Ecawsoft: A Web based Climate and Weather Data Visualization for Big Data Analysis

Fue, Kadeghe G. [α], Sanga, Camilius A. [σ] & Tumbo, Siza D. [ρ]

Abstract- Purpose: In Tanzania, data for climate and weather are normally analyzed by Meteorological Agency and then are published through TV, website and radio. Different stakeholders normally obtain the weather and climate data / information in a generalized way. This calls for a need of a system which allows data to be shared openly to different stakeholders so that they can analyze those data as per their specific needs.

Design/methodology/approach: The paper presents the overview of the developed system, ECAWsoft. Also, it gives some few interfaces showing different outputs from the system.

Findings: The goal of this paper has been attained by developing a working data visualization tool for climate and weather called ECAWsoft. The system is current operational and is providing open data for different stakeholders. It is user friendly and interactive with capability of displaying visualization of data as per fine granularity required by user. Development of open data system for data visualization has lead to a transparency system which is helping farmers, researchers, policy makers (etc.) to make informed decision on weather and climate.

Practical implications: The system presented in this paper need to be scaled up so that more data from all weather stations in Tanzania can be populated in real time.

Originality/value: The development and adoption of open systems for visualizing weather and climate data remains seriously lacking in many countries including Tanzania. This paper provides an overview of some initiative to fill such a research gap.

Keywords: *climate, weather, visualization, system, big data.*

I. Introduction

Never before in history has data been generated at such high volumes as it is today. Exploring and analyzing the vast volumes of data (i.e. big data) is becoming increasingly difficult. Information visualizetion and visual data mining can help to deal with the flood of information (Keim, 2002). Visualizations of subspaces on the World Wide Web can provide users the ability to identify relevant information from a set of Web pages, while gaining new insights or understanding of the space (Heo and Hirtle, 2001). Brodlie (1997) looked at the different players involved in the creation of a Web-based visualization service, and hence, build a reference model for Web-based visualization. Ondov et al. (2011) presented krona which is a powerful metagenomic visualization tool and a demonstration of the potential of HTML5 for highly accessible bioinformatics visualizations. Its interactive displays facilitate more infor med interpretations of metagenomic analyses, while its implementation as a browser-based application makes it extremely portable and easily adopted into existing analysis packages.

Murray (2013) presented a 3D JavaScript-based tool for loading data into a web page and generating visuals from data. Murray's study provides a better understanding to novice programmers who have little programming experience or no at all. It is a tool for non programmers in many fields including those dealing with climate and weather.

We are in the era where the effect of climate change is very noticeable. This call for a tool to analyze voluminous data generated from weather stations in real time in order to look for a pattern of effects of climate change. Climate change extremes such as flooding and seasonal drought are already undermining the economies of countries in the Horn of Africa[1], with agriculture and water resources being the most affected sectors (Rosenzweig et al., 2013).

Buja et al. (1991) presented two principles used to design interactive system to visualize different ecanerios. Two basic principles for interactive visualization of high-dimensional data-focusing and linking were discussed. Focusing techniques may involve selecting subsets, dimension reduction, or some more general manipulation of the layout information on the page or screen. A consequent of focusing is that each view only conveys partial information about the data and needs to be linked so that the information contained in individual views can be integrated into a coherent image of the data as a whole. Ladstadter et al. (2010) reported that interactive system facilitates iterative and interactive browsing of the parameter space to quickly understand the data characteristics, to identify deficiencies, to easily focus on interesting features, and to come up with new

Author α σ: Sokoine University of Agriculture.
e-mail: kadefue@sua.ac.tz
Author ρ: Centre for Agric. Mechanization and Rural Technologies Arusha, Tanzania.

[1] (retrieved from http://africanclimate.net/en/node/6080 on 2016/09/24)

hypotheses about the data. These properties extend the common statistical treatment of data, and provide a fundamentally different approach. Tomiscki et al. (2011) reported on a survey that they conducted to evaluate the application of interactive visualization methods and to identify the problems related to establishing such methods in scientific practice. The feedback from 76 participants showed clearly that state-of-the-art techniques are rarely applied and that integrating existing solutions smoothly into the scientists' workflow is problematic. They tried to illustrate how interactive visualization tools can be successfully applied to accomplish climate research tasks. They showed some examples to support that interactive sys- tems were really required. Lu et al. (2011) demonstrated the framework that has great flexibility and simplicity for end users intending to perform data analysis by aiding the integration of data and tools and enabling interactive visualization on-the-fly. The system was coupled with effective utilization of computational resources and data storage systems.

