



**Campbell**

**S C I E N T I F I C ®**

# ChinaFLUX第十八次通量观测理论与技术培训

(沈阳, 2023年8月17-18日)



---

## 通量数据和环境要素数据的补缺

杨柏 (byang@campbellsci.com)

Campbell Scientific Inc

# 导致数据缺失的常见原因

- 站点运行维护，仪器校正，供电异常，仪器故障等
- 天气原因，比如雨滴或者雾滴覆盖了开路气体分析仪的红外辐射窗口，沙尘或者冰雪蒙蔽了仪器的探头；自然灾害原因，比如林地草地火灾或者雷击；其它环境原因，比如鸟类栖息在仪器周围，昆虫在仪器探头之间穿行，各种电子信号噪音对仪器的干扰等
- 工作人员或者操作员的失误
- 低质量资料，比如由于弱湍流，大气边界层稳定层结或者逆温层而引起的；资料代表性差，比如由于观测点的不足的Fetch或者是没有足够大的源（汇）区（Footprint）

## 什么情况下需要填补资料空缺？

- 计算全年或者一个生长季的碳吸收总量 (NEE, GPP)
- 使用一些统计分析方法(如小波相干分析)时, 要求连续的时间序列 (无空缺)
- 模型的验证和校准, 模型 - 资料耦合研究

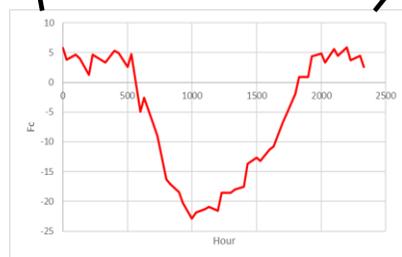
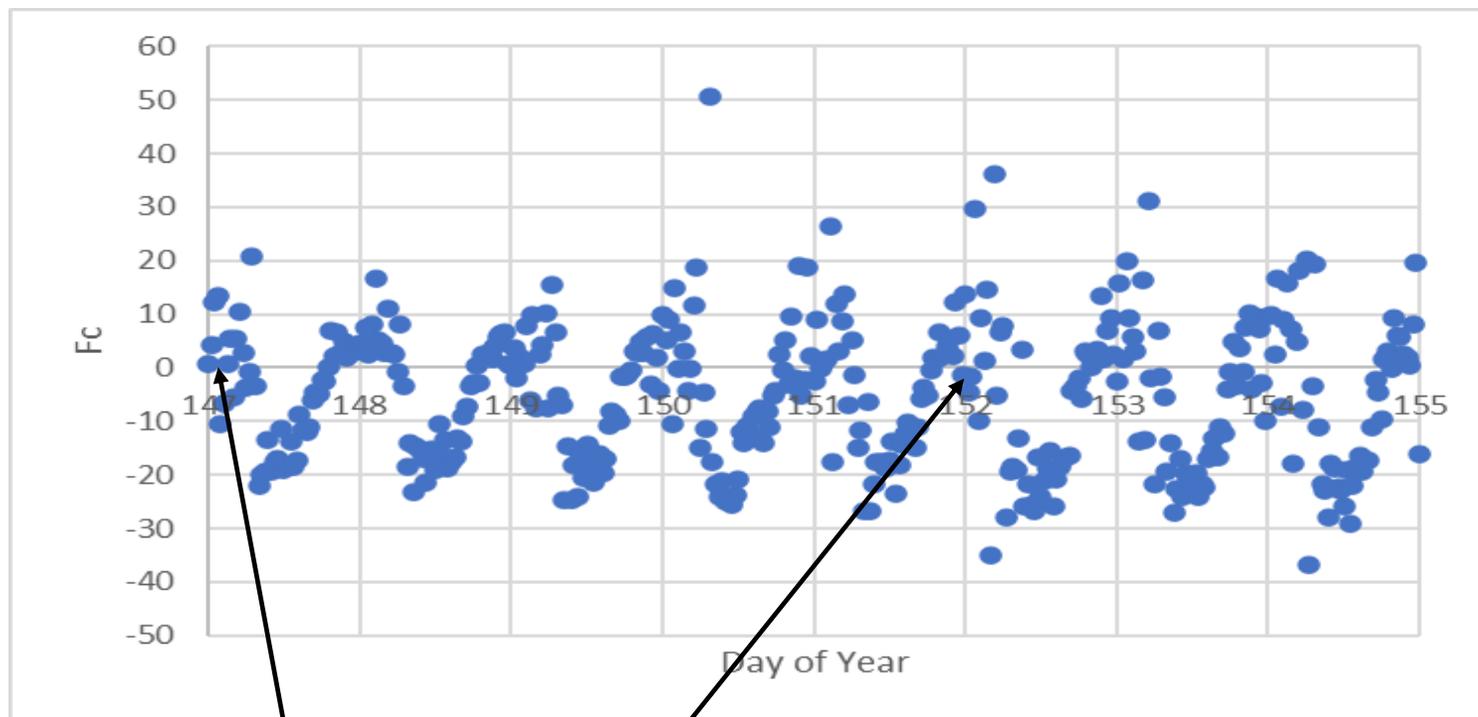
## 什么情况下不要填补资料空缺？

