Time Series Automatic Despiking for TIS Level 1 Data Products – QA/QC Algorithm Theoretical Basis Document

PREPARED BY	ORGANIZATION	DATE
Derek Smith	FIU	Mar. 2013
Stefan Metzger	FIU	Dec. 2012

APPROVALS (Name)	ORGANIZATION	APPROVAL DATE
David Tazik	CCB PROJ SCI	9 May 2013
Hanne Buur	CCB DIR SE	9 May 2013

RELEASED BY (Name)	ORGANIZATION	RELEASE DATE
Stephen Craft	CCB Admin	9 May 2013

See Configuration Management System for approval history.

Change Record

REVISION	DATE	ECO#	DESCRIPTION OF CHANGE
A	9 May 2013	ECO-01000	Initial Release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

This document specifies the time series signal despiking algorithm that will be used as part of the automated Quality Control/Quality Assurance (QA/QC) plan of observed instrument data [RD 02]. Specifically, this document outlines the automated despiking routine that will be used to create TIS L1 DPs. Further time series analyses and despiking routines will be run on higher level DPs.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

AD[01]	NEON.DOC.001113	Quality Flags and Quality Metrics for TIS Data Products
AD[02]	NEON.DOC.001069	Preprocessing for TIS Level 1 Data Products

2.2 Reference Documents

RD[01]	NEON.DOC.000008.	NEON Acronym List
RD[02]	NEON.DOC.011009	FIU Data Flow QA Plan

2.3 Acronyms

Acronym	**Explanation**
ATBD	Algorithm Theoretical Basis Document
CI	Cyber Infrastructure
DP	Data Product
FIU	Fundamental Instrument Unit
L0	Level 0
L1	Level 1
MAD	Mean Absolute Deviation
QA/QC	Quality Assurance/Quality Control

2.4 Variables

Variable	Explanation
A	Method A - Centralized observation despiking routine
B	Method B - Window wise despiking routine
b_n	MAD correction factor
F_B	Temporary place holder flag for method B
F_P	Temporary place holder flag
k	Scale factor for the MAD
MAD_{adj}	Mean Absolute Deviation for n values
$MAD\ (\ )$	The median of the absolute values for the residuals' in a window (L_1)
$MED(\ )$	Median of a finite set of values
n	Number of samples
q	MAD threshold
QF_D	Spurious spike flag
QF_I	Insufficient number of observations in window flag
QF_o	Physically feasible spike flag
s	Step size
T	Consecutive spike threshold
w	Window width (time period)
x	Observation in the time series
ω	Method B spike threshold
$\#a$	Number of assessments for method B

3 DATA PRODUCT OVERVIEW

The following algorithms are intended to be applied automatically to data, as specified in related sensor-specific ATBDs to assist in controlling the quality of Level 1 (L1) data products (DPs). These tests will be used to automatically examine data over a short timescale (*e.g.*, quasi-daily) and to determine the sanity of each individual observation. The test results calculated in this document are intended to be referenced by algorithms in other NEON documents.

For a given interval of observations statistics will be determined, defined in Section 4.2, to set thresholds for the despiking routine to identify outliers. The objective of this document is to provide a framework for a despiking routine that will contribute to the production of L1 DPs. Thus, sensor-specific details are not included; if necessary, explicit site or sensor-specific details will be defined by FIU and maintained in the CI data store. An example of such details found in Section 7.

4 TIME SERIES QUALITY CONTROL/QUALITY ASSURANCE

4.1 Overview of Despiking Routine

Despiking routines are intended to flag unphysical data points within the data stream. Spurious spikes can result from various problems, such as sensor malfunction or electrical spikes. In addition, caution must be taken to make certain that genuine spikes (i.e. unusual but physical trends or ramps) are not mistakenly identified and flagged. Often despiking involves an "eyes on" approach, in which an experienced person can distinguish between plausible and implausible data. However, an "eyes on" approach is time consuming if not impractical for large data streams due to its high resource demand. Thus, in order to process data on the scale that will be generated by NEON, automation of the despiking algorithm is required, with a focus on utilizing robust algorithms and minimizing computational costs. As discussed in Section 4.2, any spike identified as spurious in an observation period will be flagged as such.

