May, 2014

Short Term Scientific Mission

Boulder, CO, USA

Bentorey Hernández Cruz
AGENCIA ESTATAL DE METEOROLOGÍA

Overview

Located in the city of Boulder in the outskirts of Denver, Colorado, the National Oceanic and Atmospheric Administration (NOAA) of the United States of America dispose of the David Skaggs Research Center. In this center, a team leaded by Irina Petropavlovskikh works in the development and maintenance of the NeuBrew Network.

About 8.000 kilometers from Boulder, in front of the coast of Morocco, in the Izaña Atmospheric Research Center and under the direction of Alberto Redondas, a spanish team works in a similar Project, the EuBrewNet.

Index

Introduction

The motivation of this Short Term Scientific Mission was the establishment of a comprehensive vision of the problem that is going to be solved: the development and maintenance of a whole European network that will receive data from Brewers distributed in several meteorological stations along Europe and its neighbourhood.

Eubrewnet database will be the answer to this problem. A fully integrated database with capacities of processing Brewer files in real time and generation of different level products.

For this purpose it is necessary to study the actual solutions trying to understand the issues that will be found.

In this document, two descriptions of the Networks that are now working are included: Iberonesia and NeuBrew.

Finally, a comparative study of the networks and the conclusions where a short description of the solution adopted by the workgroup can be found.

Iberonesia

Iberonesia is a Network of Brewer Spectroradiometers covering the North Macaronesian and the Iberian Peninsula area. Continuous monitoring and study of spectral UV and total Ozone column as well as satellite validation are the main purposes of this project. The area spans from 30W to 3E in longitude and 28N to 44N in Latitude, covering more than 60000 km2

Three institutions from Spain and Portugal collaborate in the creation of this project and today is the base of the calibrations of the RBCC-E and host regularly Brewer Spectroradiometers from Argelia, Morocco, Portugal, Spain and Uruguay.

This system actually consists of a number of subsystems or components of varying complexity that interact to perform tasks that include among others:

- Data Import: The system receives data from outside and stores them in a file schema for fast retrieval, even as history file.
- Processing.- The files are processed and divided in more elaborated data for their automated processing in a database.
- Error Handling.- The data are examined in case of error. It is an interactive process, with the possibility to repeat it the times needed to get a better refinement of the data.
- Product Generation.- Various tasks (some of them automated, others on demand) are in charge of generating products (elaborated information for analytical and graphical purpose) as system's output.

System Description

The Figure 1 shows a simplified diagram of Iberonesia. The system works using Hypertext Transfer Protocol (HTTP), which most external layer is a HTTP proxy. This proxy attends external requests and redirects them to the appropriate subsystem:

- A wiki: Which contains information about the system, user support, events, news, etc…
- The website of Iberonesia: Which attends the user requests and shows visual information, as well as perform the products generation tasks on demand.
- The monitoring system: Which takes care of the HTTP services. In case of failure it tries to recover (Iberonesia App and Proxy) and notifies it by mail to the administrators.
- The tasks scheduler: including, among others, importing information (if applicable), products generation and database backups.

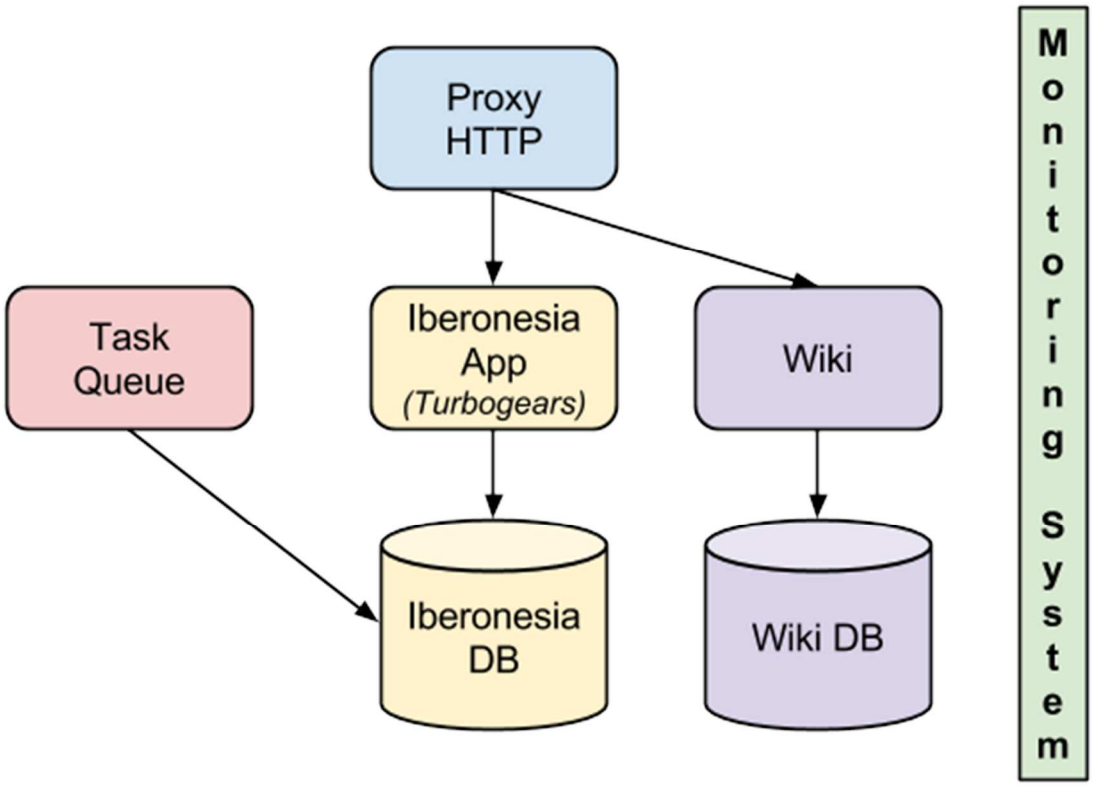

FIGURE 1: SIMPLIFIED DIAGRAM OF IBERONESIA 2.0

Iberonesia network is working in a cloud server. It means that there is not a physical machine running it, but a group of high performance servers are in charge of it. This kind of architecture provides of lots of benefits, letting maintenance problems to the remote servers administrators.