- 研究通量和环境因子的控制关系

# 资料补缺的常用方法

- 统计方法：线性回归（内插），非线性回归（多项式函数，指数或者对数函数，Spline Fitting等），日变化平均值(mean diurnal variation; MDV)，查表法(look-up table; LUT), Marginal Distribution Sampling (MDS)，等
- 机理模型：光响应曲线，土壤呼吸（Q10）方程
- 数学模型：机器深度学习(machine learning), 人工神经网络(artificial neural network)

# 日变化平均值(MDV)



- 在一个移动窗口里 (5 - 10天) 获取日变化平均曲线, 用平均值去填补在窗口内各日同一时间的空缺
- 在逐日变化较大时 (比如生长季初期或者末期), 使用较小的移动窗口; 在逐日变化较小时 (比如生长季中期), 使用较大的移动窗口
- 补缺过程不需要使用环境气象资料, 或者说假设窗口内每日同一时间的环境气象是一致的

# 查表法(LUT)

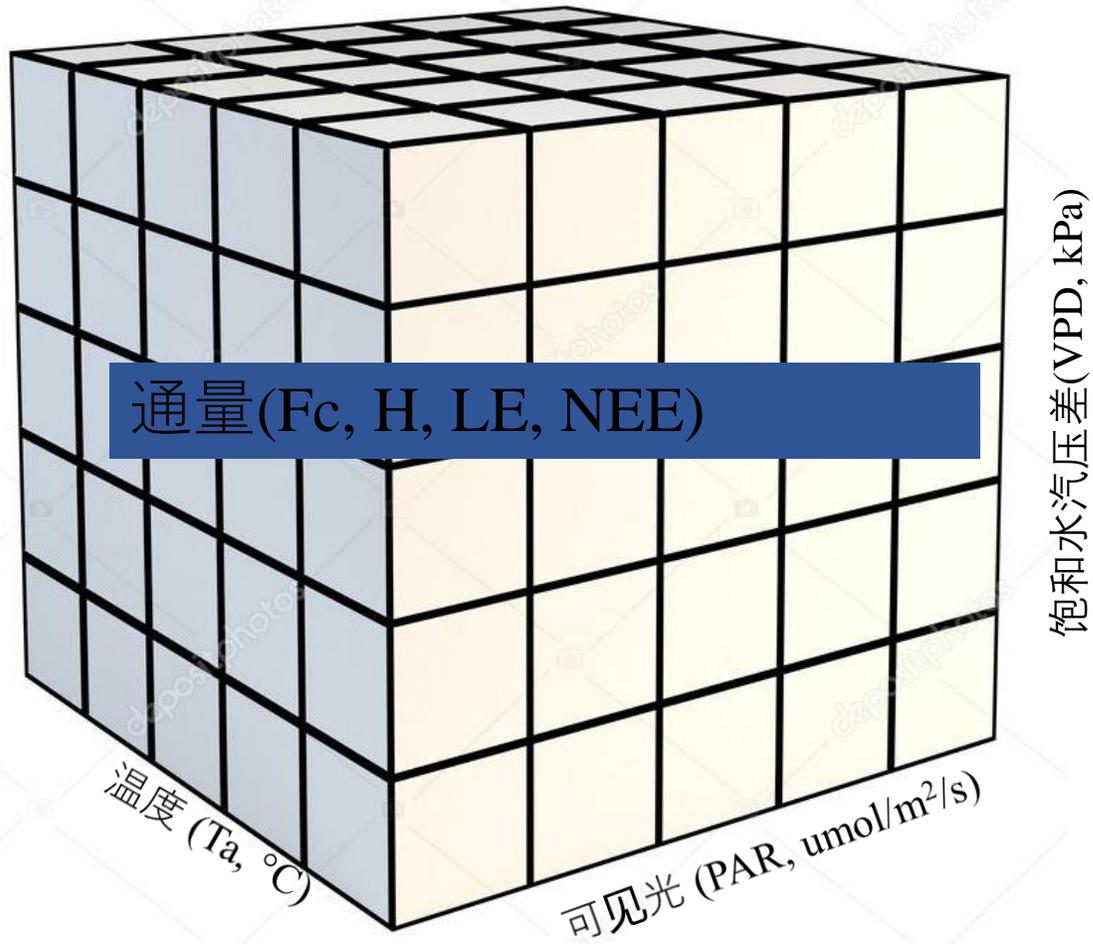
温度 ( $T_a$ , °C)

可见光 (PAR, $\mu\text{mol}/\text{m}^2/\text{s}$ )	温度 ( $T_a$ , °C)								
	0-5	5-10	10-15	15-20	20-25	...	...	...	...
0-250									
250-500									
500-750									
750-1000									
...									
...									
...									

