ALGORITHM THEORETICAL BASIS DOCUMENT: QUALITY FLAGS AND QUALITY METRICS FOR IS DATA PRODUCTS

PREPARED BY	ORGANIZATION	DATE
Derek Smith	FIU	07/29/2013
Stefan Metzger	FIU	07/29/2013
Cove Sturtevant	IS	03/31/2020

APPROVALS (Name)	ORGANIZATION	APPROVAL DATE
Kate Thibault	SCI	04/13/2020

RELEASED BY (Name)	ORGANIZATION	RELEASE DATE
Anne Balsley	CM	04/14/2020

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by Battelle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CHANGE RECORD

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	09/17/2013	ECO-01303	Initial Release
B	04/14/2020	ECO-06425	Add science review flag, generalize document to all of IS (not just TIS) and update how quality flags and metrics are reported to the end user to reflect current practice

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

This document specifies how sensor flags and results from quality assurance and quality control (QA/QC) tests will be summarized to inform users of data quality.

1.1 Purpose

Each NEON Instrumented Systems (IS) data product (DP) has the ability to be flagged by various QA/QC tests and sensor flags. As such the need arises to assess these quality flags (QFs) and provide an overall indication of product quality to data users. QFs can either be assessed for individual measurements or a summarized as a percent of measurements, e.g. when measurements become averaged for a time period. This document describes the theoretical background and entire algorithmic process used to produce information on the quality of NEON IS data.

1.2 Scope

Information presented here relates to the data quality of NEON's IS Data Products. A final QF will be generated by NEON that determines whether a DP is *valid* or *invalid*. This final quality flag will be based on an assessment of a DP's individual QF results. Any sensor specific details are specified in a sensor's specific algorithm theoretical based document (ATBD), in the algorithm implementation section.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

Applicable documents contain information that shall be applied in the current document. Examples are higher level requirements documents, standards, rules and regulations.

AD [01]	NEON.DOC.011081	ATBD: QA/QC Plausibility Testing

2.2 Reference Documents

Reference documents contain information complementing, explaining, detailing, or otherwise supporting the information included in the current document.

RD [01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.011009	FIU Dataflow and QA Plan
RD [03]	NEON.DOC.000243	NEON Glossary of Terms

2.3 Acronyms

Acronym	**Explanation**
ATBD	Algorithm Theoretical Basis Document
DP	Data Product

L0	Level 0
L1	Level 1
QA/QC	Quality Assurance/Quality Control
QF	Quality Flag
QM	Quality Metric
IS	Instrument Systems

2.4 Variable Nomenclature

The symbols used to display the various inputs in the ATBD, e.g., calibration coefficients and uncertainty estimates, were chosen so that the equations can be easily interpreted by the reader. However, the symbols provided will not always reflect NEON's internal notation, which is relevant for NEON Cyberinfrastructure use, and/or the notation that is used to present variables on NEON's data portal. Therefore a lookup table is provided in order to distinguish what symbols specific variables can be tied to in the following document.

Symbol	Internal Notation	Description
i		Initial observation
N		Final observation
QF_{FINAL}		Final quality flag
QF_{α}		Alpha quality flag
QF_{β}		Beta quality flag
QM_{α}		Alpha quality metric
QM_{β}		Beta quality metric
QF_{j}		Quality flag for the first test in a set of quality tests
QF_{F}		Quality flag for the last test in a set of quality tests
QM_{j}		Quality metric for the first test in a set of quality tests
QM_{F}		Quality metric for the last test in a set of quality tests
QF_{Review}		Science review flag

3 DATA PRODUCT DESCRIPTION

3.1 Reported Data Products

Most IS Level 1 (L1) DPs will be accompanied by quality metrics and a final quality flag, which summarize the QF results from the QA/QC analyses and sensor flags. Results of specific QFs as they relate to individual Level 0 (L0) DPs will be presented as quality reports, but will not be presented unless requested. Detailed explainations of how quality metrics and the quality reports are presented to users can be found in section 5.3.

3.2 Input Data Products

DP inputs will consist of any L0 DP that is used to create a given L1 DP.

3.3 Product Instances

Product instances depend on the measurement in question and are detailed in each data product's ATBD.

3.4 Temporal Resolution and Extent

The temporal resolution and extent of the QFs and QMs will be dependent on the sampling frequency of the measurement in question. Relevant information is detailed in each data product's ATBD.

