A Quality Assessment Scheme for Long-term Eddy-Covariance Measurements

Matthias Mauder, Matthias Cuntz, Clemens Drüe, Alexander Graf, Corinna Rebmann, Marius Schmidt, Rainer Steinbrecher
Outline

- Introduction
- Quality assessment algorithms
  - Tests on high-frequency data
  - Tests on statistics, flux calculation and corrections
  - Quantification of errors and uncertainty estimates
  - The quality assessment scheme
  - Test data sets
- Results and discussion
  - Quality flags
  - Random errors and noise
  - Systematic error and energy balance ratio
  - Footprint analysis
  - Further tests
  - Effectiveness of the quality assessment scheme
- Conclusions
Introduction – Eddy Covariance

Objective:
- to measure the transport/flux of air properties (e.g. temperature) or air constituents (e.g. GHGs) between an ecosystem at the earth’s surface and the atmosphere

Advantages:
- non-destructive
- no disturbance of the exchange conditions
- area-averaging on the ecosystem-scale
- quasi-continuous

Assumptions (Limitations):
- Atmospheric exchange is fully turbulent, zero mean vertical wind
- Taylor hypothesis: space and time interchangeable
- Horizontal homogeneity, stationarity/steady state conditions
- All transporting eddy scales are captured
- Constant flux layer
Measurements at a single point can represent the ecosystem flux from an upwind area

(Burma and Anderson, 2010)
Introduction – Eddy Covariance

Flux Calculation (assume stationarity)

\[ F = \overline{wq} = \overline{wXq} + \overline{w'q'} \]

\[ \overline{w} = 0, \text{ assume horizontal homogeneity, } 30 \text{ – } 60 \text{ min averaging} \]
Introduction – Challenges for QA scheme

- Extensive amount of data
  - Per site @ 20 Hz: 72000 data lines per hour, 1728000 lines per day, 51840000 lines per month, $1.89216 \times 10^{10}$ lines per year
  - > 10 sites operational or planned within TERENO running for > 10 years

- Detection of instrumental failure
  - coverage by rain droplets, dew or frost

- Validity of assumptions for the EC method
  - well-developed turbulence, stationarity, zero mean vertical wind

- Quantification of measurement errors
  - noise, random, systematic

- Representativeness of the flux estimate for the targeted ecosystem
  - footprint
Introduction – State of the art

  - Tests on stationarity, well-developed turbulence, zero mean vertical wind
    => flag system 1 – 9

- Vickers and Mahrt (1997)
  - Various tests on raw data => system of hard and soft flags

- Papale et al. (2006)
  - Test on fluxes without knowledge of raw data, plus uncertainty due to corrections

- Error estimates by various authors
  - noise, random, systematic
Introduction – Goals

- Provide *comparable* flux data with *simple* quality flags and *quantitative error* estimates (random and systematic)

- Create a *comprehensive* quality assessment *scheme* of tests and algorithms, which can be *applied automatically* to long-term measurements (as strict as necessary and as simple/lean/efficient as possible)

- The selected tests and criteria should be as *fundamental* as possible to allow a *wide applicability* to different site conditions

- Provide *pre-processed additional information* about the fluxes in a *standardised* way, which may *assist the user* in deciding how to use and filter the data for his/her specific application.
Quality assessment algorithms

Quality assessment algorithms can be subdivided into three parts

- Tests on high-frequency data
- Tests on statistics, fluxes and corrections
- Quantification of error/uncertainty estimates
Algorithms - Tests on high-frequency data

- Modern micrometeorological instruments have **internal quality tests** and provide a diagnostic flag, e.g. Campbell CSAT3 or Licor LI-7500.

- **Spike test** based on Median Absolute Deviation (MAD), which is a robust measure for outlier or spike detection

- Limits for **instrumental plausibility** screening of the high-frequency data

High-frequency data points rejected by the above tests were replaced by an error code (NaN)

Alternatively for short data gaps:  
- repeat the last measured value  
- linear interpolation
Algorithms – Tests on statistics

- **Consensus about flux calculation and correction** procedures is a pre-condition for the quality assessment scheme.
  - We follow the recommendations of Lee et al. (2004), which is in accordance with procedure used in the CarboEurope-IP software comparison of Mauder et al. (2008)

- **Simple flag system**, which follows the agreement of the 2nd CarboEurope-IP QA/QC workshop (Mauder and Foken, 2004):
  - Flag 0 – high quality data, use in fundamental research possible
  - Flag 1 – moderate quality data, no restrictions for use in long-term observation programs
  - Flag 2 – low data quality, to be discarded


