Internship 2015, T-600-STAR
University of Reykjavik
Leiðbeinendur: Leifur A. Kristjánsson
og Ari Vésteinsson

MARORKA

Marorka Internship
Ocean Weather database

Þór Tómasarson
291092-3229
23. April 2015

HÁSKÓLINN Í REYKJAVÍK
REYKJAVIK UNIVERSITY

Contents

Introduction

The internship took place at Marorka's headquarters in the research and development (RAD) division. The internship started 13th of January with a work load of two 8 hour work days each week. Total work hours for the project was about 157 hours, ending on the 19th of March.

The Company

Marorka specializes in marine energy management. Marorka's headquarters is in Borgartún 26, Reykjavík along with its servers and data storage infrastructure. Marorka delivers services and guidance to vessel owners and operators to save fuel and increase profitability. The company was founded in June 2002 as a result of the PhD thesis of Jón Ágúst Þorsteinsson, founder of Marorka.

Marorka's main products are automated, on-board and online energy management systems for shipping industries. The products are all aimed at maximizing the energy efficiency and minimizing fuel consumption for vessels.

Project Description

Weather and sea state information is of great value when it comes to analyzing performance of vessel missions. The availability of this information on board the vessels is usually limited to the measurements received from the anemometer (wind force and direction) and sea state information reported once per day from each vessel. Marorka is lacking most information on ocean weather data, which is vital for post processing the vessel missions. Having information on the weather conditions the vessel sailed through enables more robust analysis on how energy was spent over the trip, and how energy efficiency might be improved.

When the vessel mission information package is retrieved, Marorka would like to use the time stamp and GPS location for that package to look up information about sea state and weather and ocean currents, and log with other information given by the measurement package. The task is either to find a service that can provide the required information, or create Marorka's own weather service.

Stated requirements for the service:

- There should be available data for most parts of the ocean.
- The data set should be collected as early as possible.
- Sampling rate at least a few times a day, most preferably with less than 15 minute interval between samples.
- The data set should contain information on:
    - Wave height, period and direction
    - Ocean currents
    - Wind force and direction
- Each required sample look-up should take no longer than a few seconds.
- The data should be verified for accuracy in some manner.

- The data should be accessible with full permission at all times for Marorka.
- Accessing the data should not be expensive.

Preparation

In order to find a good solution to the need of ocean weather information, a thorough research must be done on what is currently available. The main goal is to be able to link the best known value for weather and sea state information to each retrieved data package from each vessel.

One option is to use data from NOAA (National Climate Data Center). They provide a full and open access[1] to a large dataset for weather and sea state. They have been collecting weather data for some years, and provide data for previous years and up to this day. The data NOAA provide is peer reviewed and openly available in as timely a manner as possible. The frequency of reported observations range from sub-hourly to six-hourly samples.

List of marine/ocean data NOAA provides and might fulfill the requirements[2]:

Data set	**Comment**
Global Marine Data	Costs \$.07 per station for each year. Data request takes about 1 minute to several hours.[3]
International Comprehensive Ocean-Atmosphere Data Set (ICOADS)	Data updated only once each month, even though the data is sampled much more frequently.
Extended Reconstructed Sea Surface Temperature (ERSST) v3b	Only temperature data available.
Daily $\frac{1}{4}$° Optimum Interpolation Sea Surface Temperature (OISST)	Only temperature data available.
Voluntary Observing Ship (VOS) Climate Fleet (VOSCLim)	Data updated only once each month.
Blended Sea Winds	Data for sea winds accessible in near real time (one day delay).
Surface Flux Analysis (SURFA)	Provides evaluation of forecast skills against high-quality in situ reference data. The in situ reference data is available in near real time (one day delay). See figure 1 for locations of in situ weather stations.
Marine Data Map Access	Web based user interface for Global Marine Data which costs.

