Collection and Synthesis of Historical Data from Ocean Weather Station Papa

Prepared by:

Derek Belka

Undergraduate
University of Washington
Civil and Environmental Engineering
December 2013

Table of Contents

List of Figures and Tables

Figures

Tables

Abstract

Meteorological data collection began at Ocean Weather Station Papa (50° N, 145° W) in the 1940s and continued more or less continuously until 1981 when the Canadian government defunded the Ocean Station program. Papa was revived as a permanent observation station in the mid-2000s when researchers at the National Oceanographic and Atmospheric Administration (NOAA) and University of Washington Applied Physics Laboratory (UW-APL) deployed moorings to collect meteorological and oceanographic data at the site. Stakeholders in the modern mission have been searching for a robust, quality-controlled historical dataset to use as a contextual reference for contemporary measurements. This report details the early stages of this process, primarily resolution of available data from several different sources. Ultimately, it was discovered that these seemingly disparate datasets were in fact portions of a more complete dataset archived at the University Corporation for Atmospheric Research (UCAR). Further work will involve more extensive quality controlling, developing historical baselines, and searching for climatic trends.

Introduction

As World War II escalated in the Pacific, the US military began a strategic advantage initiative to develop more accurate weather forecasts for trans-Pacific military and domestic supply lines. This initiative resulted in the establishment of two marine weather-surveying stations, one near Hawaii and the other in the Gulf of Alaska. The latter of these two was originally dubbed 'Peter', but was later changed to 'Papa' after the NATO Phonetic Alphabet. Ocean Weather Station Papa (OWS Papa, 50° N 145° W) remained under US military control until 1951 when the Navy abandoned the position and the Canadian Coast Guard took over. Ship-based meteorological measurements were taken at OWS Papa almost continuously from the 1940s until 1981 when budgetary restrictions forced the end of the Canadian Weather ship program. During OWS Papa's active years as a weather ship station, a survey along the commonly used heading to and from the mainland was also established. Later dubbed Line P, the survey was recognized as a valuable research program and Institute of Ocean Sciences (IOS) vessels continued the survey after the Canadian Weather ship program shutdown.

The observations catalogued from 1951 to 1981 represent one of the oldest, and longest, contiguous marine weather datasets available today. And yet, the data collected during this time period has not been effectively examined and interpreted, nor has the ocean climate at OWS Papa been quantitatively characterized. This makes the interpretation of modern, shorter datasets more difficult to understand. For example, there may be interdecadal trends at OWS Papa that could lead to misinterpretation of research outcomes from modern observations. Developing and quality-controlling this historical dataset will provide a much-needed contextual reference and scientific tool. This report will detail the early stages of this process that included the resolution of several data sources to a single package, reformatting that package into a more robust instrument of science, and provide an objective assessment of overall data quality. It will also express the author's opinion of what remains to be done with the dataset and the next steps in the project.

Data Sources

Meghan Cronin, NOAA

In mid-2013, Meghan Cronin supplied several files containing weather ship data collected at OWS Papa (personal communication, May 16, 2013). In a series of emails she expressed that she was uncertain of the data quality as well as its completeness. These files had come to her somewhat circuitously from Howard Freeland, an employee at IOS, who had obtained them from a former staff member's hard drive. In his own search for the ship-based observations, Howard had also contacted Bill Large of UCAR, who directed Howard to his archived data, though it did not seem to be accessible. Additionally, another IOS employee, who also had a copy of the weather ship data, contacted Meghan in March 2013 but indicated that this was probably the same copy that Howard had retrieved from the retiree.

The data came in several different formats, but lacked documentation about quality control. Initially, it was believed that the lack of quality control documentation indicated a relatively raw set of data. A colleague of Meghan's completed some fairly extensive research and found several articles about the weather ships themselves and distilled them into relevant notes that may prove to be useful interpretation tools. This data is referred to as 'NOAA data' for the remainder of this report.

