



UNIVERSITY OF BAYREUTH

Department of Micrometeorology

**GaFiR: a gap-filling package for
ecosystem-atmosphere carbon dioxide flux and
evapotranspiration data**

Peng Zhao

Johannes Lüers

Thomas Foken

**Arbeitsergebnisse
Nr. 59
Bayreuth, June 2014**

Arbeitsergebnisse, Universität Bayreuth, Abt. Mikrometeorologie
Work Reports, University of Bayreuth, Dept. of Micrometeorology
Print, ISSN 1614-8916; Internet, ISSN 1614-8924

<http://www.bayceer.uni-bayreuth.de/mm/>

https://epub.uni-bayreuth.de/view/series/Arbeitsergebnisse_2F_Universit=E4t_Bayreuth,_Abteilung_Mikrometeorologie.html

Eigenverlag: Universität Bayreuth, Abt. Mikrometeorologie
Vervielfältigung: Druckerei der Universität Bayreuth
Herausgeber: Prof. Dr. Thomas Foken

Universität Bayreuth, Abteilung Mikrometeorologie
D-95440 Bayreuth

Die Verantwortung über den Inhalt liegt bei den jeweiligen Autoren.

Contents

1. Introduction	1
1.1 History	1
1.2 Features	1
1.3 Copyright	2
2. Quick start	2
3. Overview of GaFiR	3
3.1 Modules	3
3.2 Evaluation of models	5
3.3 Files and folders	6
3.4 Required R packages	6
4. GaFiR-Evapotranspiration module	7
4.1 GaFiR-ET-PM sub-module	7
4.1.1 GaFiR-ET-PM-FAO sub-module	7
4.1.2 GaFiR-ET-PM-KP sub-module	8
4.2 GaFiR-ET-PT sub-module	8
5. GaFiR-Carbon module	8
5.1 GaFiR-C-Reco sub-module	8
5.2 GaFiR-C-GPP sub-module	9
5.2.1 GaFiR-C-GPP-Time sub-module	9
5.2.2 GaFiR-C-GPP-Temperature sub-module	9
5.2.3 GaFiR-C-Leaf sub-module	9
5.2.4 GaFiR-C-GPP-Time-VPD sub-module	9
6. Data, files and folders	10
6.1 Input data	10
6.2 Starting/configuration file	11
6.3 Program scripts	12
6.4 Output files	13
7. References	16

1. Introduction

1.1 History

Eddy-covariance technique has widely been used as a useful and effective tool to continuously and directly quantify turbulent fluxes of heat, water vapour, and carbon dioxide between land surface and atmosphere. However, data gaps inevitably exist due to system breakdown, calibration, maintenance, human activities, and unfavourable atmosphere conditions when the assumptions required by the EC technique are not fulfilled (Foken and Wichura, 1996). Averagely 35% of flux observations are reported in the literature as missing or rejected (Falge et al., 2001). Data gap-filling methods are therefore necessary to provide complete data sets for the estimation of seasonal and annual sums of energy and matter (Papale, 2012).

This manual is a documentation of the **Gap-Filling R** package GaFiR (current version 1.0), which was developed to fill data gaps in turbulent flux observations of evapotranspiration and carbon dioxide. GaFiR was based on previous experience by the Department of Micrometeorology, University of Bayreuth. In April 2005 a Matlab script package (NEE-Param and NEE-Calc) was firstly written by Christoph Thomas to fill the data gaps in flux observation at the FLUXNET-Station Waldstein-Weidenbrunnen in a spruce forest near Bayreuth. Most of those ideas in that script was presented in Ruppert et al. (2006). That package, updated by Florian Wimmer in January 2006, by Christoph Thomas in July 2007, by Johannes Lüers in January 2011, has been internally used and provides satisfactory results for the forest site. In order to adapt different features of crops for the study carried out in rapid growing croplands, Zhao and Lüers (2012) brought in new ideas to and rewrote these ideas in an R (R Development Core Team, 2012) script. Later on this R script was updated and released as GaFiR 1.0, which can be applied to data gap-filling for turbulent fluxes at forest, crop, and meadow sites. GaFiR will continuously be supported and updated by the authors in the future with new releases, which could be found on the authors' webpage on <http://www.bayceer.uni-bayreuth.de/mm>.

1.2 Features

GaFiR has the following features:

- GaFiR is powerful and efficient. It is able to not only fill data gaps in carbon dioxide and evapotranspiration (latent heat) flux in a variety of conventional and up-to-date ways for different ground surfaces, but also output plenty of interesting related information such as distribution of data gaps, statistics of model parameterizations, and goodness-of-fit

measure of models. The outputs of GaFiR are fruitful in a variety of formats such as figures, explanatory texts, and data tables, which can be directly used in scientific publications. The processing only takes up to several minutes.

- GaFiR is an open source package. All scripts in GaFiR are written in R language, which is a free popular software environment for statistical computing and graphics. Users are free to download GaFiR and copy or modify any script in it as they like for non-business use.
- GaFiR is user-friendly. It runs on Windows, UNIX/Linux, and MacOS platforms. It requires users only some basic knowledge in R. Users only have to prepare the input data and modify in a text editor one single configuration file which is easily understood even by R beginners.
- Together with TK3 (Mauder and Foken, 2011) which calculates quality-controlled turbulent flux data, and MSEF package (Lüers et al., 2014) which checks outliers in flux data, GaFiR provides full time series of flux data and completes the software family of routine flux data processing procedure in the Department of Micrometeorology, University of Bayreuth.

1.3 Copyright

GaFiR is copyrighted by its authors and Department of Micrometeorology, University of Bayreuth. Permission is granted to copy, distribute and/or modify GaFiR's codes under the terms of the Creative Commons Attribution-Sharealike 3.0 Unported License (CC-BY-SA) and, unless otherwise noted, the GNU Free Documentation License (GFDL) (unversioned, with no invariant sections, front-cover texts, or back-cover texts).

GaFiR is free for non-business use. However, it is highly recommended and appreciated that users return their feedbacks to GaFiR authors in order to control the version, improve the codes, bring in new ideas and update the package. Users with valuable contributions have the potential to be co-authors in the future version of GaFiR.

2. Quick start

In order to get an impression on what GaFiR looks like and how to use GaFiR, users can follow a quick-start procedure as follows:

1. Download R from the Comprehensive R Archive Network (CRAN, <http://cran.r-project.org/>) and install it in the computer.
2. Download 'gafir.zip' from the website of the Department of Micrometeorology,

University of Bayreuth (<http://www.bayceer.uni-bayreuth.de/mm>).

