Development of Data Visualization Tools in Support of Quality Control of Temperature Variability in the Equatorial Pacific Observed by the Tropical Atmosphere Ocean Data Buoy Array

Scholar: Elaina R. Barta

NOAA Mission Goal: Climate Adaptation and Mitigation

Abstract

El Niño Southern Oscillation (ENSO) is characterized by a near abeyance in the trade winds in the equatorial Pacific, leading to unusually warm sea surface temperatures off the coast of Peru and Ecuador. This elongation of warm water at the ocean's surface influences the hydrologic cycle on a global scale and modifies worldwide precipitation patterns. Monitoring changes in ocean surface and subsurface temperatures at the equator is crucial to anticipating the development and intensity of these ENSO events. The Tropical Atmosphere Ocean (TAO) array consists of 55 buoys positioned in the equatorial Pacific whose primary purpose is to collect *in situ* temperature observations and relay the data back to the National Data Buoy Center (NDBC) in real-time where data analysts work round-the-clock to quality control this data and make it available to researchers, forecasters, and the general public. However, the dynamic variations in ocean temperature inherent to the equatorial Pacific make it challenging for data analysts to set thresholds to quantify and exclude bad data; therefore, NDBC has a need for a statistical analysis of TAO data over the past 10 years that will be used to create an automated quality control algorithm as well as data visualization tools to assist data analysts in identifying erroneous data. This project builds upon an existing statistical analysis of TAO data over the past 10 years to create and develop preliminary data visualization tools to support quality control and explores other potentially useful statistical parameters that might aid NDBC data analysts in their quality control process.

The Question

Can results from statistical analysis studies on ocean temperature in the equatorial Pacific be used to create tools that will help data analysts visualize incoming data on temporal and spatial scales and automate portions of the data quality control process?

Introduction

I completed my internship at the National Data Buoy Center (NDBC) at the John C. Stennis Space Center in Mississippi. I chose this site and project because I knew little about NDBC's roles and responsibilities prior to researching the project and saw an internship there to be an opportunity to see how a singular side of NOAA operates. Secondly, I learned about ENSO in one of my meteorology lectures as I was applying for internships and found it intensely interesting.

I relied on prior knowledge of ENSO, advice and information from my mentor, and results from a statistical study on historical TAO data conducted by the National Gulf Institute (NGI) (Fitzpatrick, 2013) to found a basis for this project. Additionally, I explored the NDBC TAO website to become familiar with the buoys, sensors, and moorings. To further enhance my understanding through hands-on experience, I assisted NDBC lab technicians with temperature sensor testing and attended TAO data management meetings to see first-hand the challenges of collecting, filtering, and publishing the massive amounts of data associated with this array.

Prediction

It was predicted that a statistical analysis of TAO sea surface and subsurface temperature data over the past 10 years for each of the 55 stations would provide parameters that could be superimposed on real-time data to make erroneous data more obvious. Eventually this process of quality controlling the data could be automated. The calculated parameters could then be applied to tools that would help analysts to better visualize the incoming data on both spatial and temporal scales, which would enhance their understanding and ability to identify inaccurate data and to ultimately reduce the man hours used for data QC, enhance quality of data, and expedite data delivery to users.

Materials

Because the project consisted primarily of data analysis, materials used were minimal and entirely virtual:

- 10 years worth of data from each of the 55 TAO stations (per data availability)
- Documentation and Excel spreadsheets from National Gulf Institute study on TAO array (Fitzpatrick, 2013)
- MATLAB R2013b
- Ocean Data View (ODV)

Procedure & Methods

The procedure followed for this project consisted of two main steps and accompanying sub-tasks:

I. Determine mean ocean temperature of the Tropical Equatorial Pacific for the last 10 years.
   a. Download TAO data from NOAA websites for the years 2003-2013.
      i. To make the process of finding monthly averages more efficient, data files were sorted into directories, first by month and then by station.

   b. Write MATLAB script that will filter data, extract desired variables, and calculate a weighted 10 year average and first and second standard deviations for each of the 55 stations.
      i. Data was filtered according to the same algorithm used in the National Gulf Institute TAO buoy study: observations within 1.5 times the interquartile range above the third quartile and below the first quartile were used; all other observations were excluded (Fitzpatrick, 2013).
      ii. In calculating the mean temperature for a given station, weight was assigned to each year according to the number of its observations; years with more observations were given more weight than years with fewer observations. This method attempted to account for any bias that might have been present as a result of missing data.
      iii. The temperature standard deviation of a given station was calculated as the square root of the variance, which was calculated using the weighted mean temperature.

   c. Check calculated values against established values.
      i. Calculated means, minimum and maximum values, and standard deviations were compared to values found in NGI report 12-NGI12-24, *Assessing statistical climate variability form the TAO buoy array*, as a sanity check.