Therefore, this paper presents a technology developed to allow easy visualization and interaction with the Agricultural Model Inter-comparison and Improvement Project (AgMIP) output data that contain climatic information of several locations in Tanzania. AgMIP is a major international effort linking the climate, crop, and economic modeling communities with cutting-edge information technology to produce improved crop and economic models and the next generation of climate impact projections for the agricultural sector (Rosenzweig et al., 2013; Sanga et al., 2013). The technology that was developed for visualization of AgMIP output is known as ECAWsoft. This is the short form of Enhancing Climate Change Adaptation in Agri- culture and Water Resources in the Greater Horn of Africa (ECAW) Software. The data sets obtained from the ECAW project that have been generated using models were used in the development of this system. The coding was done mainly using PHP, HTML5 and JavaScript. The famous JavaScript libraries like jquery and bootstrap were used to integrate interactive features of the visualization system. The system is web based and uses open source tools that cost almost nothing because of the free license behind their innovations. The interaction system is user friendly. The code integrated in this system might seem to be complex but it was made so to achieve the best human to machine interaction and experiences. The system is designed for visual presentation of the data based on the region-oriented metaphor that includes visualization levels and aggregation or fusion features of the graphs. The system is able to present comparison of up to six locations of the information in the interactive manner that allows the user to granulate or aggregate the data presented.

II. Materials and Methods

This system used the Unified Modeling Language (UML) for analysis and design of the system components. The system was designed to get text files from the folder and then process the data dynamically and draw the dynamic graphs (Figure 1).

a) ECAWsoft flowchart

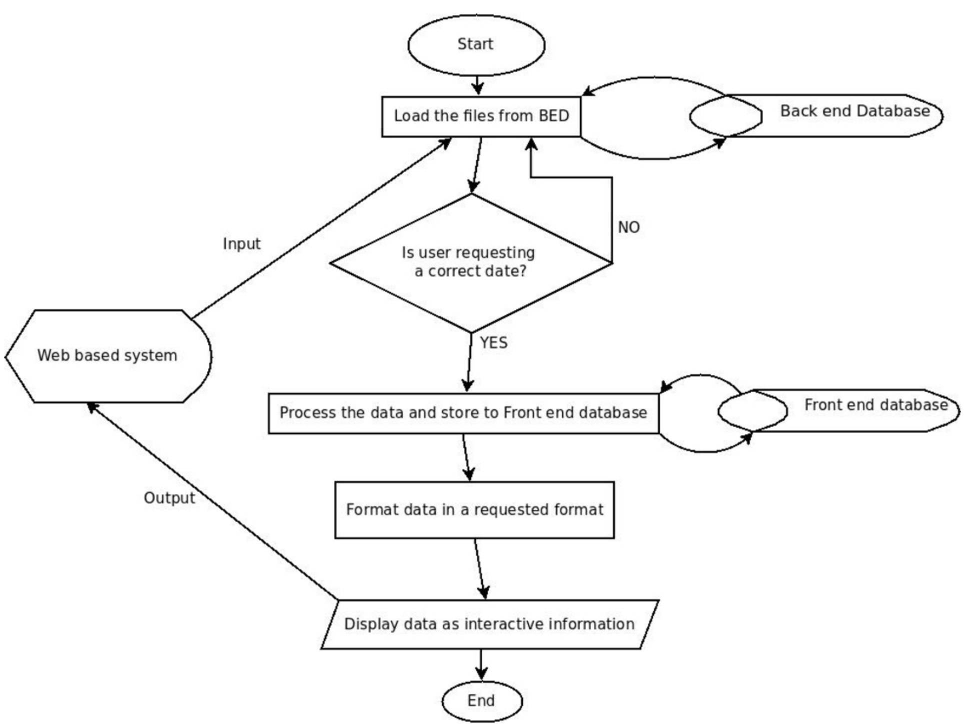

Figure 1: ECAWsoft flowchart

Global Journal of Computer Science and Technology (H) Volume XVII Issue I Version I Year 2017

Figure 1 presents the logic flowchart of the ECAWsoft. The system reads input and creates a dynamic back end database then loads the information to the frontend which processes and present it output to the users. The information is presented as interactive information that allows front end database to be upda- ted with information from the database as requested by the user while navigating the system.

