2020-09-23

Windowed Inspection of Stationarity & Quality (WISQ): An Algorithm For Eddy Covariance Sample Quality Control and Assessment

Ortiz-Suslow, David G.; Kalogiros, John; Yamaguchi, Ryan; Wang, Qing

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/65815

NPS-MR-20-001

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

WINDOWED INSPECTION OF STATIONARITY & QUALITY (WISQ): AN ALGORITHM FOR EDDY COVARIANCE SAMPLE QUALITY CONTROL AND ASSESSMENT

by

David G. Ortiz-Suslow
John Kalogiros
Ryan Yamaguchi
Qing Wang

September 23, 2020

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

VADM (ret) Ann Rondeau President	Dr. Robert F. Dell Acting Provost

This report was prepared by:

David G. Ortiz-Suslow	John Kalogiros
Ryan Yamaguchi	Qing Wang

Reviewed by:	Released by:
Wendell Nuss Chairman of Meteorology	Jeffrey D. Paduan Vice President and Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
23–9–2020	Technical Report	2017-09-22—2020-07-10

4. TITLE AND SUBTITLE

Windowed Inspection of Stationarity & Quality (WISQ): An Algorithm For Eddy Covariance Sample Quality Control and Assessment

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

David G. Ortiz-Suslow, John Kalogiros, Ryan Yamaguchi, Qing Wang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER

NPS-MR-20-001

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Department of the Navy

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

The views expressed in this report are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT

Measuring atmospheric fluxes requires various steps of measurement quality control, in addition to experimental design and post-processing corrections, in order to provide robust and high-fidelity data for wider use. However, within the measurement community, methods for these control steps are still applied *ad hoc*. Regardless of the availability of several comprehensive references texts available in the literature and licensed software programs.The theoretical and technical design of an algorithm for eddy covariance flux sample quality control and assessment is presented. This algorithm, WISQ, is robust and efficient and can be readily incorporated into existing processing experimental software packages. The goal of this algorithm is to output a flagging system that can be used to judge the quality of individual flux samples, with the option for outputting more detailed information. WISQ is unique in that it directly and automatically assess the sample flux accumulation and convergence. WISQ is also a general method that can be utilized for flux measurement outside of the realm of meteorology and is open-sourced for ease in development and innovation.

15. SUBJECT TERMS

NPS Technical Report

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	UU	32	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

ABSTRACT

Measuring atmospheric fluxes requires various steps of measurement quality control, in addition to experimental design and post-processing corrections, in order to provide robust and high-fidelity data for wider use. However, within the measurement community, methods for these control steps are still applied *ad hoc*. Regardless of the availability of several comprehensive references texts available in the literature and licensed software programs.The theoretical and technical design of an algorithm for eddy covariance flux sample quality control and assessment is presented. This algorithm, WISQ, is robust and efficient and can be readily incorporated into existing processing experimental software packages. The goal of this algorithm is to output a flagging system that can be used to judge the quality of individual flux samples, with the option for outputting more detailed information. WISQ is unique in that it directly and automatically assess the sample flux accumulation and convergence. WISQ is also a general method that can be utilized for flux measurement outside of the realm of meteorology and is open-sourced for ease in development and innovation.

A. Motivation

In the second quarter of the twentieth century and into the first decades of this century, micrometeorological techniques have greatly matured for measuring the bulk and turbulence characteristics of the atmospheric surface layer over land and water. Developing theses experimental practices has largely been undertaken to quantify the flux of energy and material at the interface between the lower atmosphere and the surface; then to develop empirical models (parameterizations) relating the flux to the bulk variance; and finally, to incorporate these empirical relations into numerical models that cannot directly resolve the fine-scale turbulence and exchange processes. While micrometeorological study can be focused on topics outside of this framework, the drive to understand the flux at the base of the atmosphere has nonetheless been, and persists to be, an underlying objective across a large swathe of the discipline. Like other fluid physics disciplines, this effort involves theoretical developments bolstered by significant experimental efforts and field observations remain central to the discipline for both basic research and monitoring campaigns.

Currently, investigators seeking to quantify the flux overwhelmingly rely on the eddy covariance method[i] [2]. This approach is well-suited to three-dimensional anemometry, explicitly estimates the wind stress (i.e., local friction velocity) without making spectral assumptions, and is mathematically elegant. As with any technique, practically employing EC is an involved process with many quality control steps from set-up through data processing to ensure robust measurements [2, 4]. While many of these steps have become more-or-less standard practice, there are some aspects of the general EC methodology that seem to be applied only sporadically. In particular, it is rarely reported in the literature how, or if, an investigator confirmed that each analyzed EC sample (i.e., discrete averaging/processing window) satisfied the condition of *stationarity*, which is an essential prerequisite for using EC. When this analysis is conducted, the approach is idiosyncratic and a standard method has not emerged. An exception to the above is the advent of analysis software (e.g. EasyFLux® DL or EddyPro® from Campbell Scientific), which do incorporate stationarity assessment. These softwares are becoming more common [1, 3], but they are not ubiquitous.

In this report, the design of the windowed inspection of stationarity & quality (WISQ) algorithm is described. WISQ is an efficient and robust tool that can be readily incorporated into

[i]Synonyms include: eddy covariance technique, eddy covariance (no modifier), direct covariance method, and direct method. Occasionally, eddy *correlation* is used to describe this method. Herein, we will use eddy covariance method or in abbreviation, EC.

EC data processing routines and provides critical information to the investigator on EC analysis window quality, including stationarity, homogeneity, and that all relevant turbulence scales are resolved. Here, we will demonstrate the WISQ method for EC sampling and provide examples of its implementation from a marine meteorological field campaign.

B. A Brief Summary of the Eddy Covariance Method's Statistical Basis and the Implications for Stationarity Analysis

Here we will provide a brief summary of the mathematical and statistical basis for using EC to quantify the turbulence and fluxes within the atmospheric surface layer (ASL). This material can be found in several reference texts for micrometeorology [2, 4][ii] and, on a more purely physical level, for turbulence [13]. The original concept for EC came out of mathematical and physical arguments put forward in the World War II era for the purpose of better understanding near-surface atmospheric turbulence and a desire to move beyond Prandtl's restrictive mixing length theory (e.g., see [6, 11]). The reader is directed to these references for further details, with the textbook by Foken, Aubinet, & Leuning being the most updated and relevant to EC. We should note that the discussion herein is concerned with the typical method of quantifying the covariance between the vertical turbulent (or perturbation) wind velocity, w', and some other turbulent variable, α'. This is only one component of a *generalized eddy covariance method* for scalar budget analysis [2].

A time series of a random variable, α(t), can be decomposed using Reynolds averaging:

$$\alpha(t) = \bar{A} + \alpha'(t), \tag{1}$$

where

$$\bar{A} = \frac{1}{T_\alpha}\int_{t_0}^{t_0+T_\alpha} \alpha(\tau)d\tau, \tag{2}$$

and α' is the perturbation component of α, i.e., this holds the turbulence information. Here, T_α is the analysis time for EC, which has several acceptable monickers but will be referred to within this report as the *averaging window*, and t_0 is the first time step within T_α. An over bar, $(\bar{\ })$, indicates that the operator (2) has been applied.

