Determining the Optimal Sampling Rate of a Sonic Anemometer Based on the Shannon-Nyquist Sampling Theorem

Andrew Mahre*

National Oceanic and Atmospheric Administration

Research Experiences for Undergraduates (REU) Intern

Norman, Oklahoma

Gerry Creager

National Oceanic and Atmospheric Administration

National Severe Storms Laboratory

Forecast Research and Development Division

Norman, Oklahoma

ABSTRACT

While sonic anemometers have been in use for nearly 50 years, there is no literature which investigates the optimal sampling rate for sonic anemometers based on the Shannon-Nyquist Sampling Theorem. In this experiment, wind is treated as a wavelet, so that sonic anemometer data with multiple sampling rates can be analyzed using spectral analysis techniques. From the power spectrum, it is then possible to determine the minimum frequency at which a sonic anemometer must sample in order to maximize the amount of information gathered from the wavelet, while minimizing the amount of data stored. Using data from the Oklahoma Mesonet and data collected on-site, no obvious peak is present in any resulting power spectra that can be definitively be considered viable. This result suggests a nearly random power distribution among frequencies, which is better-suited for averaging and integrating data collection processes.

1. Introduction

a. Motivation

This experiment was conducted to attempt to establish a set, optimal sampling rate based on the frequency and power content of the wind's 'signal'. Currently, sonic anemometer data is averaged over a two minute time window. This technique, described further in section 1.b, has been carried over from older types of anemometers, and may not be as well suited for non-mechanical anemometers. While a temporal average may give a better estimate of the general wind pattern for a given location, it does not provide as much information for those interested in modeling the boundary layer, who may be interested in more specific details of the wind and its properties.

b. Background

Prior to the invention of the sonic anemometer, anemometers had used mechanical moving parts in their measurement of wind speed. Because of the momentum associated with moving parts of finite mass, the wind speed which is measured at a given point in time with one of these anemometers is not instantaneous; there is some amount of time that it takes the anemometer to accelerate and decelerate. It is then necessary to take a running temporal average to account for this start-up time and slow-down time. However, for sonic anemometers, the wind speeds which are measured are nearly instantaneous measurements of the wind at that particular time. Temporal averaging for sonic anemometers was adopted in part to allow direct comparison between datasets which use mechanical anemometers and those which use sonic anemometers; there is no scientific basis for the 2-minute average on sonic anemometer data.

c. Prior Research

No studies regarding the optimization of the sampling rate of a sonic anemometer have been conducted to date. Previous papers in the atmospheric sciences have used techniques of wavelet analysis to study surface turbulence during hurricane landfall and to analyze meteorological variables under the winter thunderclouds. General signal processing and spectral analysis techniques were used for this study.

2. Methods

a. Data Used

For data analysis in this project, multiple datasets were acquired. The primary dataset was from an experiment by the Oklahoma Mesonet on July 30, 2003, using a sonic anemometer (unknown manufacturer) sampling at a rate of 10 Hz. Due to concern about the sampling rate being too fast, decimated versions of the data were also analyzed, at sampling rates of 5 Hz and 2.5 Hz. In addition, daily data was collected and analyzed from a local station of the Oklahoma Mesonet, which samples every 3 seconds. While data analysis on these data was being conducted, data were being collected simultaneously by two Vaisala anemometers on-site, both with a sampling rate of 1 Hz. Looking at data at various sampling rates was helpful, as data which samples at too high of a frequency will introduce noise in addition to the signal; if the sampling rate is too low, then higher frequencies (higher than the Nyquist Frequency) will not be analyzable, and signal aliasing becomes an issue. The concepts of Nyquist Frequency and aliasing will be described further in Section 2.b.

b. Spectral Analysis Techniques

Before any analysis was conducted, the data were first subjected to a number of quality control checks. The first check was to throw out any data which was flagged as 'bad' by the anemometer. The second check was to conduct a simple 'sanity check'; if there were any data points that obviously were erroneous, these were discarded as well. The final check was conducted through an algorithm which computed a running average of acceleration for all points, and if the instantaneous acceleration between two points was at or above a certain threshold and significantly differed from the running average, then it was flagged for further evaluation. This helped to identify multiple regions of data where the raw velocity did not immediately appear suspicious, but where it was determined that the instrument could have introduced a bias.

