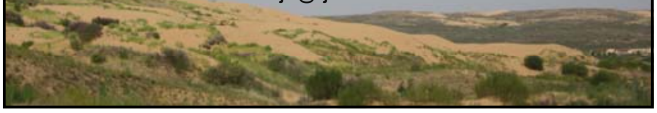
中国通量观测研究联盟
ChinaFLUX

生态系统通量观测缺失数据插补理论与技术

贾昕（博士、教授）
北京林业大学，水土保持学院
xinjia@bjfu.edu.cn



数据插补不是万能药!!!

- 再好的插补算法也不能完全补救数据缺失的损失
- 并不是所有变量都能够被合理插补（如反射辐射、土壤水分，重复的重要性！）
- 减小数据插补造成的不确定性（站点维护的重要性！）

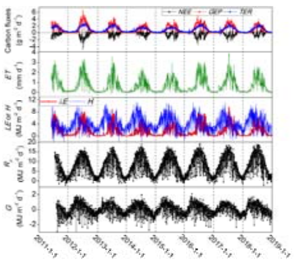
数据缺失的原因

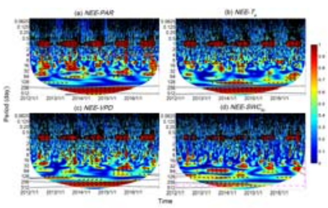
- 仪器故障
- 供电故障
- 不利天气
- 仪器校准
- 人为因素
- 质量控制

EC平均年数据覆盖约60-80% (Falge et al., 2001a,b; Moffat et al., 2007)

为什么要插补缺失值？

1. 估算一段时间（如一年）的通量累计值；2. 进行特定数据分析（如小波分析）





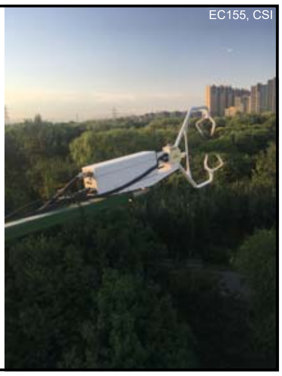
Jia et al. 2018 AFM

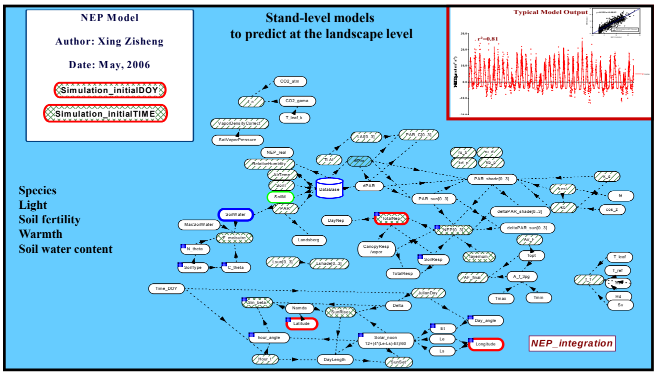
数据插补理论

if Gap > 2h

- 根据内在变化规律，如MDV/MDC/MDT, LUT, Marginal Distribution Sampling (MDS) (Falge et al. 2001a,b; Reichstein et al. 2005; Moffat et al. 2007; Wutzler et al. 2018)
- 经验模型，如NLR、ANN (Amiro et al. 2006; Moffat et al. 2007; Jia et al. 2014, 2016a,b, 2018a,b)
- 机理模型，如Xing et al. 2007, 2008a,b