b) Characteristics of the Input data

The data sets are provided in the form of text files with extensions “agMIP” (Mourice et al., 2017). The datasets were collected from the weather stations located in Tanzania and other missing data were generated automatically using weather modeling algorithms and software. The data sets include a form that text fields are separated by space. The algorithm goes through the data set to establish database that can be visualized using HTML5 capable browser. In Figure 2, the focus is on 7 types of data: Solar Radiation (SRAD), Maximum Temperature (TMAX), Minimum Temperature (TMIN), Rainfall (RAIN), Relative Humidity (RHUM), Wind (WIND) and Dew Point (DEWP) (Wambura et al., 2015). All the given data sets are presented using interactive line graphs except rainfall datasets which are presented using bar charts. The data axis is presented against time -series. Time series is the first column with @DATE as name. The file has important information such as LAT for Latitude and LONG for Longitude of the weather station.

The user is given an opportunity to decide when to display the information and which data sets to include and which places of that data set is desired for comparisons. In fact, the user can choose up to six places to display. The user may decide to add new dataset that can be displayed automatically by the system. These datasets are categorized in two parts; baseline simple-scenario data and Coupled Model Inter-comparison Project phase 5 (CIMP5) generated data (Msongaleli et al., 2015).

*WEATHER DATA : USAMCAXA - baseline dates maintained for leap year consistency

@ INSI	LAT	LONG	ELEV	TAV	AMP	REFHT	WNDHT				
USAM	42.017	-93.750	329	11.2	14.6	-99.0	-99.0				
@DATE	YYYY	MM	DD	SRAD	TMAX	TMIN	RAIN	WIND	DEWP	VPRS	RHUM
1980001	1980	1	1	1.2	1.3	-1.5	0.0	3.1	-0.3	6.0	89
1980002	1980	1	2	4.7	-0.3	-2.6	0.0	4.9	-7.6	3.5	58
1980003	1980	1	3	1.9	-0.3	-4.8	0.0	4.3	-9.0	3.1	52
1980004	1980	1	4	3.8	0.2	-2.6	0.0	4.1	-5.2	4.2	67
1980005	1980	1	5	1.0	0.2	-3.2	1.5	3.4	-2.5	5.1	82
1980006	1980	1	6	8.5	1.9	-7.0	2.1	9.1	-0.8	5.7	82
1980007	1980	1	7	6.7	-6.4	-15.9	0.0	8.2	-13.4	2.2	58
1980008	1980	1	8	6.7	-8.7	-16.5	0.0	2.9	-17.1	1.6	51
1980009	1980	1	9	2.2	-11.4	-20.4	1.5	3.7	-16.2	1.8	68
1980010	1980	1	10	8.0	6.3	-13.7	0.0	8.0	5.3	8.9	93
1980011	1980	1	11	4.0	11.3	-9.8	0.0	11.9	7.3	10.2	76
1980012	1980	1	12	8.6	0.8	-13.7	0.0	6.1	-10.4	2.8	43
1980013	1980	1	13	8.6	12.5	1.3	0.0	6.7	2.4	7.3	50
1980014	1980	1	14	2.1	6.9	-3.2	0.0	4.1	-1.0	5.7	57
1980015	1980	1	15	1.0	8.6	4.6	0.0	4.1	7.2	10.2	91
1980016	1980	1	16	1.9	7.5	2.4	24.4	5.6	7.4	10.3	99
1980017	1980	1	17	3.7	2.5	-2.6	1.2	5.6	-0.6	5.9	80

Figure 2: agMIP output file which is the ECAWsoft input file

c) Implementation of the data sets and system

The visualization tool has integrated HTML5 features, PHP programming language version 5.6 and JavaScript scripting language to achieve a web-based visualization and interaction system. The system can easily be installed in Apache 2 web server. The system has been maximized to use google-chrome browser and in fact, it works very fine with other famous browsers like Firefox and Internet explorer. The on-line tool has been made easier for any user to install it even in a local machine with Apache 2 and PHP version 5 installed.