Equation 1 mathematically argues that α(t) is separable into distinct components $\bar{A}$ and α' over T_α, which is only strictly possible when the average is defined using ensembles (i.e.,

[ii] Specifically: Chapter 7 in Kaimal & Finnigan 1994 and Chapters 1 & 4 in the Foken et al. text.

ensemble averaging). An ensemble average, $\hat{A}$, is a mean taken over numerous independent realizations of α(t) for identical system conditions. This *may* be achievable in the laboratory, but is unachievable and unphysical in the real atmosphere. This challenge is typically circumvented by calling upon the *ergodicity postulate*: that under stationary conditions, $\bar{A} = \hat{A}$. Here, stationarity indicates that the statistics over the averaging window are not dependent of your selection of t_0 within the record spanned by T_α. To emphasize, using eqn. 1 in the real ASL requires invoking ergodicity[iii] which is only a valid proposition *if* the ASL during T_α is stationary. For $\bar{A} = \hat{A}$ over T_α, requires not only stationarity in the mean, but in the higher order moments, e.g., variance, as well. Confirming an individual EC sample observed locally stationary conditions is essential to employing the mathematical basis of the EC method and making a robust flux measurement.

The length of T_α is central to the issue of stationarity in EC flux calculations and analysis because this determines the variance contributing to $\bar{A}$. Fundamentally, T_α is directly related to sample altitude into the boundary layer and the flow stability. The goal of EC is to quantify the local turbulence and flux conditions, using a T_α that is too long for the conditions represents entraining local and macroscopic variance into a single α(t), which can manifest as a trend or local maxima/minima in the time record that creates non-stationarity. If T_α is too short, then the sample does not capture all of the low frequency contributions to the turbulence and the total local flux will be underestimated. For typical daytime, convective surface layer conditions, $\sim$30 minutes is a suitable T_α, though this should shorten (lengthen) as conditions become more convective (stratified). However, given different study aims, especially where spatial heterogeneity is expected or platform translation is a factor, shorter intervals have been used (e.g., [7, 8, 10]) to quantify the flux, with the acknowledgement that the *total* flux from the atmosphere to the surface may not have been captured for every sample.

The literature is replete with various lengths of T_α in both terrestrial and marine studies and there is no particular "standard", though 30 minutes has emerged as a typical window length. It is important to note that while theoretically T_α should be adaptively prescribed for the local turbulence and altitude, the authors are not aware of any study where this was systematically done in the course of processing a complete flux data set for the purposes of optimizing the physical robustness of the EC flux estimates. If T_α is varied, it is typically done as part of methodological studies or quality control analyses (e.g., [9]). Statistically, the optimal window

---

[iii] Whether knowingly, or not.

length can be estimated as:

$$T_\alpha \simeq 2\frac{\sigma_\alpha^2 \tau_\alpha}{\bar{A}^2 \varepsilon}, \tag{3}$$

where σ_a^2 is the time averaged variance, τ_α is the integral time scale determined from the integration of the autocorrelation of α(t) over T_α, and ε is the acceptable error level, etc. 0.02, 0.05. Specifically for flux analysis, qualitatively assessing T_α uses the ogive, Γ. In general, Γ is defined as the normalized cumulative integral of the scaled co-spectrum between variables x and y:

$$\Gamma(n_i) = \int_\infty^{n_i} C_{xy}(n) dlog(n), \tag{4}$$

$$C_{xy} = -nS_{xy}/\gamma^2, \tag{5}$$

where n_i is the i^{th} frequency bin of an FFT of x and y with length $N/2+1$, S_{xy} is the co-spectrum (e.g., x and y are the vertical (w) and horizontal (u) wind velocity, respectively), and γ is the relevant turbulent scaling (e.g., u_*). $\Gamma(n)$ is typically integrated from high frequencies (usually the Nyquist) to low frequencies (ideally $1/T_\alpha$). In completely stationary and homogeneous flux conditions, $\Gamma(n)$ should follow a smooth spline that has an initial plateau around 0 for high frequencies, rolls off to a quasi-linear slope over the mid-frequency subrange, and finally converges on -1 at low frequencies. The ogive is a powerful tool in analyzing the quality of a particular EC sample, but to the authors' knowledge it remains largely absent as a quality control step in the majority of flux analysis presented in the literature. We believe this is primarily due to the difficulty of incorporating the ogive into an automatic processing algorithm.

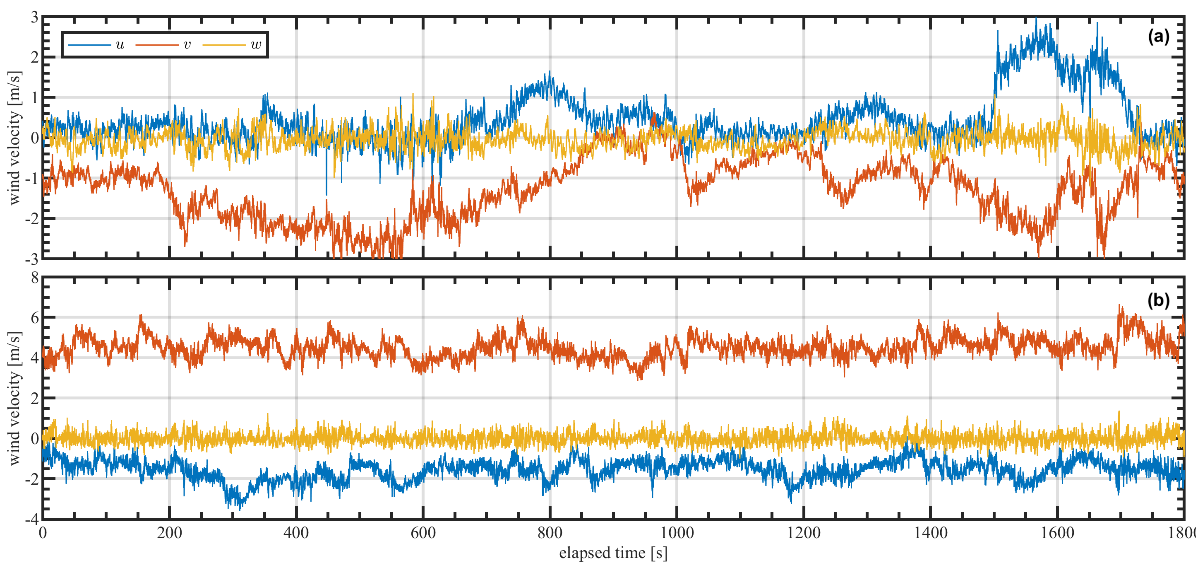

Figure 1: **(a)** EC window from CASPER-West that failed every WISQ test; **(b)** an EC window that passed every WISQ test. These are referred to as **A** and **B** and were observed on 9/30 20:30 UTC and 10/03 16:45 UTC, respectively. The velocity components are the $x \rightarrow u$, $y \rightarrow v$, and $z \rightarrow w$, where u, v, and w are physically the eastward, northward, and upward velocities, respectively.