It is common for spike detection methods use the average and standard deviation of a time series to define thresholds for spike detection (EddyPro, 2011; Mauder, 2011; Vickers, 1997; Hojstrup, 1993). However, due to the effects outliers can have on Gaussian metrics, iteration through the selected data is typically required. This can substantially increase computational costs as many routines rely on setting a threshold based on the standard deviation of the time series, multiplied by an empirical adjustment factor. With each iteration, the multiplication factor increases by a defined value until no additional spikes are found. Iteration can be avoided and computational time can be greatly reduced by relying on more robust statistics, specifically the Median Absolute Deviation (MAD) (Mauder et al., 2013; Metzger et al., 2012; Papale et al., 2006). In addition to reducing computational time, the MAD offers a more robust test statistic than relying on the mean and standard deviation. While the mean is an ideal estimator for the location of a normal distribution, it is problematic for distributions that even slightly deviate from normal. The median is known to be "resistant to gross errors" whereas the mean is not; thus the MAD provides greater confidence, which is crucial in spike identification (Venables and Ripley, 2002).

4.1.1 Overview of Quality Flags

A summary of the flags (*F*) and quality flags (*QF*) that are generated during the despiking routine are displayed in Table 1. Each QF can be set to one of three states, 1, 0, or -1 (i.e., high, low, and NA [not able to be run due to a lack of ancillary data]). A brief description of their purpose is also presented. Criteria for how flags are defined can be found below in Section 4.4. Information regarding the calculation of quality metrics from quality flags for one- and thirty-minute averages can be found in AD[01].

Flags	Purpose	Values
F_B	Spike place holder flag for method B despiking analysis	1 or 0 or -1
F_P	Place holder flag so that physically feasible spikes can be identified	1 or 0 or -1
QF_I	Insufficient reliable data available	1 or 0 or -1
QF_O	Identifies a physically feasible spike	1 or 0 or -1
QF_D	Identifies a spurious spike	1 or 0 or -1

Table 1: Quality flags generated from the time series despiking routine.

4.2 Automated Despiking Algorithm

The despiking routine will be applied to the preprocessed data stream for a particular sensor. Preprocessing will occur according to AD[02]. Observations will be assessed within a sliding window of width, *w*, which will be sensor specific. The window width, defined as a number of observations, includes both actual and missing, i.e., NA, observations. Window width is directly related to variation dampening among observations. Thus, large window sizes will result in a greater degree of dampening. Alternatively, smaller windows are better suited to retain of variation among observations, i.e., measurements that are highly variable. Therefore, window size will be subjective to each specific sensor measurement. The despiking window will step through the time series by a given number of points, which will also be sensor-specific parameter. All sensor specific test parameters will be defined by FIU and maintained in the CI data store and unless explicitly stated calculated values will be rounded down to the nearest integer. An example of how window and step size are defined can be found in Figure 1.