For questionable wind speed values in the 10 Hz dataset, the wind speed data was compared against the meteogram for the closest Mesonet stations to the data collection, to check for possible real meteorological conditions causing a sudden change in wind speed. Because of the spatial distance between the data collection site and the Mesonet site, the same downburst might not hit the same location, but the Mesonet data can be used to see what sort of weather was in the area that day.

Once the data were run through the proper quality control checks, analysis of the data was conducted. To observe the frequencies at which the signal was obtaining most of its power, a power spectrum was plotted for the wind speed. The first step involved processing the speed data through a window function. This function helps to eliminate end effects of the data by 'weighting' data near the center of the time interval being analyzed. End effects can result in spectral leakage, making the peak or peaks of interest more difficult to see. For this case, a Hamming window function was used, given by the equation

$$w(n) = \alpha - (1 - \alpha)\cos\left(\frac{2\pi n}{N-1}\right) \qquad (1)$$

where n = index number, going from 0 to N - 1, N = total number of samples, and the standard $\alpha = 0.54$ level was used. The α level was adjusted to 0.50 for some of the data, as described in Section 3. Figure 1 shows the subtle difference in these two window functions. These 'windowed' data were run through a Fast Fourier Transform (FFT), yielding N complex outputs, with each output corresponding to a different frequency. The first half of the outputs correspond to frequencies up to the Nyquist Frequency, while the 2nd half correspond to frequencies between the negative Nyquist Frequency and 0. The Nyquist Frequency, which is the highest frequency that can be observed by the power spectrum, is given by the equation

$$f_{Nyquist} = 0.5(f_s) \qquad (2)$$

where f_s is the sampling rate. Only positive values of frequency were used in the analysis, as the negative frequencies should contain the same information as the positive frequencies.

Because the power spectrum can only see up to the Nyquist Frequency, aliasing becomes a concern. Aliasing is a phenomenon that occurs when the signal's frequency of interest is higher than the Nyquist Frequency. The power from this signal becomes folded back into the visible power

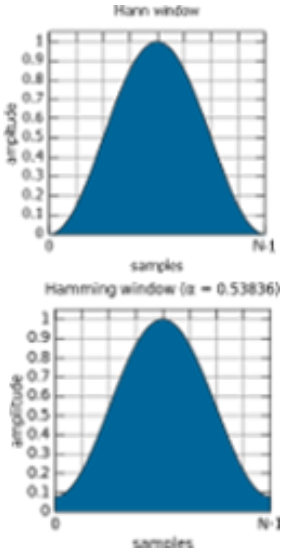

FIG. 1. This figure shows the two window functions used; the lower graph is the Hamming, which uses $\alpha = 0.54$. The upper graph shows the Hann, a subtle variant of the Hamming, which uses $\alpha = 0.50$. The Hann, unlike the Hamming, sets the datapoints on the ends to 0, rather than to 8% of their value.

spectrum, and can look as if the power is coming from other frequencies. In order to ensure that the power spikes that are seen are not being aliased, a higher sampling rate is necessary, which is mentioned in Section 5.

The magnitude of each output was obtained by multiplying each output by its complex conjugate and taking the square root. Thus, for an output of $a + bi$, the magnitude is given by $\sqrt{a^2 + b^2}$. The magnitude represents the amount of the signal's power given by that particular frequency. For analysis, the power spectra were run through a boxcar smoothing function, which attempts to eliminate noise in the spectrum. This function works by computing a running average over a certain number of points; for a 17 point boxcar, as was used here, the i^{th} point in the smoothed graph was computed as the unweighted average of the $(i - 8)$ to $(i + 8)$ points. Other smoothing functions (triangular, etc.) could be used, but were found to yield similar graphs as the unweighted boxcar smoothing function. The number of points in the boxcar smoothing function was chosen by trial and error; 17 eliminates a significant portion of the noise, but using more points has little effect on the power spectrum noise level.