通量值( $F_c$ ,  $H$ ,  $LE$ ,  $NEE$ )

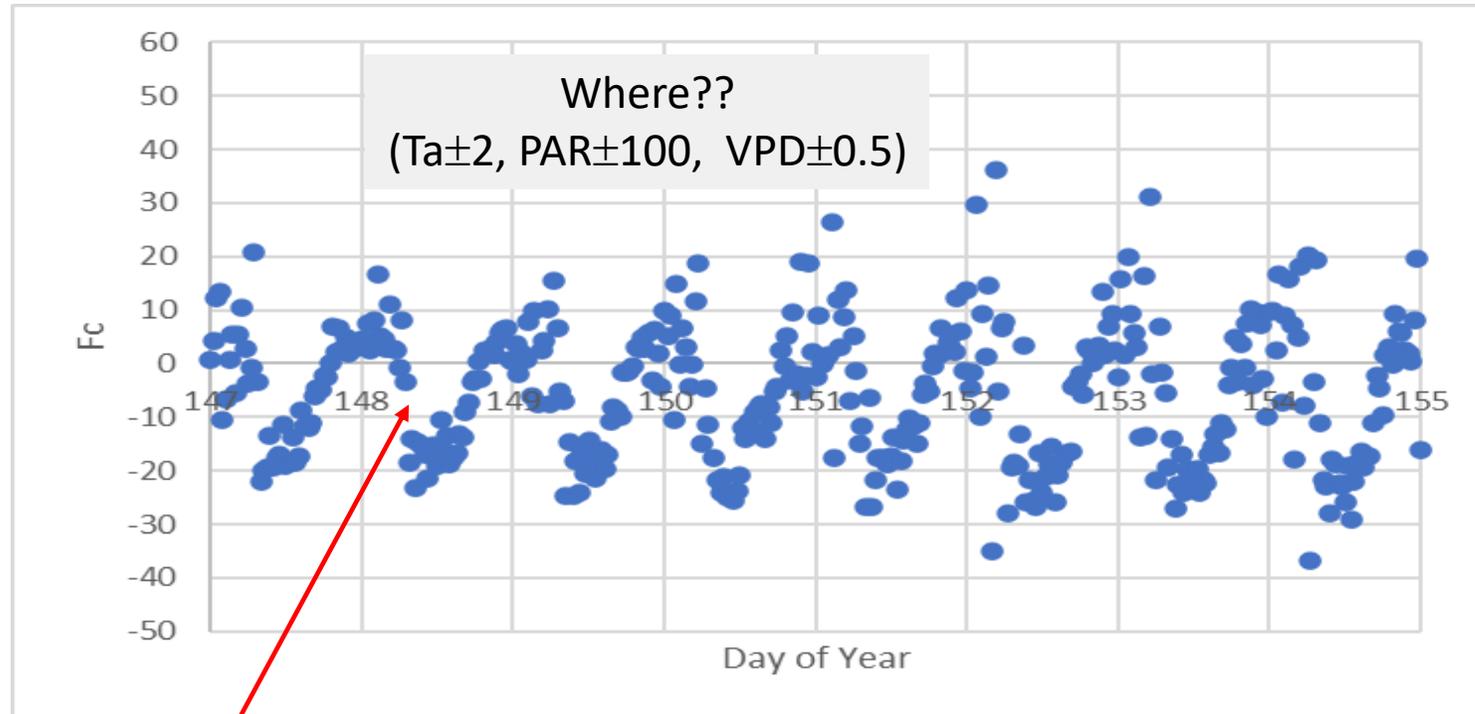
- 可以在一个月的时间段内使用（相对较准确，但未必有足够的观测资料建表），也可以在一个生长季节内使用（有足够的资料建表，但准确性相对较差）
- 依赖于环境气象资料，需要完整的气象资料或者先补缺气象资料

# 查表法(LUT)



- 比起二维表格，三维表格方法更准确（考虑了更多的影响因子），但建表需要数据量大，过程略复杂

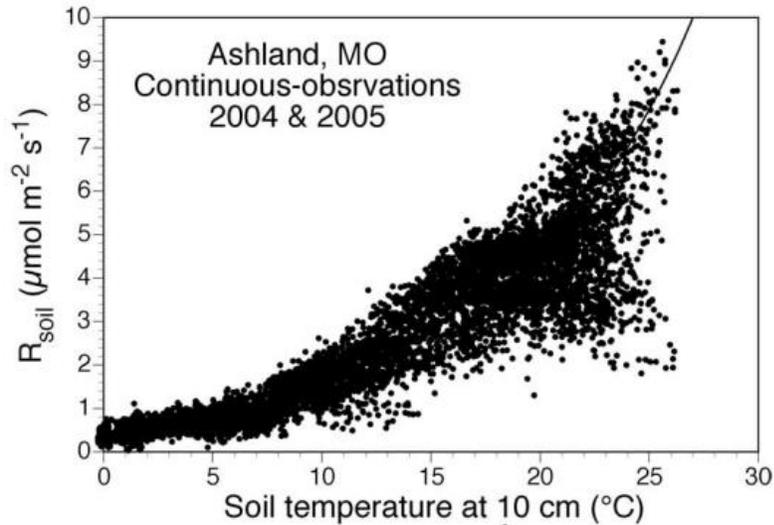
# Marginal distribution sampling (MDS)