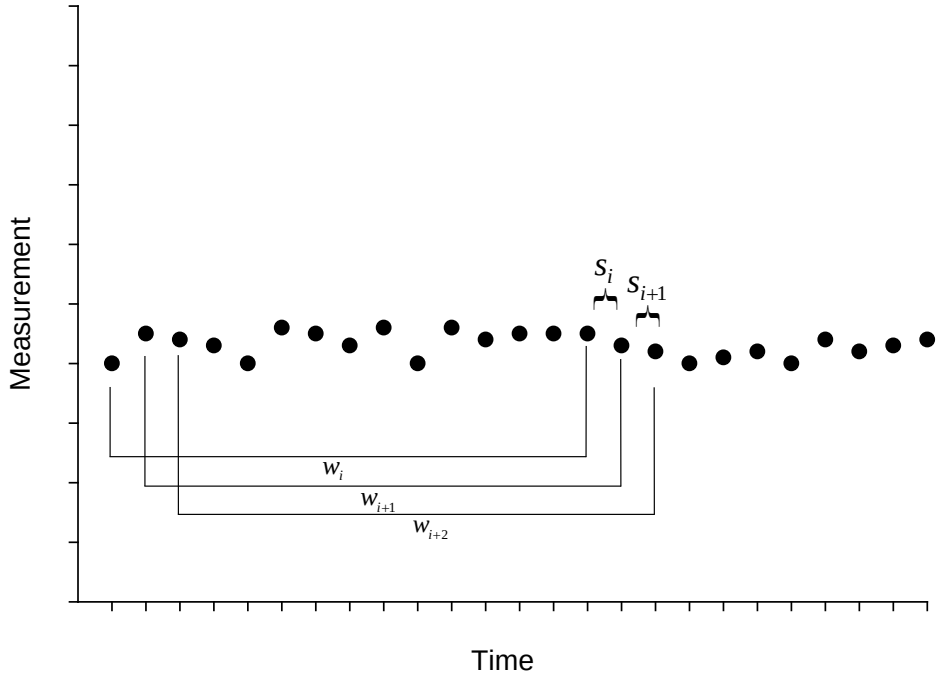

Figure 1: How windows step though a time series of observations. Here the window, w, is defined as 15 observations and the step size, s, is 1.

Before calculating the MAD for a window, we define $MED(\,)$ as the function for calculating the median. The median for a finite set of values can be defined by the two following steps: 1) Order observations values from smallest to largest. 2) Take the middle value from the set of ordered observations. In the event an even number of observations exist, the arithmetic average of the two middle values is taken. Next, for a window, w, of observations, x_i, where $i = (0, n)$, the MAD will be computed, defined as a function by $MAD(\,)$. The MAD is computed for a window with n observations by computing the median of the residuals' absolute values accordingly:

$$MAD = MED(|x_i - MED(w)|) \quad (1)$$

Next, to use the MAD as a consistent estimation of the standard deviation we define k as a constant scale factor. The constant k is defined as follows:

$$k = \frac{1}{\Phi^{-1}(3/4)} \approx 1.4826 \quad (2)$$

where k is equal to the reciprocal of the quantile function (i.e., inverse of the cumulative distribution function for a normal distribution), Φ^{-1}, evaluated at a probability of (3/4). Thus, it is expected that k times the MAD for a set of samples with a Gaussian distribution is approximately equal to the population standard deviation (Ruppert, 2010).

Lastly, following the work of Croux and Rousseeuw (1992), we define a correction factor for the MAD to reduce bias induced by varying window sizes. For windows with more than nine points, we define the correction factor, b_n, as:

$$b_n = \frac{n}{n - 0.8} \quad (3)$$

where n is the number of actual observations, i.e. non-NA observations, within the window.

If possible, windows should always have more than nine points. If the number of points within a widow is < 10, Table 2 displays the correction factor values that will be used (Croux and Rousseeuw, 1992). Additionally, Rousseeuw and Verboven (2002) states that when $n < 4$, it is no longer possible to estimate scale robustly. Therefore, if there are less than four points in a window, the despiking routine will not be run and the quality flags QF_D, QF_O, and QF_I will be set to NA, i.e., -1, to reflect results from the tests are not available.

n	4	5	6	7	8	9
b_n	1.363	1.206	1.200	1.140	1.129	1.107

Table 2: Correction factor values for windows with less than 10 samples.