3. Unzip `gafir.zip` and get a folder named 'GaFiR' with three sub-folders and a file named 'gafir-start.r' in it.
4. Open `gafir-start.r` with a text editor and find the following code line:

```
wd <- 'C:/GaFiR'
```

Change this path into the real path of GaFiR in user's computer and save the file.

5. Run `gafir-start.r` in R console. Data in an example file named 'demo.csv' located in the 'input' folder is then processed and the web browser will automatically open report files in several minutes. Users could see all results in the 'output' folder.

3. Overview of GaFiR

3.1 Modules

The general structure of GaFiR contains two modules (Fig. 1): the GaFiR-Evapotranspiration module, which is used for ecosystem evapotranspiration (ET) data, and the GaFiR-Carbon module, which is used for gap-filling of carbon dioxide flux data. These two modules are independent on each other. Users can easily switch either of them on or off according to their purposes. Details of these modules are described in Chapters 4 and 5. This section gives a brief introduction.

The GaFiR-Evapotranspiration module contains two independent sub-modules. The first is the GaFiR-ET-PM sub-module, which applies the Penman-Monteith model (Monteith, 1965) either with the approach abbreviated as PM-FAO suggested by the Food and Agriculture Organization (Allen et al., 1998) or with the approach abbreviated as PM-KP proposed by Katerji and Perrier (1983). The second is the GaFiR-ET-PT sub-module, which applies the Priestley-Taylor model (Priestley and Taylor, 1972).

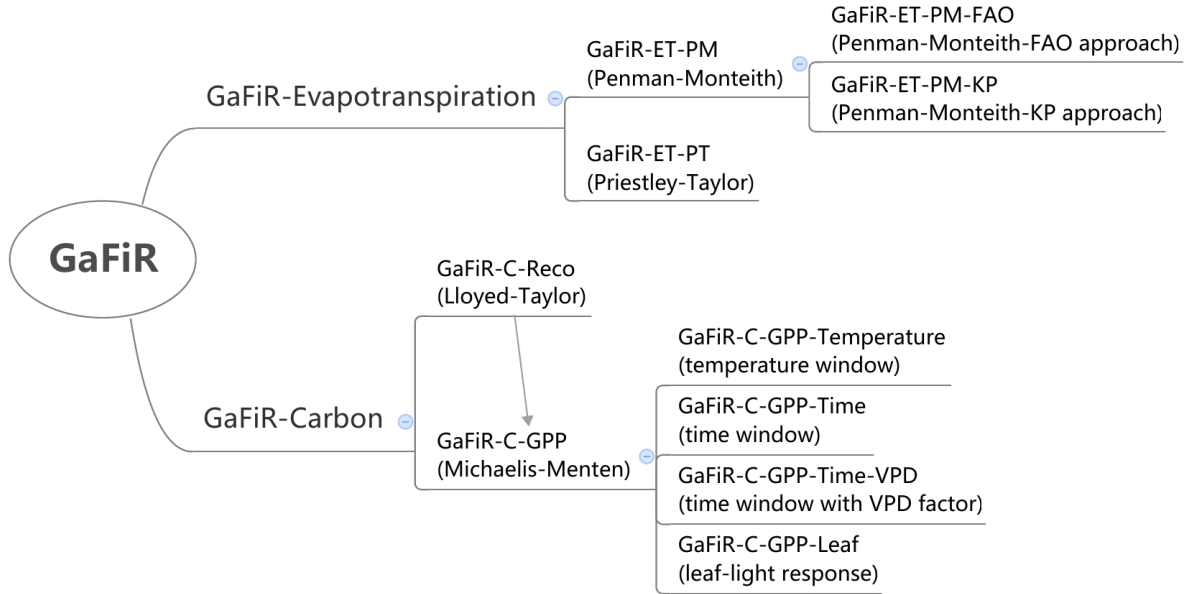


Fig. 1. Modules in GaFiR

The GaFiR-Carbon module contains two steps in sequence. In the first step (abbreviated as GaFiR-C-Reco), measured carbon dioxide flux in night-time is ecosystem respiration (Reco), which is used for the parameterization of Lloyd-Taylor function (Lloyd and Taylor, 1994) and the fitted parameters are used not only to fill the gaps in night-time Reco data but also to extrapolate to daytime Reco estimation. In the second step (abbreviated as GaFiR-C-GPP), according to the partitioning strategy by Falge et al. (2001) and Reichstein et al. (2005), gross primary productivity (GPP) in the daytime is derived as:

$$GPP = NEE - R_{eco} \quad 1$$

where NEE is the net ecosystem exchange of carbon dioxide, i.e. the carbon dioxide flux measured by eddy-covariance technique. The signs follow the conventional meteorological definitions that the flux away from the surface is positive. This derived GPP is used for the parameterization of Michaelis-Menten function (Michaelis and Menten, 1913) and the fitted parameters are used to fill the gaps in daytime GPP data.

The GaFiR-C-GPP step contains four independent sub-modules: (1) GaFiR-C-GPP-Temperature sub-module, which groups GPP data into temperature classes to determine individual light response parameters in each class in order to test the influence of temperature on GPP (Falge et al., 2001; Ruppert et al., 2006), (2) GaFiR-C-GPP-Time sub-module, which

segments the GPP data into sub-periods by a certain time window and determine individual light response parameters in each sub-period, (3) GaFiR-C-GPP-Time-VPD sub-module, which additionally introduces the influence of vapour pressure deficit (Lasslop et al., 2010) into the time window approach, and (4) GaFiR-C-GPP-Leaf sub-module, which applies a leaf-light response function proposed by Zhao (2014) for fast growing croplands.

3.2 Evaluation of models

GaFiR evaluates the performance of the gap-filling models by the comparison between field observation and model simulation in a variety of goodness-of-fit measures (Table 1). The evaluations are written in report files when GaFiR has been successfully run. Users are free to pick out those measures they need. For simplicity, the overall performances of the simulations can be evaluated by ranking either Nash-Sutcliffe model efficiency coefficient (NSE, Nash and Sutcliffe, 1970) or the index of agreement (d , Willmott, 1981). NSE varies between minus infinity to 1, and d varies between 0 to 1, both with larger values indicating better performance of the simulation. A value of $NSE = 1$ or $d = 1$ indicates a perfect agreement between observation and simulation. $NSE = 0$ means that the model is as good as the mean observation, while $d = 0$ means a complete disagreement. The best simulation is the one which has a NSE or d closest to unity and with the lowest RMSE.