UCAR Archive

Initially, poor organization and ambiguous file identification prevented the pursuit of this data as a source, but the seemingly improper tampering of the NOAA data, discussed in subsequent sections, led to revisiting the UCAR archive. The data, supplied with documentation but still no quality control information, is stored in large, ASCII-formatted text files containing measurements from roughly half of the ocean stations. Finding, and isolating, OWS Papa from this was relatively easy, albeit time consuming. Distilling the 34 individual variables required parsing each of 93,023 lines of coded observations using the data format described in the National Climatic Data Center's documentation for DSI-1129 (NCDC, 2003).

Modern Data

Rather than ship-based observation, modern wave data collection relies on a Datawell directional waverider buoy continuously deployed at OWS Papa and owned by the University of Washington Applied Physics Laboratory (UW-APL). This data is available online through the National Data Buoy Center (NDBC, www.ndbc.noaa.gov) under Buoy 46246. Wind data was supplied by a nearby buoy owned by NOAA and the Pacific Marine Environment Laboratory, also available online through the NDBC webpage under Buoy 48400.

Resolution and Quality Control of Historical Data

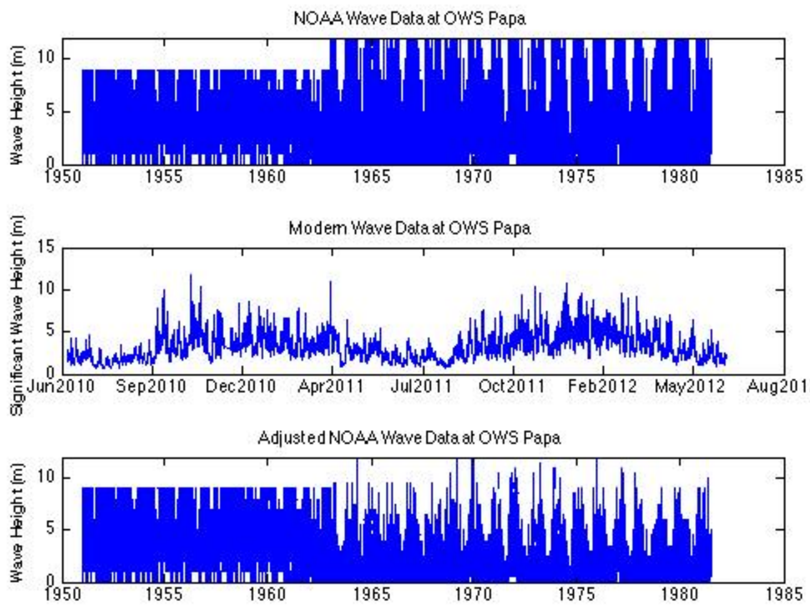
As previously stated, the primary goal of this study is to develop an implementable historical OWS Papa dataset. Cross-examination of the datasets should help determine what overlaps, if any, exist and how these might aid in the determination of the most appropriate data source. Additionally, some datasets may contain different categories of values that may prove useful for interpreting values from other datasets. Initial work has consisted of determining the best raw sources of data and compiling them into a complete, quality controlled data package. It is worth mentioning that the disparate datasets from NOAA were found to be identical data contained in different file formats.

Wave Height

The NOAA data source indicated a predilection to the belief that the wave height values contained in the provided dataset were suspect and would provide a good starting point for analysis (Cronin, personal communication, May 16, 2013). A preliminary plot of the complete wave height series (Figure 1) showed serious discrepancies when compared to the modern set. The most notable and obvious of these discrepancies is the significant change of magnitude in 1962. A small magnitude of change might be expected in 1968 as wave height measurements prior to that were artificially capped at 9.5-meters because of pre-1968 observation recording techniques. However, this adjustment occurs 6 years prior to its expected date. The dubious doubling of magnitude prior to 1968 also seemed to indicate that these earlier values had been adjusted while the later values had not. The pre-1962 values seemed to be similar in magnitude to modern wave heights, further supporting the conclusion that they had been adjusted at some point. The third plot in Figure 1 shows a correction to the wave height values from 1962 onward, and seems to be more reflective of the modern data, at least in terms of the magnitude. However, there also appears to be a maximum value of 9-meters in the pre-1962 data and not the expected 9.5-meters, indicating there was never a 9.5-meter, or greater, wave height recorded until after 1962 despite 9.5-meter, or greater, wave heights being relatively frequent occurrences in the modern data and adjusted data. The expected 9.5-meter cap notably appears on data from 1962 through 1967, although there was not a single instance of a half-meter value in the pre-1962 data indicating that no half-meter values were ever recorded in this period.