if Gap <= 2h
then 线性插值



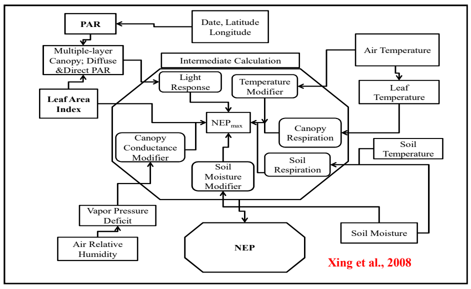


Evolution of NEP model

initially intended for gap filling

Ecological Modelling papers

Full Model



Xing et al., 2008

Model parameters

Model Parameter	Function	Value derivation	Values	Range[a]
Ecoresp_beta	Ecosystem respiration curve		0.017825	n/a
Ecoresp_alpha	Ecosystem respiration curve	Set according	308	n/a
optimum_soilwater	Field capacity for sandy cla	to the literature	0.32	n/a
optimum_temperature	Tree response to air temperature		17.5	n/a
alpha	Light use efficiency for multi-layered canopy	Species-specific and	0.0016[b]	0-1
NEP_max	Maximum possible NEP ass	set based on NWL 2006 data	34.4	0-100
NEPgama	Light compensation point of balsam fir		202.9	0-300
krt	Influence of temperature on photosynthesis		-[c]	0-3000
pco2	Rate of CO_2 exchange between the atmosphere and forest canopy	Determined by fitting NEP model to data	-	0-5
kvapord	Relates stomatal conductance	from both sites and for all years	-	0-5
R_{eco}	Maximum ecosystem respiration		-	0-20

Process-specific parameters

Species-specific parameters

Site-specific parameters

n/a= not applicable
[a] the range is provided to NEP model to constrain feasible solution
[b] determined through pre-calibration of the NEP model (see Results and Discussion section); alpha is in µmol CO_2 (µmol PAR)$^{-1}$; NEP_max, in µmol CO_2 m^{-2} s^{-1}; and NEPgama, in µmol PAR m^{-2} s^{-1}
[c] variable parameter values to be set according to a non-linear fit of NEP model to NEP and eco-hydrometeorological data from all sites and years

Gap filling strategies for defensible annual sums of net ecosystem exchange

We review several methods of gap filling and apply them to data sets available from the EUROFLUX and AmeriFlux databases. The methods are based on mean diurnal variation (MDV), look-up tables (LookUp), and nonlinear regressions (Regr.) and the impact of different gap filling methods on the annual sum of F_{NEE} is investigated. The difference between annual F_{NEE} filled by MDV compared to F_{NEE} filled by Regr. ranged from −45 to +200 g C m^{-2} per year (MDV−Regr.). Comparing LookUp and Regr. methods resulted in a difference (LookUp−Regr.) ranging from −30 to +150 g C m^{-2} per year.

We also investigated the impact of replacing measurements at night, when turbulent mixing is insufficient. The nighttime correction for low friction velocities (u_*) shifted annual F_{NEE} on average by +77 g C m^{-2} per year, but in certain cases as much as +185 g C m^{-2} per year.

Our results emphasize the need to standardize gap filling-methods for improving the comparability of flux data products.

Falge et al. 2001a AFM

Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes

statistical metrics. The performance of the gap filling varied among sites and depended on the level of aggregation (native half-hourly time step versus daily), long gaps were more difficult to fill than short gaps, and differences among the techniques were more pronounced during the day than at night.

The non-linear regression techniques (NLRs), the look-up table (LUT), marginal distribution sampling (MDS), and the semi-parametric model (SPM) generally showed good overall performance. The artificial neural network based techniques (ANNs) were generally, if only slightly, superior to the other techniques. The simple interpolation technique of mean diurnal variation (MDV) showed a moderate but consistent performance. Several sophisticated techniques, the dual unscented Kalman filter (UKF), the multiple imputation method (MIM), the terrestrial biosphere model (BETHY), but also one of the ANNs and one of the NLRs showed high biases which resulted in a low reliability of the annual sums, indicating that additional development might be needed. An uncertainty analysis comparing the estimated random error in the 10 benchmark datasets with the artificial gap residuals suggested that the techniques are already at or very close to the noise limit of the measurements. Based on the techniques and site data examined here, the effect of gap filling on the annual sums of NEE is modest, with most techniques falling within a range of ±25 g C m^{-2} year^{-1}.