Table 1 Goodness of fit measures in GaFiR
(P is model prediction, O is observation, σ is standard deviation)

Goodness of fit	Calculation
Mean Error (ME)	$ME = \frac{1}{N} \sum_{i=1}^N P_i - O_i $
Mean Absolute Error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)$
Mean Squared Error (MSE)	$MSE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{MSE}$
Normalized Root Mean Square Error (NRMSE)	$NRMSE = 100 \frac{RMSE}{O_{\max} - O_{\min}}$
Percent Bias (PBIAS)	$PBIAS = 100 \frac{\sum_{i=1}^n (P_i - O_i)}{\sum_{i=1}^n O_i}$
Ratio of RMSE to the Standard Deviation of the Observations (RSR)	$RSR = \frac{RMSE}{\sigma_O}$
Ratio of Standard Deviations (rSD)	$rSD = \frac{\sigma_P}{\sigma_O}$
Nash-Sutcliffe Efficiency (NSE)	$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$
Modified Nash-Sutcliffe Efficiency (mNSE)	$mNSE = 1 - \frac{\sum_{i=1}^n O_i - P_i ^j}{\sum_{i=1}^n O_i - \bar{O} ^j}$

Relative Nash-Sutcliffe Efficiency (rNSE)	$\text{rNSE} = 1 - \frac{\sum_{i=1}^n \left(\frac{O_i - P_i}{\bar{O}}\right)^2}{\sum_{i=1}^n \left(\frac{O_i - \bar{O}}{\bar{O}}\right)^2}$
Index of Agreement (d)	$d = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - O_i + P_i + O_i)^2}$
Modified Index of Agreement (md)	$md = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^j}{\sum_{i=1}^n (P_i - O_i + P_i + O_i)^j}$
Relative Index of Agreement (rd)	$rd = 1 - \frac{\sum_{i=1}^n \left(\frac{P_i - O_i}{\bar{O}}\right)^2}{\sum_{i=1}^n \left(\frac{ P_i - O_i + P_i + O_i }{\bar{O}}\right)^2}$
Persistence Index (cp)	$\text{cp} = 1 - \frac{\sum_{i=2}^n (O_i - P_i)^2}{\sum_{i=1}^{n-1} (O_{i+1} - O_i)^2}$
Pearson Correlation coefficient (R)	$R = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}}$
Coefficient of Determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2$
R^2 multiplied by the slope (b) of the regression line between P and O (bR2)	$\text{bR2} = bR^2$

3.3 Files and folders

GaFiR package is compressed as a .zip file named ‘gafir.zip’ for downloading. After unzip gafir.zip, users get a folder containing an R script file named ‘gafir-start.r’ and three sub-folders named ‘program’, ‘input’, and ‘output’. Users should firstly in a text editor configure gafir-start.r, which will later read users’ data (a .csv file with header) in the input folder, call those scripts in the program folder, and write the results into the output folder. Details of the files and folders are described in Chapter 6.

3.4 Required R packages

GaFiR requires some packages from CRAN besides basic R installation. GaFiR checks whether they have been installed in users’ computer. If not, GaFiR will download, install, and load them automatically. The required R packages include:

- nls2 (Grothendieck, 2012), for non-linear regression with the brute-force method,
- R2HTML (Lecoutre, 2003), for producing .html report files,.
- hydroGOF (Zambrano-Bigiarini, 2014), for calculation of goodness-of-fit measures between simulation and observation, and
- lattice (Sarkar, 2008), for plotting 3D figures.

4. GaFiR-Evapotranspiration module

4.1 GaFiR-ET-PM sub-module

Penman-Monteith (PM) model (Monteith, 1965) to estimate the latent heat flux (Q_E) is written as:

$$Q_E^{\text{PM}} = \frac{s_c(-R_n - Q_G) + \frac{\rho c_p(e_s - e_a)}{r_a}}{s_c + \gamma(1 + \frac{r_s}{r_a})} \quad 2$$

where R_n is net radiation; Q_G is ground heat flux; e_s is the saturated vapour pressure in hPa, calculated as a function of air temperature t in $^{\circ}\text{C}$ as $e_s = 6.112 \exp \frac{17.62t}{243.12+t}$ over water at -45 to 60 $^{\circ}\text{C}$ (Sonntag, 1990); s_c is the slope of the saturation vapour-pressure curve, $s_c = \frac{4284e_s}{(243.12+t)^2}$; e_a is the partial vapour pressure of the air, and thus $e_s - e_a$ is VPD; ρ is the air density; γ is the psychrometric constant, 0.667 hPa K^{-1} for air pressure of 1000 hPa and temperature of 20 $^{\circ}\text{C}$; r_s is the stomatal resistance of the entire "big leaf", or simply the canopy resistance; r_a is the aerodynamic resistance. The estimation of sensible heat flux by PM method is simply calculated as the difference between the available energy (Q_A , the sum of R_n and Q_G) and modeled latent heat flux ($Q_H^{\text{PM}} = -Q_A - Q_E^{\text{PM}}$).

The estimation of Q_E by Eq. 2 requires the parameterization of r_a and r_s . The estimation of r_a can be performed as

$$r_a = \frac{\ln \frac{z-d}{z_{om}} \ln \frac{z-d}{z_{oh}}}{\kappa^2 u}, \quad 3$$

where z is height at which wind speed is measured; d is displacement height, estimated as $\frac{2}{3}$ of the vegetation height (h); κ is Von-Kármán constant, equal to 0.40; and u is horizontal wind speed at sensor height z ; z_{om} is roughness height for momentum, approximated as $0.1h$; z_{oh} is roughness height for water vapour, approximated as $0.1z_{om}$.

The estimation of r_s can be performed either by FAO approach or by KP approach.

4.1.1 GaFiR-ET-PM-FAO sub-module

FAO approach estimates r_s as

$$r_s = \frac{r_{si}}{\text{LAI}_{\text{active}}}, \quad 4$$

where r_{si} is the stomatal resistance of a single well-illuminated leaf with a value between 70 to 80 s m^{-1} for estimation of hourly or shorter-time-based Q_E for agricultural crops (Allen, 2005), and $\text{LAI}_{\text{active}}$ is the LAI of the active sunlit leaves, which is generally the upper part of

the canopy and can be estimated as $LAI_{\text{active}} = 0.5LAI$.