Lastly in order to identify outlier observations we define, q, as the MAD threshold value, which will be sensor-specific. Combining these steps, we calculate the adjusted MAD, i.e. MAD_{adj} for a window accordingly:

$$MAD_{adj} = b_n * q * k * MAD. \quad (4)$$

Once MAD_{adj} has been obtained for a window, an observation, x_i, within the window is flagged as a spike if it is outside the following range:

$$MED(w) - MAD_{adj} \leq x_i \leq MED(w) + MAD_{adj} \quad (5)$$

4.3 Despiking Routine Implementation

The despiking routine will identify spikes using one of two procedures, defined below and maintained in the CI data store. The basic difference between the two methods is that method Aassesses only the central most observation of the window before it steps to the next observation. Alternatively, method B assesses all observations within a window before it steps to the next window.

4.3.1 Despiking Methods

Method A

The first procedure, A- the default, is to assess only the central most observation for a window using Eq. (6). For example, if a window contains 13 observations, the central observation will be assessed using the six observations on both sides of the observation, i.e. the six preceding and succeeding observations. The window would then step through the time series where the next central point in the window would be assessed. For this option, the step size will always be one and the window set for an odd number of observations to be present, which includes both actual and NA observations.

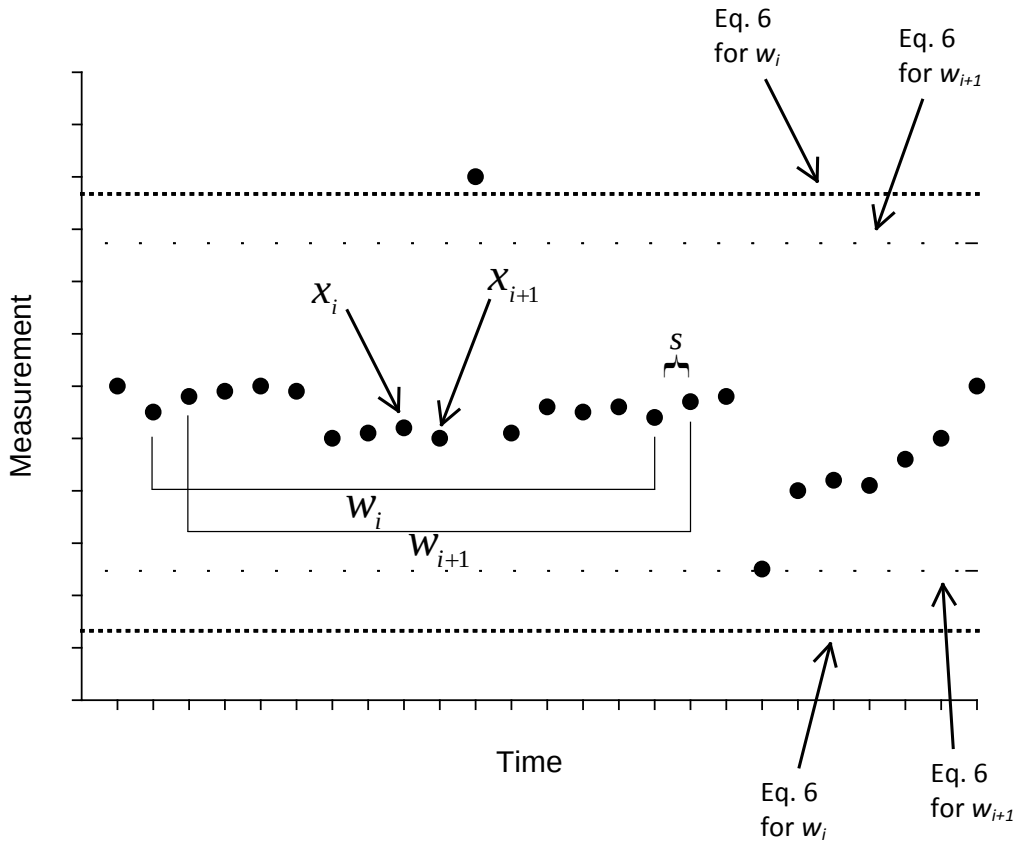

Figure 2: Representation of how method A assesses a time series. Dashed and dotted lines represent the MAD range, Eq. (6) for the two different windows, w_i and $w_{(i+1)}$.