Moffat et al. 2007 AFM

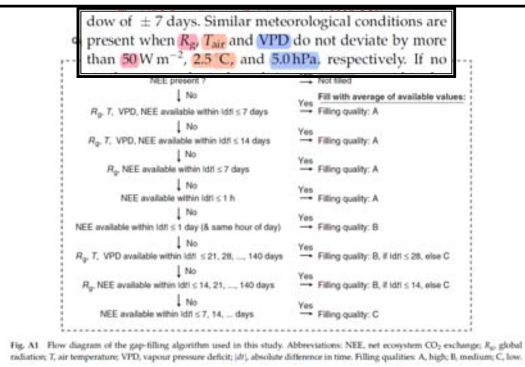
Gap filling strategies for long term energy flux data sets

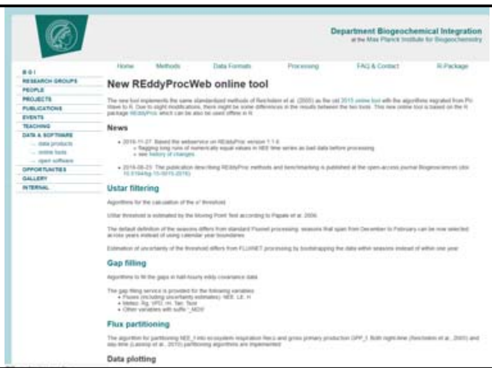
At present a network of over 100 field sites are measuring carbon dioxide, water vapor and sensible heat fluxes between the biosphere and atmosphere, on a nearly continuous basis. Gaps in the long term measurements of evaporation and sensible heat flux must be filled before these data can be used for hydrological and meteorological applications. We adapted methods of gap filling for NEE (net ecosystem exchange of carbon) to energy fluxes and applied them to data sets available from the EUROFLUX and AmeriFlux eddy covariance databases. The average data coverage for the sites selected was 69% and 75% for latent heat (λE) and sensible heat (H). The methods were based on mean diurnal variations (half-hourly binned means of fluxes based on previous and subsequent days, MDV) and look-up tables for fluxes during assorted meteorological conditions (LookUp), and the impact of different gap filling methods on the annual sum of λE and H is investigated. The difference between annual λE filled by MDV and λE filled by LookUp ranged from −120 to 210 MJ m^{-2} per year, i.e. −48 to +86 mm per year, or −13 to +39% of the annual sum. For annual sums of H differences between −140 and +140 MJ m^{-2} per year or −12 to +19% of the annual sum were found.

Falge et al. 2001b AFM

数据插补工具

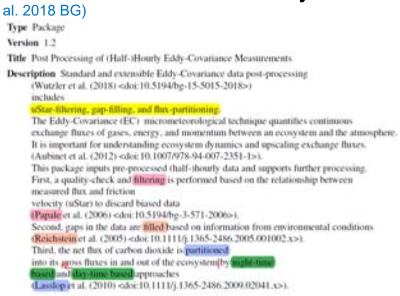
- 研究人员独立开发代码（MATLAB, Python, Fortran）
- 网页工具 (https://www.bgcjena.mpg.de/bgi/index.php/Services/REddyProcWeb)
- R软件REddyProc包





Basic and extensible post-processing of eddy covariance flux data with REddyProc

(Wutzler et al. 2018 BG)



Type Package

Version 1.2

Title Post Processing of (Half-)Hourly Eddy-Covariance Measurements

Description Standard and extensible Eddy-Covariance data post-processing (Wutzler et al. (2018) <doi:10.5194/bg-15-5015-2018>) includes uStar-filtering, gap-filling, and flux-partitioning. The Eddy-Covariance (EC) micrometeorological technique quantifies continuous exchange fluxes of gases, energy, and momentum between an ecosystem and the atmosphere. It is important for understanding ecosystem dynamics and upscaling exchange fluxes. (Aubinet et al. (2012) <doi:10.1007/978-94-007-2351-1>). 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 (uStar) to discard biased data (Papale et al. (2006) <doi: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>). 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>).