[1] http://www.ncdc.noaa.gov/oa/about/open-access-climate-data-policy.pdf
[2] http://www.ncdc.noaa.gov/data-access/marineocean-data
[3] http://cdo.ncdc.noaa.gov/cdo/cdoprice.html

The data sets that do not fail on the stated requirements are SURFA, VOSCLim and ICOADS. VOSCLim and ICOADS are updated once a month while SURFA is updated daily. Thus the best choice would probably be to use a combination of the in situ data ground truth values (they are sparse like figure 1 shows) from SURFA and European Centre for Medium-Range Weather Forecasts (ECMWF) from SURFA. The usage of SURFA would require a download of two encoded and compressed 30 MB sized files for each sample for optimum accuracy. SURFA provides collected values for measurements in addition to their weather forecasts from 2007 and up to this day.

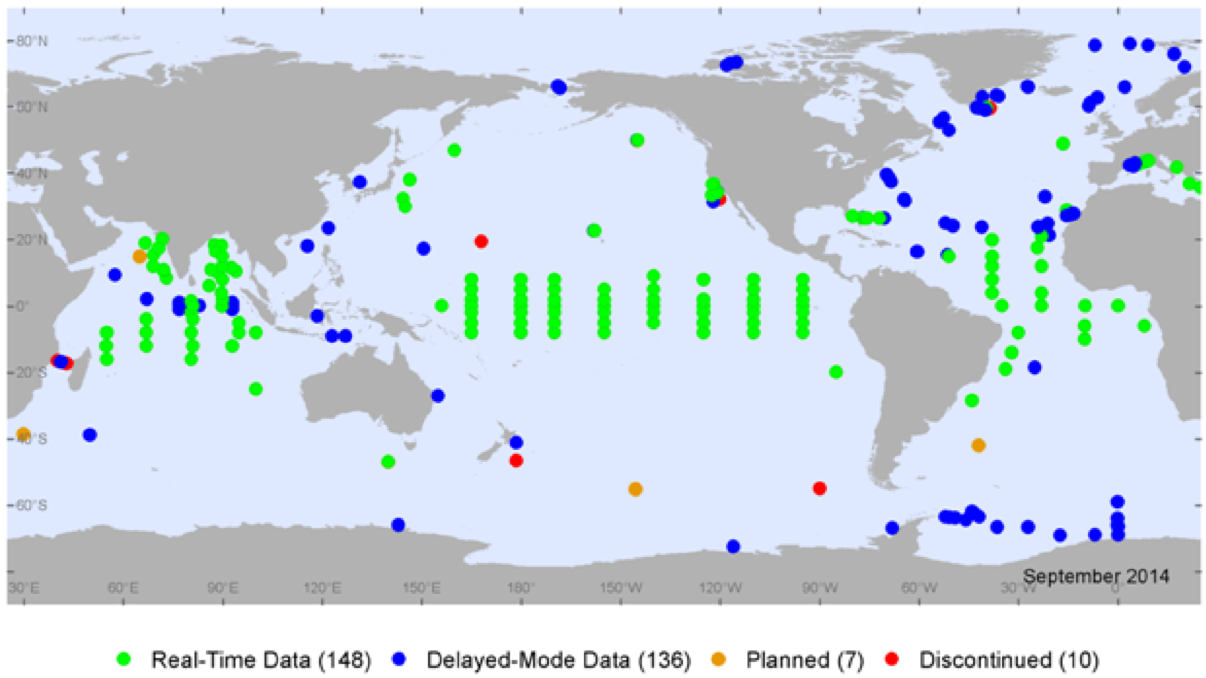

Figure 1: OceanSITES in situ weather stations

Another option is to use the [National Weather Service](#) (NWS) which provides real time data with now-casts one day ahead[4], and stores hind-casts from July 1999 to January 2014[5]. A good choice might be to use the hind-cast to update old missions. The real time now-cast would then be used to update weather information as each data is retrieved for a mission. Figure 2 displays vectors for wave height and direction data gathered from the NOAA Wavewatch III (NWWIII).

National Weather Service provides:	**Description:**
Ocean – Global RTOFS	A hybrid coordinate, 1/12° global ocean model, run once a day. Each run starts with 48 hours of hind-casting, and produces forecasts every 3 hours (surface values only) and daily full-volume forecasts from the initial time (0Z) out to 144 hours (6 days).
Waves – NOAA Wavewatch III	A third generation wave model run four

[4] http://nomads.ncep.noaa.gov:9090/dods/
[5] ftp://polar.ncep.noaa.gov/pub/history/waves/

	times a day (00Z, 06Z, 12Z, and 18Z). Each run starts with 9-, 6- and 3-hour hind-casts and produces forecasts of every 3 hours from the initial time out to 180 hours (84 hours for the Great Lakes).