3.5 Spatial Resolution and Extent

The spatial resolution and extent of the QFs and QMs will be dependent on the location of the measurement in question. Relevant information is detailed in each data product's ATBD.

4 SCIENTIFIC CONTEXT

Data quality has always been paramount to ecology; however, historically the need to automate its assessment has not been. This has primarily been due to the quantity of data that has been needed to answer research questions. Yet, there is an increasing need for "big data" to answer the research questions of tomorrow. As a result, there is a growing demand for automated approaches to assess data in order to replace time consuming manual assessments. Our goal was to develop an automated framework that allows the QA/QC and sensor flag results to be summarized in a transparent and easily interpretable way for IS L1 DPs.

We derived techniques from standard timeseries analysis, which are widely used in eddy covariance research due to their need to routinely assess large data sets and scrutinize data quality. As such, we drew concepts from the following works: Mauder et al., 2013, Mauder, 2011, and Gockede et al., 2004. These approaches were designed to determine the quality of a data product that was produced using multiple other data products. Briefly, these schemes define criteria to rank the quality of a data product's inputs and then use that information to determine the quality of the final data product. A result is that these rank based approaches tend to be subjective and predominantly measurement specific. In addition, these approaches were not designed to be used for a multitude of measurements and as such are not easily transferable. Therefore, these approaches could not be directly transferred for NEON, but instead their concepts were drawn from to create our quality assessment scheme.

While other quality flag systems for climatological data were examined, the sheer magnitude of the data that NEON collects inhibits the use of many preexisting quality flag schemes. For example, many of NEON's TIS sensors collect data at 1 Hz and consequentially many of the QA/QC and sensor flags are generated at the same rate. Thus, for flags generated at 1 Hz, there will be over 86,000 outcomes for each flag in a day and over 600,000 in a week. The ability to digest and interpret that quantity of data can quickly become overwhelming. Therefore, a framework was developed to objectively assess data quality, while remaining transparent and transferable to all of NEON's instrumented products.

4.1 Theory of Algorithm

4.1.1 Quality Flags

Quality flags are generated by a number of QA/QC analyses as well as sensor flags. For example, QFs for L1 DPs include flags produced by plausibility, despiking, and sensor flags. While L0 DPs flagged by some of these tests will result in a datum not being used to create a L1 DP, this is not true for all QFs. Information on which QFs will result in a datum being excluded when calculating a L1 DP is included in each data product's ATBD. Each QF can be set to one of three states as shown in Eq. (1). For specific details on a QF please refer to their corresponding ATBDs.

$$\mathrm{QF} = \begin{cases} 1 \text{ if the quality test failed} \\ 0 \text{ if the quality test passed} \\ -1 \text{ if NA i.e. not able to be run due to a lack of ancillary data} \end{cases} \quad (1)$$

4.2 Quality Metrics

Since each L1 DP is composed of multiple L0 DPs, any QFs that were applied to L0 DPs need to be summarized. Thus, a DP consisting of multiple observations will have three QMs associated with each QF. The three QMs will summarize as a percent of the total number of observations, used to create a DP, where a QF was set to 1, 0, and -1. QMs will always be rounded half up to a whole percentage. QMs are defined as follows:

$$\mathrm{QM}_{j,1} = \frac{\sum_{i=1}^{N}(\mathrm{QF}_{j,i} \equiv 1)}{N} * 100 \quad (2)$$

and

$$QM_{j,0} = \frac{\sum_{i=1}^{N}(QF_{j,i} \equiv 0)}{N} * 100 \quad (3)$$

and

$$QM_{,-1} = \frac{\left|\sum_{i=1}^{N}(QF_{j,i} \equiv -1)\right|}{N} * 100 \quad (4)$$

Where:

$QM_{j,1}$	= Quality metric associated with QF_j for the percent of tests set high
$QM_{j,0}$	= Quality metric associated with QF_j for the percent of tests set low
$QM_{j,-1}$	= Quality metric associated with QF_j for the percent of NA tests
QF_j	= Results of test, j, for a L0 DP
i	= Running index over sample size
N	= Sample size

In reality one only needs to compute two of the three QMs and the third QM can be derived as follows:

$$100\% - \left(QM_{j,1} + QM_{j,-1}\right) = QM_{j,0} \quad (5)$$

Table 1 illustrates how QMs for a L1 DP, where $\{QF_F | QF_j \in QF_F\}$ and $\{N | i \in N\}$, are determined according to Eq. (2), (3), and (4).