(Ta, PAR, VPD)

- 介于日变化平均值方法和查表法之间，或者说是二者的混合体
- 利用了查表法的原理，但不用建立一个完整的表，只在一个较小的移动窗口里（5-10天）应用
- 类似于MDV方法，因为相邻日的同一时间环境气象条件非常接近

# 生态系统呼吸简单模型（夜间碳通量补缺方法）



$$R_{eco} = a * e^{bT}$$

$R_{eco}$  – 生态系统呼吸量  
 $T$  – 空气或者土壤温度

$$R_{eco} = a * e^{bT}$$

使用  $R_{eco}$  at 10 °C 作为参考值,  $R_{ref} = R_{T=10} = a * e^{10b}$

定义  $Q_{10} = e^{10b}$

$$R_{eco} = R_{ref} * Q_{10}^{\frac{T-10}{10}}$$

$T=10, R_{eco} = R_{ref}$ ;  $T=20, R_{eco} = R_{ref} * Q_{10}$ ;  $T=30, R_{eco} = R_{ref} * (Q_{10})^2$  .....

温度每增加10度,  $R_{eco}$  增加  $Q_{10}$  倍

## 生态系统呼吸模型之二

$$R_{\text{eco}}(T) = R_{\text{Ref}} \exp \left[ E_0 \left( \frac{1}{T_{\text{Ref}} - T_0} - \frac{1}{T - T_0} \right) \right],$$

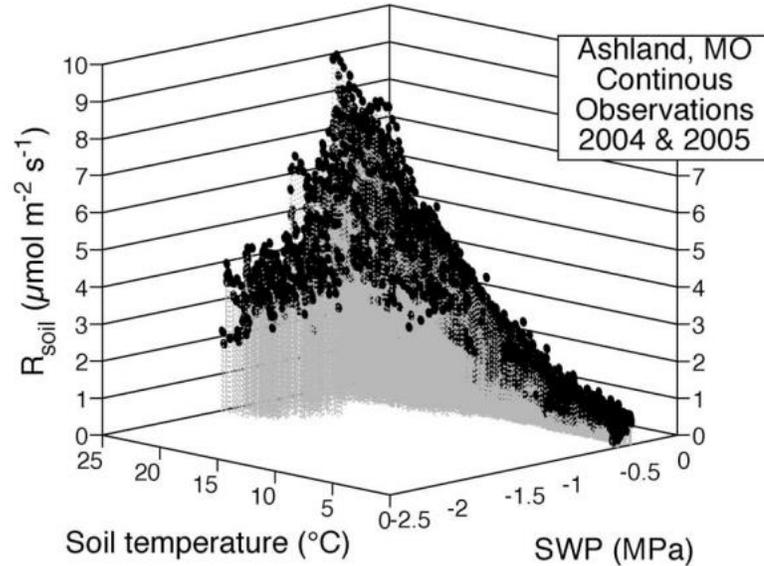
$R_{\text{Ref}}$  --- ecosystem respiration at reference  
temperature  $T_{\text{Ref}}$

$E_0$  --- temperature sensitivity

$T_0$  --- constant

Lloyd and Taylor (1994)

# 生态系统呼吸模型之三



$$R_{eco} = R_{ref} * \left( \frac{\psi_{max} - \psi_{obs}}{\psi_{max}} \right) * Q_{10}^{\frac{T-10}{10}}$$

$Q_{10}$  --- the amount of change in  $R_{eco}$  for every 10  $^{\circ}\text{C}$  change in temperature

T --- air or soil temperature

$R_{ref}$  --- ecosystem respiration at reference temperature, 10  $^{\circ}\text{C}$

$\psi_{max}$  --- maximum soil water potential (MPa) at the complete inhibition of  $R_{eco}$

$\psi_{obs}$  --- observed soil water potential (MPa) in the top soil layers

# GPP模型（白天NEE补缺方法）

Rectangular Hyperbolic light-response curve

$$GPP = \frac{\alpha * \beta * PAR}{\alpha * PAR + \beta}$$

$\alpha$  --- 光利用系数 [ $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  per unit of PAR]

$\beta$  --- 光强极大时，植物对CO<sub>2</sub>的最大吸收率 ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ )

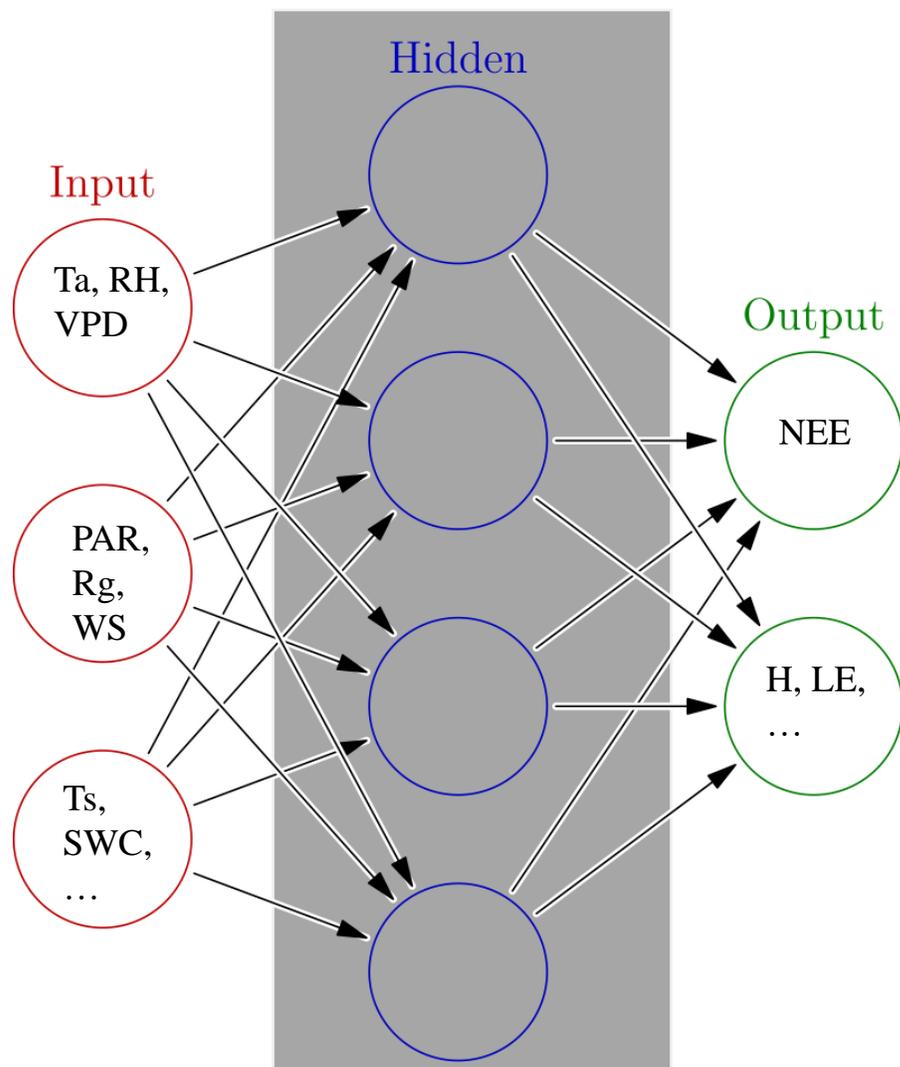
PAR --- 可见光强度 (PAR,  $\mu\text{mol photon m}^{-2} \text{ s}^{-1}$ )

$$\beta = \begin{cases} \beta_0 \exp[-k(\text{VPD} - \text{VPD}_0)] & \text{if VPD} > 10 \text{ hPa} \\ \beta_0 & \text{otherwise} \end{cases}$$

VPD --- 饱和水汽压差 (KPa)

$$NEE = R_{\text{eco}} - GPP$$

# 人工神经网络或者机器深度学习方法



- 比较复杂的数学“黑箱”模型，不需要了解太多的生态系统的机理过程，通过对观测资料的自我学习来建立输入因子和输出的关系。
- 需要大量（数年）的资料来建模（训练模型）。

# 各种方法的优缺点

Method	Strength	Weakness
线性回归	方法简单，易于操作; 不需要环境或者气象要素作控制变量	只适合补缺小的空缺(1-2 小时); 方法只能在较小的时间段窗口(数个小时内，通量变化近似线性) 应用
非线性回归	方法简单，易于操作	适合补缺小的空缺(数个小时内); 需要环境或者气象要素作控制变量; 只能在较小的时间段窗口 (一至几天) 应用
平均日变化 (MDV) ， marginal distribution sampling (MDS)	方法简单，易于操作; 比查表法 (LUT) 要准确	适合补缺小或者中等的空缺; 需要环境或者气象要素作控制变量; 只能在较小的或者中等的时间段窗口应用 (7至10天); 在一个生长季度或者一年内，补缺过程需要多次重复

# 各种方法的优缺点

Method	Strength	Weakness
查表法 (LUT)	补缺较大空缺; 在几个月内或者一个生长季节内甚至一年内建一个表	没有MDV或者MDS方法准确; 如果选用多个气象变量作控制因子的话, 表会变得表较复杂
生态系统机理模型	补缺较大空缺; 模型建立在比较大的时间段窗口 (几个星期到几个月)	需要完整的环境或者气象资料序列来驱动模型; 补缺后的资料不适合用来作其它模型的验证或者校正
人工神经网络或者机器深度学习方法	可以来补缺任意大小的空缺; 在几年的时间段内只需建一个模型	需要大量资料 (2-3年, 包含各种气候背景的年型) 来建立或者训练模型

