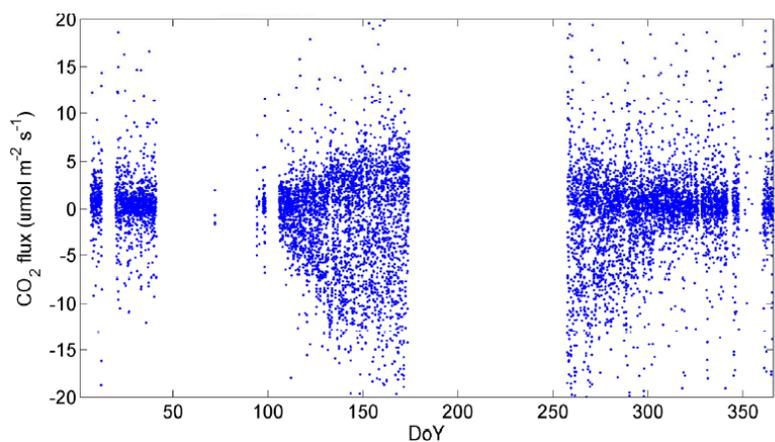
站点: CPEC N40					
标题	状态	值	阈值	信息	勾选 动作
No Data	✔				<input type="checkbox"/> ⚙
Watchdog Errors	⚠	6	value > 0	Watch dog errors (6) have been detected	<input type="checkbox"/> ⚙
Card Storage Running Low	✖	-5.463903	value < 7	There are less than seven days of card storage on the datalogger.	<input type="checkbox"/> ⚙
FC OK	✔	-1.688884e-06	value = NAN		<input type="checkbox"/> ⚙
LE OK	✔	-2.343969e-06	value = NAN		<input type="checkbox"/> ⚙
H OK	✔	-3.004075e-06	value = NAN		<input type="checkbox"/> ⚙
TAU OK	✔	2.233485e-08	value = NAN		<input type="checkbox"/> ⚙
USTAR OK	✔	0.07817449	value = NAN		<input type="checkbox"/> ⚙
TA_1_1_1 OK	✔	0.6096894	value = NAN		<input type="checkbox"/> ⚙
U OK	✔	0.7488447	value = NAN		<input type="checkbox"/> ⚙
V OK	✔	0	value = NAN		<input type="checkbox"/> ⚙
W OK	✔	3.72529e-09	value = NAN		<input type="checkbox"/> ⚙
T_Sonic OK	✔	1.026994	value = NAN		<input type="checkbox"/> ⚙
CO ₂ OK	✔	439.0398	value = NAN		<input type="checkbox"/> ⚙
H ₂ O OK	✔	4.804107	value = NAN		<input type="checkbox"/> ⚙



提高数据质量的方法 (5)

Fill the gaps

- 硬件故障
- 环境因素：风沙、降雨或结露等
- 供电限制：长时间雨雪天气
- 质量控制：质量等级为9的数据通常需要剔除
- ...



常见的数据插补方法：

- 平均昼夜变化法 (MDV)
- 查表法 (LUT)
- 边际分布采样法 (MDS)
- 人工神经网络法 (ANN)
- 非线性回归法 (NLR)

$$NEE_{day} = \frac{\alpha \times PAR \times P_{max}}{\alpha \times PAR + P_{max}} - RE_{daytime}$$

Michaelis-Menten 光响应方程



$$RE_{night} = A \times e^{(BT)}$$

Lloyd & Taylor 呼吸方程



提高数据质量的方法 (6)

▶ 及时维修仪器故障



▶ 备份观测仪器



Campbell Scientific 在北京提供工厂级别售后服务, 有效降低设备维修周期