3. Results

Once the magnitude was calculated, it was plotted against frequency to determine the frequencies at which there are large, sudden increases in power. A 'spike' in power at a given frequency, X, would indicate that the sampling rate must be at least 2X. After analyzing the data from the Mesonet Joint Urban experiment at the full 10 Hz and decimated versions at 5 Hz and 2.5 Hz, as well as daily Mesonet data at $\frac{1}{3}$ Hz and two on-site datasets at 1 Hz, a spike was observed only in the Joint Urban data. The spike in power in the power spectrum came only from a small section of data, which lasted less than an hour, where the wind speed reached values that could possibly be unphysical (up to 62 m/s). Outside of this section of the Joint Urban data, no spikes were observed, as seen in Figure 2. The data were all analyzed in the same manner as described in Section 2b, with the exception of the on-site data. These data were analyzed with a special case of the Hamming window, called the Hann window, which uses $\alpha = 0.50$ instead of $\alpha = 0.54$, in order to further negate end effects. For the on-site data, the collection program terminated unexpectedly, shortly after a sharp rise in wind speed. Because it is unclear whether or not the data near the end are good, the Hann window was used instead of the Hamming, in order to avoid having to make this judgment call. Aside from the different treatment of the data near the beginning and end of the dataset, there is little difference between the Hann and Hamming window functions, and all datasets were treated equally otherwise.

The power spectrum was plotted for each dataset on both a linear set of axes, and on a log-y axis. The log-y plot takes the log of the power and multiplies by 10 to obtain the power in decibels.

4. Conclusions

These preliminary results disprove the initial hypothesis that there is a defined spike in power which would clearly define the optimal sampling rate. While initially a spike was observed in the power spectrum, this contribution to the power came only from when the equipment's heating element was enabled; after further analysis, this spike at 0.1 Hz (and repeating spikes at 0.3, 0.5, and 0.7 Hz) is believed to be an artifact of the instrument, and not the wind. An autocorrelation function, which is a measure of the correlation between neighboring datapoints, also shows this phenomenon. During the meteorological events which caused the increase in wind speed, the heating element on the anemometer was activated during a few short periods of time for an unknown reason. For further analysis purposes, a new dataset was created where wind speed value in the data entries where the heating element was activated