C. WISQ: Windowed inspection of stationarity & quality

The concept and impetus behind WISQ are not novel. Foken et al. (2012) dedicate the entire Chapter 4 as a reference guide for assuring data quality in EC data. WISQ as an algorithm, nor this report, can supersede their substantial effort. However, the algorithm we describe here provides an efficient and robust approach to testing stationarity for an individual EC sample. WISQ can be readily incorporated into an investigator's or system's data processing algorithm. In particular, WISQ conducts a direct test of the ogive and provides critical information on the homogeneity (or steady state) and convergence of the flux over the prescribed averaging window. For our purposes, we will define an individual EC sample as the discrete time series of variables with temporal extent T_α that is used for the flux calculations. Herein, we will use the term *window* to define the temporal extent of the sample. WISQ is designed around three tests: (1) checking the window is ergodic, (2) checking the flux accumulation over the window is homogeneous, and (3) checking that the integral flux scale is resolved. The outcomes of these tests are denoted E, X, and O, respectively. The results of the three tests are categorical values $[0,1]$, except for O, which includes values $[0,0.5,1]$. For all tests, a value of 0 (1) indicates that the evaluation passed (failed). For O, 0.5 is used as intermediate conclusion, which can be considered a *low pass*. WISQ can also output more detailed information that could be useful for quality control analysis. The details regarding each test are discussed below.

WISQ operates on individual windows (a time series beginning at t_0 and ending at $t_0 + T_\alpha$) of the three orthogonal components of the wind velocity $[u,v,w]$, where w is always the vertical component[iv]. WISQ is an empirical method and therefore literally tests for the user's acceptable amount of *non-stationarity*. Thus, we define ε as the critical limit of non-stationarity. In Kaimal and Finnigan, their equivalent ε is set to 0.02, which is most likely too restrictive for field data. By default, WISQ uses the limit $\varepsilon = 0.1$. The functionality of WISQ will be demonstrated using two averaging windows of data from the CASPER-West field study (see [9]), a window

[iv]A scalar flux version has been developed: WISQS, that applies similar analysis but only requiring w and a scalar record s as input

that failed all the tests ($E = X = O = 1$) and one that passed all of them ($E = X = O = 0$). These will be referred to as windows **A** and **B**, respectively. Both of these samples come from a Campbell Scientific IRGASON system mounted ~5 meters above the ocean surface.

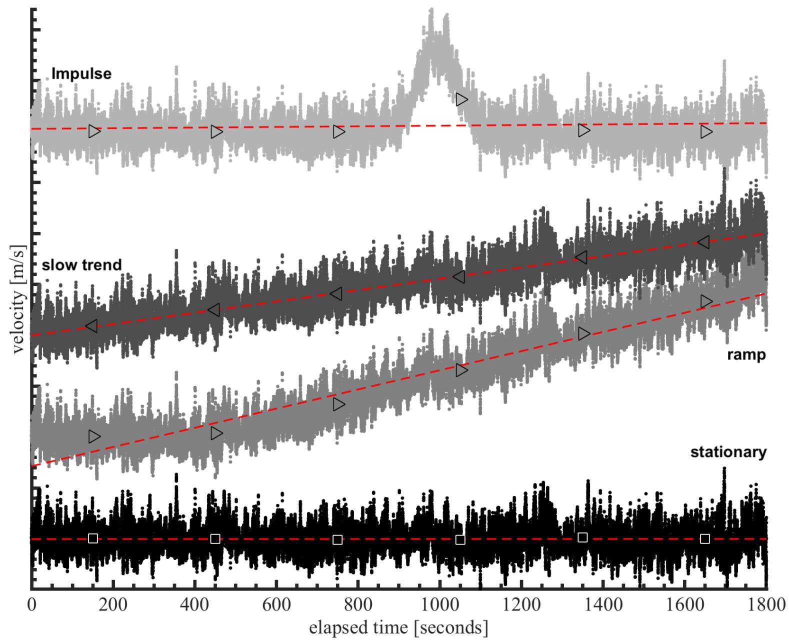

Figure 2: Demonstration of four idealized EC sample phenotypes. Markers denote ensemble averages ($\hat{A}_i$) and the red-dashed line marks a linear trend of ensemble members. Here, the samples have been scaled and translated for easier comparison.

E: Testing for Sample Ergodicity

Ensemble statistics are calculated for both horizontal components $[u, v]$ using discrete sub-windows spread continuously across T_α. The number of realizations within each sub-window is defined as:

$$\hat{n} = \frac{N\delta t}{N_i}, \tag{6}$$

where N is the number of realizations within T_α, δt is the time step (inverse of sampling frequency, n_s), and N_i is the number of sub-windows (user-defined, e.g., 10). If necessary, $\hat{n}$ is the nearest integer value closest to zero. For an ergodic window we assume that,

$$\frac{\sum_{j=1}^{N_i} \hat{A}_j}{N_i} \simeq \bar{A}, \tag{7}$$

again, where $\bar{A}$ is the time-average mean. However, simply confirming (7) over T_α is not sufficient for ergodicity. Figure 2 provides an example of four idealized time series highlighting various conditions found in typical EC data. For example, there is a window that clearly satisfies (7); a window with a slow, mean trend over T_α; a case with a ramp beginning part-way through the record such that $\hat{A}_i \neq \bar{A}$; and a final window with a distinct "impulse" or event halfway through T_α. In this final case, $\hat{A}_i \approx \bar{A}$ and (7) would have been validated to within ε, but qualitatively and intuitively, this is not an ergodic sample.

WISQ evaluates ergodicity by testing the hypothesis that the statistics of the sample over T_α converges on the statistics of N_i ensembles within T_α (i.e., $\hat{A} \rightarrow \bar{A}$). This is done using three sub-tests: (1) that $\hat{A}_i$ are normally distributed, (2) the root-mean-square deviation of $\hat{A}_i$ from $\bar{A}$ is less than ε, and (3) the root-mean-square deviation of $\hat{\sigma}_i$ from $\bar{\sigma}$ is less than ε. The first test (T) is categorical true/false ($\{0,1\}$) determined using a Shapiro-Wilk test algorithm [12][v] and performed on the N_i ensembles. For this test, the confidence interval is fixed at 0.05. For the other two tests, the deviations are defined as:

$$\mu = \sqrt{\frac{\sum_{i=1}^{N_i} (\hat{A}_i - \bar{A})^2}{N_i}}, \tag{8}$$

and,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N_i} (\hat{\sigma}_i - \bar{\sigma})^2}{N_i}}, \tag{9}$$

for the means (μ) and variance (σ^2), respectively. These tests are applied independently to the horizontal velocity components $[u, v]$ and the parameters are denoted by a subscript: T_u, μ_u, etc. From these sub-tests, component-wise indices are defined:

$$I_u = \omega_u \times (T_u + \mu_u + \sigma_u) \tag{10}$$

$$I_v = \omega_v \times (T_v + \mu_v + \sigma_v), \tag{11}$$

where $\omega_{u,v}$ are weighting functions for the relative contribution of the component to the overall flow magnitude, i.e., $\omega_u = |\bar{u}/\sqrt{u^2 + v^2}|$. This weighting is done because WISQ can analyze any orthogonal velocity components, not necessarily those in a stream orientation. The results of E do not want to give equal importance to the cross-stream component as the along-stream

[v] Ahmed BenSaïda (2020). Shapiro-Wilk and Shapiro-Francia normality tests. (https://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests), MATLAB Central File Exchange. Retrieved June 23, 2020.

component, where the former is more susceptible to non-stationarity. In order to pass E, the condition $I_{u,v} < 2\omega_{u,v}\varepsilon$ must be satisfied (a 2 is used because an ergodic sample would have $T = 0$.)