Algorithms – Tests on statistics

- Minimum of 90% of the raw data available per averaging period

- Test the assumptions of the EC method (simplified after FW96):
  - Stationarity (>30% = \textit{flag} 1; >75% = \textit{flag} 2)
  - Well-developed turbulence (>30% then \textit{flag} 1; >75% = \textit{flag} 2)
  - Zero mean vertical wind velocity
  - (>0.10 m s\(^{-1}\) then former \textit{flag} +1; >0.15 then \textit{flag} 2)

- Flux conversions and corrections cause an interdependence of the simultaneously measured flux estimates:
  - if \textit{flag}\(\lambda\)E == 2 then former \textit{flag}H + 1
  - else if \textit{flag}H == 2 then former \textit{flag}\(\lambda\)E + 1
  - else if \textit{flag}\(\lambda\)E == 2 or \textit{flag}H == 2 then former \textit{flag}NEE + 1

- All these thresholds may appear somewhat arbitrary but are based on many years of experience of FW96, the authors of this work, and in part have been applied by >500 users of TK2/TK3 worldwide
Algorithms – Quantification of errors and uncertainty

1. Instrumental noise

2. Random error

3. Systematic error

4. Representativity – Source area
Algorithms – Quantification of errors and uncertainty

- **Instrumental noise**
  - Spectral contribution in the high-frequency range
  - Non-correlated within time series

- Modified after Lenschow et al. (2000)

  - The noise error is **only present** in the first term of auto-covariance functions **but not** in the first term of cross-covariance functions. Thus, the noise error of the covariance is calculated by **error propagation**.
Algorithms –
Quantification of errors and uncertainty

- Random error
  - Generally: \( \sim \frac{1}{\sqrt{\# \text{ independent observations}}} \)
  - Beware: \( \# \text{ independent observations} \neq \# \text{ observations} \), because the time series are auto- and cross-correlated.

- Modified after Finkelstein & Sims (2001)
    - The statistical variance of a covariance can be expressed as function of its auto-covariances and cross-covariance
    - An autocovariance is only meaningful for data without trend
      \( \Rightarrow \text{detrending} \) through high-pass filter \textbf{before} calculation of random error
Algorithms –
Quantification of errors and uncertainty

- **Systematic error**
  - In the presence of **large eddies**, the covariance at a single point **does not represent** the total surface flux
  - Why?
    - averaging interval too short
    - large eddies often “attached” to surface heterogeneities and do not propagate with the mean wind

- Capturing all relevant scales of biosphere-atmosphere exchange is subject of current research (among others: Helmholtz Young Investigators Group)

- Systematic error can only be determined indirectly
Algorithms –
Quantification of errors and uncertainty

Energy balance closure problem

\[ R_n - G = \lambda E + H \]

In-situ measurements worldwide show: underestimation of the turbulent transport \((\lambda E + H)\) by 10-30%

Strong evidence from 180 sites (Stoy and Mauder, 2011):
transport on large scales \((10^2 - 10^4 \text{ m})\) not captured by measurements

Helmholtz Young Investigators Group

Integrated research approach which interlinking modeling and measurements:

- **Tower-measurements**
  - Helmholtz-TERENO Observatories

- **Aircraft-measurements**
  - NRC Twin Otter (Canada)

- **Ground-based remote sensing**
  - Doppler LiDAR

- **Modeling**
  - Large-Eddy-Simulation LES
    (Uni Hannover, DLR)

**Goals**

- **Quantification/parameterization** of the missing fluxes
- Improve the **understanding of the effects** of large-scale transport on flux measurements
Algorithms –
Quantification of errors and uncertainty

- Systematic error

- Energy balance ratio \( EBR \)

\[
EBR = \frac{\sum_{i=1}^{K} (H_i + \lambda E_i)}{\sum_{i=1}^{K} (R_n - G_i - J_i)}
\]

- only applicable for daytime \( (R_g > 20 \text{ W m}^{-2}) \) and for 1-day averaging

- the error is defined as

\[
\sigma_{F}^{sys} = F \cdot \left( \frac{1}{EBR} - 1 \right) \quad R_g > 20 \text{ W m}^{-2}
\]

\[
\sigma_{F}^{sys} = \frac{n}{a} \quad R_g \leq 20 \text{ W m}^{-2}
\]
Footprint = spatial filter, “field of view”

\[ F(x) = \int \int \mathcal{Q}_s(x') \cdot f(x-x') \cdot dx' = \mathcal{Q}_s \ast f \]

(convolution of the source distribution, \( \mathcal{Q}_s \), with the footprint, \( f \))

Inputs:
- \( z_m \)
- \( u* \)
- \( z_0 \)
- \( \sigma_w \)
- \( h \)
- \( \sigma_v \)