# On-line tool: <https://bgc.iwww.mpg.de/5622399/REddyProc>



MAX PLANCK INSTITUTE  
FOR BIOGEOCHEMISTRY



BIOGEOCHEMICAL INTEGRATION | PROF. REICHSTEIN

DEUTSCH

Search



HOME DEPARTMENT | RESEARCH | PEOPLE | PUBLICATIONS | EVENTS | TEACHING | SOFTWARE | CAREER

[Back to Home Page](#)

[Home](#) > [Departments](#) > [Biogeochemical Integration | Prof. Reichstein](#) > [Data and Software](#) > [REddyProcWeb](#)

## REddyProcWeb online tool

The REddyProc online tool implements the standardized methods for processing Eddy-Covariance data. This online tool is based on the [REddyProc R package](#) which can be used offline in R with more options and flexibility.

Citation: Wutzler T, Lucas-Moffat A, Migliavacca M, Knauer J, Sickel K, Sigut, Menzer O & Reichstein M (2018) Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences, Copernicus, 15, doi: [10.5194/bg-15-5015-2018](https://doi.org/10.5194/bg-15-5015-2018)

### News

- 2023-01-20: We transition to another web presentation.  
The REddyProc webpages are planned to come online **again Fri, 27th of Jan.**  
You can still process your data here: <https://www.bgc-jena.mpg.de/REddyProc/brew/REddyProc.rhtml>
- 2022-09-21: Migrated web-service to a new compute cluster,
  - now using recent REddyProc v.1.3.2



[No Title]

# REddyProc: basic and extensible post-processing of eddy covariance flux data

Wutzler et al. (2018) <[doi:10.5194/bg-15-5015-2018](https://doi.org/10.5194/bg-15-5015-2018)>

Version: 1.3.2

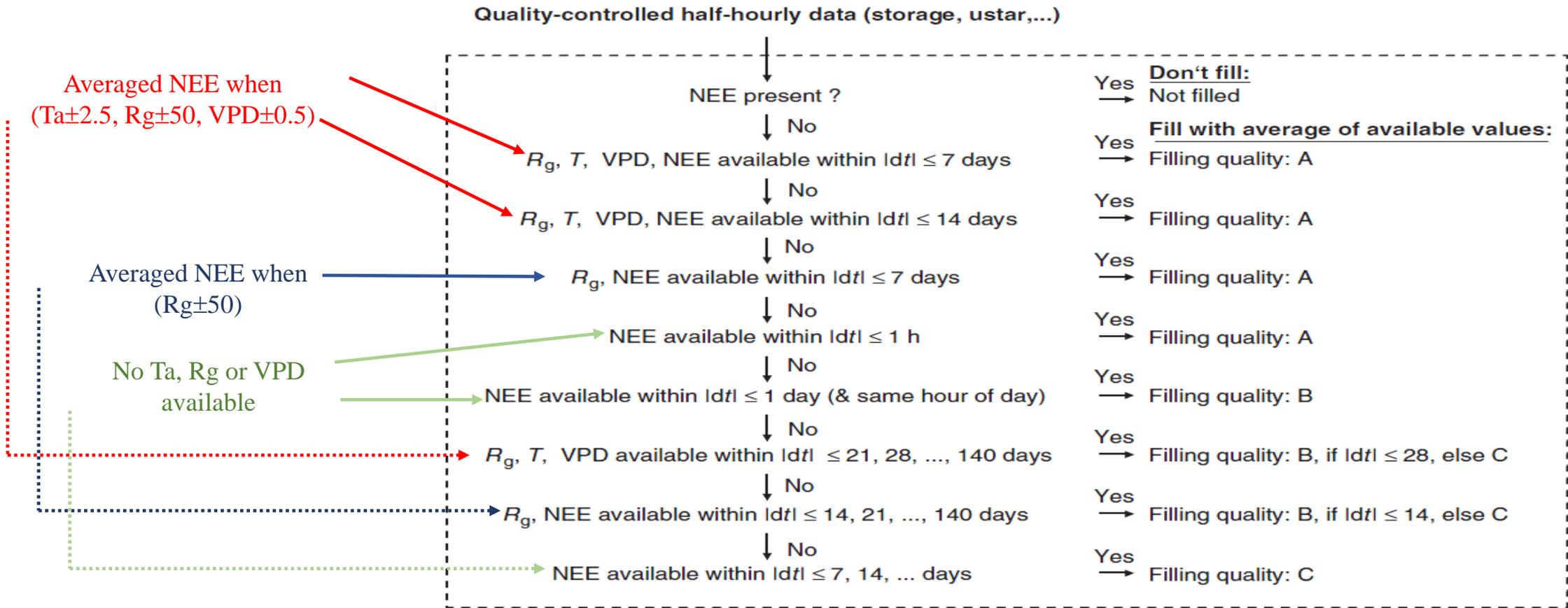
Depends: R ( $\geq 3.0.0$ )

This package inputs pre-processed (half-)hourly data and supports further processing.