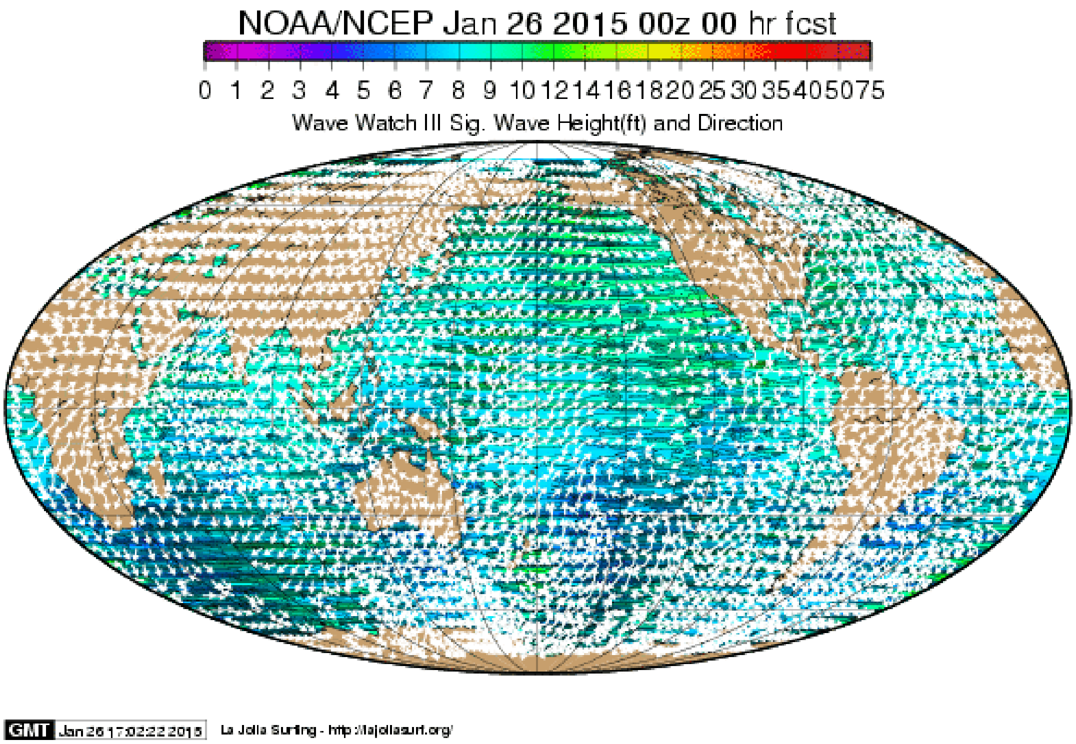

Figure 2: Sig. Wave Heigt(ft) and Direction, 00hr forecast with NWW3[6]

Other useful sources are listed here below.

Additional data:	**Description:**
[OSCAR – Ocean Surface Current Analyses Real-time]()	OSCAR (Ocean Surface Current Analysis Real-time) contains near-surface ocean current estimates, derived using quasi-linear and steady flow momentum equations. The horizontal velocity is directly estimated from sea surface height, surface vector wind and sea surface temperature. These data were collected from the various satellites and in situ instruments. The model formulation combines geostrophic, Ekman and Stommel shear dynamics, and a complementary term from the surface buoyancy gradient. Data are on a 1/3 degree

[6] http://www.lajollasurf.org/gblpac.html

	grid with a 5 day resolution. OSCAR is generated by Earth Space Research (ESR) http://www.esr.org/oscar_index.html
Global Forecast System (GFS)	GFS provides forecasts on a 6 hour basis and 3 hour basis up to 7 days ahead. Period of recording is stated to be online Jan 22, 2014 and offline Feb 15, 2005 up to this date. The data base GFS provides does not include ocean currents or wave height and direction, but GFS do plan on expanding there repository for ocean models[7].
HYCOM	HYCOM.org provides access to near real time global HYCOM + NCODA based ocean prediction system output. The ocean prediction system runs daily at the Navy DoD Supercomputing Resource Center. Upon completion, these data are remotely copied to their local HYCOM.org servers for additional processing. Daily data is typically accessible within 48-hrs of the initial runtime.