Observation	QF_j	…	QF_F
i=1	$1 \vee 0 \vee -1$	⋱	$1 \vee 0 \vee -1$
i=2	⋮	⋱	⋮
⋮	⋮	⋱	⋮
⋮	⋮	⋱	⋮
⋮	⋮	⋱	⋮
⋮	⋮	⋱	⋮
⋮	⋮	⋱	⋮
N	⋮	⋱	⋮

$QM_{j,1}$	$QM_{j,0}$	$QM_{j,-1}$		$QM_{F,1}$	$QM_{F,0}$	$QM_{F,-1}$
$\frac{\sum_{i=1}^{N}(QF_{j,i} \equiv 1)}{N} * 100$	$\frac{\sum_{i=1}^{N}(QF_{j,i} \equiv 0)}{N} * 100$	$\frac{\lvert\sum_{i=1}^{N}(QF_{j,i} \equiv -1)\rvert}{N} * 100$	…	$\frac{\sum_{i=1}^{N}(QF_{F,i} \equiv 1)}{N} * 100$	$\frac{\sum_{i=1}^{N}(QF_{F,i} \equiv 0)}{N} * 100$	$\frac{\lvert\sum_{i=1}^{N}(QF_{F,i} \equiv -1)\rvert}{N} * 100$

Table 1. Illustration of how QFs are summarized into QMs for a L1 DP.

5 ALGORITHM IMPLEMENTATION

5.1 α and β Quality Flags and Metrics

In order to assess the overall quality of a particular L1 DP, it is necessary to summarize the number of flagged observations among all L0 DPs that were used in its calculation. Thus, we define alpha (α) and beta (β) QFs and QMs, which will incorporate the outcomes of several QFs in order to assess the L1 DP's quality. The following defines how α and β QFs and QMs are determined.

α and β QFs will be calculated based on the outcomes of the QFs from QA/QC analyses as well as sensor flags. What QFs will be used to calculate α and β QFs is DP-specific and must be specified in its corresponding ATBD. The calculation of α and β QFs is very similar except that QF_α will determine for a subset of QFs (defined in a sensor-specific ATBD) whether or not at least one QF was set to 1 for an observation. Likewise, QF_β will determine for a subset of QFs, whether or not at least one QF was set to -1 for an observation. The calculation of QF_α and QF_β are shown Eq. (6) and (7).

$$QF_{\alpha} = \begin{cases} 1 \text{ if } \sum_{i=j}^{F} (QF_i \equiv 1) > 0 \\ 0 \text{ otherwise} \end{cases} \quad (6)$$

and

$$QF_{\beta} = \begin{cases} 1 \text{ if } \sum_{i=j}^{F} (QF_i \equiv -1) > 0 \\ 0 \text{ otherwise} \end{cases} \quad (7)$$

QFs in Eq. (6) and (7) will be used to calculate α and β QMs. QM_{α} and QM_{β} will be calculated according to Eq. (2). Let it be noted that for a set of QFs, $\{QF_F | QF_j \in QF_F\}$, QF_{α} and QF_{β} can both be set high for the same observation in the event one QF has been set high and another QF could not be computed due to a lack of ancillary data. Therefore, while the QMs for an individual QF will always sum to 100 %, QM_{α} and QM_{β} will always be less than or equal to 200 %.

Below, Table 2 summarizes how all QFs and QMs, including α and β QFs and QMs, are calculated for a particular DP.