Global Journal of Computer Science and Technology (H) Volume XVII Issue I Version I Year 2017

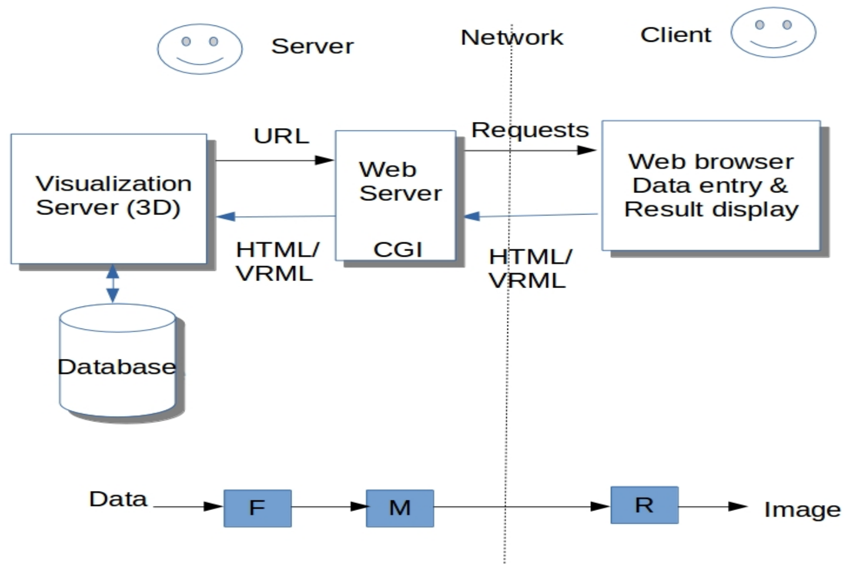

Figure 3: Server-side visualization using CGI and VRML. Note: the happy face means light load and the unhappy face means heavy load. F, M and R stand for Filter, Map and Render, respectively. (Huang, B., 2003)

In a server-side application, a web browser is used to generate requests, send them to the application server, and display the results. The application server, connected with the web server via the Common Gateway Interface (CGI), processes the requests and delivers the result in a standard Web format (e.g. HTML) back to the client. In such an application, the client is usually an HTML page containing forms connected with the application server, while all the software as well as the databases resides on the server that is administrated by the deploying organisation (Huang, 2003).

III. Results and Discussions

a) *Visualization and interaction system*

In this section, the discussion is on the visual parts of the system used for visualization and interaction approaches for visual analysis of the input datasets against time-series.

The main view of the visualization tool has two sidebars; left and right sidebars. The left sidebar show Home, Baseline data, Present Stations (These are Tanzania regions specific data) and help as presented in Figure 5 and Figure 6.