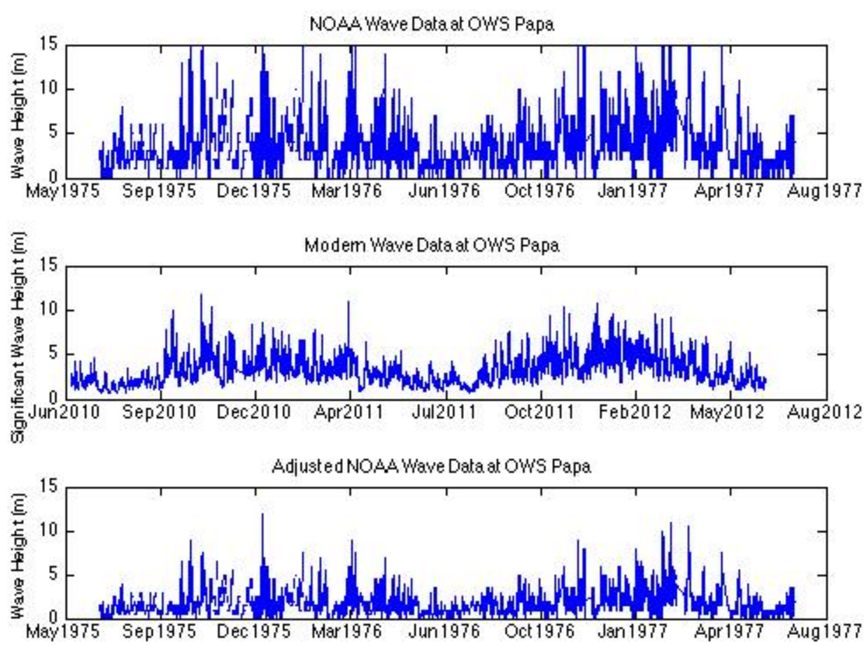
While this comparison is revealing in that it exposes several possibly erroneous quality control activities, it does not show a correlation between historical data and modern data characteristics other than local minima and maxima. Figure 2 shows a plot of the NOAA data compared to the modern data on a similar time scale. By isolating a similar time period, it becomes clear that the unadjusted NOAA data is in fact roughly double the modern data. For the remainder of the report, only these corrected values will be considered. A supplemental figure in the appendices shows a similar comparison for pre-1962 and modern data.

Figure 1. Full Time Series Wave Data Magnitude Comparison



The full time series of the datasets shows that unadjusted historical wave data is dissimilar in magnitude to modern values. However, it does become similar when the later years are halved.

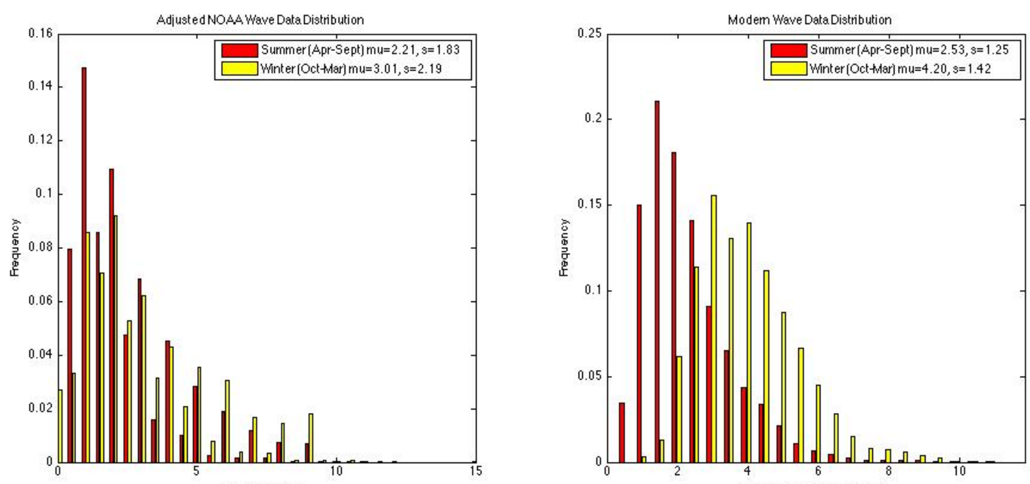
Figure 2. Comparison of Post-1968 Wave Data Magnitudes