The evaluation of E heavily weights the results of T, because if $T = 1$ then $I_{u,v}$ will always be $> 2\omega_{u,v}\varepsilon$ (for $\varepsilon < 0$). In doing this, we directly test the normality of the EC sample, which is fundamental to the overall application of Reynolds decomposition. This test is limited in that it cannot discern between the stationary and slow, quasi-linear trend examples in Figure 2 (the impulse and partial ramp all *failed* this test). The other two tests directly evaluate whether the ensemble statistics have converged on the sample statistics to a sufficient threshold and are more sensitive to subtle trends in the sample. For example, in Figure 2, the slow trend example has an $I_u = (0+0.29+0.19) = 0.48$, which is greater than the default threshold ($2\times\varepsilon = 0.2$), assuming $\omega_u = 1$ for our demonstration.

The ergodicity evaluation (E) in WISQ provides information on the mean over T_α[vi]. While this is fundamentally important to EC and the Reynolds decomposition of field data, it is not necessarily indicative of the stationarity of the perturbations, or *turbulence*, across T_α. Figure 2 provides a key example to this point. In the Reynolds decomposition (eqn. 1) α' is extracted from α by demeaning *and* detrending. This process effectively removes slow, quasi-linear trends in α. From Figure 2, detrending cannot adequately remove the instance of an impulse or ramp localized to some time within T_α. More sophisticated methods for removing slow, non-linear trends (see [5]), but they are not incorporated into WISQ at this time. It is important to note that while the results of E may not directly pertain to the turbulence statistics, the results of E may provide insight into the overall conditions within the sample window.

[vi]An additional component to E has been added to test the vertical velocity, w. These tests are very similar to those described above, except that only σ is evaluated, such that there is an unweighted index $I_w = T_w + \sigma_w$. The w test is preempted by a check that $|\bar{w}/\sqrt{u^2+v^2}| > 0.01$, and if **true**, then the sample is assumed to be non-ergodic. Exceeding 0.01 is probably caused by very low winds and/or the presence of a mean tilt off the vertical in $[u,v,w]$.

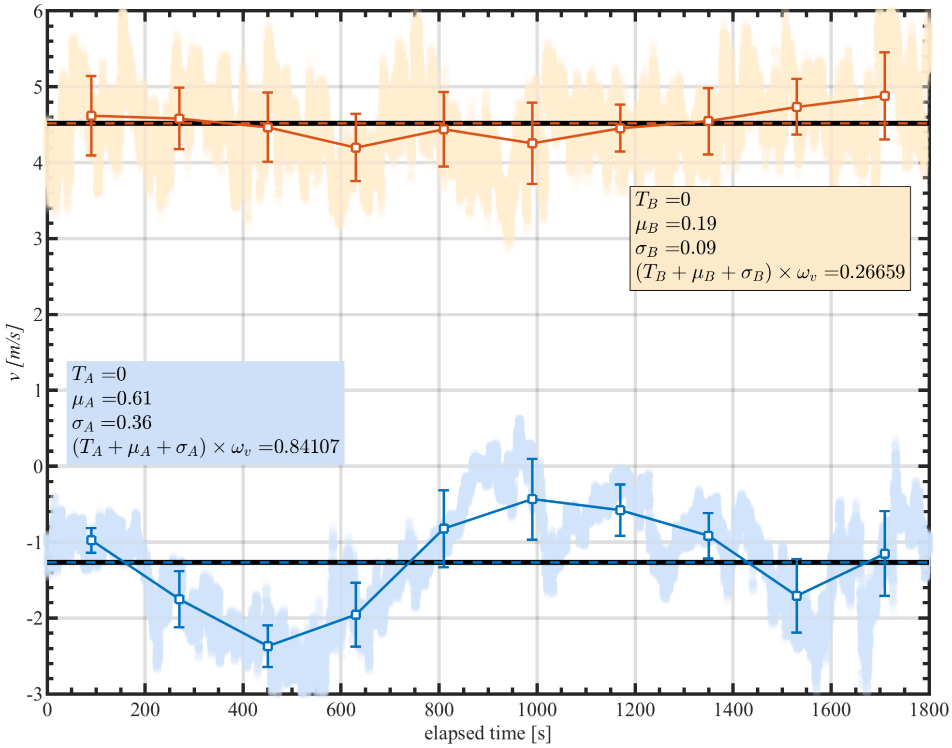

Figure 3: v-component of **A** and **B**, showing the observed (thick lines) and ensemble (error bars) values. The three parameters used to determine E are also presented. The critical values, using $\varepsilon = 0.2$, for these samples are 0.35 and 0.38, respectively. Here we can see that $\hat{v}_i$ from **A** and **B** are normally distributed ($T_{A,B} = 0$), but $I_{v,A}$ is nearly $3\times$ the critical value.

X: Testing for Homogeneous Flux Accumulation

The following two tests, X and O, directly address the stationarity of the flux over T_α. Particularly, the instantaneous vertical flux of stream-wise momentum, $wu(t) = \langle w - \bar{w} \rangle \langle U - \bar{U} \rangle$, where the $\langle \rangle$ indicates a window-wide detrending and demeaning. Here, U is the stream-wise velocity calculated from $[u, v]$ following,

$$\theta = \tan^{-1}(u/v) + \pi \tag{12}$$

$$U = -v\cos\theta - u\sin\theta. \tag{13}$$

The stream-wise component is tested because it is assumed this holds the majority of the flux and is most sensitive to the energy containing scales. This assumption falters in the case of low winds or strong convection—cases, in general, where stationarity itself fails. For scalar fluxes, *WISQS* is recommended because the stationarity results can be quite divergent from the momentum fluxes. X tests the homogeneity of the flux using the normalized cumulative summation, χ. In a sense, χ is the fractional accumulation of flux, which for stationary conditions, should be homogeneously distributed across T_α. The time-dependent accumulation for the stream-wise flux is defined as:

$$\chi_{wu}(t_i) = -\frac{\sum_{i=t_0}^{t_i} wu}{\int_{i=t_0}^{N} wu}, \tag{14}$$

An ideal EC sample would exhibit $\chi \approx \chi_0 = -i/N$, which is simply a direct linear proportionality with a zero intercept and a slope of unity. Here, i/N is the fraction of run, i being the i^{th} realization in T_α and N being the total number of realizations. We will refer to the ideal accumulation function as χ_0. The numerator in (14) is carried out using discrete summation, while the denominator is calculated using trapezoidal numerical integration. Figure 4 compares χ_{wu} for **A** and **B**. The quality of χ_{wu} is evaluated in WISQ using the standardized residual between χ_{wu} and χ_0:

$$R_\chi = \chi_{wu} - \chi_0 \tag{15}$$

$$\sigma_\chi = -\sqrt{\frac{\sum_{i=1}^{N} [R_\chi(t_i) - \bar{R}_\chi]^2}{N}}, \tag{16}$$

where σ_χ cannot exceed the threshold, $\sigma_c = \varepsilon$. Figure 5 provides the time-dependent distributions of R_χ for the two samples from CASPER-West. While **B** does have some $R_\chi \geq \sigma_c$, σ_χ^B for this window is $\sim$0.06.