Observation	**QF_j**	…	**QF_F**	**QF_{α}**	**QF_{β}**
i=1	$1 \vee 0 \vee -1$	⋱	$1 \vee 0 \vee -1$	If $(\sum_{i=f}^{F} QF_i \equiv 1) > 0 \rightarrow QF_{\alpha} = 1$ otherwise 0	If $(\sum_{i=f}^{F} QF_i \equiv -1) > 0 \rightarrow QF_{\beta} = 1$ otherwise 0
i=2	⋮	⋱	⋮	⋮	⋮
⋮	⋮	⋱	⋮	⋮	⋮
⋮	⋮	⋱	⋮	⋮	⋮
⋮	⋮	⋱	⋮	⋮	⋮
⋮	⋮	⋱	⋮	⋮	⋮
⋮	⋮	⋱	⋮	⋮	⋮
N	⋮	⋱	⋮	⋮	⋮

$QM_{j,1}$	**$QM_{j,0}$**	**$QM_{j,-1}$**		**$QM_{F,1}$**	**$QM_{F,0}$**	**$QM_{F,-1}$**	**QM_{α}**	**QM_{β}**
Eq. (2)	Eq. (3)	Eq. (4)	…	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (2)	Eq. (2)

Table 2. Illustration of how α and β QFs and QMs are determined for a DP with *N* observations and *F* QFs.

5.2 NEON Data Product Quality Assessment

When L0 DPs are used to generate a L1 DP, that L1 DP will be assessed by NEON on its data quality. A final QF, QF_{FINAL}, will be set depending on whether the DP has passed or failed NEON's assessment. The threshold for QF_{FINAL} is based on the results of QM_{α} and QM_{β}, which establishes a limit for an acceptable amount of data that can either fail specific quality tests and/or the tests could not be run due a lack of ancillary data. If a DP reaches or exceeds this threshold it is flagged as invalid data, i.e., QF_{FINAL} = 1, and valid, i.e., QF_{FINAL} = 0, if it does not. It is envisioned that the threshold for QF_{FINAL} will change over time and may vary for different DPs. Here we present QF_{FINAL} as a 2:1 ratio of QM_{β} to QM_{α} with maximums of 20% for QM_{β} and 10% QM_{α}, Eq. (8). Figure 3 represents the ratio of QM_{β} to QM_{α} in a graphical form.

$$QF_{NEON} = \begin{matrix} 1 \text{ if } QM_{\beta} + (2 * QM_{\alpha}) \geq 20 \\ 0 \text{ otherwise} \end{matrix} \quad (8)$$