For describing the network it's important to speak about our definition of server and client. The server is the system where the web service is running and the database and the backup copies are stored. The clients are the computers which control the Brewers.

The server is running an *Ubuntu Server* which is a Linux distribution. There is no need of a specific operative system for the clients, but it is necessary to have installed a python interpreter. Python is a dynamic, general-purpose, high-level, interpreted language that provides useful tools that are independent of the computer where the interpreter is running.

Thanks to the cloud server architecture, the problems related to the backups are managed by the server administrators. Several incremental copies are made from the data stored and the web service each day at night. Once a week, a full backup of the information is made as well as an incremental one of the system. Once a month there is a backup of the whole server.

Iberonesia is oriented to real time reception of the data. Each client (having independence from the operative system) has a programmed task that try to send data to the server each fifteen or twenty minutes. This reception is processed and after that the new data are accessible from the internet.

Modules

The tasks that are described before, are performed by different modules:

- **Brewer files acquisition module:** Takes care of storing the files on the server generated by each brewer. On each computer connected to the brewer, client.py has to be installed. This program connects to the server on each run, getting local brewer files that have changed from the files on the server (either be new or be modified). Once the list is transferred to the server, the contents of these files are transferred, checking md5 checksums. Each file transferred to the server is stored in the input directory of each brewer. Both the client and server programs are implemented in python. Communication between client and server is via the xmlrpc protocol.
- **Import module:** This module is the basis of the entire project. In essence, is responsible for verifying the validity of the brewer files obtained in the acquisition module, and store their information in the database. The module is divided into three sub-modules:
    - **Import sub-module:** Importing files is a periodic task, which checks the input by all Brewers directories. The corresponding parser is called for each Brewer file, the status of each file is stored and the file is moved to its final location. The master plan (implemented as periodic task) is implemented in python (*import_all.py*). The program that imports the files from a particular Brewer is implemented in perl.
    - **Parser sub-module:** Each file type has a corresponding parser. Each parser is responsible for verify the correction of a file, and insert it into the database. It is implemented in perl.
    - **Gendefs sub-module:** Because of the varied structure of the brewer files between each type, even between different brewers for the same file type, file parsers require quite complex definition. The gendefs module generates these files from a simple description files. Implemented in C / lex / yacc.

When a file is correct and inserted into the database successfully, it moves to the directory tree of data files. Otherwise, it moves to a defined directory error, such that the operator can decide (edit) what to do with it.

Data Transmission

Data are sent to the server from each client several times per hour. It depends on the client when data are going to raise the server. Using an own protocol based on xml-rpc (a protocol of remote communication access that use XML for codifying data and HTTP as message transfer protocol) a request based on the last changes in the local files of the brewer.

Data are transmitted to the server attending to:

- New files that are not stored in the server.
- Last change in the files.
- Change in a checksum based in md5 (an algorithm that is used to verify data integrity through the creation of a 128-bit message digest from data input).

It's important to mention that the information that is sent to the server is only the new information. When it arrives, is added to the specified file and stored in the database if there are not errors.

All the information that comes from the files is stored twice:

- File system: where the raw data files can be accessible.
- Database: where the web service is going to look for information.

Database

The technology used for implementing the database is MySQL.

It is important to speak about some data issues:

- For an appropriate data processing there is stored the brewers information that indicates where the brewers are and some constants and values related to its position.
- Code error messages stored in a table for processing different errors during reception and processing of the brewer files.
- Log table with messages emitted during processing and communication.
- File status information stored in database for control purpose. The status of each file from brewers is stored in the database for controlling the different errors that can be produced during file processing.
- Refresh information for each file received from brewers. The size, date time and md5 checksum are stored for each received file. Used in communications for filtering data.
- Calibration and configuration of the brewers stored in the configuration and instruments constants files. It is necessary to upload the icf file (Calibration)and the Martin Stanek's O3brewer software (how the data is processed).

- Neighborhood/Satellite information for product comparison and warning system.

Web Service

The technologies used during development of the different services involved in Iberonesia Network are Php, Python, Matlab, Shell Scripts, Perl, MySQL, Turbogears… It was necessary this big amount of technologies to try a simple way of solving the issues found during Analysis Stage.

Acceding www.iberonesia.net will redirect automatically to the wiki where all information is stored. A link can be found there that allows access to the data. First sight shows the list of Brewers and the status of each of them.

After login, the interface shows a different number of actions allowed, including file status view and checklist and lamps management.