4.1.2 GaFiR-ET-PM-KP sub-module

KP approach establishes a linear relationship between $\frac{r_s}{r_a}$ and $\frac{r^*}{r_a}$ as

$$\frac{r_s}{r_a} = a \frac{r^*}{r_a} + b \quad 5$$

where a and b are regression coefficients. r^* is determined as

$$r^* = \frac{(s_c + \gamma)\rho c_p(e_s - e_a)}{s_c\gamma(-R_n - Q_G)} \quad 6$$

and r_s determined from the observation by inverting the PM equation (Eq. 2):

$$r_s = \frac{r_a s_c(-R_n - Q_G) + \rho c_p(e_s - e_a) - r_a Q_E(s_c + \gamma)}{\gamma Q_E} \quad 7$$

4.2 GaFiR-ET-PT sub-module

Priestley-Taylor (PT) model to estimate potential evapotranspiration Q_{Ep} is written as

$$Q_{Ep}^{\text{PT}} = \alpha_{\text{PT}} \frac{s_c}{s_c + \gamma} (-Q_A), \quad 8$$

where α_{PT} is the PT coefficient, an empirically determined dimensionless correction. The mean value of α_{PT} for several vegetated surfaces and large water bodies is suggested to be 1.26 with a range between 1.08 and 1.34 (Priestley and Taylor, 1972).

The original PT function is valid for conditions of potential ET. In the case of actual ET, a linear relationship is assumed as

$$Q_{Ea} = aQ_{Ep} + b. \quad 9$$

5. GaFiR-Carbon module

5.1 GaFiR-C-Reco sub-module

Night-time Reco was used for the parameterization of the temperature dependence of Reco in Lloyd-Taylor function (Lloyd and Taylor, 1994; Falge et al., 2001):

$$R_{\text{eco}} = R_{\text{ref}} e^{E_0 \left(\frac{1}{T_{\text{ref}} - T_0} - \frac{1}{T - T_0} \right)}, \quad 10$$

where R_{ref} ($\mu\text{molm}^{-2}\text{s}^{-1}$) is the respiration at a reference temperature (T_{ref} , set as 283.15K);

E_0 (K) is the temperature sensitivity, assumed to be constant over the entire observation period; T_0 (K) is a constant value of 227.13 K.

5.2 GaFiR-C-GPP sub-module

The relationship between GPP and solar radiation (R_g) is expressed by the Michaelis-Menten function (Michaelis and Menten, 1913):

$$\text{GPP} = \frac{\alpha R_g \beta}{\alpha R_g + \beta}, \quad 11$$

where α ($\mu\text{mols}^{-1}\text{W}^{-1}$) is the initial slope of the curve, and β ($\mu\text{molm}^{-2}\text{s}^{-1}$) is the saturated carbon dioxide uptake rate when R_g is close to infinity.

5.2.1 GaFiR-C-GPP-Time sub-module

In GaFiR-C-GPP-Time sub-module, users can specify a width of the time window (t_w , in day). GaFiR segments the whole input dataset into data subsets according to t_w . The i -th subset contains the data from the day $(i-1)t_w + 1$ to the day it_w . Then the parameterization for Eq. 11 is applied individually for each subset and used for the estimation of GPP within the sub-period.

5.2.2 GaFiR-C-GPP-Temperature sub-module

In GaFiR-C-GPP-Temperature sub-module, users can specify an interval of temperature (T_e , in K). GaFiR segments the whole input dataset into data subsets according to T_e . The i -th subset contains the data from $T_{\min} + (i-1)T_e$ to $T_{\min} + iT_e$, where T_{\min} is the minimal temperature in the dataset. Then the parameterization for Eq. 11 is applied individually for each subset and used for the estimation of GPP within the range of temperature.

5.2.3 GaFiR-C-Leaf sub-module

The leaf-light response function is written as

$$\text{GPP} = \text{LAI}_{\text{act}} \frac{\alpha' R_g \beta'}{\alpha' R_g + \beta'}, \quad 12$$

where LAI_{act} is the mean LAI which is active in photosynthesis. The parameters α' and β' can be defined as specific light use efficiency and specific saturated GPP.

5.2.4 GaFiR-C-GPP-Time-VPD sub-module

A VPD factor is introduced to account for the stomatal response to dry air conditions. The saturated GPP in the light response function can be expressed by an exponential function of

VPD (Lasslop et al., 2010):

$$\beta = \begin{cases} \beta_0 e^{-k_\beta(\text{VPD}-\text{VPD}_0)} & , \text{VPD} > \text{VPD}_0 \\ \beta_0 & , \text{VPD} \leq \text{VPD}_0 \end{cases} \quad 13$$

where the threshold VPD_0 (hPa) is set to 10hPa.

6. Data, files and folders

6.1 Input data

GaFiR reads input data from the input folder (Fig. 2). If there are multiple files in the input folder, GaFiR only reads the first one. In order not to be confused, users are recommended to put only one single file in the input folder. GaFiR assumes that the input data should be reliable and quality-controlled, thus all the input data could be used for model parameterization.