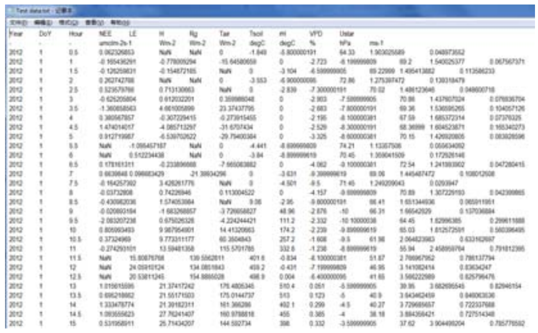
```
install.packages("REddyProc")
library(REddyProc)
?REddyProc()
```

```
install.packages("REddyProc")
library(REddyProc)
?REddyProc()
```

install.packages("REddyProc")
library(REddyProc)
vignette("DEGebExample")

Crop example demonstrating multiple years and user defined uStar-Seasons

The data file (.txt)

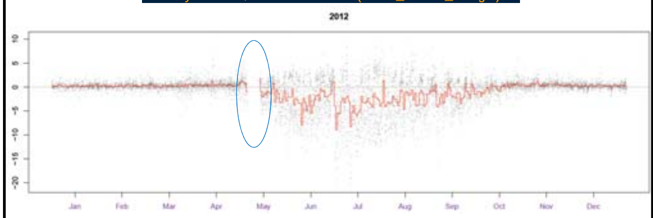


A code template

```r
library(REddyProc)
library(dplyr)
setwd("C:/Users/hp/Documents")
EddyData.F<-fLoadTXTIntoDataframe("Test data.txt")
EddyDataWithPosix.F<-fConvertTimeToPosix(EddyData.F,'YDH',Year='Year',Day='DoY',Hour='Hour')
EddyProc.C<-sEddyProc$new('yc1', EddyDataWithPosix.F,c('NEE','Rg','Tair','VPD','Ustar'))
uStarTh<-(EddyProc.C$sEstUstarThreshold()$uStarTh)
select(uStarTh, -seasonYear)
uStarThAnnual<-usGetAnnualSeasonUStarMap(uStarTh)
uStarThAnnual
uStarSuffixes<-colnames(uStarThAnnual)[-1]
uStarSuffixes
EddyProc.C$sMDSGapFillAfterUstar(fluxVar="NEE",uStarVar="Ustar",uStarTh=uStarThAnnual,uStarSuffix=uStarSuffixes,FillAll=TRUE)
grep("NEE_.*_f$",names(EddyProc.C$sExportResults()),value=TRUE)
grep("NEE_.*_fsd$",names(EddyProc.C$sExportResults()),value=TRUE)
#EddyProc.C$sPlotFingerprintY('NEE_uStar_f', Year=2012)
EddyProc.C$sSetLocationInfo(LatDeg=37.68,LongDeg=107.23,TimeZoneHour=8)
EddyProc.C$sMDSGapFill('Tair',FillAll=FALSE)
EddyProc.C$sMDSGapFill('VPD',FillAll=FALSE)
EddyProc.C$sMDSGapFill('Rg',FillAll=FALSE)
EddyProc.C$sMRFluxPartition(Suffix=uStarSuffixes)
grep("GPP.*_f$|Reco",names(EddyProc.C$sExportResults()),value=TRUE)
#EddyProc.C$sPlotFingerprintY('GPP_uStar_f', Year=2012)
FilledEddyData.F<-EddyProc.C$sExportResults()
FilledEddyData.F
CombinedData.F<-cbind(EddyData.F, FilledEddyData.F)
fWriteDataframeToFile(CombinedData.F,'yc1(2012).xls',Dir="")
```

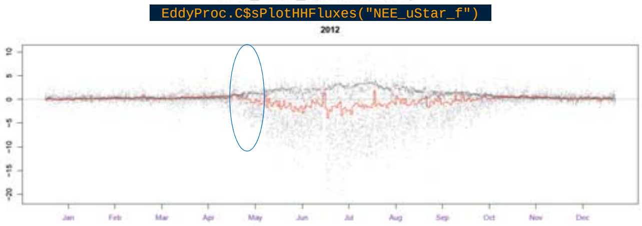
Half−hourly fluxes and daily means at yc1:
NEE_uStar_orig (umolm−2s−1)