The data access methods for GFS are nice to work with. Data request can be done for a single coordinate point with specific data or for a specific range of coordinates and the relevant data. Thus GFS might come to be useful when/if they expand their database with ocean models.

There seems to be no solution available that fulfills all the stated requirements. In order to have data for ocean currents, waves and wind available for each measurement package there seems to be a need for Marorka's own weather service.

NWWIII is the leading provider of wave analysis and is capable of fulfilling most of the stated requirements, except NWWIII does not provide ocean current measurements. RTOFS provides current analyses with frequent sample interval but is lacking historical data, which HYCOM provides. The conclusion is to use NWWIII for wave data and HYCOM for current analysis. It seems to be a good option, and these two are currently being used as forecast providers for Marorka onboard models.

Process

A weather service that is responsive and provides current, wave and wind measurements for the world's oceans needs to be carefully constructed. Large amount of data follows each sample since each sample consists of the entire globe. The two weather systems NWWIII and HYCOM provide their measurements over an FTP link. The file types available over the FTP link are different, NWWIII

[7] http://nomads.ncdc.noaa.gov/data.php?name=inventory#hires_weather_datasets

stores there data in Grib2 files whereas HYCOM stores data in NetCDF files. Both Grib2 and NetCDF are common file types for meteorology data. Grib2 is a concise data format, a collection of self-contained records of 2D data. NetCDF is machine-independent data format that supports access of array-oriented scientific data. Both file types are self-describing. Each record has two components, the part that describes the record (the header) often referred to as metadata, and the actual binary data itself.

There were three setups tested for the weather service database layout. The first setup was to store all the data from NWWIII and HYCOM in NetCDF file library hierarchy, splitting folders on the data type stored (see figure 3 for the layout). Once the data would be stored, an API (application program interface) would be constructed to manage the front end of the weather service, see figure 4 for the workflow.

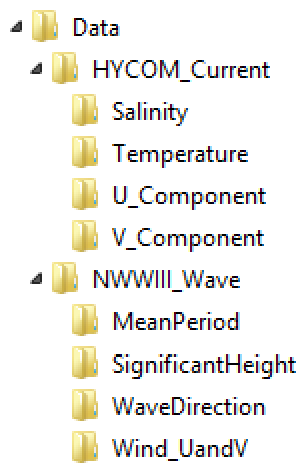

Figure 3: Marorka's weather service forlders

This setup was achieved with python script and the netCDF4 module provided by Unidata[8]. The NetCDF files are relatively nice to work with, and are easier to visualize than the 2D concise Grib2 files. Storing the data in NetCDF files is quite space consuming compared to the Grib2 format. NWWIII one month Grib2 file takes about 186MB, but when converted to NetCDF it takes 2.8GB of space. Thus if the weather service where to store five years of weather data with HYCOM and NWWIII in NetCDF format it would take $5_{years} \cdot \frac{12_{months}}{year} \cdot (2.8GB_{NWWIII} + 2.5GB_{HYCOM}) = 318GB$ of space. If the same data where stored in Grib2 file format it would take about $5_{years} \cdot \frac{12_{months}}{year} \cdot (186MB_{NWWIII} + 200MB_{HYCOM}) \cong 24GB$ of space.