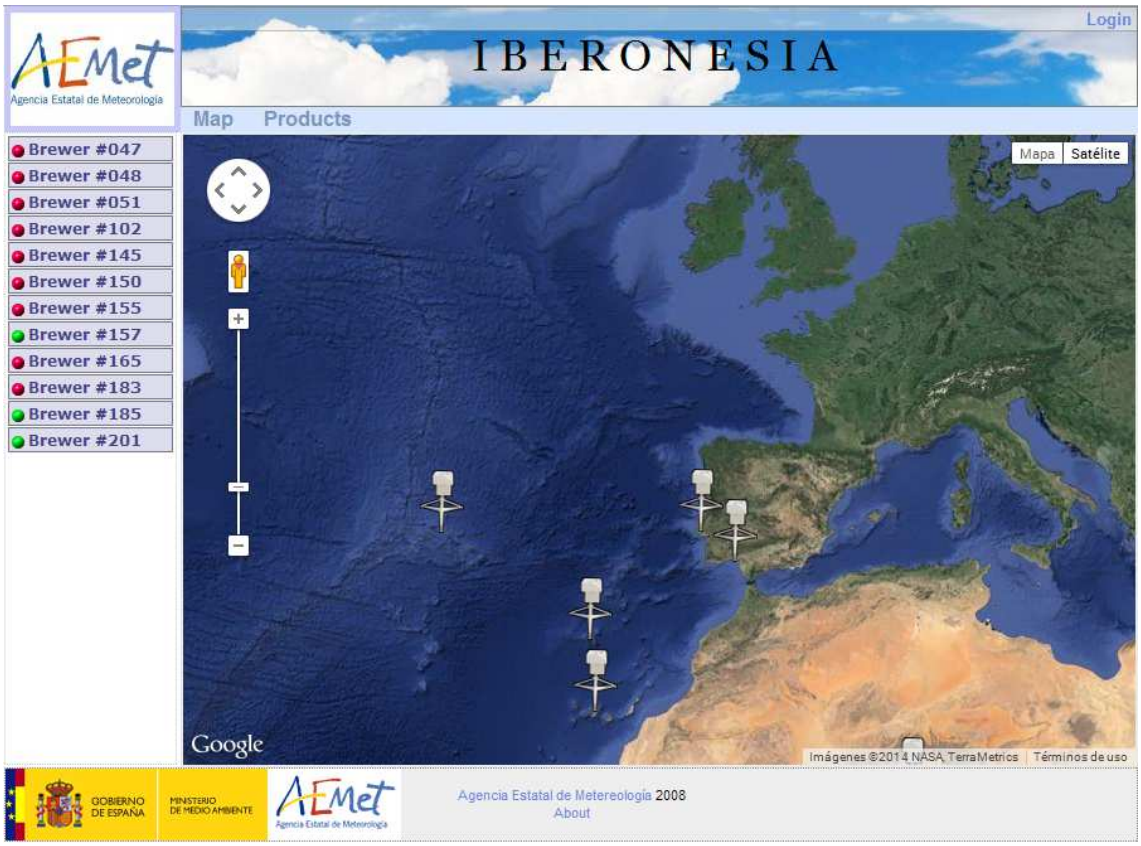

FIGURE 2: IBERONESIA 2.0

Products Generation

The generation of different products related to Ozone and UV measurements is still under development. The solution adopted is an implementation of several scripts in Matlab that perform a list of images that are stored in the file system. After running the scripts, the images can be accessible from the website. Due to space problems, now the products generation that was available is not working.