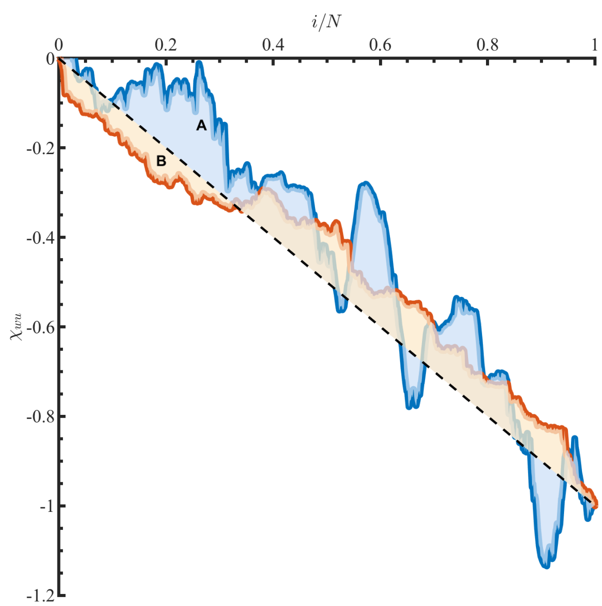

Figure 4: Flux accumulation, χ for **A** and **B**, as a function of fractional segment time i/N. The ideal $\chi_0 = -i/N$ is shown as the dashed line.

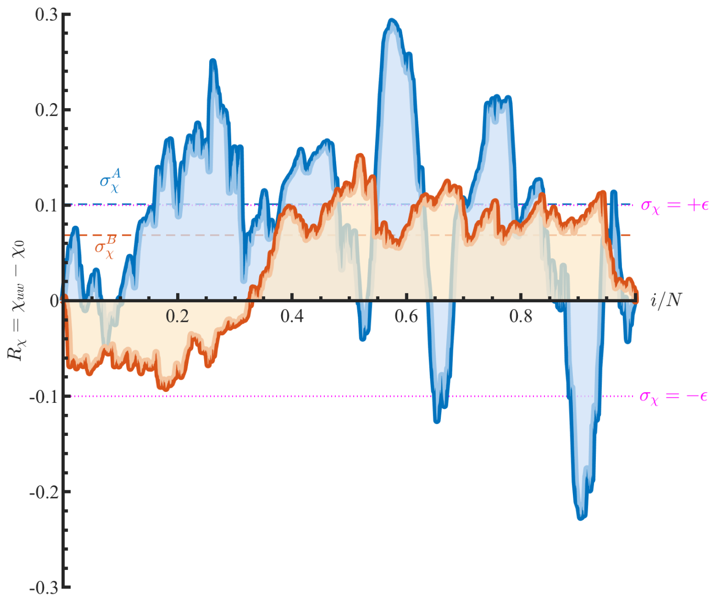

Figure 5: Time-dependent residual, $R_\chi = \chi_{wu} + \chi_0$ with the standard deviations (σ_χ) of both **A** (blue) and **B** (orange). The cutoff standardized residual, $\sigma_c = \varepsilon$, are shown as magenta dotted lines. $\sigma_\chi^A = 0.101$ and $\sigma_\chi^B = 0.069$

O: Testing for the Convergence of the Ogive

The ogive, Γ, is the spectral equivalent to χ and provides critical information on the stationarity and completeness of the flux sample over T_α. In particular, an ogive that asymptotes to $|1|$ at low frequency indicates that the integral scale of turbulence is resolved and thus the given sample is a complete estimate of the local flux. In this sense, the asymptotic behavior—if present—directly relates to the suitability of the prescribed window length. The integral scale tends to shift to higher (lower) frequencies in convective (stratified) conditions. Fundamentally for EC, T_α should be determined iteratively for each local condition in order to achieve the most representative sample of the local flux. However, this is rarely done in practice and it is generally assumed that a single T_α is sufficient—which may be more acceptable in the marine environment where stability tends to be weaker and near neutral.

The co-spectrum, $\Re(F_{wu}) = S_{wu}$, is calculated using the windowed (Hamming or Blackman-Harris) fast Fourier Transform (FFT) scaled into the variance spectrum. As part of the spectral processing, F_{wu} is smoothed using log-uniform averaging routine that has a tunable smooth parameter, s, that increases (decreases) the amount of smoothing when s decreases (increases). Specifically, s controls the slope of the distribution of un-smoothed samples within the smoothed frequency spectrum. This slope, m, is inversely proportional to s, $m \propto s^{-1}$. For example, $s = 4$ increases the number of additional amplitudes averaged together per increase in frequency, twice as quickly as $s = 8$. Due to this relationship, s must be > 0 and saturates in effective smoothing at around $s \sim 24$, therefore a value [2,24] is recommended. Figure 6 provides an example for the impact of s on a sample spectrum from CASPER-West and includes a comparison of the results from a Blackman-Harris and Hamming windowing of the FFT.

Using $s = 20$ and a Blackman-Harris window, the ogives (Γ) for **A** and **B** were compared (Figure 7a). As a reference, WISQ assumes that an ideal C_{wu} (the non-dimensional S_{wu}) should approximately follow a Gaussian distribution. Using the observed $\Gamma \approx 0.5$ and the variance of $\log n$ as scaling parameters, WISQ builds the reference ogive, Γ_0. In this way, Γ_0 is an adaptive reference for Γ. For each window, Γ is tested within each of the major sub-regions of the ogive:

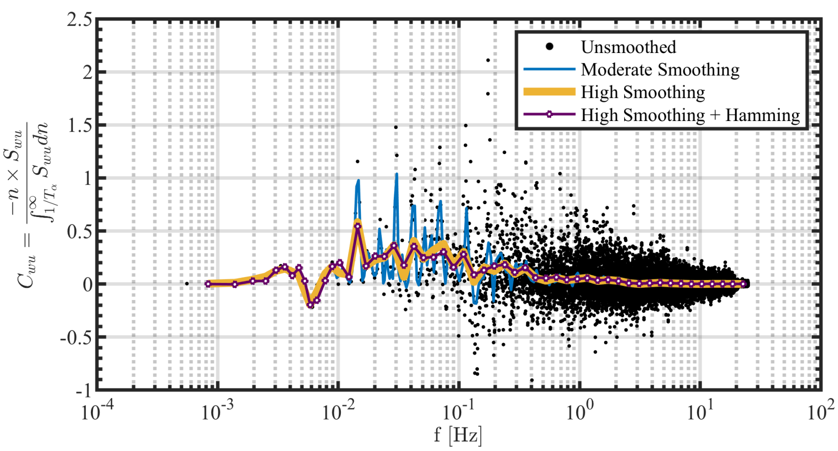

Figure 6: Examples showing varying s in WISQ. For all except one, the Blackman-Harris FFT windowing method was used. Here, moderate and high smoothing are defined as $s = 16$ and $s = 4$, respectively.