[8] https://github.com/Unidata/netcdf4-python

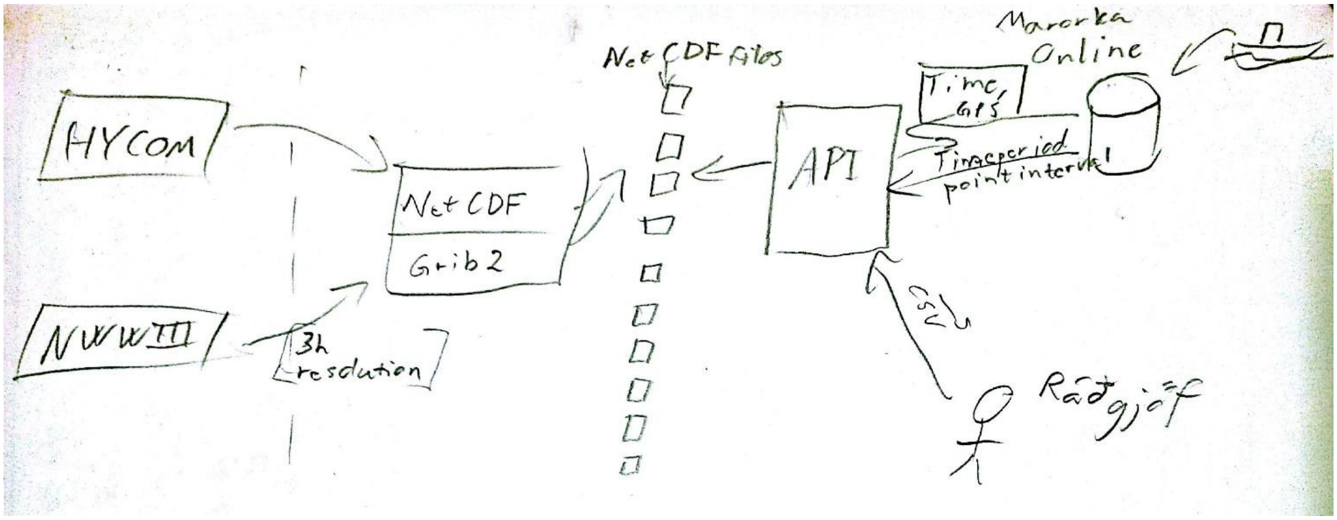

Figure 4: Workflow for the weather service

With regard to space usage further development on the first setup was halted. The second setup tested was to store all the weather data in MS SQL database, the proposed columns for the SQL table can be seen in figure 5.

	lat	lon	time_stamp	data_type	value

Figure 5: Column arrangement for the database

In order to store the weather data in the database, each data file would need to be copied, one measurement value at a time. Considerable effort was put into setting up one C# script that handles both Grib2 and NetCDF and writes them each measurement point at a time to the SQL database. Once the script was able to locate and download a given time period of weather data, a process was started to store the measurement points in the SQL database. The Process of storing one month of NWWIII data took about four hours. In one month of measurements from NWWIII, which are sampled at 3 hour interval and with a grid resolution of 30 arc-minutes, are about 55 million measurement points. If one month takes about four hours to load to the database then five years would take $5_{years} \cdot \frac{12_{months}}{year} \cdot \frac{4_{hours}}{month} \cdot (1_{HYCOM} + 1_{NWWIII}) = 480_{hours} = 20_{days}$. This seems to be unsatisfying, and thus the third option was tested.

The third setup tested was to store the data in Grib2 folder structure, similar to the one proposed for the NetCDF files in figure 3. In order to work with and manage the compact Grib2 files, the Climate Data Operators (CDO)[9] tool set is the recommended option. CDO is an open source tool set for working with meteorology data, it has multiple command line operations for NetCDF and Grib files. CDO has very small memory requirements and can process files larger than the physical memory of the computer. One downside to the CDO tool set is that in order for the NetCDF support to work the CDO must be setup on a Cygwin shell and run from there.

A C# script was written that accepts year as a parameter and searches for available data from

[9] https://code.zmaw.de/projects/cdo

HYCOM and NWWIII. The script gathers a list of FTP download links for the specific year, then downloads and stores one file at a time, editing each file if needed to store only the desired data. The CDO tool set provides NetCDF to Grib2 conversion, when applied the NetCDF file is converted to Grib2 but the needed metadata for the new Grib2 file is not created. Thus there is a need to figure out what the correct metadata should be and add it to all the HYCOM files when converted.

Another C# script was written to serve as the interface for the Grib2 folder database, the web front layout can be seen in figure 6.