Comparing a section of the historical data similar in time scale to the modern data enables a better comparison of the need to correct the NOAA wave data. The third plot shows this correction and is noticeably more similar to the modern data.

These figures are useful only in that it is possible to see the correlation in magnitude between the historical and modern data. However, it is not clear from the plotted data whether or not they are similarly distributed as one might expect. In order to determine if these measurements are in fact similar in character, it is more useful to use statistical methods. Rayleigh distributions are particularly well suited to determining statistical similarities in distributions that contain only positive values, such as wave heights. In the case of relatively large datasets such as these, one would expect similar standard deviations and means for data that is relatively similar in character. The statistical analysis in Figure 3 shows that both data sets are similarly distributed about their means, although the means are not similar. This may be an artifact of bias in the individual datasets. One notable difference is that the modern data does not contain any zero values nor does it have a high frequency of values close to zero. The modern data also shows a clear difference between seasonal wave characteristics. By comparison, the adjusted NOAA data has a high frequency of zero values and values near zero with relatively similar seasonal distributions. These differences could explain the dissimilarities between historical and modern data means.

Figure 3. Comparison of Wave Data Distributions

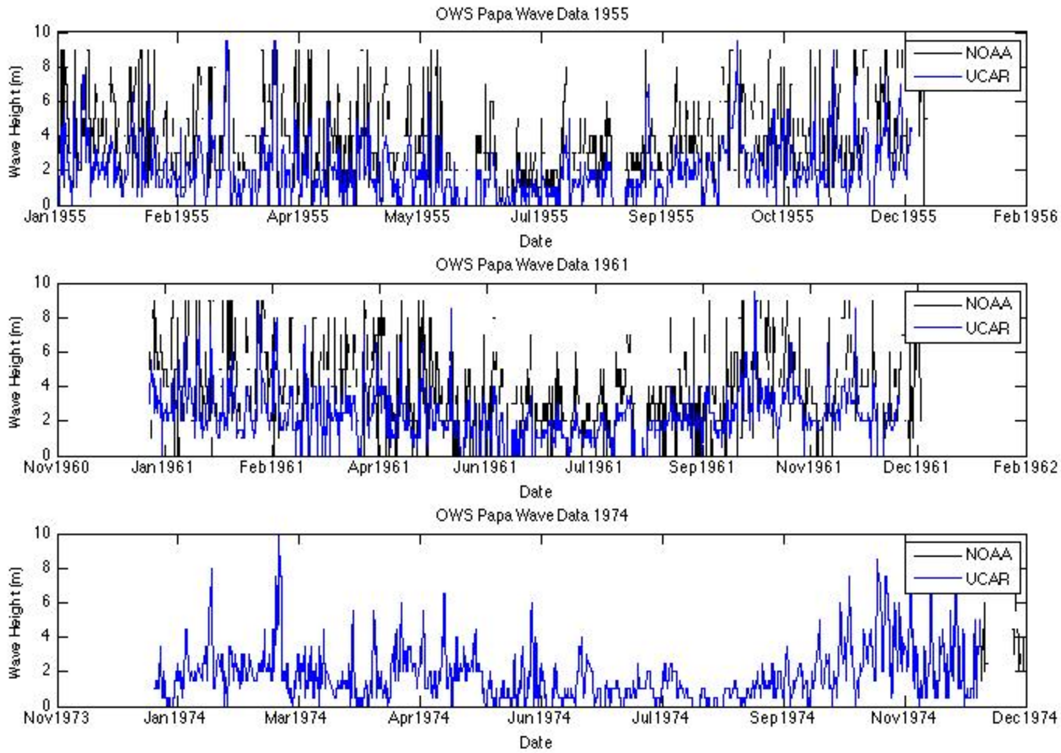


While the distributions appear similar, the modern data is notably more severe in the winter than the adjusted NOAA data. This may be due to observational and/or statistical biases.