II. Apply calculations from step I to develop graphical methods for data visualization to assist with quality control of sea surface and subsurface temperature.
   a. Write script that graphs real-time sea surface temperature (SST) against calculated values in a time series.
   b. Apply time series visualization capabilities to subsurface temperatures.
   c. Develop methods of spatial temperature visualization.

Results

I. **Figures 1.1-1.4** are examples of a statistical analysis of real-time sea surface temperature (SST) data graphed in a temporal domain. Ideally, up-to-the-minute incoming data would be superimposed over the appropriate calculated seasonal or monthly temperature averages and first and second standard deviations for that depth. Here, sea surface SST data from 17 June 2014 thru 4 July 2014 is graphed simultaneously with mean sea surface temperature and first and second standard deviations that were calculated using historic data for the month of June at stations on the 95 west longitude line of TAO.

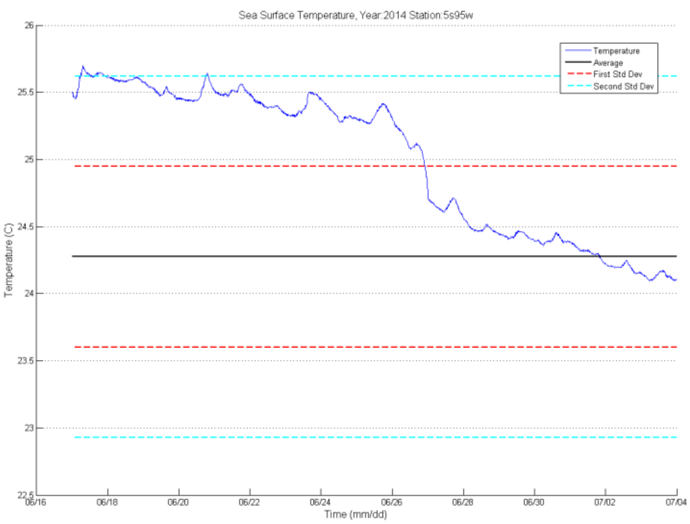

Figure 1.1 *SST for 5s95w*

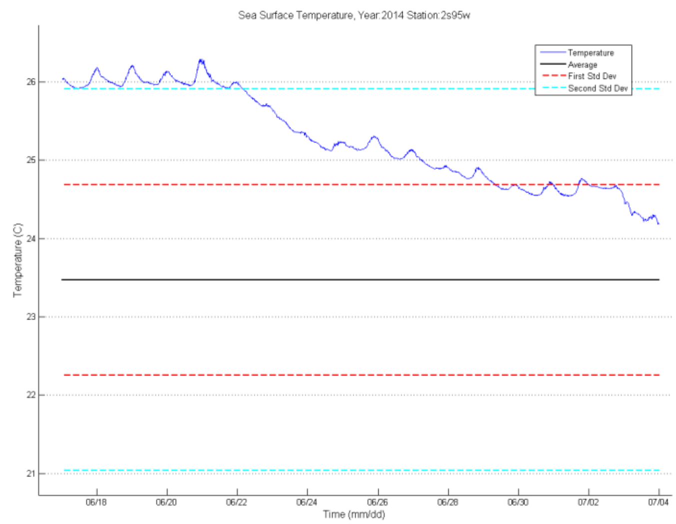

Figure 1.2 *SST for 2s95w*

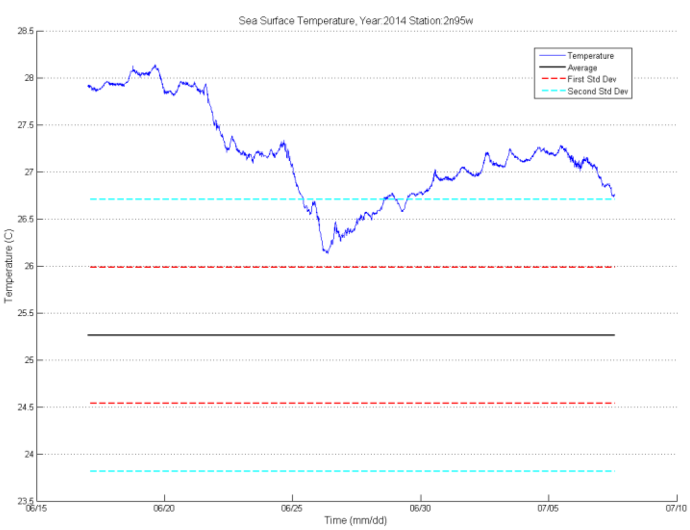

Figure 1.3 *SST for 2n95w*

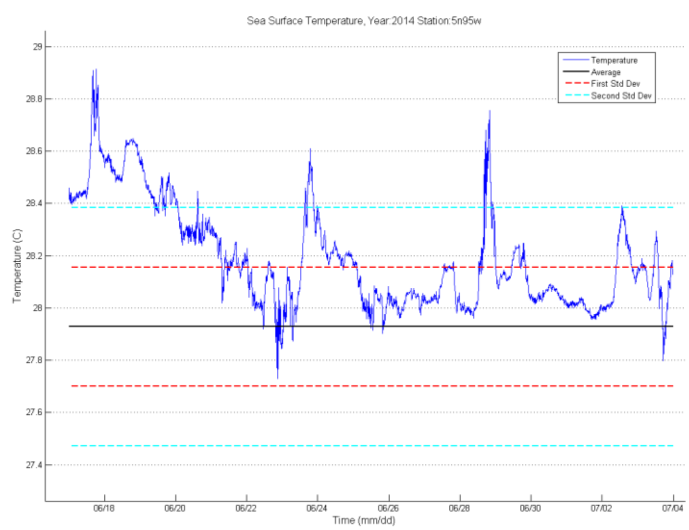

Figure 1.4 *SST for 5n95w*

Results (cont.)

II. Following a similar methodology, **Figures 2.1-2.6** are time series graphs of temperature data at all recorded depths sorted into three ocean temperature sections: ‘surface’, ‘thermocline’, and ‘at depth’. A simple MATLAB algorithm sorts depths into these sections by taking the temperature range of the data and dividing it by 3 to set thresholds for each section. Again, these graphs are using data from the 95 west longitude line of TAO.