(1) the roll-off, $-0.2 < O < 0$ (Figure 7b); (2) the transition, $-0.8 < O < -0.2$(Figure 7c); and (3) the convergence zone, $-1 < O < -0.8$ (Figure 7d). For the roll-off (1), WISQ tests O against an exponential function of the form: ae^{bn}. For this test, it is required that the adjusted r^2 between the observed and model function be above $1 - \varepsilon/2$ *and* that $b < 0$. For the transition (2), we expect that O will follow a quasi-linear function and the only parameter tested for this sub-region is the r^2 between the observed and linear fit. As a preemptive, if the correlation between Γ and Γ_0 is below $1 - \varepsilon$, i.e., an ogive that is grossly malformed, then WISQ assumes

that tests (1) and (2) would fail and skips to (3).

For the convergence zone (3), Γ is compared directly to Γ_0. WISQ determines the number of Γ amplitudes within $\pm\varepsilon/2$ of the corresponding G amplitudes. If the number of amplitudes exceeds 20% of the total number of samples within the convergence zone, then Γ has satisfactorily converged. Otherwise, the observed ogive did *not* converge. Figure 7b-d demonstrate the results of this three-step analysis of Γ for **B**.

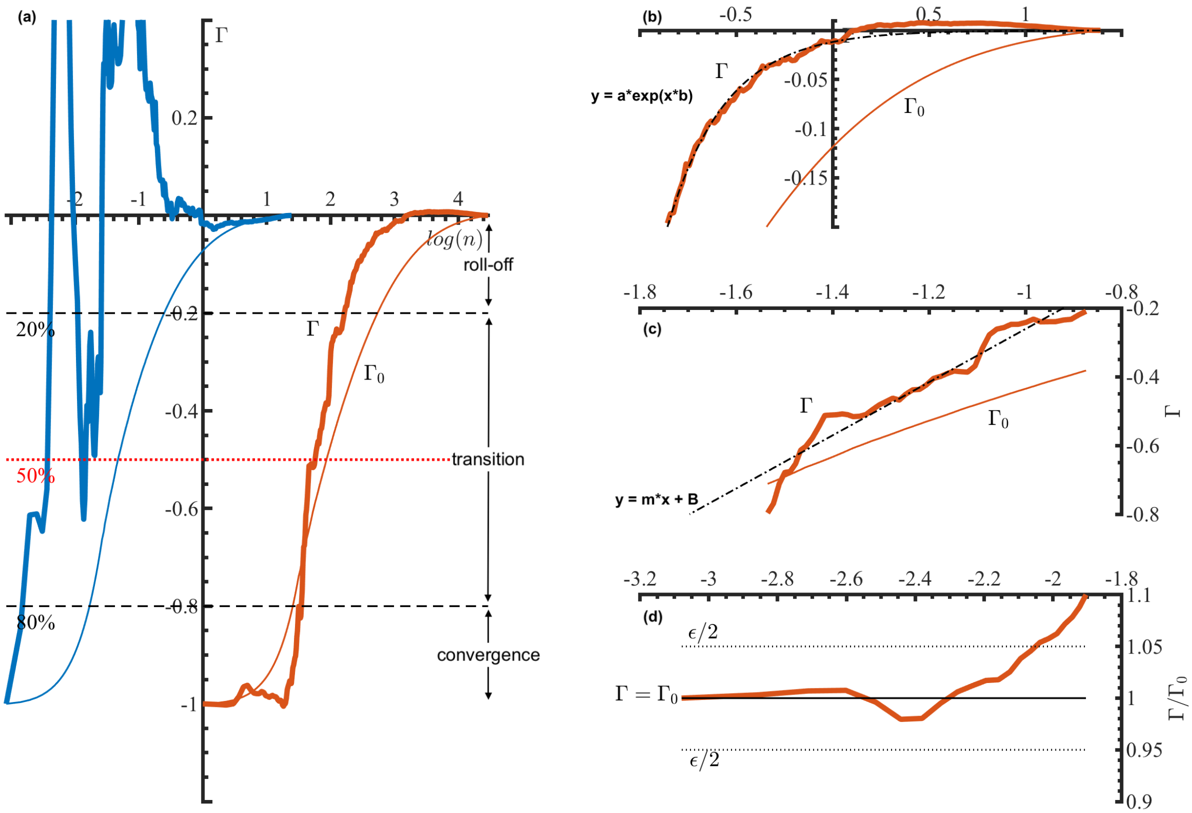

Figure 7: Results of the three-part ogive test applied by WISQ. **(a)** The ogives for **A** (blue) and **B** (orange) are shown. They have been shifted along the abscissa for comparison and are given along-side their respective Γ_0. **(b)** The evaluation of the roll-off against an exponential model, ae^{bn} (dot-dashed black line). For this case, $b = -3.3$ and $r^2 = 0.98$, indicating strong agreement with the expected roll-off behavior. **(c)** Analysis of the transition region, which is expected to be quasi-linear and for this case exhibited an $r^2 = 0.94$. **(d)** Analysis of the convergence zone, showing Γ/Γ_0 as a function of $\log n$. For this case, $\sim$40% of Γ is within $\pm\varepsilon/2 = 0.05$ of Γ_0, which meets the criteria for convergence within WISQ. For reference, the time-to-convergence was $\sim$65 seconds. **b-d** only show results for **B** and the horizontal scale is $log(n)$ for each axis.

D. Results from a field dataset: CASPER-West

WISQ was applied to the CASPER-West observations using a T_α of 30 minutes and analyzing flux sensors at two levels, approximately 5 and 16 m above the mean water level, respectively. A total of 2,185 windows were analyzed for each flux level, but only 1,704 (78%) are analyzed here because of flow interaction with the *FLIP* superstructure and mast. WISQ was systematically run to evaluate the impact of two critical input parameters: N_i and ε, which were applied over the sets $\{4,6,10,20\}$ and $\{0.05,0.1,0.2,0.3\}$, respectively. Of course, E is the only test sensitive to N_i. This analysis will help to demonstrate the sensitivity of the algorithm to different criteria and environmental conditions. In terms of bulk results, E and X were the two tests that failed the most often and the rate of failure was strongly dependent on ε (Figure 8ab). Regardless of N_i, if $\varepsilon = 0.05$, then practically 100% of the CASPER-West samples at both elevations failed ergodicity ($E = 1$). For larger values of ε, increasing N_i tended to increase the prevalence of $E = 1$. Using $N_i = 4$ (the default value for WISQ), about 60% of windows were flagged $E = 1$ (for $\varepsilon = 0.2$), this increased to over 80% if using $N_i = 20$. We would expect these results to be sensitive to T_α and the results of the O test, but this was not evaluated systematically here. We also found a slight z-dependence in these results for $0.1 < \varepsilon < 0.3$, in that about 5% more windows failed this test at 16m versus 5 m. These changes were partially driven by increased prevalence of $T = 1$ with increasing N_i—for $N_i = 4$ only 93 windows failed this criteria ($T = 1$), for $N_i = 20$ this increased to 224, a 2.4-fold increase. However, the predominant cause of the changes in Figure 8a was the relationship between N_i, μ, σ, and ε.