- First, a quality-check and filtering is performed based on the relationship between measured flux and friction velocity ( $u_{\text{Star}}$ ) to discard biased data (Papale et al. (2006) <[doi:10.5194/bg-3-571-2006](https://doi.org/10.5194/bg-3-571-2006)>).
- Second, gaps in the data are filled based on information from environmental conditions (Reichstein et al. (2005) <[doi:10.1111/j.1365-2486.2005.001002.x](https://doi.org/10.1111/j.1365-2486.2005.001002.x)>).
- Third, the net flux of carbon dioxide is partitioned into its gross fluxes in and out of the ecosystem by night-time based and day-time based approaches (Lasslop et al. (2010) <[doi:10.1111/j.1365-2486.2009.02041.x](https://doi.org/10.1111/j.1365-2486.2009.02041.x)>).

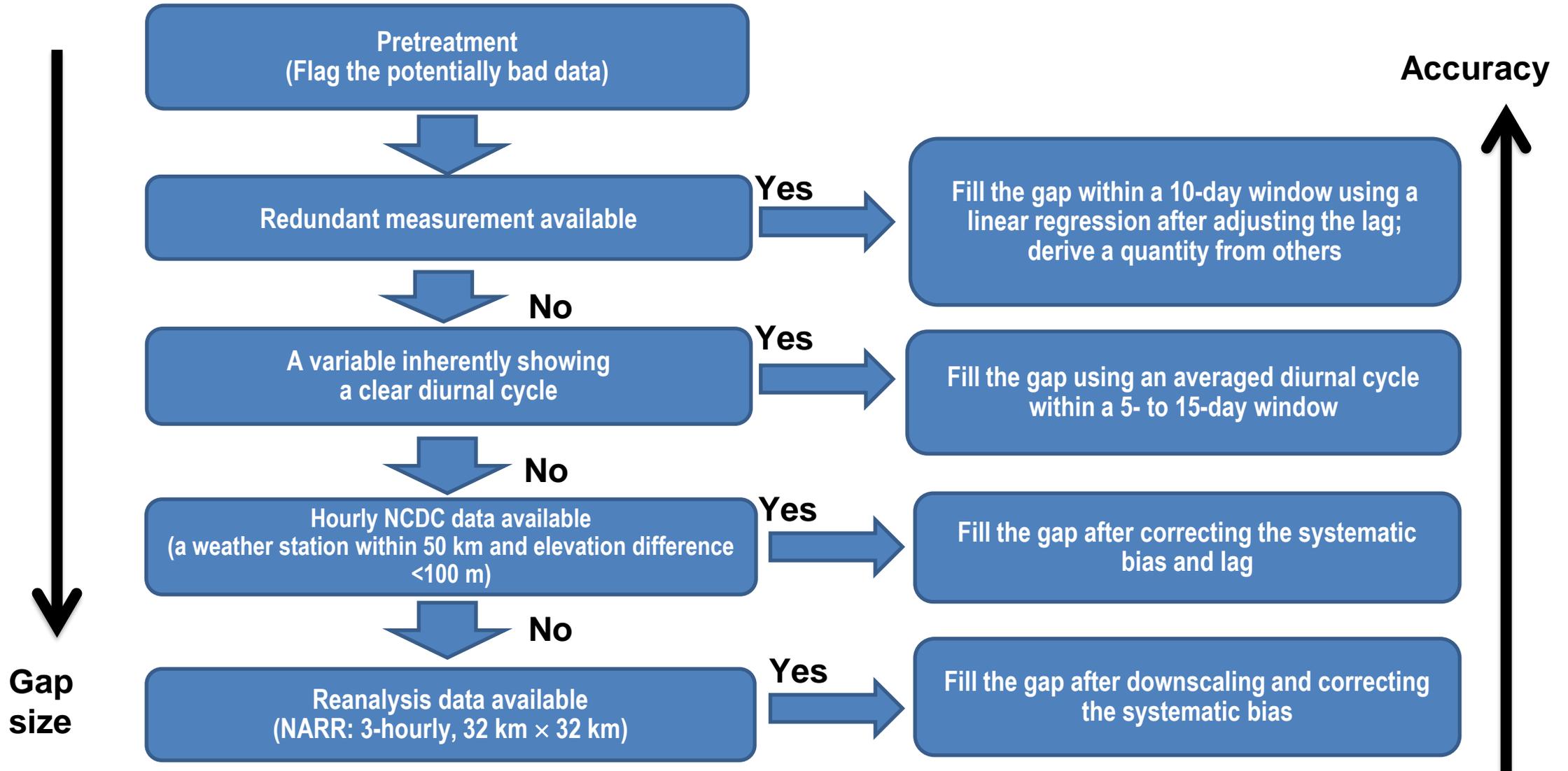
# How are data gaps filled in REddyProc?