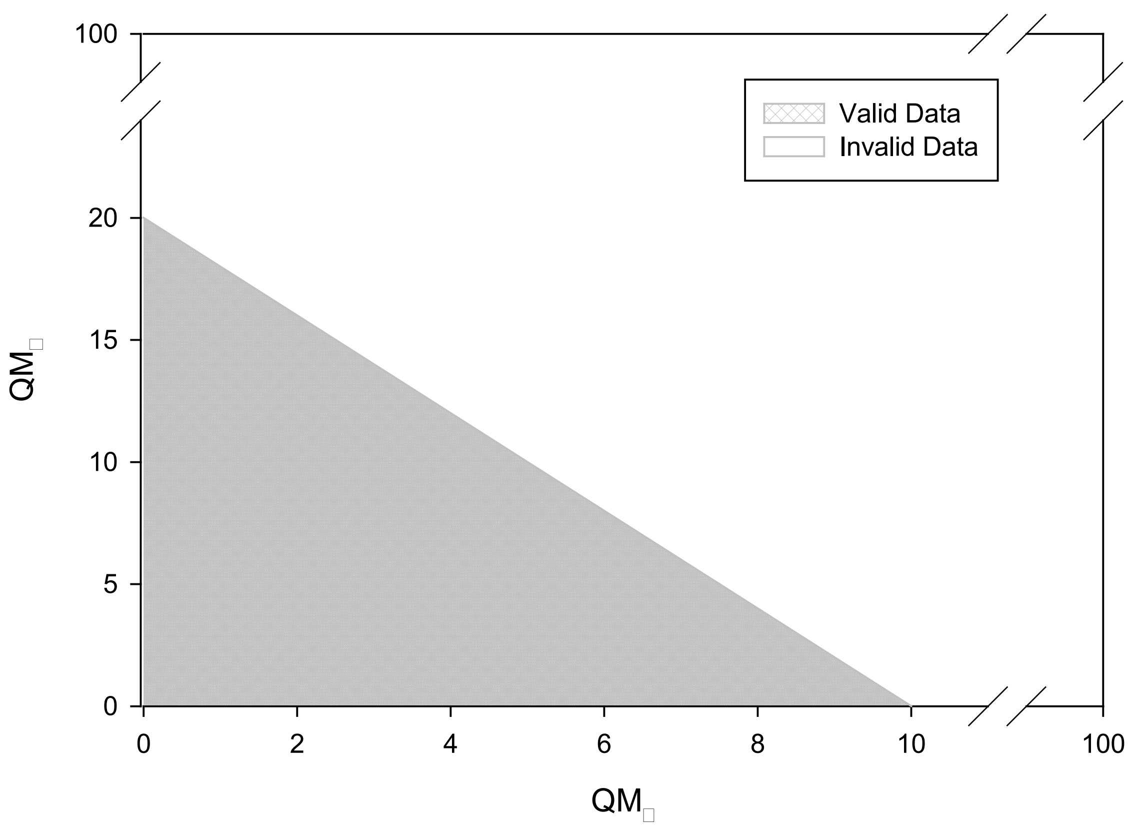

Figure 1. A graphical representation of QF_{FINAL}, using Eq. 8, to determine whether a L1 DP is valid or not.

5.2.1 Missing Data

Each L0 datum shall first be assessed by the null test to determine whether data is missing for a given timestamp, see AD[01] for details. If the null test fails, thus indicating that data is missing for primary variable/s the null flag shall be triggered, the gap test assessed, and processing shall end for that data point (i.e., no other QA/QC or algorithm processing will occur). Therefore, the only inputs to alpha and beta quality flags and subsequently the final quality flag will be the results of the null and gap test. Primary variable/s are defined as the inputs that directly feed into the creation of the level one data product, as opposed to ancillary variables which are used in data QA/QC and sensor health monitoring. For example infrared biological temperature has two primary variables, thermistor resistance and thermopile voltage. These two variables are used in conjunction to create the L1 infrared biological temperature DP. Single aspirated air temperature has one primary variable, the platinum resistance thermometer output, while it has multiple ancillary variables that include heater status and flow rate.

5.2.2 Science Review Flag

Not all quality issues can be detected with automated tests. Field personnel often encounter measurement interference in which a sensor is operating properly and within range but is not measuring the intended target. For example, the throughfall precipitation collector is prone to blockage from leaves and other debris. The sensor cannot detect when it is blocked, it simply records no rain. Therefore, field reports of sensor blockage must be reviewed and the impacted data flagged manually. The Science Review flag will act as a catch-all for communicating quality issues not captured by the automated quality flagging routines (as specified in the ATBD for each data product).

The science review flag is raised manually after expert review to communicate that data are suspect. In extreme cases where the data are determined unusable for any foreseeable use case the related data values are removed from the published dataset. In all circumstances, the original data are retained in internal NEON data storage, and the reason for manually flagging data values is provided with the published data. Computation of this science review flag, QF_{Review}, is as follows:

$QF_{Review} =$

2	If, after expert review, the data are determined unusable in all foreseeable use cases and should be removed from published data. Setting $QF_{Review} = 2$ triggers $QF_{FINAL} = 1$ and removes data values from published data.
1	If, after expert review, the data are determined to be suspect. Setting $QF_{Review} = 1$ triggers $QF_{FINAL} = 1$. Data values are retained in published data.
0	A previous quality issue was resolved (e.g. by reprocessing), and the data are no longer considered suspect.
Blank	No information regarding expert quality review is available

QF_{Review} is generated at the L1 and above (e.g. L1, L2, L3, etc.) data product levels after automated quality tests have been performed. A database of the data products, locations, time periods, and reasons for flagging with QF_{Review} is jointly maintained by NEON Science and Cyberinfrastructure teams.

Note that QF_{Review} is implemented differently depending on whether a final quality flag is computed for the data product. For data products with a final quality flag (the typical case), $QF_{Review} = 1$ triggers $\mathrm{QF}_{\mathrm{FINAL}} = 1$ as well. This scenario can be identified when the term name corresponding to QF_{Review} is the same term name as that of $\mathrm{QF}_{\mathrm{FINAL}}$ except with 'SciRvw' appended to the end. For example, if the term name for $\mathrm{QF}_{\mathrm{FINAL}}$ is *gWatTempFinalQF*, the term name for QF_{Review} will be *gWatTempFinalQFSciRvw*. For data products without a final quality flag, the science review flag is a standalone flag that is simply published as another column in the dataset. The standalone term name for QF_{Review} is formatted with 'QF' at the end (e.g. *gWatTempSciRvwQF*).