For X and O, the results were insensitive to N_i because ensemble members were not used in their evaluations (Figure 8bc). Predictably, windows more often passed the homogeneity and convergence tests as ε increased to 0.3, changed by about 20% at every interval. There was a significant difference between the proportion of X and O that failed, for a given ε. For example, when $\varepsilon = 0.05$, over 80% of windows failed the homogeneity test, whereas only about 30% of windows did not converge—note that Figure 8c only includes the proportion of ogives where *no* integral length scale was resolved. However, as expected, flux convergence was strongly z-dependent, with the two levels shown here separated by over 10%, the largest margin due to z for any test in WISQ.

The WISQ output also enables investigating the trends within the various quality control metrics. A key parameter for flux analysis is T_α. The time scale of convergence of the ogive is related to energy containing eddy frequency, which must be resolved within T_α for a *complete* flux sample. Figure 9 shows the distribution of this time scale for CASPER-West. Specifically, this distribution reflects the windows where $O \neq 1$ and the time scale was calculated from the frequency of the local minimum of Γ. As expected, for ogives that converged, the majority of time scales approached 20-30 minutes, or T_α was sufficient to resolve all of the relevant turbulent motion for the given z and stability. However, we did also find that nearly >20% of windows converged at time scales less than 6 minutes, or $1/5$ of T_α. These cases coincided with relatively strong convection within the surface layer. For strongly stable cases, which were rare in this data set, we would expect the ogive to *not* converge and, thus, these windows would not be reflected in Figure 9.

Comparable to E, X revealed a drastic change in retention rates (i.e., windows that passed the criteria $X = 0$) with changing ε. From Figure 8, there is a nearly 60% change in the proportion of windows passing this test as ε varies from 0.05 to 0.3. X is strongly related to the mean wind speed and exhibited a persistent diurnal pattern (Figure 10). For $\varepsilon = 0.1$, the most prevalently homogeneous windows typically come from the local maxima associated with the diurnal sea breeze. We can also see that the transitions in wind speed regime do not typically pass criteria for homogeneity; also, while it is rare, low wind periods can be adequatetly homogeneous at the $\varepsilon = 0.1$ cut-off (see inset in Figure 10).

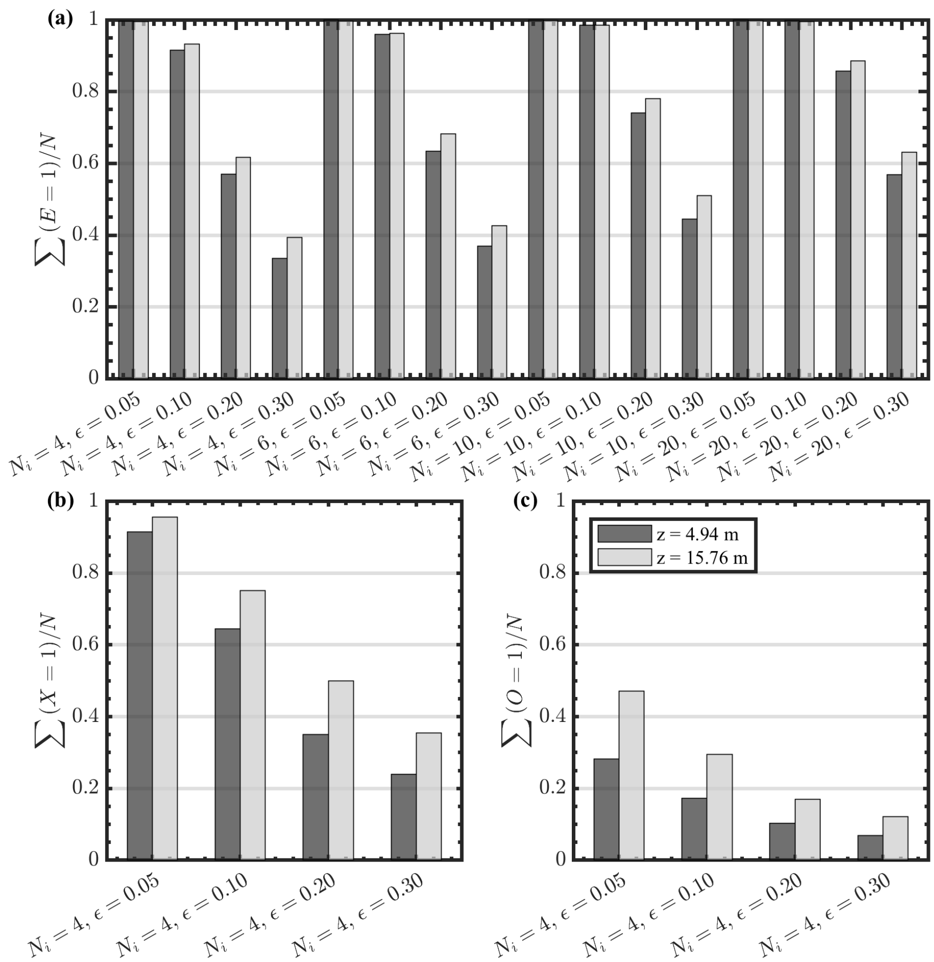

Figure 8: The proportion of windows that failed the three primary evaluations in WISQ: E (**a**), X (**b**), and E (**c**). The latter two are independent of the number of ensembles used N_i. For O, this only includes ogives that did not converge, not ogives that were malformed. The results of two different measurement heights are compared. ε is the tolerance or threshold used to assess the individual averaging windows; N is the total number of windows analyzed (this analysis, and N, does not include windows impacted by flow distortion).

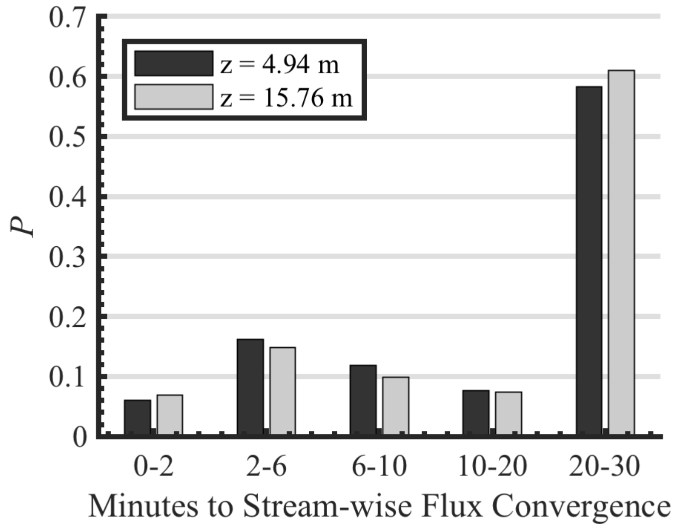

Figure 9: Probability distribution (P) of the time to convergence (in minutes) determined from WISQ analysis of the stream-wise ogive. For the CASPER-West analysis, T_α was set to 30-minutes, therefore, the bin $20-30$ minutes indicates that this interval was appropriate for approximately 60% of the windows where the ogive converged.