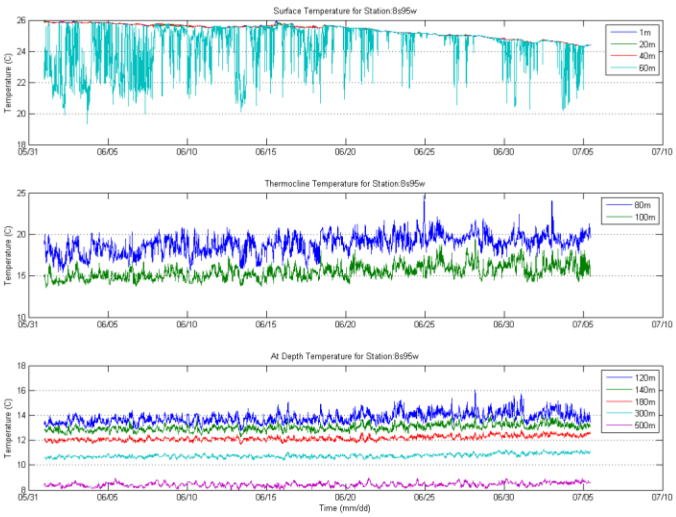

Figure 2.1 *Temperature at Multiple Depths for 8s95w, 1 June 2014 - 5 July 2014*

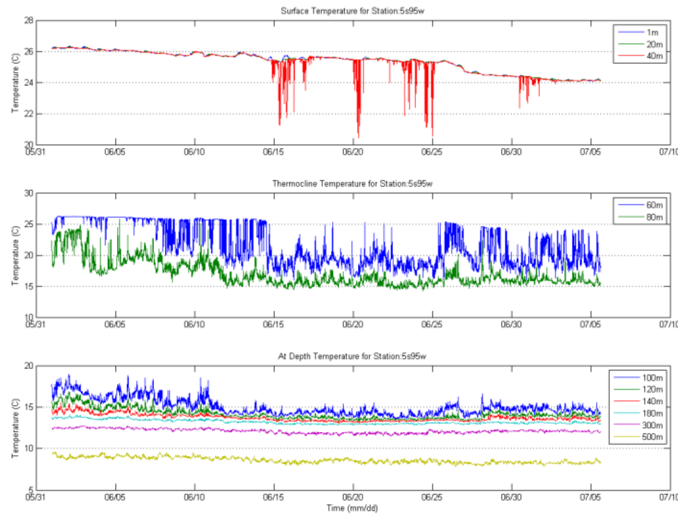

Figure 2.2 *Temperature at Multiple Depths for 5s95w, 1 June 2014 – 5 July 2014*

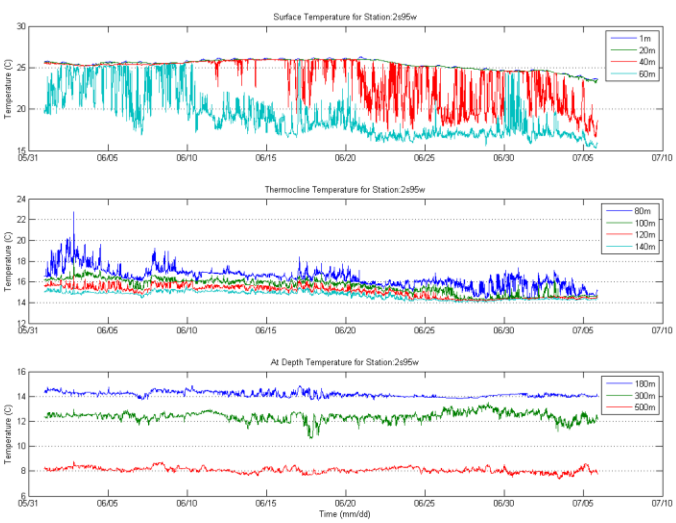