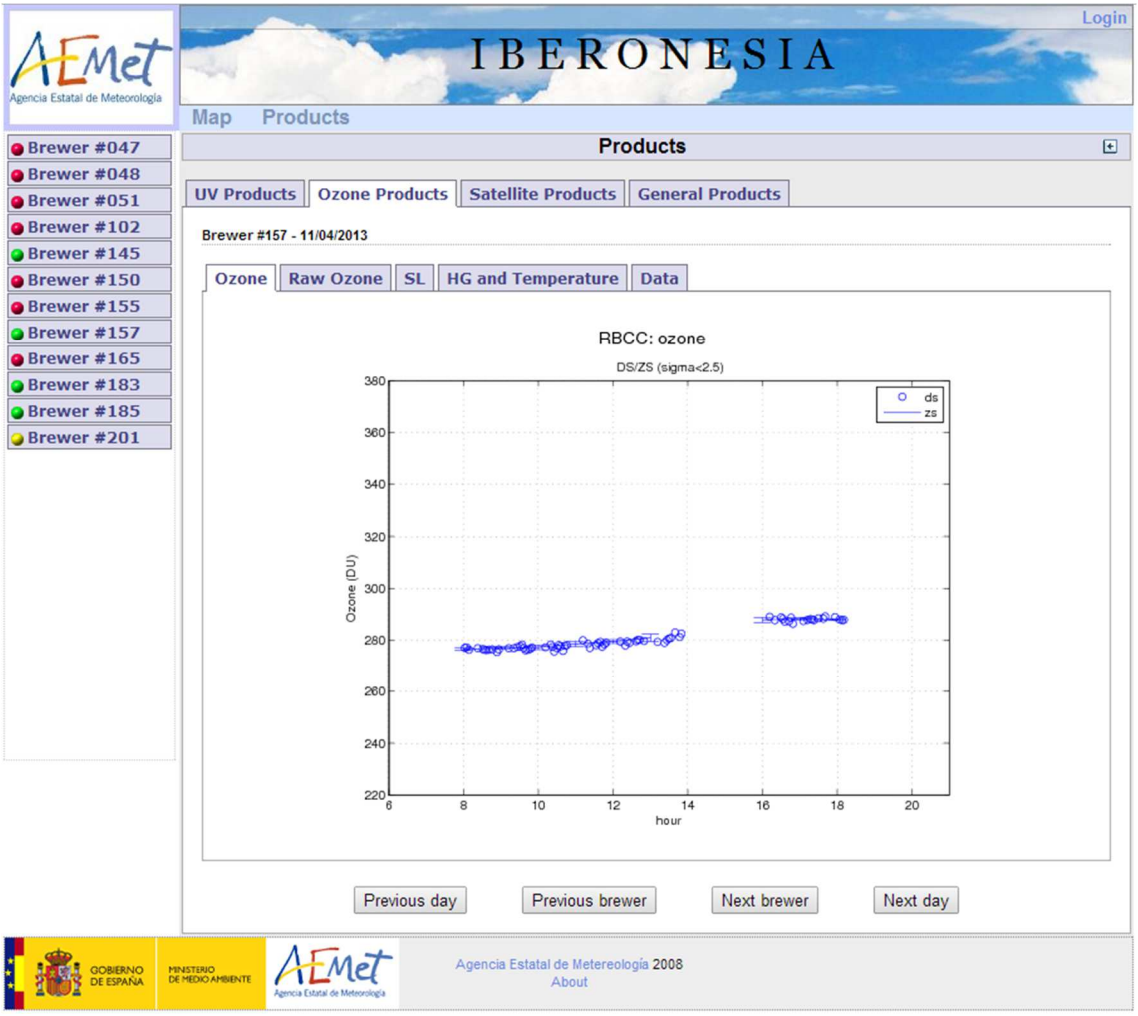

FIGURE 3: IBERONESIA PRODUCTS

Quality Assurance

There are several tools that help the Iberonesia's Administrator in his tasks:

- Real time connection control: A colour based tool that shows information about the last reception of data coming from the clients.
  - Green: Data received in less than fifteen minutes.

- Yellow: Data received in less than one hour.
- Red: Data received more than an hour ago.

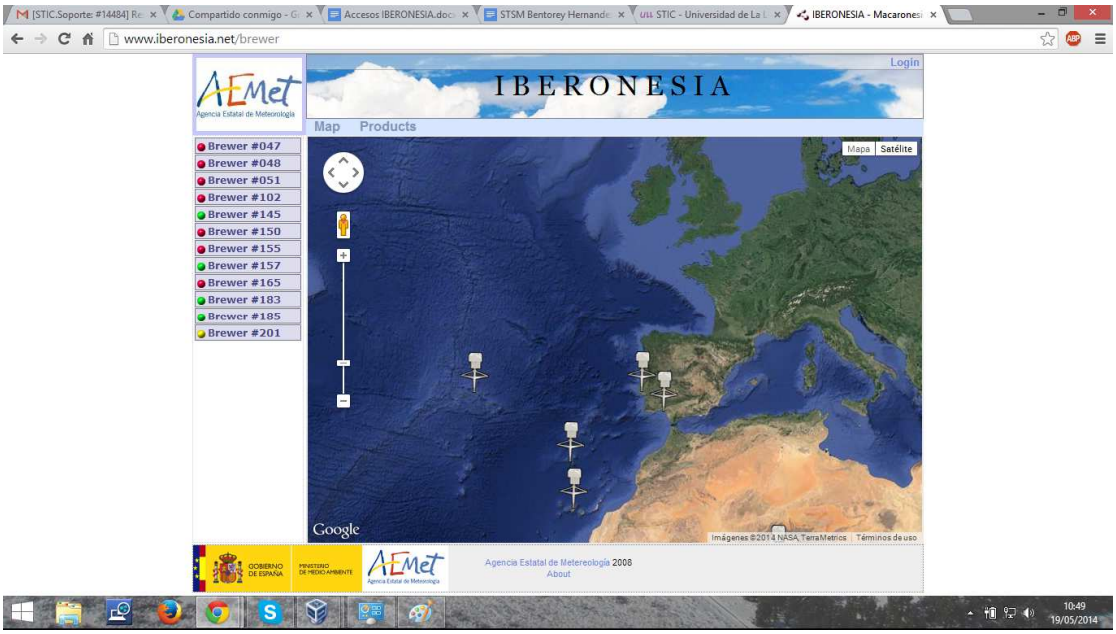

FIGURE 4: MONITORING BREWER TRANSMISSION

- Management of the brewer's calibration and configuration files by the administrator as an error control procedure. Is a task of the administrator to keep an actualized version of the icf.
- File status view after login in the web service. A comprehensive message of fails in the reception of files can be listed filtering by dates, brewers and type of message.