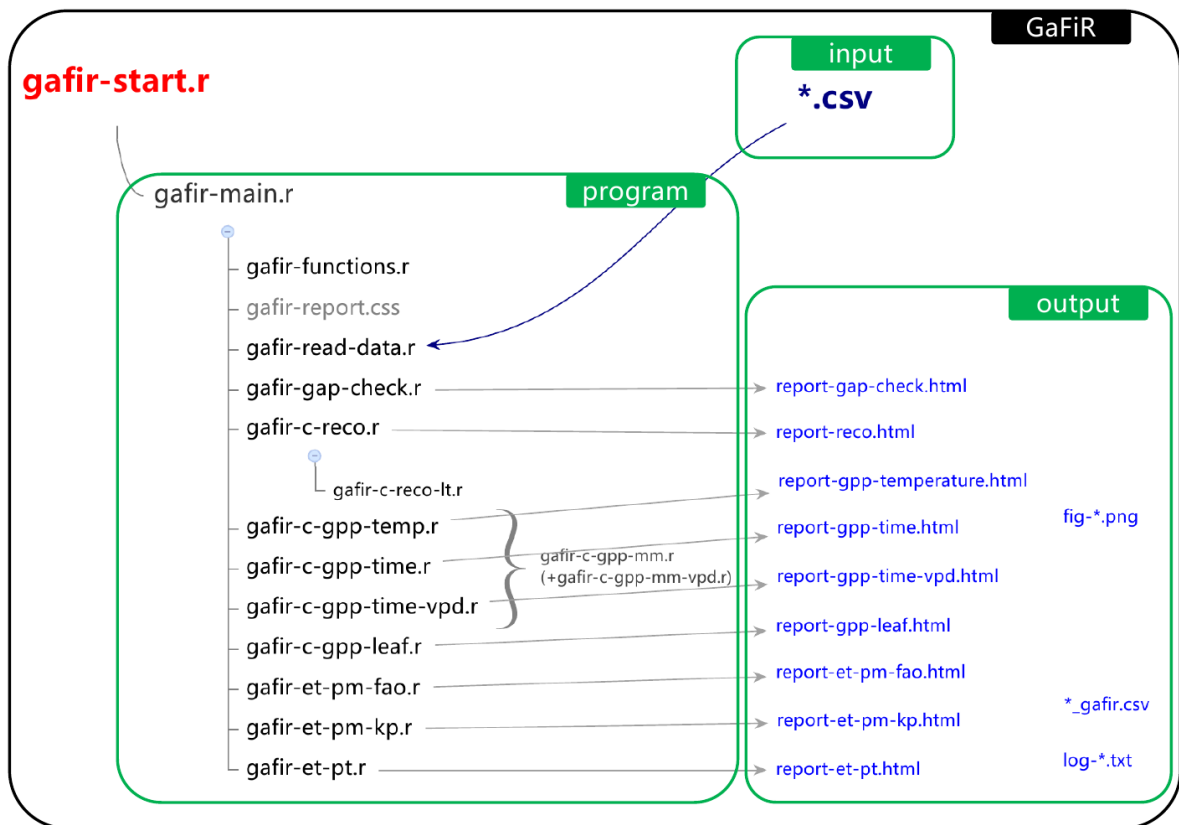


Fig. 2. Files and folders of GaFiR

Requirement of input variables depends on users' purpose. For instance, night-time respiration (equal to night-time NEE) and air temperature appear in Lloyd-Taylor function, therefore these two variables are required to parameterize Lloyd-Taylor function. Moreover,

time stamps are additionally required to indicate night-time. Required variables for different modules are listed in Table 2. Users can easily specify which modules to run. If some variable required by the chosen modules is missing, GaFiR will return error messages.

Table 2. Required variables in the input data file by parameterization in modules of GaFiR

Modules	Required variables
GaFiR-C-Reco	Time, temperature, NEE.
GaFiR-C-GPP-Temperature	Time, temperature, NEE, solar radiation
GaFiR-C-GPP-Time	Time, NEE, solar radiation
GaFiR-C-GPP-Time-VPD	Time, NEE, solar radiation, VPD
GaFiR-C-GPP-Leaf	Time, NEE, solar radiation, LAI
GaFiR-ET-PM-FAO	Time, net radiation, ground heat flux, temperature, canopy height, wind speed, VPD
GaFiR-ET-PM-KP	Time, net radiation, ground heat flux, temperature, canopy height, wind speed, VPD
GaFiR-ET-PT	Time, net radiation, ground heat flux, temperature, evapotranspiration

The input file must be a .csv file (i.e. ‘comma separated value’ file) with the first line as the header (i.e. column names). Users must strictly name the columns and use the required formats or units following Table x, because GaFiR recognizes variables by column names, and process input data as default formats or units. Letters in column names must all be in lower-case. Missing values An example file (demo.csv) could be found in the input folder for users to easily understand.

Table 3. Column names and formats or units of input variables.

Variable	column name	format or unit	Example
time	time	string, YYYY-mm-dd HH:MM:SS	2010-06-28 12:30:00
air temperature	temperature	numeric, °C	28
solar radiation	rg	numeric, W m ⁻²	-300
net ecosystem exchange of CO ₂	nee	numeric, any unit	0.048
net radiation	rn	numeric, W m ⁻²	-200
wind speed	u	numeric, m s ⁻¹	0.9
vapour pressure deficit	vpd	numeric, Pa	1000
canopy height	h	numeric, m	2.5
leaf area index	lai	numeric, demensionless	2
ground heat flux	g	numeric, W m ⁻²	50
evapotranspiration (latent heat flux)	qe	numeric, W m ⁻²	50

6.2 Starting/configuration file

GaFiR starts running from gafir-start.r. Besides the input data, gafir-start.r is normally the only file that users have to edit. In gafir-start.r, users can do the following:

- Configure the working directory of GaFiR. It should be the path of the GaFiR folder.

- Specify the strings or numbers which indicate missing records in the input data.
- Specify how to separate daytime and night-time. GaFiR uses the value of solar radiation above which the relative records will be processed as daytime data.
- Specify which modules to be run.
- Specify gap distribution of which variables would be checked.
- Specify widths of time windows or temperature windows, initial estimation of model parameters for non-linear regression, and other related information.

The information mentioned above is required by subsequent parameterization and gap-filling. The script `gafir-start.r` is self-explanatory with adequate explanatory comments to guide users. At the end of `gafir-start.r`, a program file named ‘`gafir-main.r`’ is called as an entrance to all the program files.

6.3 Program scripts

All program scripts are located in the ‘`program`’ folder. The role that each script file plays is as follows.

- **`gafir-main.r`** is the main or ‘trunk’ script, which calls all other ‘branch’ scripts and shows the main structure of GaFiR package. This script
 - creates a log file to record the processing procedure for future check and debugging,
 - sets the style file for .html reports,
 - installs and loads R packages and functions,
 - reads input data,
 - calls user-specified modules, and
 - saves the gap-filled data into a file named ‘*_gafir.csv’ the ‘output’ folder.
- **`gafir-funtions.r`** is a file containing functions used in GaFiR.
- **`gafir-report.css`** is a Cascading Style Sheets (CSS) file, which define the style how to display HTML elements such as colours, fonts, and tables, in reports.
- **`gafir-read-data.r`** reads the input data into R and do some pre-calculations.
- **`gafir-gap-check.r`** counts the number of data gaps exists in user’s required variables and

the length of each gap in the input data. It writes the result into a report file named ‘report-gap-check.html’.

- **gafir-c-reco.r** calls GaFiR-C-Reco sub-module. It parameterizes Lloyd-Taylor function (written in **gafir-c-reco-lt.r**) and fill the data-gaps in Reco. It writes results into a report file named ‘report-reco.html’.
- **gafir-c-gpp-temp.r**, **gafir-c-gpp-time.r**, and **gafir-c-gpp-time-vpd.r** call GaFiR-C-GPP-Temperature sub-module, GaFiR-C-GPP-Time sub-module, and GaFiR-C-GPP-Time-VPD sub-module, respectively. They parameterize Michaelis-Menten function (written in **gafir-c-gpp-mm.r**, and with VPD factor in **gafir-c-gpp-mm-vpd.r**) and fill the data-gaps in GPP. They write results into report files named ‘report-gpp-temperature.html’, ‘report-gpp-time.html’, and ‘report-time-vpd.html’ in the output folder.
- **gafir-c-gpp-leaf.r** calls GaFiR-C-GPP-Leaf sub-module, which parameterize the leaf-light response function and fill data-gaps in GPP. It writes results into a report file named as ‘report-gpp-leaf.html’.
- **gafir-et-pm-fao.r** calls GaFiR-ET-PM-FAO sub-module. It fills evapotranspiration data gaps by Penman-Monteith-FAO approach. It writes results into a report file named ‘report-et-pm-fao.html’.
- **gafir-et-pm-kp.r** calls GaFiR-ET-PM-KP sub-module. It fills evapotranspiration data gaps by Penman-Monteith-KP approach. It writes results into a report file named ‘report-et-pm-kp.html’.
- **gafir-et-pt.r** calls GaFiR-ET-PT sub-module. It fills evapotranspiration data gaps by Priestley-Taylor approach. It writes results into a report file named ‘report-et-pm-kp.html’.