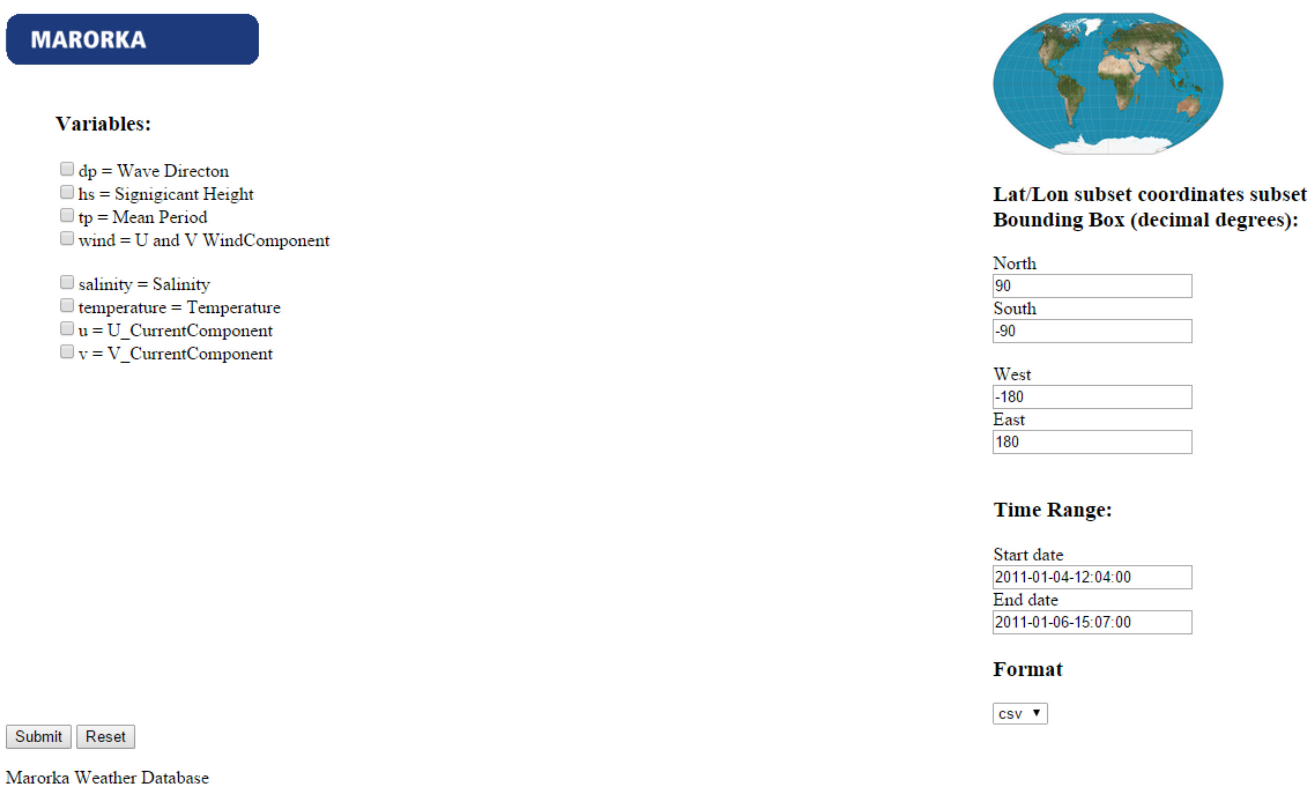

Figure 6: Web front for Marorka's weather server

Results

Going over the requirement list stated before, the current solution seems to be capable of fulfilling all the requirements to some extent. Collecting data from NWWIII and HYCOM provides data for most parts of the ocean (see figure 7 for the missing parts). The sampling rate has a three hour interval for NWWIII data but only once a day for HYCOM. The CDO tool set provides good tools for interpolating for the desired time and place. NWWIII and HYCOM do provide data for wave height, period and direction, ocean currents, wind force and direction. Depending on the size of the request, CDO is capable of sorting together multiple values from multiple files in a reasonable time. One simple sample request with CDO typically takes only a part of a second. With the weather service database located at Marorka's head office there should not be any down time, and thus the weather service should be accessible to Marorka at all times. To sum up the current status of Marorka's weather service. The weather service is still missing a module that could take care of updating the database for each day. The current C# script for updating the database takes in one year as a parameter and overwrites duplicate files, if there are any. The HYCOM NetCDF file are currently stored as NetCDF files, they should be converted to Grib2 in order for the web front to work properly.