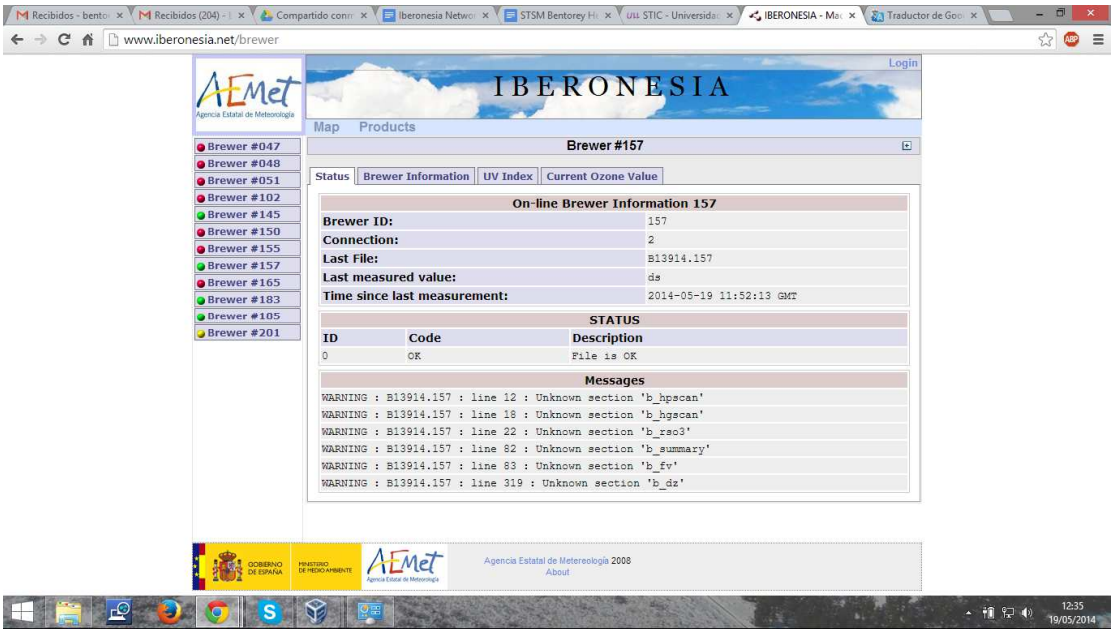

FIGURE 5: BREWER STATUS

NeuBrew Network

The NOAA/EPA Brewer Spectrophotometer Network consists of six stations located in the western, central and eastern United States. Brewer instruments provide daily UltraViolet (UV) Radiation and Total-Column Ozone measurements. Many Brewers are co-located at NOAA SURFRAD stations equipped with Total Surface Radiation Budget instrumentation, and Total Sky Imagers.

System description

NeuBrew network is working in a centralized server. It's located in the NOAA facilities in Boulder, Colorado.

About the characteristics of the system we can mention that it's running in a Quad core machine with 16 Gigabytes of RAM. It's really important to mention that it's a material machine, so there's a computer that is running the service. For data storage they had reserved a hard drive with 1.8 Terabytes of space.

All the computers working in the network are running under Centos, the computers which control the brewers and the server. By this way, the complexity of configuration of networking and even the process of data has been reduced.

The backups of the systems are taken every night, copied in an external drive. Several scripts using rsync for data, system configuration and web service files are executed after processing data. The database has its own system for synchronizing.

It's important to think about that Neubrew Network is not oriented to real time, so the data transmission and calculations are processed at night, after the devices stop their measures. In contrast to Iberonesia where the data are received in near real time which allows a quality check of the network also in near real time.

Data transmission

Data are requested from the server for each brewer that is working. Using secure shell communication protocol, the server gets a list of all files that are located in each client. After that, it looks for the recent files which are not stored in the file system and database and sends requests for a compressed one including all the new data.

It's important to mention that the requests of new data are made in a sequential mode. It means that the server is going to ask each brewer one by one for the new data for storage. This kind of data transmission is allowed because they have not real time as a requirement. Waiting for night for the reception of data is a way to get a secure, efficient and integral way of data transmissions.

The network has some communications problems related to modem access, due to the need of communication with remote and poorly communicated areas. This kind of problems had to be solved during the development and installation of the network.

Database

The technology used for implementing the database is PostgreSQL. Working with the pgAdmin III and some useful web tools for maintenance and management.

It is important to speak about some data issues:

- Introduction of the concept **"Station"**: In the database and for elaboration of products there is a table that contains data from the places where it is normal to locate brewers. These places are the ambiental stations where the brewers take measurements and the calibration centers. When there is the need of obtain values like latitude and longitude of a Brewer for calculations, depending on the timestamp these values are taken from the corresponding station.
- Use of very descriptive column names: Use of complete name for column names in tables instead of using acronyms. This politic make easier the way to understand and operate with data.
- Use of tables for logic behaviour: Database contents include some tables related to access of information (for filtering data from brewers or dates), accessible methods (products generation available), storage of coefficients and constants (involved in calculations related to products generation)...
- Configuration of communications for different Brewers: Because of the heterogeneity of the communication ways, a table is filled with the data related to configure and proceed with each case.
- Products storage and different data levels: Products generation are stored in the database after calculations to avoid time waiting so as the different data levels.
- Contacts: Information related to operators of the Brewers if there is the need to establish a communication.

Web Service

After getting data and processing it, there is a time to show results. The user's interface has been programmed using JAVA and HTML. The server

is running an Apache Proxy that waits for connections. After processing them, the Tomcat server runs the servlet that is going to show the requested information.