Figure 2.3 *Temperature at Multiple Depths for 2s95w, 1 June 2014 – 5 July 2014*

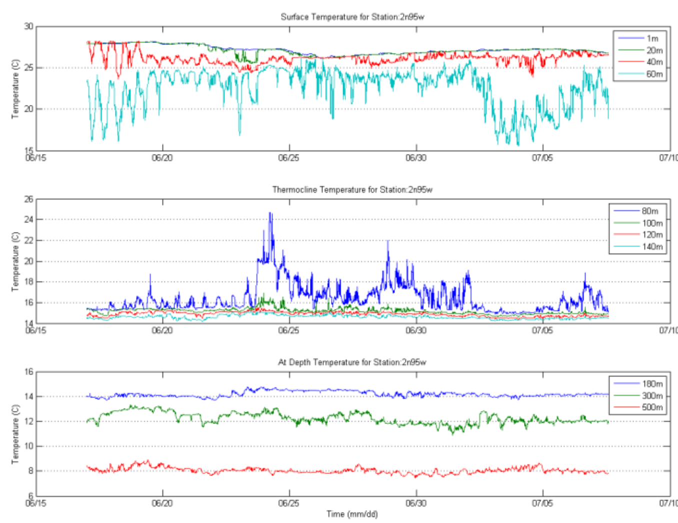

Figure 2.4 *Temperature at Multiple Depths for 2n95w, 17 June 2014 – 7 July 2014*

Results II (cont.)

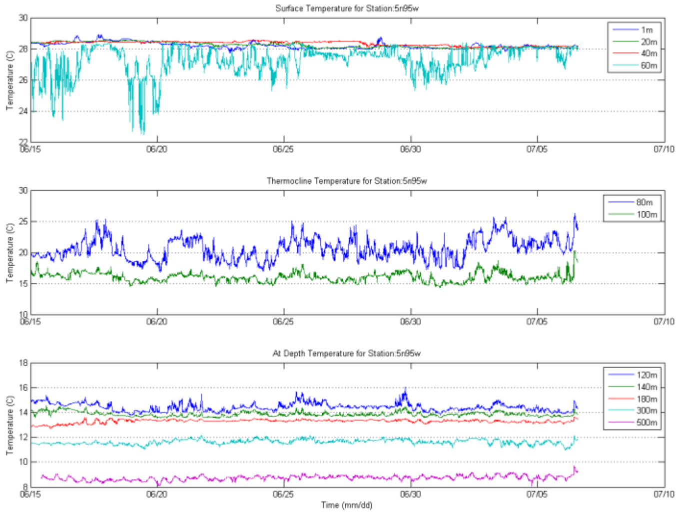

Figure 2.5 *Temperature at Multiple Depths for 5n95w, 15 June 2014 – 6 July 2014*