`EddyProc.C$sPlotHHFluxes("NEE_uStar_orig")`



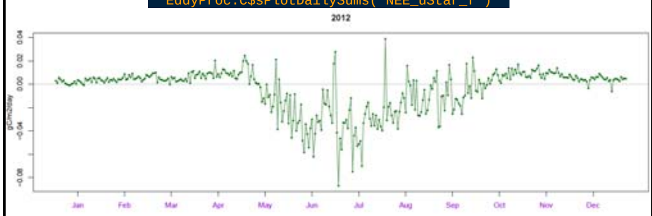
Half−hourly fluxes and daily means at yc1:
NEE_uStar_f (umolm−2s−1)

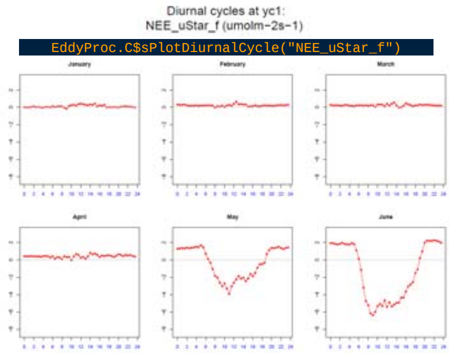
`EddyProc.C$sPlotHHFluxes("NEE_uStar_f")`

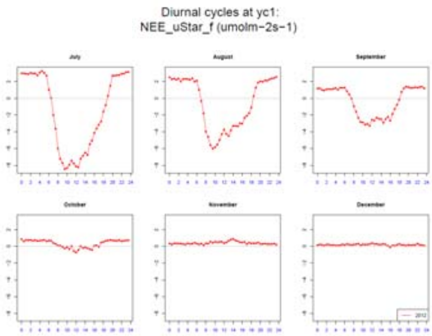


Daily sums at yc1:
NEE_uStar_f (gC/m2/day)

`EddyProc.C$sPlotDailySums("NEE_uStar_f")`

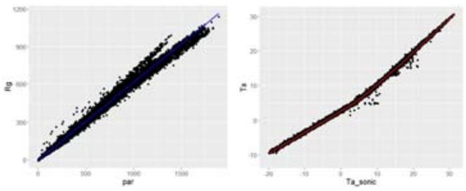






如何插补微气象数据

- “Gaps in environmental variables (e.g., R_g and T_a) were filled either using complementary measurements at the site (e.g., photosynthetically active radiation or soil temperature), or with the MDS method.”(Jia et al. 2019 manuscript)



参考文献

- Falge E et al. 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. AFM 107, 43-69.
- Falge E et al. 2001. Gap filling strategies for long term energy flux data sets. AFM 107, 71-77.
- Reichstein M et al. 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. GCB 11, 1424-1439.
- Moffat AM et al. 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. AFM 147, 209-232.
- Wutzler T et al. 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15, 5015-5030.
- Jia X et al. 2014. Biophysical controls on net ecosystem CO_2 exchange over a semiarid shrubland in northwest China. Biogeosciences 11, 4679-4693.
- Jia X et al. 2016a. Carbon and water exchange over a temperate semi-arid shrubland during three years of contrasting precipitation and soil moisture patterns. AFM 228:120-129.
- Jia X et al. 2016b. Energy partitioning over a semi-arid shrubland in northern China. Hydrological Processes. 30, 972-985.
- Jia X et al. 2018a. Multi-scale dynamics and environmental controls on net ecosystem CO_2 exchange over a temperate semiarid shrubland. AFM 259, 250-259.
- Jia X et al. 2018b. Canopy photosynthesis modulates soil respiration in a temperate semi-arid shrubland at multiple timescales. Plant and Soil 432, 437-450.
- Amiro BD et al. 2006. Carbon, energy and water fluxes at mature and disturbed forest sites, Saskatchewan, Canada. AFM 136, 237-251.
- Lasslop G et al. 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. GCB 16, 187-208.
- Xing Z et al. 2007. A simple net ecosystem productivity model for gap filling of tower-based fluxes: an extension of Landsberg’s equation with modifications to the light interception term. Ecological Modelling 206, 250-262.
- Xing Z et al. 2008. Modification of an ecosystem model for filling medium-sized gaps in tower-based estimates of net ecosystem productivity. Ecological Modelling 213, 86-97.
- Xing Z et al., 2008. A process-based model designed for filling of large data gaps in tower-based measurements of net ecosystem productivity. Ecological Modelling 213, 165-179.