5.3 Presentation of results to end user

Each L1 DP will have its QF results from the QA/QC analyses and sensor tests presented through two separate schemes; quality reports and quality metrics. A quality report for a L1 DP will present the results of a specific QF as they relate individual L0 DPs, which includes quality reports for α and β QFs. For example, the range test quality report for the L1 DP of one-minute mean air temperature (sampled at a rate of 1 Hz) allows users to differentiate the 60 different outcomes of the range test as they relate to the L0 DPs that went into that one-minute average. Each L1 DP will only have quality reports generated for its smallest time aggregated value (generally one-minute averages). The quality report information for a L1 DP's larger time aggregated value (generally thirty-minute averages) can be obtained by concatenating the quality reports from the smaller time aggregated value. Moreover, there is a one-to-one relationship between the number of flags that accompany a measurement and the number of quality reports that are produced. Using Figure 4 as an example, say the 15 observations are averaged to create a L1 DP. The quality report for QF_1 that accompanies this L1 DP would be the 15 outcomes of QF_1 shown in rows 1-15.

Alternatively, L1 DPs will have quality metrics generated for both time aggregated values, which are not a one-to-one relationship to the number of QFs that exist. Instead a quality metric is produced for each state that a quality flag can take. Generally, three states will exist for a flag as shown in Eq. (1), however this number can vary. For example, the averaging flag for triple redundant air aspirated temperature has eight outcomes, and thus will have eight associated quality metrics, see NEON.DOC.000654 for detail. The final QF will also be associated with each L1 DP. By presenting QF information in this manor, several levels of detail that pertain to the quality of sensor data are retained. This process will facilitate data transparency and usability, as well as enable sensor related issues to be backtracked.

Quality Report Information

$Observation$	QF_1	QF_2	QF_3	QF_α	QF_β
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	1	1	0
9	1	0	1	1	0
10	1	0	0	1	0
11	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	-1	-1	0	1

$QM_{1,1}$	$QM_{1,0}$	$QM_{1,-1}$	$QM_{2,1}$	$QM_{2,0}$	$QM_{2,-1}$	$QM_{3,1}$	$QM_{3,0}$	$QM_{3,-1}$	QM_α	QM_β
13%	87%	0%	0%	93%	7%	13%	80%	7%	20%	7%

$QM_\alpha, QM_\beta \rightarrow finalQF = 1$

Quality Metrics

Figure 2. Example of the information that will accompany a L1 DP. Shown here for 15 observations and three arbitrary QFs. Each QF will have a quality report that accompanies a L1 DP's smallest time aggregated value. QMs are shown here for QFs with three states (pass = 0, fail = 1, could not be run = -1).

6 ALGORITHM VERIFICATION

Verification of the algorithms disclosed in this ATBD shall follow the procedures outlined in AD[01].

7 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

While the framework presented here is applied only in the context of NEON IS DPs, it could also be easily transferred to other data sets. The quality of large data sets can often be difficult to interpret and the application of these algorithms is intended to balance information with accessibility. In addition, this framework was developed to cater to users with varying backgrounds and levels of expertise.

8 FUTURE PLANS AND MODIFICATIONS

This ATBD is version controlled, i.e. future developments might results in modifications to this ATBD, which will be documented accordingly.

9 BIBLIOGRAPHY

EPA. (2006). Data quality assessment: A reviewer's guide (QA/G-9R). U.S. EPA. pp. 61 [Online]. Avaliable: http://www.epa.gov/QUALITY/dqa.html [June, 2013].

Göckede, M., Rebmann, C., and T. Foken. (2004). A combination of quality assessment tools for eddy covariance measurements with footprint modeling for the characterization of complex sites. *Agricultural and Forest Meteorology*. 127, 175-188.

Mauder, M. and T. Foken. (2011). Documentation and instruction manual of the eddy-covariance software package TK3. pp. 60 [Online]. Available: http://opus.ub.uni-bayreuth.de/volltexte/2011/866/pdf/ARBERG046.pdf [June, 2013].

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H. P., Shmidt, M., and R. Steinbrecher. (2013). A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements. *Agricultural and Forest Meteorology*. 169, 122-135.

Wang, R. Y., Storey, V. C., and C. P. Firth. (1995). A framwork for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*. 7(4) 623-640.