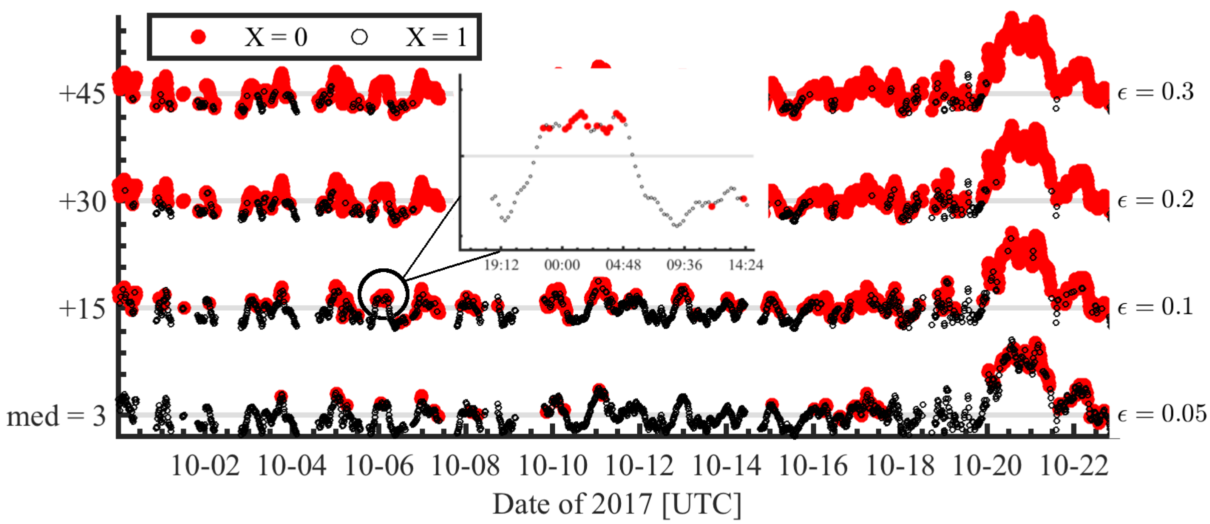

Figure 10: Time series of wind speed(vertical axis) over the CASPER-West time period with the results of the homogeneity test X marked in red-filled ($X = 0$) and open ($X = 1$) circles. The four different cases of ε are shown for comparison, the time series have been translated vertically for illustration ("med" indicates the median value). The inset highlights the characteristics over a slightly less than 24-hour period.

E. Summary

The design and implementation of a novel algorithm (WISQ) that evaluates the quality and stationarity of individual eddy covariance samples has been described. WISQ was designed with efficiency and generalization in mind and can be readily incorporated into already existing processing and quality assessment algorithms. While conceptually WISQ is not new, it is unique in that is provides an automatic evaluation of the temporal and spectral stream-wise flux, which is critical to data quality considerations and had been previously limited to visual inspections. There are always trade-offs when automatically applying an evaluation to noisy geophysical data. However, the primary advantage to these tools, such as WISQ, is they enable the efficient and quantitative analysis post-collection or in quasi-real time that is free from investigator biases and can help standardize the methods applied across a study, project, or discipline. This is especially useful for large datasets, long-term monitoring, real-time processing, and inter-platform comparison. We must emphasize that WISQ provides guidance through its flagging system, but it remains the investigator's prerogative whether or not particular flux sample is deemed high quality and stationary. While WISQ was designed using an atmospheric turbulence dataset, the method is general and applies to any sampling of turbulent flows where stationarity is a concern for data quality. The software is maintained at `https://gitlab.nps.edu/dortizsu/wisq`.

THIS PAGE INTENTIONALLY LEFT BLANK

REFERENCES

[1] Tarek S. El-Madany, F. Griessbaum, G. Fratini, J. Y. Juang, S. C. Chang, and O. Klemm. Comparison of sonic anemometer performance under foggy conditions. *Agric. For. Meteorol.*, 173:63–73, may 2013.

[2] Thomas Foken, Marc Aubinet, and Ray Leuning. The Eddy Covariance Method. In *Eddy Covariance*, pages 1–19. Springer Netherlands, 2012.

[3] Huijun Huang, Hongnian Liu, Jian Huang, Weikang Mao, and Xueyan Bi. Atmospheric Boundary Layer Structure and Turbulence during Sea Fog on the Southern China Coast. *Mon. Weather Rev.*, 143(5):1907–1923, may 2015.

[4] J.C. Kaimal and J.J. Finnigan. *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press, Oxford, 1st edition, 1994.

[5] Luís G. N. Martins, Scott D. Miller, and Otávio C. Acevedo. Using Empirical Mode Decomposition to Filter Out Non-turbulent Contributions to Air–Sea Fluxes. *Boundary-Layer Meteorol.*, 163(1):123–141, apr 2017.

[6] R. B. Montgomery. Vertical Eddy Flux of Heat in the Atmosphere. *J. Meteorol.*, 5(6):265–274, dec 1948.

[7] D. G. Ortiz-Suslow, B. K. Haus, N. J. Williams, H. C. Graber, and J. H. MacMahan. Observations of Air-Sea Momentum Flux Variability Across the Inner Shelf. *J. Geophys. Res. Ocean.*, 123(12):2018JC014348, dec 2018.

[8] D. G. Ortiz-Suslow, B. K. Haus, N. J. Williams, N. J. M. Laxague, A. J. H. M. Reniers, and H. C. Graber. The Spatial-Temporal Variability of Air-Sea Momentum Fluxes Observed at a Tidal Inlet. *J. Geophys. Res. Ocean.*, 120(2):660–676, feb 2015.

[9] David G. Ortiz-Suslow, John Kalogiros, Ryan Yamaguchi, Denny Alappattu, Kyle Franklin, Benjamin Wauer, and Qing Wang. The Data Processing and Quality Control of the Marine Atmospheric Boundary Layer Measurement Systems Deployed by the Naval

Postgraduate School during the CASPER-West Field Campaign. Technical report, Naval Postgraduate School, Monterey, CA, 2019.

[10] David G. Ortiz-Suslow, Qing Wang, John Kalogiros, Ryan Yamaguchi, Tony De Paolo, Eric Terrill, R Kipp Shearman, Pat Welch, and Ivan Savelyev. Interactions Between Nonlinear Internal Ocean Waves and the Atmosphere. *Geophys. Res. Lett.*, 2019.

[11] L. Prandtl. Bericht {\"{u}}ber Untersuchungen zur ausgebildeten Turbulenz. *Zeitschrift für Angew. Math. und Mech.*, 5(2):136–139, 1925.

[12] S. S. Shapiro and M.B. Wilk. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3-4):591–611, 1965.

[13] H. Tennekes and J.L. Lumley. *A First Course inTturbulence*. MIT Press, Cambridge, Massachusetts, 1972.

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia
2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California