6.4 Output files

GaFiR produces fruitful results in forms of explanatory text, data table, and figures. These results are well organized mainly in three types of files in the ‘output’ folder: some report files, a log file and a gap-filled data file. Although many .png figures appear in the output folder after GaFiR is successfully run, users do not have to click them one by one because they are all embedded in those .html report files.

Report files are created by related R scripts (see Chapter 6.3). With a filename extension as ‘html’, they can be opened by any web browser such as IE, Chrome, and Firefox. An example of a report file is shown in Fig. 3. Users can find in report files detailed information of the fitted parameters of models, the goodness-of-fit measures, and related figures. Data tables in

the report files can be easily copied and pasted into other spreadsheet software such as MS Excel and OpenOffice Calc.

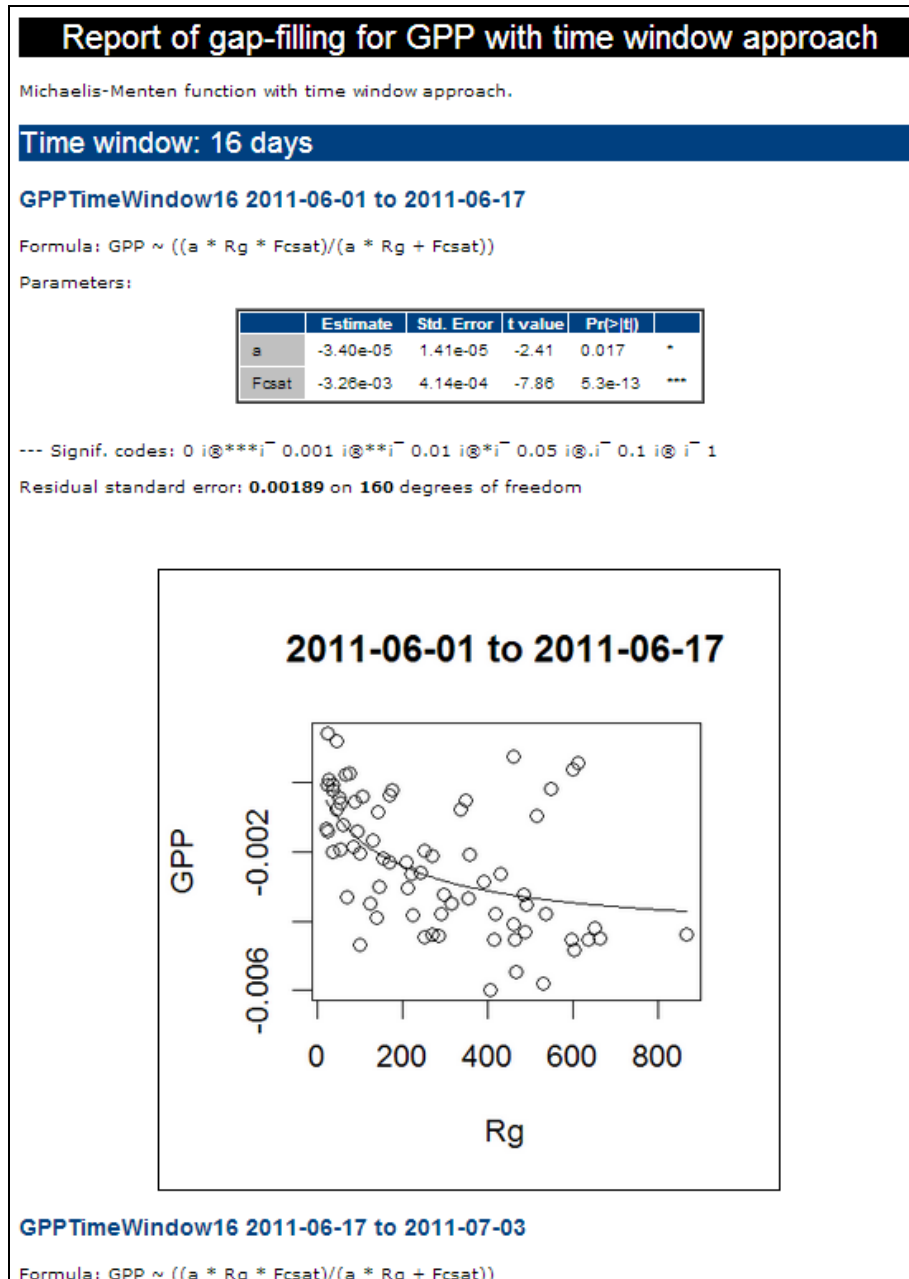


Fig. 3. Screenshot of an example of GaFiR reports

The log file contains all screenshot output in R console (Fig. 4). The file name begins with 'log' followed by a time stamp. It can be opened by any text editor. It is helpful for users to track the behavior of GaFiR, finding errors, and debugging.

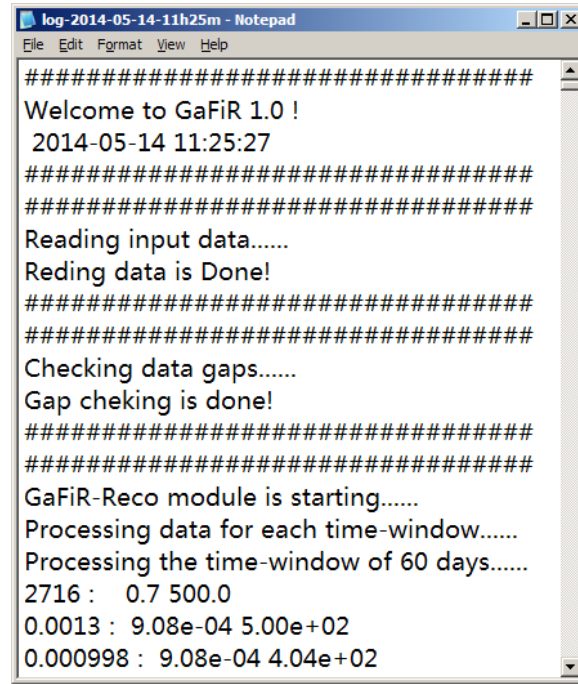


Fig. 4. Screenshot of an example of the log file

Gap-filled flux data are appended to the input data frame and saved in a .csv file with a filename beginning with the name of the input file followed by ‘_gafir’. It can be opened by any text editor or spreadsheet software. Column names and meanings of variables in ‘*_gafir.csv’ are listed in Table 4.

