PREPROSESSING FOR TIS LEVEL 1 DATA PRODUCTS – QA/QC ALGORITHM THEORETICAL BASIS DOCUMENT

PREPARED BY	ORGANIZATION	DATE
Derek Smith	FIU	07/03/2013

APPROVALS (Name)	ORGANIZATION	DATE
Dave Tazik	SCI	07/11/2014
Hanne Buur	PSE	12/03/2014

RELEASED BY (Name)	ORGANIZATION	DATE
Judy Salazar	CM	12/05/2014

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	12/05/2014	ECO-01052	Initial Release

TABLE OF CONTENTS

1 DESCRIPTION

This document specifies the preprocessing approach 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 how data will be preprocessed, if required, to implement QA/QC processes that will be used to create TIS L1 DPs. Details on whether it is necessary to preprocess data prior to QA/QC and or other analyses will be specified either in a sensor's ATBD in the algorithm implementation section or be explicitly stated in a QA/QC ATBD.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Reference Documents

RD[01]	NEON.DOC.000008 NEON Acronym List
RD[02]	NEON.DOC.011009 NEON FIU Dataflow and QA Plan
RD[03]	NEON DOORS Requirements Database

2.2 Acronyms

Acronym	**Explanation**
ATBD	Algorithm Theoretical Basis Document
DAS	Data Acquisition System
DP	Data Product
L0	Level 0
L1	Level 1
QA/QC	Quality Assurance/Quality Control
TIS	Terrestrial Instrument System

2.3 Variables

Variable	**Explanation**
f	Frequency
n	Number of samples
t	Time
u_{DAS}	DAS Uncertainty

3 PREPROCESSING

Only actual sensor observations will be output by the DAS. Therefore, any interruption in the data stream that results in missing observations is captured as a jump between sample timestamps. However, the QA/QC algorithms used to process TIS sensor data require that all data is complete with respect to a sensor's sampling frequency, f. Therefore, the number of samples, n, for a time period should be as follows:

$$n = \Delta t * f \quad (1)$$

Where: n = Number of samples
Δt = Time period
f = Sample frequency

In order to identify missing observation times, an observation's timestamp will first be compared to the timestamp of the next consecutive observation:

$$\Delta t = t_{x+1} - t_x \quad (2)$$

Where: Δt = Time interval between consecutive observations in the time series.
t_x = Time of the observation being assessed.
t_{x+1} = Time of the next consecutive observation in the time series.

Next, the difference that exists between the two timestamps is compared to the sampling frequency of the sensor as well as the timestamp uncertainty associated with data acquisition system (DAS), u_{DAS}. If $\Delta t \leq \frac{1}{f} + 2 * |u_{DAS}|$, then preprocessing is complete for the observation at t_x and preprocessing will progress to the next observation in the time series. If $\Delta t > \frac{1}{f} + 2 * |u_{DAS}|$, then Eq. (3) will be used to determine the number of number of missing samples (n) for a given time period (Δt).

Note: Time units for Δt and $\frac{1}{f}$ must be the same and information on the uncertainty associated with different sampling frequencies can be found in RD[03].

$$n = (\Delta t * f) - 1 \quad (3)$$

Where: n = Number of missing samples for the time interval being assessed.
Δt = Time interval between consecutive samples in the time series.
f = Nominal sample frequency

The results from Eq. (3) indicate the number of time-value pairs, where the value is NA and the time corresponds to the timestamp of the **expected** datum, that need to be inserted between t_x and t_{x+1}. Time-value pairs will be inserted starting with the next timestamp, according to the sensor's nominal sampling frequency, after t_x. This will make the time series complete with respect to the sampling

frequency of the sensor. The number of time-value pairs inserted in the time series, n, will always be *rounded half up* to the nearest integer. Once the time series has been preprocessed, it will proceed to the next step of the algorithm implementation process as specified in the sensor-specific ATBD. Let it be noted that the preprocessed time series will only be used for QA/QC purposes and that reported data products will only be composed of actual observations as recorded by the DAS.