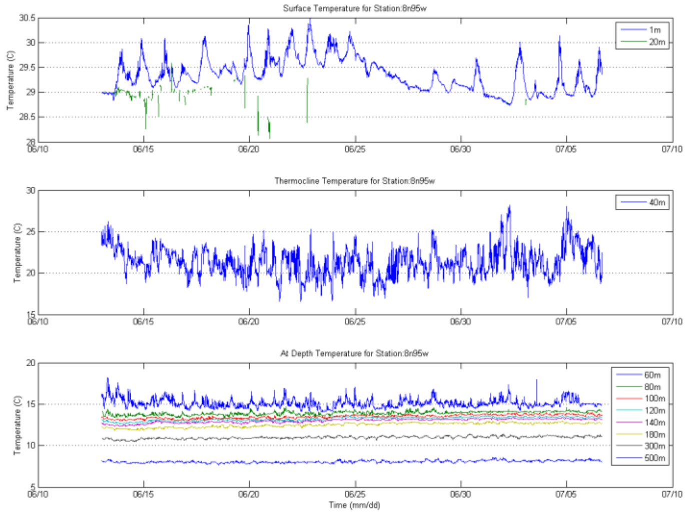

Figure 2.6 *Temperature at Multiple Depths for 8n95w, 13 June 2014 – 6 July 2014*

III. The third and final example of data visualization is a transect of ocean temperature at 00Z on 1 July 2014 for the 95w line on a spatial scale. It was produced using Ocean Data View, but it still stands as an example of the spatial visualization that would eventually be provided.