Table 4. Column names and meanings of output variables.

column name	example	meaning
‘reco-’ followed by a number n	reco-15	Respiration simulated with a time window of n days
‘gpp’	gpp	Daytime GPP calculated using observed NEE and simulated Reco
‘gpp-te-’ followed by a number n	gpp-te-4	Daytime GPP calculated with a temperature window of n K
‘gpp-tw-’ followed by a number n	gpp-tw-2	Daytime GPP calculated with a time window of n days
‘gpp-tw-vpd-’ followed by a number n	gpp-tw-vpd-4	Daytime GPP calculated with the VPD factor and a time window of n K
‘gpp-leaf’	gpp-leaf	Daytime GPP calculated by leaf-light response function
‘qe-pm-fao’	qe-pm-fao	Evapotranspiration simulated by PM-FAO method
‘qe-pm-kp-tw-’ followed by a number n	qe-pm-kp-tw-4	Evapotranspiration simulated by PM-KP method with a time window of n days
‘qe-pt’	qe-pt	Evapotranspiration simulated by PT method

7. References

- Allen, R., 2005. Penman-Monteith equation, in: Hillel, D. (Ed.), *Encyclopedia SoilsEnvironment*. Elsevier, Oxford, 180–188.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. *Crop evapotranspiration: Guidelines for computing crop water requirements*. FAO Irrigation and drainage Paper 56. FAO, Rome, 300 pp.
- Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology* 107, 43–69.
- Foken, T., Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. *Agricultural and Forest Meteorology* 78, 83–105.
- Grothendieck, G., 2012. nls2: Non-linear regression with brute force. <http://CRAN.R-project.org/package=nls2>.
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biology* 16, 187–208.
- Lecoutre, E., 2003. The R2HTML Package. *R News* 3, 33–36.
- Lloyd, J., Taylor, J., 1994. On the temperature dependence of soil respiration. *Functional Ecology* 315–323.
- Lüers, J., Detsch, F., Zhao, P., 2014. Application of a multi-step error filter for post-processing atmospheric flux and meteorological basic data. *Universität Bayreuth, Abt. Mikrometeorologie*, in print, ISSN 1614-8916, Arbeitsergebnisse 58, 22 pp.
- Mauder, M., Foken, T., 2011. Documentation and instruction manual of the eddy covariance software package TK3. *Universität Bayreuth, Abt. Mikrometeorologie*, Print, ISSN 1614-8916, Arbeitsergebnisse 46. 58 pp.
- Michaelis, L., Menten, M.L., 1913. Die kinetik der invertinwirkung. *Biochemische Zeitschrift* 49, 352.
- Monteith, J., 1965. Evaporation and environment, in: *Symposia SocietyExperimentalBiology*. 205 pp.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I — A discussion of principles. *Journal of Hydrology* 10, 282–290.
- Papale, D., 2012. Data Gap Filling, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), *EddyCovariance*, SpringerAtmosphericSciences. Springer, Dordrecht, 159-172.
- Priestley, C., Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* 100, 81–92.
- R Development Core Team, 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havr ánkov á K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology* 11, 1424–1439.

- Ruppert, J., Mauder, M., Thomas, C., Lüers, J., 2006. Innovative gap-filling strategy for annual sums of CO₂ net ecosystem exchange. *Agricultural and Forest Meteorology* 138, 5–18.
- Sarkar, D., 2008. *Lattice: Multivariate Data Visualization with R*. Springer, New York.
- Sonntag, D., 1990. Important new values of the physical constants of 1986, vapour pressure formulations based on the ITS-90, and psychrometer formulae. *Zeitschrift für Meteorologie* 70, 340–344.
- Willmott, C.J., 1981. On the validation of models. *Physical Geography* 2, 184–194.
- Zambrano-Bigiarini, M., 2014. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series.
- Zhao, P., Lüers, J., 2012. Improved determination of daytime net ecosystem exchange of carbon dioxide at croplands. *Biogeosciences Discussions* 9, 2883–2919.

Work report list

Volumes in the series, University of Bayreuth, Department of Micrometeorology

Nr	Author(s)	Title	Year
01	Foken	Der Bayreuther Turbulenzknecht	01/1999
02	Foken	Methode zur Bestimmung der trockenen Deposition von Bor	02/1999
03	Liu	Error analysis of the modified Bowen ratio method	02/1999
04	Foken et al.	Nachfrostgefährdung des ÖBG	03/1999
05	Hierteis	Dokumentation des Experimentes Dlouhá Louka	03/1999
06	Mangold	Dokumentation des Experimentes am Standort Weidenbrunnen, Juli/August 1998	07/1999
07	Heinz et al.	Strukturanalyse der atmosphärischen Turbulenz mittels Wavelet-Verfahren zur Bestimmung von Austauschprozessen über dem antarktischen Schelfeis	07/1999
08	Foken	Comparison of the sonic anemometer Young Model 81000 during VOITEX-99	10/1999
09	Foken et al.	Lufthygienisch-bioklimatische Kennzeichnung des oberen Egertales, Zwischenbericht 1999	11/1999
10	Sodemann	Stationsdatenbank zum BStMLU-Projekt Lufthygienisch-bioklimatische Kennzeichnung des oberen Egertales	03/2000
11	Neuner	Dokumentation zur Erstellung der meteorologischen Eingabedaten für das Modell BEKLIMA	10/2000
12	Foken et al.	Dokumentation des Experimentes VOITEX-99	10/2000
13	Bruckmeier et al.	Documentation of the experiment EBEX-2000, July 20 to August 24, 2000	01/2001
14	Foken et al.	Lufthygienisch-bioklimatische Kennzeichnung des oberen Egertales	02/2001
15	Göckede	Die Verwendung des Footprint-Modells nach Schmid (1997) zur stabilitätsabhängigen Bestimmung der Rauigkeitslänge	03/2001
16	Neuner	Berechnung der Evaporation im ÖBG (Universität Bayreuth) mit dem SVAT-Modell BEKLIMA	05/2001
17	Sodemann	Dokumentation der Software zur Bearbeitung der FINTUREX-Daten	08/2002
18	Göckede et al.	Dokumentation des Experiments STINHO-1	08/2002
19	Göckede et al.	Dokumentation des Experiments STINHO-2	12/2002
20	Göckede et al.	Characterisation of a complex measuring site for flux measurements	12/2002
21	Liebenthal	Strahlungsmessgerätevergleich während des Experiments STINHO-1	01/2003
22	Mauder et al.	Dokumentation des Experiments EVA_GRIPS	03/2003
23	Mauder et al.	Dokumentation des Experimentes LITFASS-2003, Dokumentation des Experimentes GRASATEM-2003	12/2003
24	Thomas et al.	Documentation of the WALDATEM-2003 Experiment	05/2004
25	Göckede et al.	Qualitätsbegutachtung komplexer mikrometeorologischer Messstationen im Rahmen des VERTIKO-Projekts	11/2004
26	Mauder & Foken	Documentation and instruction manual of the eddy covariance software package TK2	12/2004
27	Herold et al.	The OP-2 open path infrared gas analyser for CO ₂ and H ₂ O	01/2005
28	Ruppert	ATEM software for atmospheric turbulent exchange measurements using eddy covariance and relaxed eddy accumulation systems and Bayreuth whole-air REA system setup	04/2005
29	Foken (Ed.)	Klimatologische und mikrometeorologische Forschungen im Rahmen des Bayreuther Institutes für Terrestrische Ökosystemforschung (BITÖK), 1989-2004	06/2005
30	Siebeke & Serafimovich	Ultraschallanemometer-Überprüfung im Windkanal der TU Dresden 2007	04/2007
31	Lüers & Bareiss	The Arctic Turbulence Experiment 2006 PART 1: Technical documentation of the ARCTEX 2006 campaign, May, 2nd to May, 20th 2006	07/2007
32	Lüers & Bareiss	The Arctic Turbulence Experiment 2006 PART 2: Visualization of near surface measurements during the ARCTEX 2006 campaign, May, 2nd to May, 20th 2006	07/2007
33	Bareiss & Lüers	The Arctic Turbulence Experiment 2006 PART 3: Aerological measurements during the ARCTEX 2006 campaign, May, 2nd to May, 20th 2006	07/2007
34	Metzger & Foken et al.	COPS experiment, Convective and orographically induced precipitation study, 01 June 2007 – 31 August 2007, Documentation	09/2007