After examining these divergences it became clear that the NOAA data, while somewhat similar in character, had been altered in a statistically significant manner before it was acquired for this project. This conclusion led to the pursuit of an alternate data source rather than attempting to backtrack the alterations. It had been previously indicated that another dataset existed on UCAR's website. A similar analysis was conducted on data obtained through UCAR's data access portal (rda.ucar.edu/datasets/ds535.0/) and compared to the modern and NOAA data.

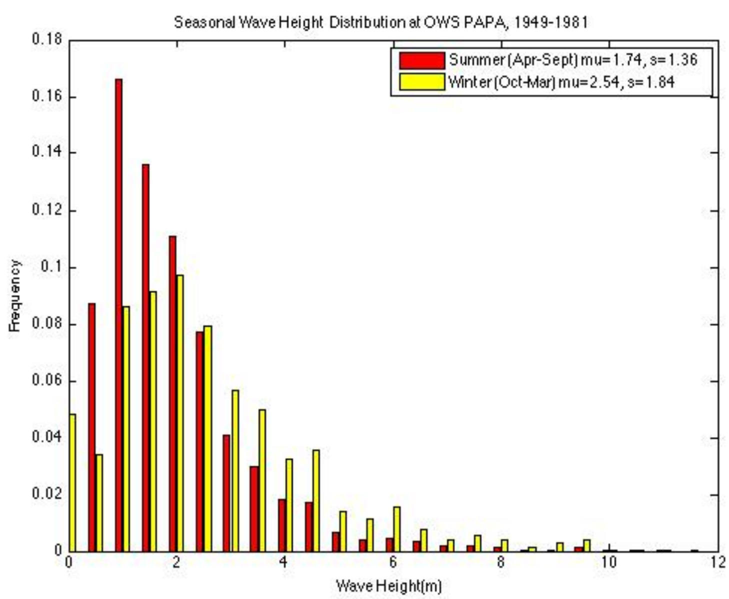
The next task was to determine if the UCAR data and NOAA data were different datasets or if the NOAA set had in fact been derived from the UCAR set as initially predicted. Parsed wave height values were divided by two, consistent with the convention that wave heights were originally recorded in half-meters by marking an entry in a table of visual observations onboard the weather ships. The preliminary examinations shown in Figure 4 demonstrate that the wave height observations differ significantly in the early part of the time series. Strikingly, the later part of the time series is a perfect match, indicating that the NOAA data is an altered subset of the UCAR data. This discovery led to the decision to pursue the use of the UCAR data as the primary source for historical observations, given its less altered appearance. A cursory examination of the UCAR wave distribution (Figure 5) indicated that statistical biases, similar to the NOAA biases, still remained, despite more similar distribution characteristics to the modern data. The reduction of bias will be important for future comparisons, but was not a part of this analysis.

Figure 4. Year-By-Year Comparison of NOAA and UCAR Datasets



The first two plots show a significant difference between the NOAA and UCAR data in the early part of the time series. However, data in the latter part of the time series appears to be identical.

Figure 5. UCAR Wave Data Statistics



The same statistical biases of the NOAA dataset exist in the UCAR dataset. However, the distribution appears to have a more similar shape to the modern data with the winter appearing notably more severe.