Footprint-Model


\[ f = \frac{1}{\Gamma(\mu)} \frac{\xi^\mu}{x^{1+\mu}} e^{-\xi/x} \]

\[ \xi(z) = \frac{U z^r}{r^2 \kappa}, \]  
flux length scale

\[ \mu = \frac{1 + m}{r}, \]  
constant

\[ r = 2 + m - n \]  
shape factor

- Analytical
- Numerically robust (log-profiles into power laws)
- Computationally fast and reliable enough to be applied for every 30-min interval in a long-term measurement programme
The complete quality assessment scheme

Tests on high-frequency data
- Instrument diagnostic flags (CSAT3 0-63, LI7500 240-251)
- Instrumental/plausibility limits (site-specific)
- Spike-detection with MAD-test, z = 7

Tests on statistics, flux calculation + corrections
- Maximum number of missing values: ≤ 10%: flag = 0, >10%: flag = 2
- Calculate averages, variances and covariances on 30 min and 5 min basis with automatic lag correction by maximising covariances
- Stationarity test covariances (FW96, < 30%: flag = 0, < 75%: flag = 1)
- Flux calculation, conversions and corrections applied: Planar fit, Schotanus, spectral losses (Moore or Eugster&Senn), WPL, iteration
- ITC test (FW96, MF04, < 30%: flag = 0, < 100%: flag = 1)
- Combination of flags according to MF04
  - w after Planar fit > 0.10 m s⁻¹: former flag +1, > 0.15 m s⁻¹: flag = 2
  - Interdependence of flags due to corrections/conversions:
    if flagE == 2 then former flagH + 1
    else if flagH == 2 then former flagLE + 1
    else if flagLE == 2 or flagH == 2 then former flagNEE + 1

Quantification of errors/uncertainty estimates
- Random error: use Finkelstein&Sims (2001), but on high-pass filtered (detrended) time series and on full period. Consider Gaussian error propagation for u.; take relative error of covariances and multiply with corrected fluxes (neglecting error propagation through corrections)
- Noise error after Lenschow et al. (2000)
- Systematic error: Very low-frequency eddies and non-propagating eddies lead to a flux underestimation and consequent lack of energy balance closure, only applicable for $R_g > 20$ W m⁻² for one day; the error is \( \text{flux}^* (1/\text{EBR} - 1) \)
- Footprint: Kormann&Meixner (2001) using measured \( \langle u \rangle \) instead of \( z_0 \)-dependent wind profile; calculate percentage of flux contribution for several targets of interest; recommended percentage threshold 70% depending on user’s requirements

31.01.2012

Matthias Mauder – A Quality Assessment Scheme for Long-term Eddy-Covariance Measurements

Atmospheric Environmental Research KIT/IMK-IFU
The test data sets

Overview of the test data sets

<table>
<thead>
<tr>
<th>Site name</th>
<th>Operator</th>
<th>Ecosystem</th>
<th>Measurement height (a.g.l.*)</th>
<th>Sensor combination</th>
<th>Data period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selhausen</td>
<td>FZJ</td>
<td>agricultural land, sugar beet</td>
<td>2.5 m</td>
<td>CSAT3/LI-7500</td>
<td>01/06/2011 – 30/06/2011</td>
</tr>
<tr>
<td>Wetzstein</td>
<td>MPI-BGC</td>
<td>Spruce forest on low mountain range</td>
<td>30.0 m</td>
<td>Solent-R3/LI-6262</td>
<td>15/07/2006 – 13/08/2006</td>
</tr>
</tbody>
</table>

* a.g.l.: above ground level
Results - Percentage of available data with flag 0, 1 and 2

![Graph showing the percentage of available data with flags 0, 1, and 2 for different locations.](image-url)

Legend:
- Flag 0
- Flag 1
- Flag 2

Locations:
- Fendt
- Graswang
- Lackenberg
- Selhausen
- Wetzstein

Percentage of data (%)

- 0
- 20
- 40
- 60
- 80
- 100

Flags represented by different patterns for each location.
Results –
Relative random flux error (%) vs. quality flag

- \( \tau \)
- \( H \)
- \( \lambda E \)
- NEE

Fendt, Graswang, Lackenberg, Selhausen, Weizstein
<table>
<thead>
<tr>
<th></th>
<th>Fendt</th>
<th>Graswang</th>
<th>Lackenberg</th>
<th>Selhausen</th>
<th>Wetzstein</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ</td>
<td>0.39</td>
<td>0.38</td>
<td>0.12</td>
<td>0.34</td>
<td>0.08</td>
</tr>
<tr>
<td>H</td>
<td>0.54</td>
<td>0.64</td>
<td>0.31</td>
<td>0.95</td>
<td>0.30</td>
</tr>
<tr>
<td>λE</td>
<td>0.45</td>
<td>0.46</td>
<td>0.35</td>
<td>0.67</td>
<td>0.53</td>
</tr>
<tr>
<td>NEE</td>
<td>0.61</td>
<td>0.47</td>
<td>0.48</td>
<td>0.79</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Even, if this error may be small in relation to the magnitude of the flux on average, this can be different for single estimates, e.g. the maximum relative noise error that we found for the five test data sets was 9%.