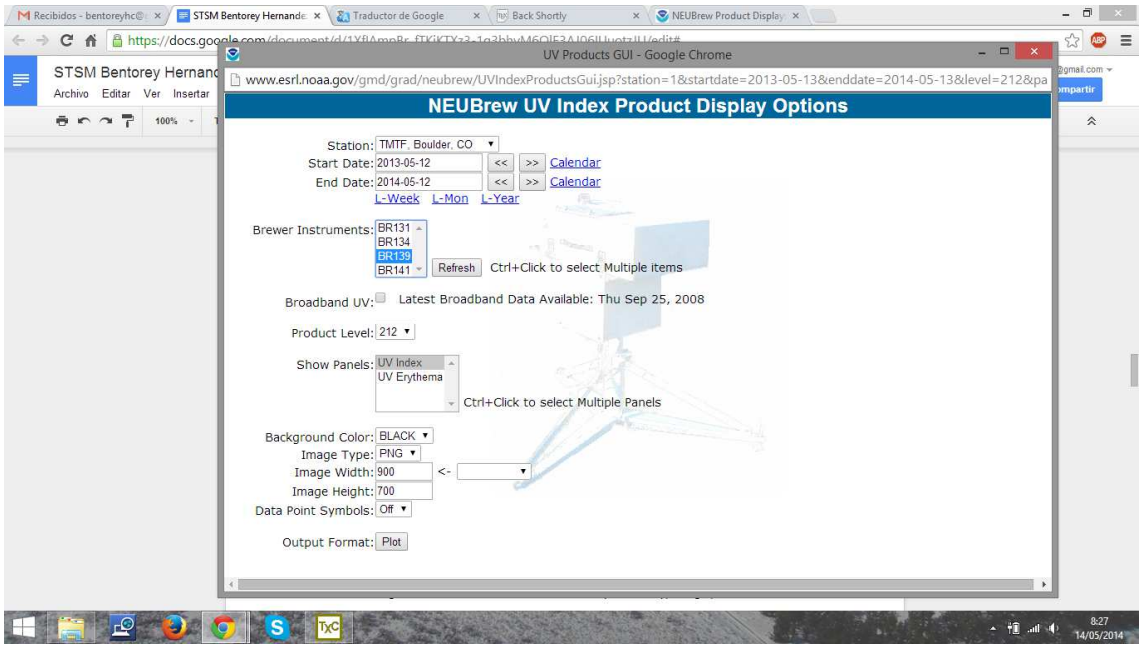

FIGURE 6: UV INDEX DISPLAY OPTIONS

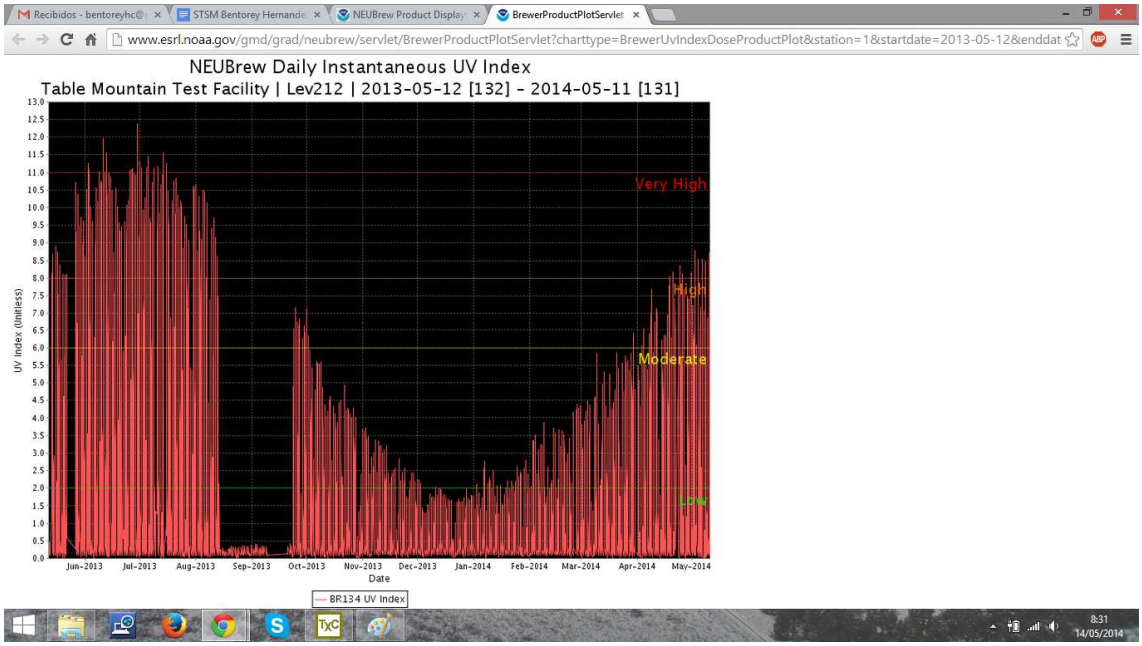

FIGURE 7: SERVLET SHOWING GRAPH REQUEST

Products generation

The products provided by NeuBrew include:

- UV Index and UV Erythema: Brewer UV Instantaneous Index and Daily UV Erythema Time Series allow plot durations ranging a single day to multiple years. There is a comparison between the UV Index by NeuBrew and one provided by Climate Prediction Center for each station, too.
- UV Irradiance: Brewer UV Irradiance plotted as a daily time series. Seven discrete wavelengths are plotted separately, 300nm - 360nm in 10nm increments.
- UV Spectral Scans: Brewer Daily UV Spectral Scan overlays multiple UV scans for a single day. Scans can be filtered for AM-Only, PM-Only, or both AM/PM scans.
- Ozone Time Serials: Brewer Daily Total-Column Ozone Time Series with the option to plot against TOMS/OMI Ozone Level 3e satellite data.
- Tropospheric Ozone time series: Tropospheric Ozone values are derived from the Umkehr Vertical Profile dataset.
- Ozone Vertical Profiles: Ozone vertical profiles derived from Umkehr scans.

At least it is necessary speak about Langley and Unkehr's profile calculation:

Langley

For calculating the Langley Regression the system use an algorithm developed by Peter Kiedrom.

Repeat
Make a linear regression of MS9 vs Airmass
Remove the farthest outlayer
until (|ETC| < 8) or (Number of points remaining < 5)
Discard ETC when data remaining are not representative

TABLE 1: LANGLEY CALCULATION ALGORITHM

Umkehr

Ozone vertical profiles are calculated in an automatized way daily. At night, as part of the script that is going to store data into the database and calculate the different data levels, software developed by Martin Stepanek in collaboration with Irina Petropavlovskikh is run using wine. In the database there are the coefficients and constants needed to run the software. The configuration file of the application is modified using these values for a correct calculation.