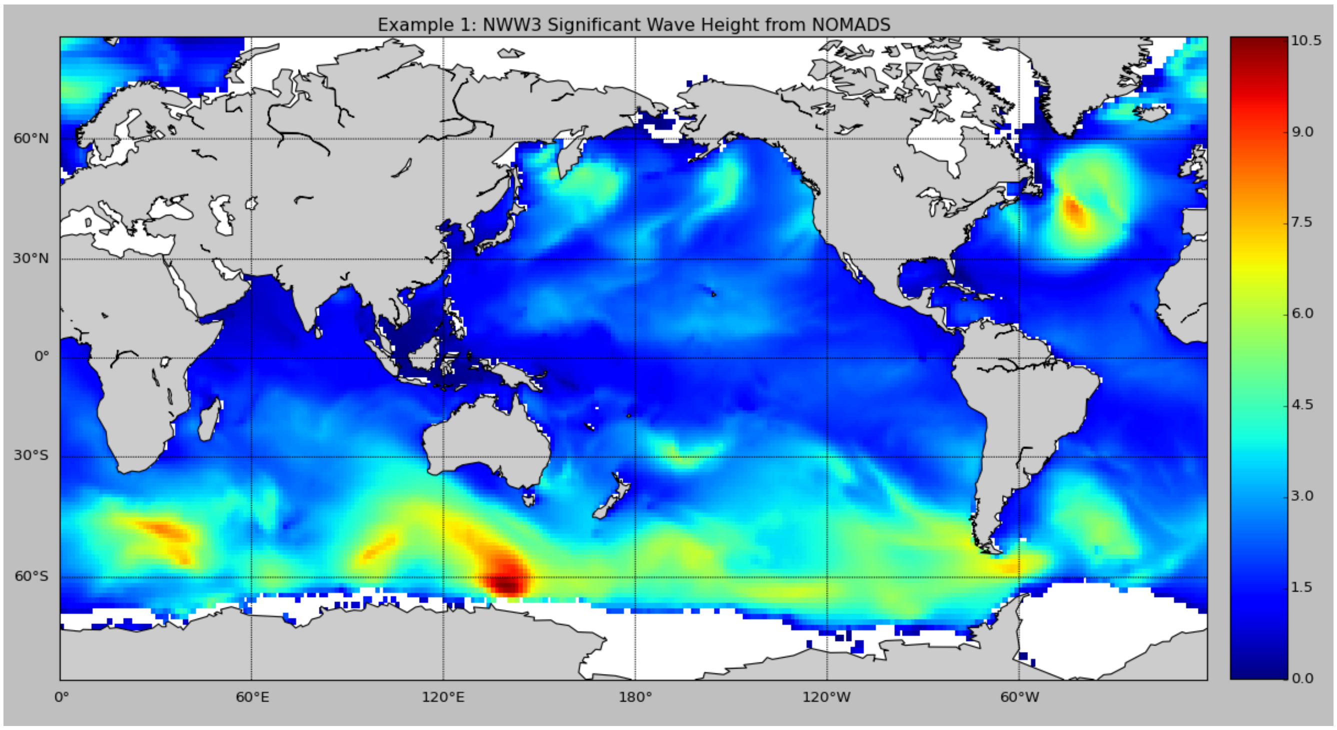

Figure 7: NWWIII Significant Wave Height plot on world map

Conclusion

The Marorka's weather station is a set of two C# scripts, one for retrieving and storing weather data, and the other for interacting with the data. The interface has some restrictions; the HYCOM data is not accessible in the current setup, and the requested time range is limited to days within the same month. The missing HYCOM data is due to the fact that it has not been converted to Grib2 file with the correct metadata. The limitation on time range can be improved with better combination of CDO executions.

Although there were great many obstacles to overcome and the learning curve was quite steep for many of the tested tool sets, the overall process of the internship went well.