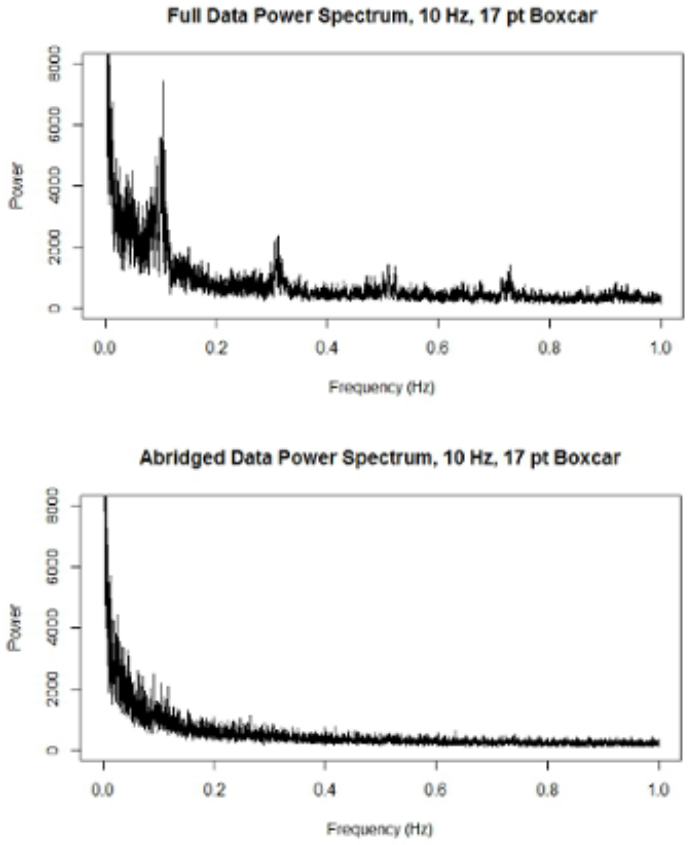

FIG. 2. The top graph represents the power spectrum for the full 10 Hz wind speed dataset, while the bottom dataset has excised 50000 datapoints surrounding a possible meteorological event. Note the presence of power spikes in the top graph, and their absence in the lower graph. Both the x and y axes are linear.

was replaced with a running average of wind speed values in surrounding data entries; this dataset is generally referred to as being 'smoothed'. The wind speeds for the full, abridged, and smoothed datasets are shown in Figure 3.

This method eliminated the spike in power in the power spectrum, leading to the conclusion that the instrument could have produced the power spike. A demonstration of this phenomenon is shown in Figure 4.

5. Future Work

While there is no defined spike in the power at a given frequency, an optimal sampling rate can be determined by calculating the frequency at which the signal's power levels off to a constant value. While an initial, qualitative analysis of the power spectra has yielded an approximate value of 0.3 to 0.5 Hz where the power levels off, a more quantitative analysis is necessary to determine the exact sampling rate at which the power levels off. At greater frequencies, the sampling does not contribute more to the signal power, and can add to the noise of the signal.

Acceleration data was assessed with the same general techniques used for the wind speed, but further analysis must be conducted to determine if there is a spike in a viable acceleration power spectrum. From an initial periodogram, it appears that any spikes in the speed power spectrum match up with peaks in the acceleration power spectrum.

It has also been proposed that 10 Hz is not a sufficiently high sampling rate for observing the spike in the power spectrum. For future research on this topic, an anemometer with a sampling rate of 60 Hz would be ideal, as decimated versions of the data could be used if 60 Hz is too high of a sampling rate.

Future analyses will also look into the color of

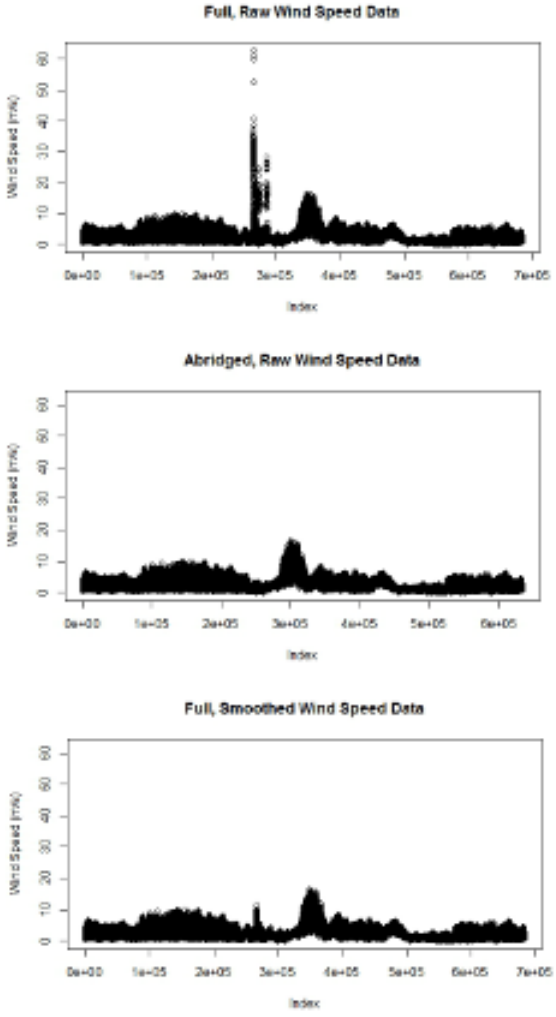

FIG. 3. The 3 graphs above show the wind speed in m/s from the full dataset, the abridged dataset, and the smoothed dataset.