Results –
Systematic errors in percent (%)

- Fendt, monthly 24%
- Lackenberg, monthly 2%
- Graswang, monthly 39%
Results – Footprint: flux contribution from target

Flux contribution from target land cover (%)

<table>
<thead>
<tr>
<th>Location</th>
<th>Fendt</th>
<th>Graswang</th>
<th>Lackenberg</th>
<th>Selhausen</th>
<th>Wetzstein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Possible further tests, not implemented in the quality assessment scheme

- Stationarity-test based on detrending vs. block-averaging

![Graph showing relative differences in various parameters (τ, H, λE, NEE) for flag 0 and flag 1.](image-url)
Possible further tests, not implemented in the quality assessment scheme

- Skewness and Kurtosis

Excerpt of CO₂ flux time series at the site Selhausen with visually identified strongly suspect fluxes (encircled).


Possible further tests, not implemented in the quality assessment scheme

Skewness and Kurtosis

Performance of the quality assessment scheme presented in this paper and several potential additional flagging schemes on visually determined suspect fluxes.

<table>
<thead>
<tr>
<th></th>
<th>presented scheme</th>
<th>skewness and kurtosis flags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a) Flag 1 + 2</td>
<td>c) VM97</td>
</tr>
<tr>
<td></td>
<td>b) Flag 2</td>
<td>d) bimodal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>e) K &gt; β + 3</td>
</tr>
<tr>
<td>α: detections</td>
<td>3.0%</td>
<td>1.7%</td>
</tr>
<tr>
<td>β: missed</td>
<td>0.6%</td>
<td>1.9%</td>
</tr>
<tr>
<td>γ: false detection</td>
<td>24.4%</td>
<td>30.8%</td>
</tr>
<tr>
<td>δ: correct null</td>
<td>71.9%</td>
<td>65.6%</td>
</tr>
<tr>
<td>TSS: true skill score</td>
<td>0.59</td>
<td>0.16</td>
</tr>
</tbody>
</table>

α: detections 3.0% 1.5% 1.7% 0.9% 0.3% 3.1%
β: missed 0.6% 2.1% 1.9% 2.7% 3.3% 0.5%
γ: false detection 24.4% 7.3% 30.8% 5.3% 8.2% 26.7%
δ: correct null 71.9% 89.1% 65.6% 91.1% 88.2% 69.6%
TSS: true skill score 0.59 0.33 0.16 0.20 0.01 0.59
Effectiveness of the quality assessment scheme

Results of the MAD-based outlier test (Papale et al. 2006) after application of the quality assessment scheme (number of detected values/available data after QC). Before QC, 1440 data were available per test data set.

<table>
<thead>
<tr>
<th></th>
<th>Fendt</th>
<th>Graswang</th>
<th>Lackenberg</th>
<th>Selhausen</th>
<th>Wetzstein</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>1/1277</td>
<td>5/1348</td>
<td>0/1044</td>
<td>1/1383</td>
<td>2/1395</td>
</tr>
<tr>
<td>$H$</td>
<td>1/916</td>
<td>7/1121</td>
<td>21/882</td>
<td>9/1262</td>
<td>19/1153</td>
</tr>
<tr>
<td>$\lambda E$</td>
<td>2/820</td>
<td>5/850</td>
<td>7/762</td>
<td>13/1127</td>
<td>18/1059</td>
</tr>
<tr>
<td>$NEE$</td>
<td>3/757</td>
<td>9/888</td>
<td>8/765</td>
<td>7/1113</td>
<td>2/1064</td>
</tr>
</tbody>
</table>

Conclusions

- An effective automatic quality assessment of long-term EC measurements was presented.
- This scheme assesses the data quality based on *fundamental* information contained in the measured data directly: no *a-posteriori* checks required.
- Regular/daily visual inspection of the data is still required.
- Expert knowledge can help to retain more data from the original data set.
- The presented quality assessment scheme will be the standard for EC-measurements within TERENO. However, further development is possible when new scientific findings become evident.
- The algorithms are newly implemented in the software TK3.1 (freeware) and will be available as a pre- and postprocessor for the software package ECpack and as part of EddySoft.
Thank you for your attention!