Wind Speed

OWS Papa has an interesting historical importance for wind speed measurements and wind-wave relations. Large and Pond (1981) used data gathered aboard OSW Papa stationed vessels to develop an empirical method for calculating wind stress at the ocean surface from wind speed measurements. Specifically, they developed a relation between wind speeds, standardized to a height of 10-meters, U_{10}, and the surface drag. Subsequent researchers have frequently used this relationship to develop and refine models of wind-generated waves in the open ocean (e.g. Juszko, Marsden, & Waddell, 1995; Thomson et al., in press). Correcting any measured values to the standard U_{10} will be essential for an effective comparison of historical and modern data.

Wind speed adjustments can be made, similar to those described in Donelan (1993), utilizing a simplified version of the equations originally developed by Large and Pond. Specifically, the wind profile law (Eq. 1) and a theory developed by Charnock (1955, Eq. 2) can be combined and solved iteratively using MATLAB to provide estimates of U_{10} wind speed.

$$U_N(z) = \frac{u_*}{k}\ln\left(\frac{z}{z_0}\right) \qquad \textit{Eq. 1}$$

$$z_0 = \frac{\alpha u_*^2}{g} \qquad \textit{Eq. 2}$$

where U_N is wind speed in m s^{-1}, z is the height at which the wind was measured, k is the von Karman constant (assumed to be 0.41), u^* is the friction velocity in m s^{-1}, z_0 is the roughness length, α is taken to be 0.011 , and g is the acceleration due to gravity.

In order to develop more accurate estimates, the fact that more than one weather ship was deployed at OWS Papa had to be considered. Fortunately, historical documents (Garrett, 2006; Freeland, 2007) sourced by a NOAA employee provide information about the characteristics of the different vessels that occupied OWS Papa. At this point, the intersection of the NOAA and UCAR datasets becomes particularly important. The UCAR dataset does not indicate which ships occupied OWS Papa at which times, while the NOAA dataset does. This is probably due to the fact that the NOAA data had been acquired from former IOS staff that likely had access to more specific ship logs than those who synthesized the UCAR data into its current DSI-1129 format. By cross-referencing the NOAA ship IDs (shown in Table 1) with the dates they correspond to and historical references (Garrett, 2006) the anemometer height for each measurement can be determined. Some resolution difficulties still exist, however, as the NOAA dataset is slightly shorter that the UCAR dataset. It is likely that additional data points may have been introduced during the synthesis of multiple datasets. Finding the intersecting dates of the two datasets should allow straightforward reconciliation of the ship IDs to the wind speed measurements. This process has not been completed as of yet, but will be included in subsequent reports. For now, only the modern data has been corrected for height.

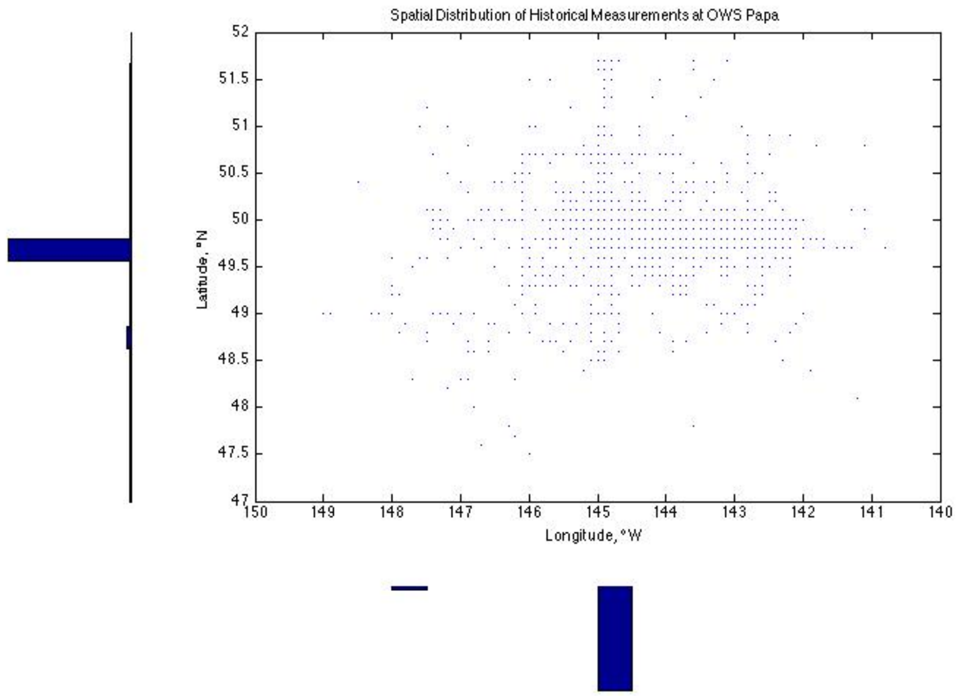
Table 1. NOAA Ship ID Anemometer Information