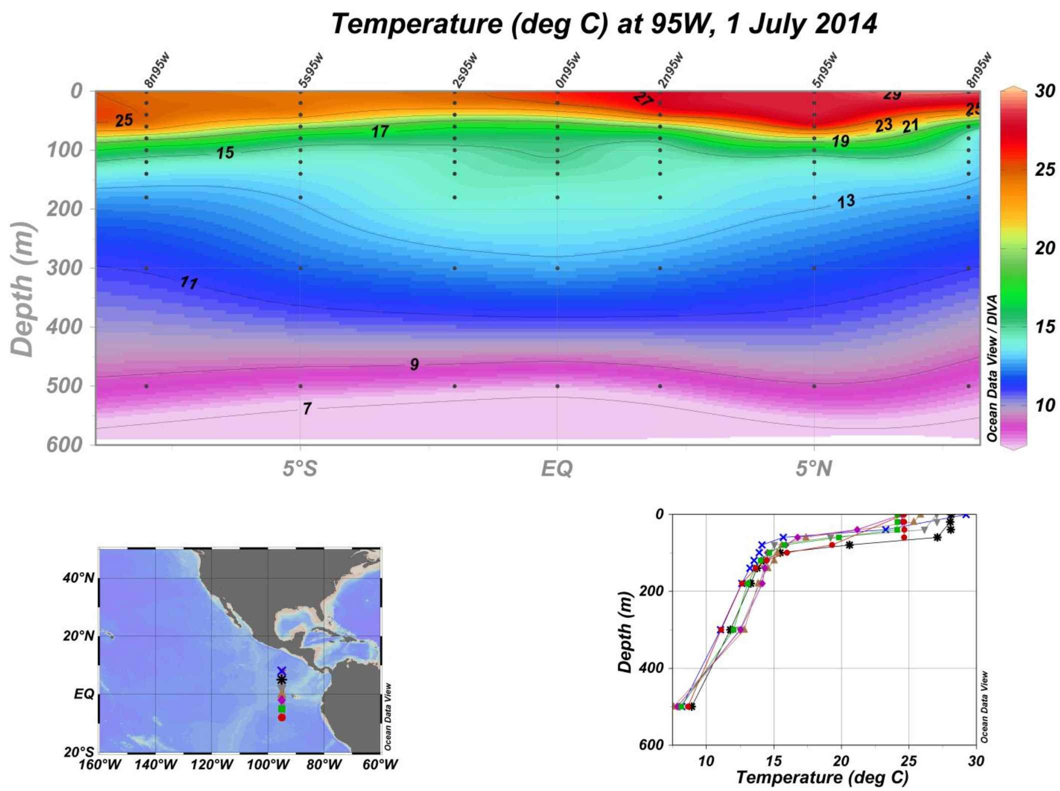

Figure 3 Example of temperature visualization on a spatial scale

Conclusion

The outcome of this project is that results from statistical analysis studies on ocean temperature in the equatorial Pacific can be used to create data visualization tools on both temporal and spatial scales. From **Figures 1.1-1.4** it is evident that, while recorded temperatures seem to be warmer than they were in the past, observations stay within a range that is consistent with past observations, so any observation that fell outside of this range would be apparent and simple to remove. **Figures 2.1-2.6** could be potentially helpful in identifying observations that are incorrect relative to the observations deeper and shallower than it. Graphs like **Figure 3** could be used for identifying erroneous observations at a single point in time and would likely be even more helpful if the transect was taken along a line of latitude, which would facilitate locating the thermocline and determining if erratic observations are due to temperature variation within the thermocline or are simply bad data.

So far it seems that the original prediction that data visualization tools would likely benefit data analysts is correct, however the process of putting the visualization tools into practice, determining their impact on the quality control process, and turning the tools into an automated QC program would need to be the focus of future work. Because the results of such a study would be largely subjective, the effect of the visualization tools on the QC process could likely be quantified by the number of man hours saved by the data analysts or the amount of time it takes for the data to travel from the field to the NDBC public website and Global Telecommunications System.

Future Work

There are several other avenues for future work, including developing a more robust algorithm that would separate the depths in graphs like **Figures 2.1-2.6** according to temperature variability, which would account for the changing level of the thermocline at each station.

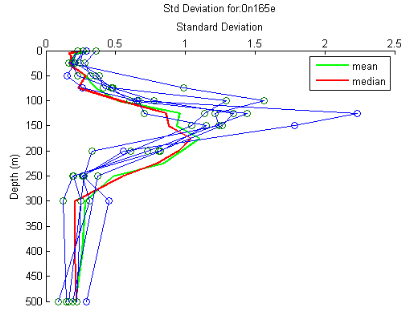

Figure 4.1 Temperature standard deviation on the western-most side of the array (at equator)

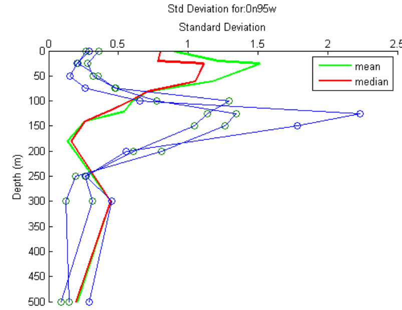

Figure 4.2 Temperature standard deviation on the eastern-most side of the array (at equator)

Another possible avenue to explore would be developing a visualization of a statistical analysis on a spatial domain. For example, **Figures 4.1-4.2** show temperature standard deviations calculated from data over the past 10 years at stations along the equator on the extreme ends of the TAO array. From these are calculated June mean and median temperature standard deviations at these stations for the past 10 years, which could be superimposed on a profile of real-time temperature data from these stations. Because standard deviation is a proxy for variability, this technique would highlight the level of the thermocline, since the depth with the greatest temperature variability is, by definition, the thermocline. Data analysts might then be able to see if the depth with the greatest temperature variability on the real-time profile matched the expected depth of the thermocline. If the depths did not match, further analysis could be done to see if the difference was due to bad data or internal waves.

Acknowledgements

National Weather Service, National Data Buoy Center, Helmut Portmann, Joseph Swaykos, and Karen Grissom (mentor)

References:

Fitzpatrick, P., 2013: Assessing statistical climate variability from the TAO buoy array. National Gulf Institute, Mississippi State University. NGI Project File Number: 12-NGI2-24, 13 pp.