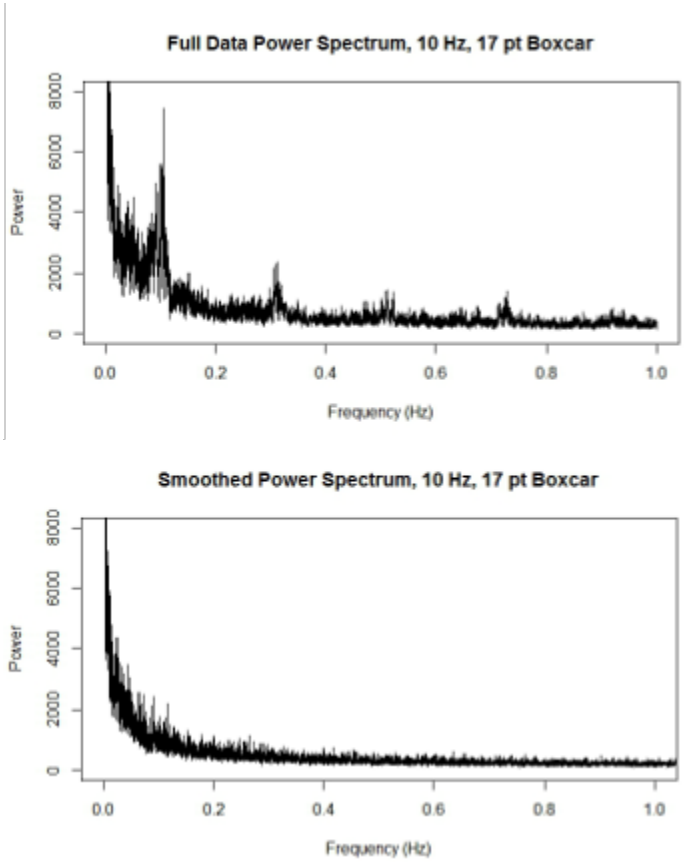

FIG. 4. The top graph represents the power spectrum for the raw, unaltered wind speed data; the bottom is the power spectrum from the set of wind speed values where a running average replaces any datapoint where the instrument's heating element has been enabled.

noise that is present on the power spectra. Initial work has shown that a red (Brownian) noise phenomenon appears to be present in the power spectrum, which results in high correlation between one data point and its neighbors, which could be the result of oversampling. An autocorrelation function, used in Section 4, shows high correlation between datapoints close together, yielding the belief that red noise may be present.

Acknowledgments.

I would like to thank my mentor, Gerry Creager, for his help and guidance on this project. I would also like to thank Dr. Chris Fiebrich, for allowing us to use the 2003 10 Hz data, as well as the daily data from the Oklahoma Mesonet. I would like to thank the National Science Foundation for funding the grant (NSF-AGS 1062932). I would like to thank Dr. Kim Elmore and Matt Carney for their assistance in the project, as well as the input from Dr. Michael Richman.

REFERENCES

Kay, S. M., 1988: *Modern Spectral Estimation: Theory and Application*. Prentice-Hall.

Takeuchi, N., K. I. Narita, and Y. Goto, 1994: Wavelet analysis of meteorological variables under winter thunderclouds over the japan sea. *J. Geophys. Res.*, **99**, 10 751–10 757.

Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences*. 2d ed., Academic Press.

Zhu, P., J. A. Zhang, and F. J. Masters, 2010: Wavelet analyses of turbulence in the hurricane surface layer during landfalls. *J. Atmos. Sci.*, **67**, 3793–3805.