Ship ID Number	Anemometer Height	Years Active
55	28 meters	01/1951 to 02/1967
56	28 meters	01/1951 to 09/1967
500	17 meters	04/1967 to 05/1981
501	17 meters	10/1967 to 06/1981

Position

In order to ensure that the data points utilized were only in close proximity to OWS Papa's true location at 50°N, 145°W a spatial analysis was conducted. The position of the weather ships was reported in latitude and longitude, though dividing the parsed UCAR values by ten was required to correct the coded values. Additionally, the longitudinal values were all coded as positive values and had to be adjusted. By convention, °W should be negative values. Though it was noted that a wide distribution of values were present, the number of data points that deviated from OWS Papa's true location shown in Figure 6 was relatively negligible. This indicates that the removal of points beyond a threshold of ±2° latitude and longitude will not significantly impact the number of data points.

Figure 6. Spatial Distribution of UCAR Data Points



The spatial distribution of data collection in the UCAR dataset has a large range, but a small variance. This range is probably too large, and values will need to be removed. However, given the concentration of points around the true location of OWS Papa, removing these values will not compromise data integrity.

Other Corrections and Comparisons

A few other parsed variables have been corrected to their appropriate values. Values for barometric pressure, air temperature, sea surface temperature, dew point temperature, and wet bulb temperature were divided by ten in order to reintroduce the decimal points that were lost when the values were coded into DSI-1129 format. A summary table of all the parsed variables and any data adjustments is included in Appendix A. The remainder of the UCAR data relies on point scales, defined in the DSI-1129 documentation, that indicate a range of values for each point in the scale or a subjective appraisal of the conditions at the time of observation. As such, these values are not convertible to another form that is more meaningful and are left as is.

In an effort to bolster the conclusion that the NOAA data is actually a subset of the UCAR data, several other randomized comparisons were conducted across variables common to both sets in addition to a simple intersection analysis of each sets serial dates. The common variables, indexed by date, chosen for this check were wind speed, latitude, longitude, dew point temperature, wet bulb temperature, and sea surface temperature. Ten randomized index groups of ten were generated and evaluated. Since the UCAR dataset begins in 1949 and the NOAA dataset begins in 1951, the first two years of UCAR data were excluded from this analysis. Of the 87,496 NOAA date values there were 87,448 dates common to the UCAR dataset indicating significant overlap between the two in this respect. It was also found that among these intersecting date values, all ten of the index groups sampled for each of the categories list above contained matching values.

Conclusions and Future Work

Important first steps in resolving the historical data at OWS Papa have been taken. Several sources were consulted and the provided datasets compared. The NOAA provided data quality was called into question because of uncharacteristic wave height values in the early part of the time series. After comparing this data to the modern dataset, it was determined that another source needed to be found to obtain the historical wave data. Finding the UCAR data, and being able to interpret it, was a significant outcome of this project. Comparing it to modern and NOAA datasets showed that the NOAA data had been derived from the UCAR dataset and wave heights had been altered at some point. Other values appeared to be untouched. This project also sought to standardize wind speed values to the conventional U_{10} using the Charnock relation and the wind profile law combination described in Donelan (1993). While a successful standardization algorithm has been created, the ship IDs contained in the NOAA date have not been fully resolved to match the corresponding data points in the UCAR data and as such the wind speeds have not yet been corrected.