35	Staudt & Foken	Documentation of reference data for the experimental areas of the Bayreuth Centre for Ecology and Environmental Research (BayCEER) at the Waldstein site	11/2007
36	Serafimovich et al.	ExchanGE processes in mountainous Regions (EGER) - Documentation of the Intensive Observation Period (IOP1), September, 6th to October, 7th 2007	01/2008
37	Serafimovich et al.	ExchanGE processes in mountainous Regions (EGER) - Documentation of the Intensive Observation Period (IOP2), June, 1st to July, 15th 2008	10/2008
38	Siebicke	Footprint synthesis for the FLUXNET site Waldstein/Weidenbrunnen (DE-Bay) during the EGER experiment.	12/2008
39	Lüers & Foken	Jahresbericht 2008 zum Förderprojekt 01879 - Untersuchung der Veränderung der Konzentration von Luftbeimengungen und Treibhausgasen im hohen Fichtelgebirge 2007 - 2013	01/2009
40	Lüers & Foken (Eds.)	Proceedings of the International Conference of "Atmospheric Transport and Chemistry in Forest Ecosystems" Castle of Thurnau, Germany, Oct 5 to Oct 8, 2009	10/2009
41	Biermann et al.	Mesoscale circulations and energy and gas exchange over the Tibetan Plateau - Documentation of the micrometeorological experiment, Nam Tso, Tibet , 25 June 2009 - 08 August 2009	12/2009
42	Foken & Falke	Documentation and Instruction Manual for the Krypton Hygrometer Calibration Instrument	01/2010
43	Lüers & Foken	Jahresbericht 2009 zum Förderprojekt 01879 - Untersuchung der Veränderung der Konzentration von Luftbeimengungen und Treibhausgasen im hohen Fichtelgebirge 2007 - 2013	07/2010
44	Biermann & Leipold (Ed.)	Tibet Plateau Atmosphere-Ecology-Glaciology Cluster Joint <i>Kobresia</i> Ecosystem Experiment: Documentation of the first Intensive Observation Period (IOP 1) summer 2010 in Kema, Tibet	01/2011
45	Zhao et al.	Complex TERRain and ECOlogical Heterogeneity (TERRECO) WP 1-02: Spatial assessment of atmosphere-ecosystem exchanges via micrometeorological measurements, footprint modeling and mesoscale simulations - Documentation of the Observation Period, May 12th to Nov. 8th, 2010, Haeon, South Korea	03/2011
46	Mauder et al.	Documentation and Instruction Manual of the Eddy-Covariance Software Package TK3 2011	05/2011
47	Serafimovich et al.	ExchanGE processes in mountainous Regions (EGER)- Documentation of the Intensive Observation Period (IOP3) June, 13th to July, 26th 2011 2011	11/2011
48	Hübner et al.	Documentation and Instruction Manual for the Horizontal Mobile Measuring System (HMMS) 2011	12/2011
49	Lüers et al.	The Arctic Turbulence Experiment 2009: Long-term measurements of near-surface turbulent fluxes in the Arctic environment - Documentation of the additional laser Scintillometer measurement campaign 2009 at Bayelva on Svalbard, August 2009	02/2012
50	Foken	Klimawanderweg auf der Landesgartenschau in Bamberg 2012	04/2012
51	Ruppert et al.	Whole-air relaxed eddy accumulation for the measurement of isotope and trace-gas fluxes	05/2012
52	Foken	Jahresbericht 2010-11 zum Förderprojekt 01879 -Untersuchung der Veränderung der Konzentration von Luftbeimengungen und Treibhausgasen im hohen Fichtelgebirge 2007 - 2013	12/2012
53	Gerken et al.	Documentation of the Atmospheric Boundary Layer Experiment, Nam Tso, Tibet, 08th of July - 08th of August 2012	03/2013
54	Biermann (Ed.)	Tibet Plateau Atmosphere-Ecology-Glaciology Cluster Joint <i>Kobresia</i> Ecosystem Experiment: Documentation of the 2nd Intensive Observation Period (IOP 2) summer 2012 in KEMA, Tibet	05/2013
55	Babel et al.	Documentation of the EVENT-HMMS Experiment 2012 - Microclimatological effects of rain-out shelters within EVENT II	06/2013
56	Lüers et al.	160 Jahre Bayreuther Klimazeitreihe - Homogenisierung der Bayreuther Lufttemperatur- und Niederschlagsdaten	06/2014
57	Babel	An R routine for the simplified application of a footprint-based characterisation of a complex measuring site for flux measurements	06/2014
58	Lüers et al.	Application of a multi-step error filter for post-processing atmospheric flux and meteorological basic data.	06/2014
59	Zhao et al.	GaFiR: a gap-filling package for ecosystem-atmosphere carbon dioxide flux and evapotranspiration data	06/2014