Method B

The despiking routine is designed to encompass wide array of measurements. Therefore, in addition to method A, we define a second method, B, as an option for identifying spikes. Due to the variability among sensor measurements, in some situations method B may be more capable of not falsely identifying certain types of natural environmental phenomenon, e.g., ramps, as spikes. This is attributed to method B identifying spikes based on a collection of assessments for each observation and in contrast to method A where an observation is assessed based on the statistics of one window.

Functionally, method B is very similar to method A shown in Figure 2. The difference is that instead of only assessing the central most point in the window, method B assesses all observations within a window according to Eq. (6). If an observation is identified as a spike, a temporary placeholder flag, F_B, will be set high. With a step size of one, the number of times that an observation is assessed by the despiking routine is equivalent to the window size. If the $s > 1$, then $s \leq \frac{1}{2}w$ for windows with an even number of observations and $s \leq \frac{1}{2}w$ must be rounded down to the nearest integer for windows with an odd number of observations.

This will allow for the number of assessments, *#a*, for each observation to be equal to one another. The number of assessments for an observation will be equal to:

$$\#a = \frac{w}{s} \tag{6}$$

where, $\#a$ is the number of assessments for an observation, which is always rounded up to the nearest integer.

Since the statistics for a window will likely change as it moves through the time series, a point could be identified as a spike in one window and not the next. Therefore, we define a threshold value, ω, as a percent. If the number of times an observation has been identified as a spike is < ω, that observation will not be identified as a spike. Likewise, if the number of times an observation has been identified by a spike is ≥ ω, then it will be identified as a spike. The specific value for ω will be defined by FIU and maintained in the CI data store. However, by default, the value for ω will be 10 %, i.e., the number of times an observation was identified as a spike must be an order of magnitude less than the number of its assessments. Additionally, values obtained from ω will always be rounded down to the nearest integer. Figure 3 provides a visual representation of how spikes are determined using method B; the use of F_p to identify spikes is defined in Section 4.4.2.