There are more type of displays used for diagnostic purpose described on the following section.

Quality Assurance

The quality of the data received from network is processed by the administrator. Due to the dimensions of the network, there's no automatic way of checking the integrity of the data collected.

There are different graphic tools related to the data ingest, like some data transfer graphs that show the amount of files received and the size of these files. If there's some problem with data reception, the administrator can observe it because there are blanks in these graphs or for example, and using some kind of colour code for the medium size of each file, if the brewers have problems obtaining data in their measurements.

On the other hand are some visual data that can be obtained from the description of the brewers connected to the network in the main page of the project. There you can find a historic of the measurements of the last six days of the daily UV erythema, UV index and total column of Ozone. Blanks and strange measurements are located immediately in these type of graphs.

These screenshots show one correct and one with failures due to maintenance issues.

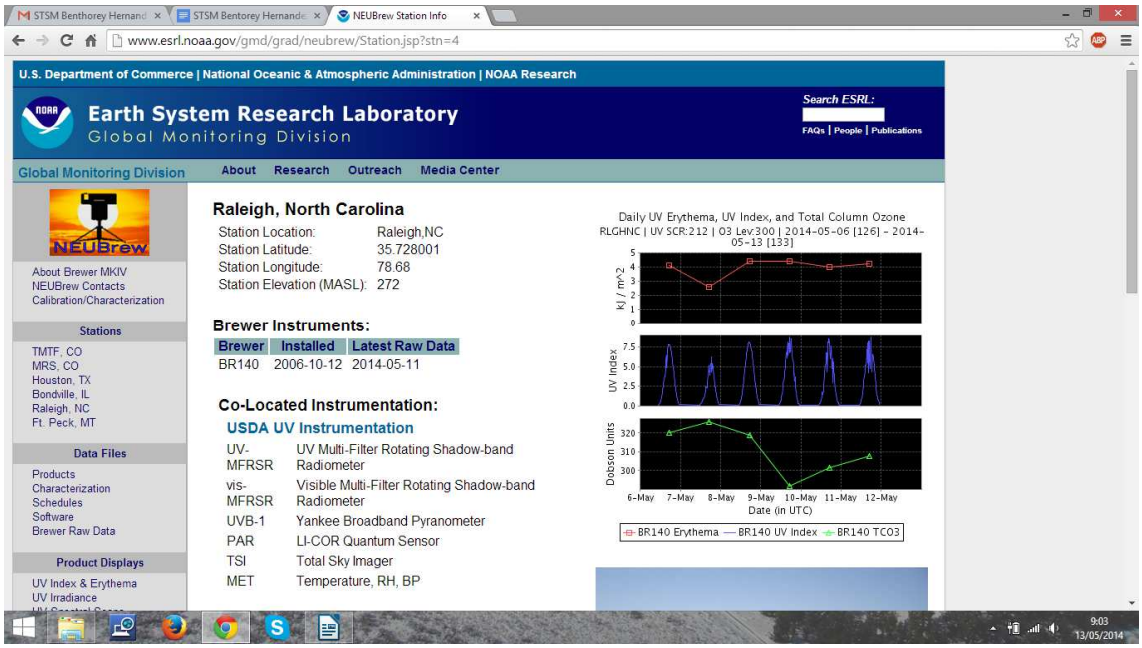

FIGURE 8: INFORMATION RELATED TO A STATION

There's another way to understand the running of the system. A comprehensive daily report is sent to administrator's mail with all the information related to the behaviour of the server. With different messages and some tables edited in plain text, and experimented administrator look for failures and problems. For example, one of the tables contains data of the time elapsed during the daily proceed: time needed to receive data (it is important due to one of the brewers need modem connection to send data) and for processing.

It's necessary to understand the process that involve the ICF (Brewer's Configuration File). A baseline was established when the project moved from the Georgia University to NOAA. From this time, considered the starting point from the measures, a relation of each configuration file has been daily stored in the database. Each day, a copy of each configuration file is pulled into the system for backup purpose. Data contained in each file is then stored in the database for later calculations.

Diagnostic

For diagnostic purpose there are several displays which show if there are problems with any data:

- Brewer O3/NO2 Langley Regression Analysis tool: Calculating Langley regression step by step.
- Brewer UV spike detection and correction (Time dependent): The percentage of spikes in UV spectra are calculated for each week of the year for posterior detection and correction.
- Brewer UV spike detection and correction (Wavelength dependent): The search now depends of each wavelength during a month.
- Brewer Event Time Series: A graphical time series of event markers depicting the starting time of the various Brewer sub-routines.
- Daily Average Ozone Time Series Comparison: This tool allows comparisons of different QC Levels of Daily Averaged Total Column Ozone measurements from Brewers and OMI instruments.
- Individual UV Scans: Time series plots of individual UV/UX scan data stored in the UV.
- B-File Comments Tool: A query tool that allows B-File comments to be searched for various criteria.
- Brewer Multi-instrumental Display: Multi-Instrument Display shows data from all instruments on a single page.
- UV/UX Daily Scan Summary: Displays a summary view of all the UX scans for a single day.
- Umkehr Scan Daily Summary: Umkehr measurements for some wavelengths are performed by the Brewer during sunrise and sunset elevations between 90-70 solar zenith angle.
- Umkehr Derived Tropospheric Ozone Time Series: Daily Average Tropospheric Ozone is derived from the lowest altitude of the Umkehr Ozone Solution Profile.