The MDS flow diagram (Reichstein et al. 2005)



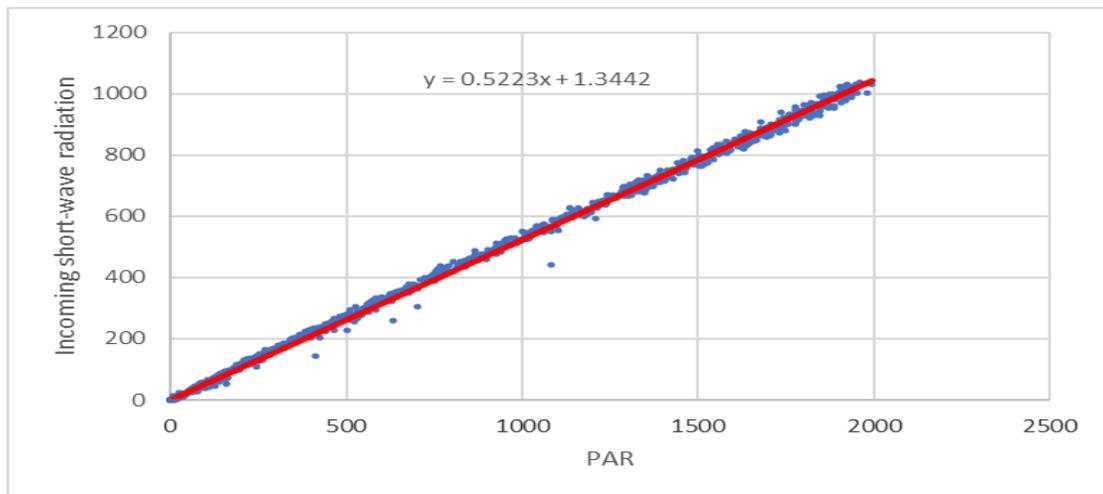
**Fig. A1** Flow diagram of the gap-filling algorithm used in this study. Abbreviations: NEE, net ecosystem  $\text{CO}_2$  exchange;  $R_g$ , global radiation;  $T$ , air temperature; VPD, vapour pressure deficit;  $|dt|$ , absolute difference in time. Filling qualities: A, high; B, medium; C, low.

# 如何补缺环境气象变量?



# 气象变量补缺个例

(利用变量的重复观测或者另一个相似的变量)



MOFlux data (June 2006):

Incoming short-wave radiation in  $W/m^2$

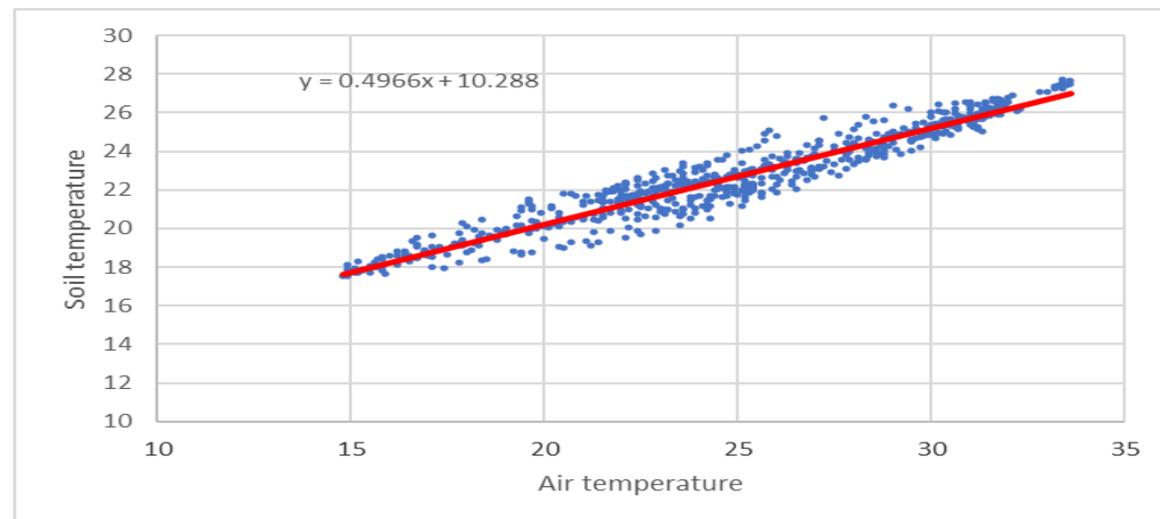
PAR (photosynthetically active radiation or visible light) in  $\mu mol/m^2/s$



MOFlux data (June 2006):

Soil temperature at 10 cm in  $^{\circ}C$

Air temperature at 30 m in  $^{\circ}C$



# 警告与结语

- 不要过分依赖于数据补缺这个手段；数据补缺应该是“没有办法的办法”
- 任何补缺方法都会有误差的；即使最好的数据补缺方案也不可能产生和直接观测一样准确并且可靠的结果
- 并不是所有变量都能够被合理插补（如反射辐射, 土壤水分, 降水量); 重复或者多套观测(redundant or replicate measurements) 非常重要
- 获得高质量资料的最好办法不是靠后期补缺，而是：维护好仪器的正常运行，经常查看资料（2-3次每周），及时发现仪器问题并且尽快修理，定期做仪器校准或者校正

# Thank you!