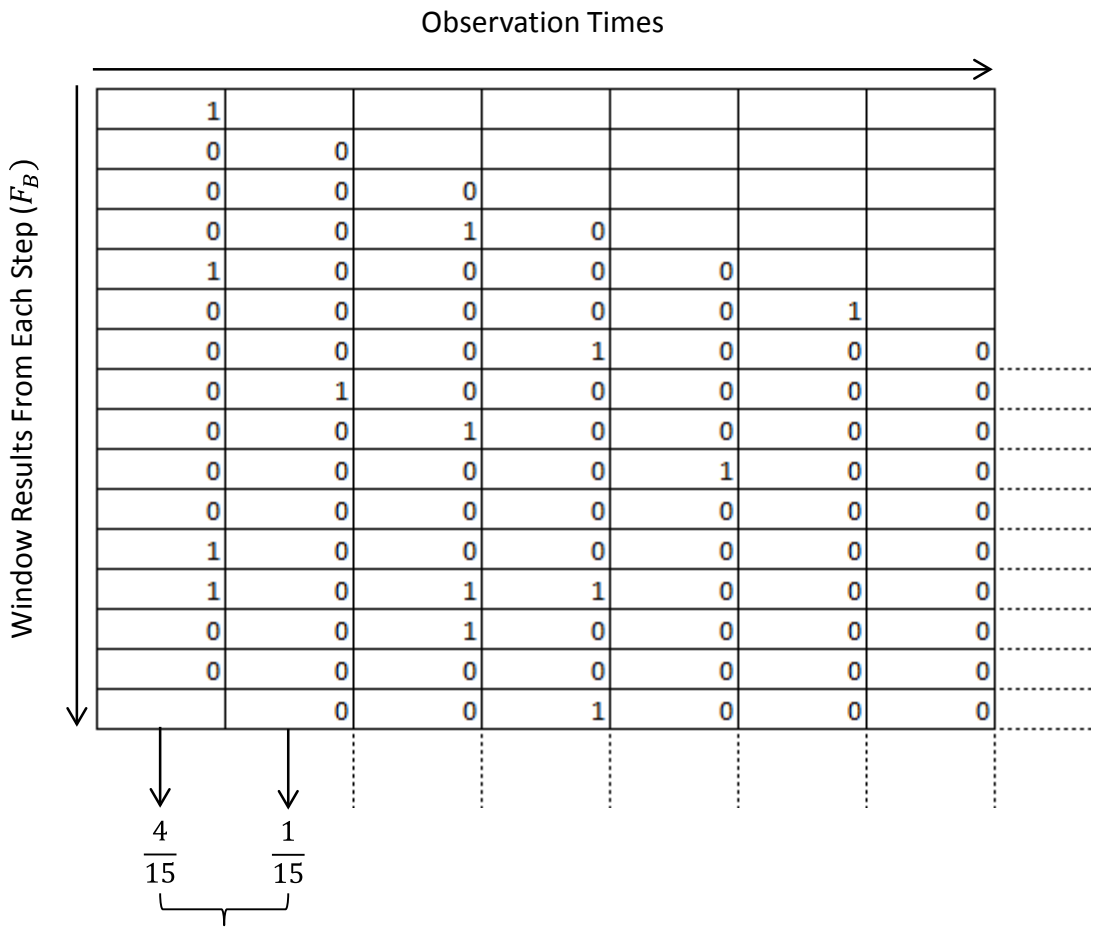

Figure 3: Criteria for spike detection using method B with a window size of 15 and a step size of 1.

4.3.2 Spike Identification and Threshold Classification

Per recommendation of Vickers (1997), in the event that consecutive spikes are identified, a threshold (T) should be set to ensure that unusual yet genuine physical trends in the data are not mistakenly classified and flagged. The threshold value for a specific sensor will depend on the characteristics of the measurement as well as the sampling frequency. Therefore, the threshold value will be sensor specific, provided by FIU, and maintained in the CI data store. While some despiking routines use a default threshold value of four, it is somewhat arbitrary as this setting simply prompts these samples, especially in spike laden records, to be visually inspected. For NEON's purposes, a quality flag, QF_O, will be set high in the event that consecutive spikes greater than the threshold value are identified. How this procedure is incorporated into the spike detection process is discussed below.

As a window passes though the time series, points that are identified as spikes within a window will have a temporary placeholder flag, F_P, applied. The placeholder flag, F_P, will be set high in the event

that a spike is detected and low otherwise. As the window steps through the time series and once it has stepped past a point by more than the threshold (*T*) value, the placeholder flag for that point will be assessed. If the number of consecutive points with a place holder flag, F_P, set high, is greater than the threshold value, then those points will have an associated quality flag, QF_O, set high. A quality flag, QF_O, set high indicates that a spike was identified within the L0 data stream, but the spike appears to be physically feasible. Points identified as a spike, F_P set high and not included in series of consecutive spikes greater than the spike threshold value, T, will have a spurious spike quality flag, QF_D, flag set high.

4.3.3 Additional Considerations

When missing points or gaps exist within the time series, it may not be possible to determine an accurate MAD for a window. Therefore, we define a threshold value stating that the amount of missing data within a window, i.e., observations assigned an NA, must be an order of magnitude less than the number of observations within the window. That is, no more than 10% of the expected observations within a window can be missing.

$$QF_I = \begin{matrix} 1\ if\ \frac{\sum_w NA}{w} \leq 0.1 * w \\ 0\ otherwise \end{matrix} \quad (7)$$

For example, if a Platinum Resistance Thermometer is sampled at 1 Hz and the window is defined as 600, i.e., 10-minutes worth of samples, the number of acceptable missing observations must be less than or equal to 60 observations. If the number of missing observations for a window exceeds this threshold, the observation or observations despiked by that window will be given a quality flag, QF_I. This quality flag, QF_I, indicates that an insufficient number of observations were present for that window.