Once the data has been satisfactorily resolved, it will be possible to begin generating the historical baselines to be used in future analyses and site characterization. These analyses will likely include examination of climatic and oceanographic trends, hind casted wind-wave models, and contextual comparison to similar outcomes in the modern data.

Appendix A – Summary of Corrections and Adjustments

Table 2. Summary of Parsed Variables and Value Adjustments

Variable	HistPapa.mat Name	Units	Quality Control Activity
Source Deck Number	cdn	N/A	None
Marsden Square	tenmar	Marsden grid number	None
Marsden Sub-Square	onemar	Marsden sub-grid number	None
Quadrant	quad	key value, see ref	None
Latitude	lat	decimal degrees	Convert to decimal degrees
Longitude	lon	decimal degrees	Convert to decimal degrees and degrees W
Year	date	MATLAB serial date	Convert to MATLAB serial date
Month			
Day			
Hour			
Wind Direction Indicator	wdi	point scale, see ref	None
Wind Direction	wd	point scale, see ref	None
Wind Speed Indicator	wsi	key value, see ref	None
Wind Speed	wspd	knots	None
Visibility Indicator	visi	key value, see ref	None
Visibility	vis	key value, see ref	None
Present Weather	presw	key value, see ref	None
Past weather	pastw	key value, see ref	None
Sea Level Pressure	seap	millibars	Divide by ten to recover decimal
Temp Indicator	tempi	key value, see ref	None
Air Temp	airt	°C	Divide by ten to recover decimal
Wet Bulb Temp	wbulbt	°C	Divide by ten to recover decimal
Dew Point Temp	dewpt	°C	Divide by ten to recover decimal
Sea Surface Temp	sst	°C	Divide by ten to recover decimal
Total Cloud Amt	tca	key value, see ref	None
Low Cloud Amt	lca	key value, see ref	None
Low Cloud Type	lct	key value, see ref	None
Cloud Height Indicator	chi	key value, see ref	None
Cloud Height	ch	point scale, see ref	None
Mid Cloud Type	mct	key value, see ref	None
High Cloud Type	hct	key value, see ref	None
Wave Direction	wavedir	point scale, see ref direction from which waves approach	None
Wave Period	waveper	point scale, see ref	None
Wave Height	waveh	meters	Convert from half to full meters
Swell Direction	swelldir	same as Wave Direction	None
Swell Period	swellper	point scale, see ref	None
Swell Height	swellh	meters	Convert from half to full meters

Data source is UCAR unless otherwise noted.

References

Charnock, H. (1955). Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society, 81,* 639-640.

Donelan, M.A., Dobson, F.W., Smith, S.D., & Anderson, R.J. (1993). On the Dependence of Sea Surface Roughness on Wave Development. *Journal of Physical Oceanography, 23,* 2143-2149.

Freeland, H. (2007). A Brief History of Ocean Station Papa and Line P. *Progress in Oceanography, 75,* 120-125.

Garrett, J. (2006). In *The Canadian Encyclopedia online*. Retrieved from http://www.thecanadianencyclopedia.ca/en/article/station-papa/

Juszko, B., Marsden, R.F., & Waddell, S.R. (1995). Wind Stress from Wave Slopes Using Phillips Equilibrium Theory. *Journal of Physical Oceanography, 25,* 185-203.

Large, W.G. & Pond, S. (1981). Open Ocean Momentum Flux Measurements in Moderate to Strong Winds. *Journal of Physical Oceanography, 11,* 324-336.

National Climatic Data Center. (2003). *Data Documentation for Data Set 1129 (DSI-1129)*. Retrieved from http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/

Thomson, J., D'Asaro, E., Cronin, M., Rogers, E., Harcourt, R., & Shcherbina, A. (in press). Waves and the Equilibrium Range at Ocean Weather Station P. *Journal of Geophysical Research.*

University Corporation for Atmospheric Research. (2013). *Ocean Station Vessel Observations.* Retrieved from http://rda.ucar.edu/datasets/ds535.0/