- PS Scan Daily Time Series: Daily time series of PS scan data contained in the Brewer Y-Files.
- CI Reference Scan: CI Reference Scan plots data contained in the brewer CI-Files against a baseline reference CI Scan.
- CI Banded Wavelength Ratios Times Series: This plot first takes the CI Scan data and calculates the ratios from the instrument's Baseline Reference CI Scan.
- CI Ratio Multi-day Overlay: Plot of the ratio of CI scan data compared to the instrument's Baseline Reference CI Scan in a timestamp.
- Dark Count Summary Time Series: This summary plots all the dark counts against one another to identify abnormal dark count levels.
- DS/ZS/SL Raw Scan Data: Raw Photon Counts, Wave Length Ratios, Filter Wheel #2 Position and Dark Count measurements are plotted.
- DS/ZS/SL Raw Scan Summary: Summaries of O3 and SO2 values, Single and Double Ratios, internal Brewer temperature, and Neutral Density Filter Positions are plotted.
- HG Calibration Data: In order to maintain accurate UV and O3 reading throughout the day, the Brewer must take several scans of its internal Mercury Lamp (HG) to compensate for temperature changes.
- Aggregate Temperatures: Display of the measure temperatures inside the Brewer's enclosure as a unified time series. If the Brewer is equipped with an internal humidity sensor, these data are also plotted.
- Ozone QC Level Diagnostic: This plot compares the differences between Ozone Data Levels.
- SL Daily Average: The Standard Lamp (SL) Average plot displays data contained in the brewer SLOAVG and SLNAVG files.
- R5, R6 and SL Daily Averages: This plot is a modified version of the SL Avg Plot, giving more resolution to the R5 and R6 values.
- RS Daily Averages: The Run/Stop (RS) Average plot displays data contained in the brewer RSOAVG and RSNAVG files.
- DT Daily Averages: The Deadtime (DT) Average plot displays data contained in the brewer DTOAVG file.
- AP Average Data: AP Average data plot provides voltage, current, temperature and humidity measurements collected by the Brewer several times over the course of the day.
- AP Average (Voltage Only) Data AP Average data focusing on internal Brewer power supply voltages.
- DUV Daily Averages. Daily averaged UV value weighted by Diffey Weighting Function.

Comparison

NeuBrew	Iberonesia
Network not prepared for solving real time issues	Network oriented for receiving and processing data in real time
Use of a physical server placed in a laboratory	Use of virtual machines and cloud computing
Designed for running in a centralized server	Designed for having the possibility to be distributed.
Very extended monitoring system for human operators	Reduced monitoring system based in logs and status messages
Graphical information generated in execution time extracting different data levels from database	Graphic information stored in file system
Use of a reduced number of different technologies	Big amount of different technologies for solving several issues
Data transferred by server requests in a scheduled way	Client oriented data transmission depending on local schedules

TABLE 2: NEUBREW AND IBERONESIA COMPARISON

Conclusions

First Steps

The main motivation of this Short Term Mission was the acquisition of the experienced knowledge which would contribute to develop a more realistic and useful network. The knowledge obtained during this visit will be presented on future COST meetings and workshops and the definition and development of the Eubrewnet database would be a reality in November 2014.

The work plan for the next months will be:

- Data reception and storage in the database and file system for the first week in September.
- Configuration and Calibrations files upload available for the first week in October.
- First generation of products in November.

Eubrewnet Database

The decisions adopted during the mission and after working in design issues are:

Due to the dates so close of the work plan and the main requirement of the network (real time data reception and processing), the more complex aspects of the network behaviour (data transmission and storage) will be inherited from Iberonesia.

An important change in the database design will be adopted. The implementation of the databases in PostgreSQL is a hard change that will modify several modules in the original network. This changes will not affect the clients that are now sending data to Iberonesia.

The website will be upgraded using Wordpress for the main page and Web2py for the visualization and processing of data. Because it's possible to develop the algorithms using Python as Programming Language, it will be an advantage choosing this model letting the news and user interface to a framework as Wordpress that is easier than other solutions solving maintenance issues and letting the database and processing issues to a framework based in a scientific oriented Programming Language.

The following table shows a comparison between the old Iberonesia and the new Database.

EuBrewNet Database	Iberonesia
Database Management System: PostgreSQL	Database Management System: MySQL
Development of a product generation oriented to JSON data interchange	Products generated by scripts and stored as images in file system
Python as programming language for product generation	Matlab as programming language for product generation

New database model for storage of the different products	Use of files for calculations and products
Use of WordPress for developing information interface	Use of PHP for developing information interface
Use of python for file parsing and checking	Use of perl and yacc/lex for file parsing and checking

It's important to mention that the boundaries of the Iberonesia network are going to be implemented as well in the new EuBrewNet Database: oriented for real time reception and processing and to have the possibility to be distributed. In fact, the new database could be considered an upgrade of the old Iberonesia network.

Discussion

Do not hesitate to express any doubt or question about the present project and its implementation. An open debate can be found in the forum in www.eubrenet.org for further considerations.

Appreciations

Finally, the experience and contacts obtained during the travel led me to introduce myself in the scientific and technical community spending an invaluable time with the NOAA ESRL Global Monitoring Division. On the other hand is the improvement of the relationship between the institutions working together on both parts of the Atlantic Ocean: AEMET and NOAA, having plans of cooperative working in the future.

I want to express my appreciations to the COST Action for giving me the opportunity of making this Short Term Scientific Mission.

Thanks to Irina Petropavlovskikh and the people of Colorado for receiving me and letting me be one more for a short time.

Specially thanks to Patrick Disterhoft for all the helpful information and the receiving and a big amount of things and to Scott Stierle for his company and for being so patient and answering all questions I had. For me it was a wonderful spent time.