Accompanying every L1 DP, will be the quality metrics listed in Table 2, which will be included in the QA/QC summary (i.e., Q_{sum}). In addition, despiking flags specific to a L0 DP will be retained and reported in the QA/QC quality report (i.e., Q_{rpt}), that accompanies every L1 DP. Once the despiking procedure has been completed for all DPs within an averaging period (i.e. 1- or 30-min), those DPs will continue to the next phase of the algorithm implementation, presented in the sensor specific ATBD.

5 REFERENCES

Croux, C. and P. J. Rousseeuw. (1992) Time-efficient algorithms for two highly robust estimators of scale. Computational Statistics. Vol. 1 pp. 411-428.

Hojstrup, J. (1993) A statistical data screening procedure. Meas. Sci. Technol. Vol. 4 pp. 153-157, doi: 10.1088/0957-0233/4/2/003

LI-COR. (2011) Eddy pro eddy covariance software version 3.0 user's guide and reference. LI-COR Inc.

pp. 200 [Online]. Available: ftp://ftp.licor.com/perm/env/EddyPro/Manual/EddyPro3_User_Guide.pdf [December, 2012].

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H. P., Schmidt, M., and R. Steinbrecher. (2013) A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements. Agricultural and Forest Meteorology. Vol. 169 pp. 122-135, doi: 10.1016/j.agrformet.2012.09.006

Mauder, M. (2011) Documentation and instruction manual of the eddy-covariance software package TK3. University of Bayreuth Department of Micrometeorology. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CDIQFjAA&url=http%3A%2F%2Fopus4.kobv.de%2Fopus4-ubbayreuth%2Ffrontdoor%2Fdeliver%2Findex%2FdocId%2F681%2Ffile%2FARBERG046.pdf&ei=R2U2UaOJA4T72QWd34GAAg&usg=AFQjCNEoyTzkcLXPE5qtAR60XVbC8BM4fw&sig2=-BOjVNxzCDG5G6TDuWmzWQ [March, 2013].

Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T. (2012) Eddy-covariance flux measurements with a weight-shift microlight aircraft. Atmos. Meas. Tech. Vol. 5 pp. 1699-1717, doi: 10.5194/amt-5-1699-2012

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and D., Yakir. (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences. Vol. 3 pp. 571-583, doi: 10.5194/bg-3-571-2006

Rousseeuw, P. J., and S. Verboven. (2002) Robust estimation in very small samples. Computational Statistics and Data Analysis. Vol. 40 pp. 741-758, doi: 10.1016/S0167-9473(02)00078-6

Ruppert, D. (2010) Statistics and data analysis for financial engineering. Springer. 1st Ed. pp. 638, ISBN 1-441-97786-4

Venables W. N. and Ripley B. D. (2002) Modern Applied Statistics with S-Plus. Springer. 4th Ed. pp. 495, ISBN 0-387-95457-0

Vickers, D and L. Mahrt. (1996) Quality control and flux sampling problems for tower and aircraft data. Journal of Atmospheric and Oceanic Technology. Vol. 14 pp. 512-526, doi: 10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2

6 APPENDIX

DATA STORE EXAMPLE

Using single aspirated air temperature and barometric pressure as examples, Table 2 displays the sensor-specific information for the despiking routine that will be provided by FIU and maintained in the CI data store. These values should not be taken as absolute; they are purely intended to illustrate what information FIU will provide to CI.

DP	Despiking Method	ω (%)	$w\ (n)$	$s\ (n)$	q	$T\ (n)$
NEON.DXX.XXX.DP1.00002.	A	NA	180	1	7	4
NEON.DXX.XXX.DP1.00004.	A	NA	300	1	7	4

Table 3: Sensor-specific information needed to run the despiking routine.