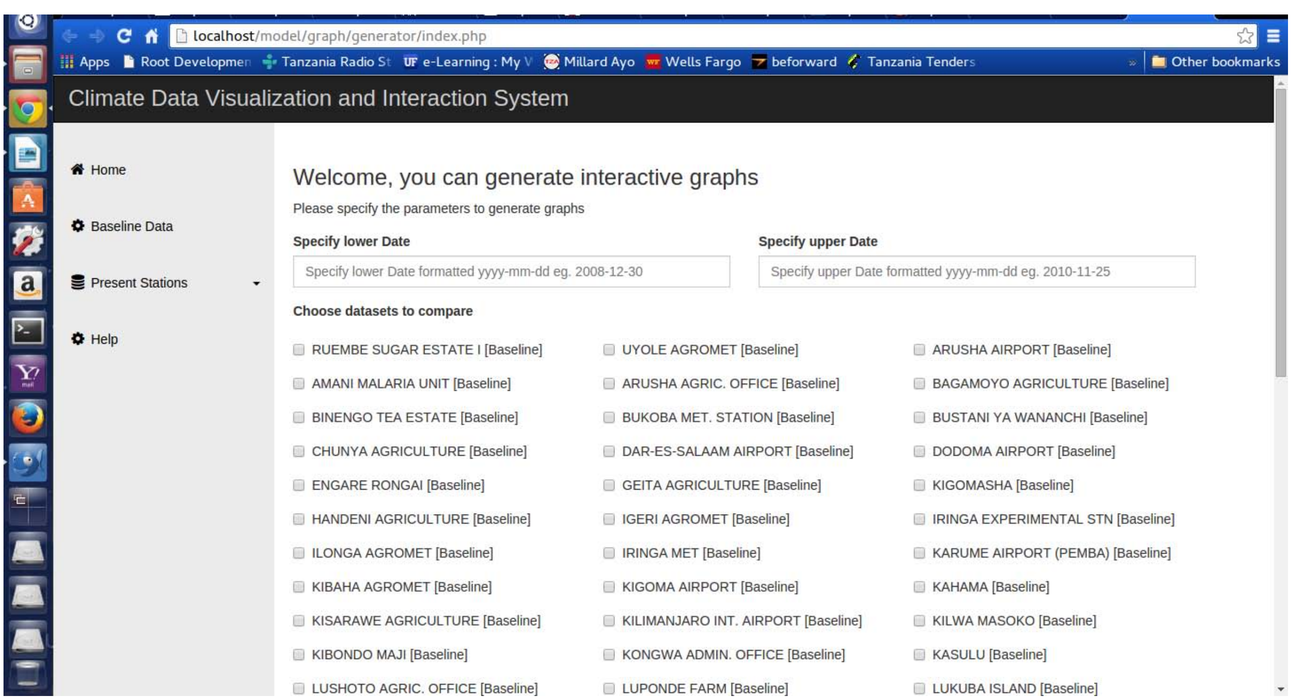

Figure 4: The Main view of the visualization Tool

The right side shows all the climate baseline dataset stations that have been read from the climate database. The user is able to choose the dates desired to be generated by this tool.

The user can then choose desired dates and places for comparison as in Figure 5, Figure 6 and Figure 7. The date panel allows user to go back years using << or monthly using < and also g forward by using the opposite signs.

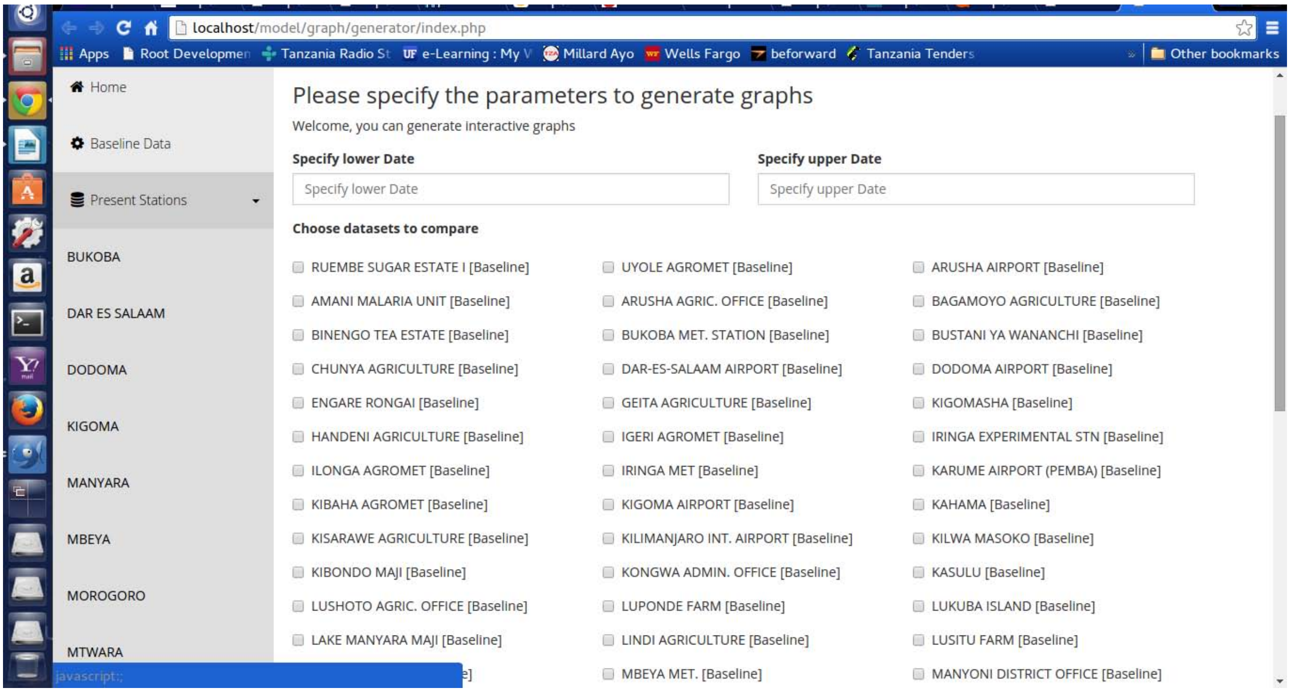

Figure 5: The Main view with present stations show on the left sidebar

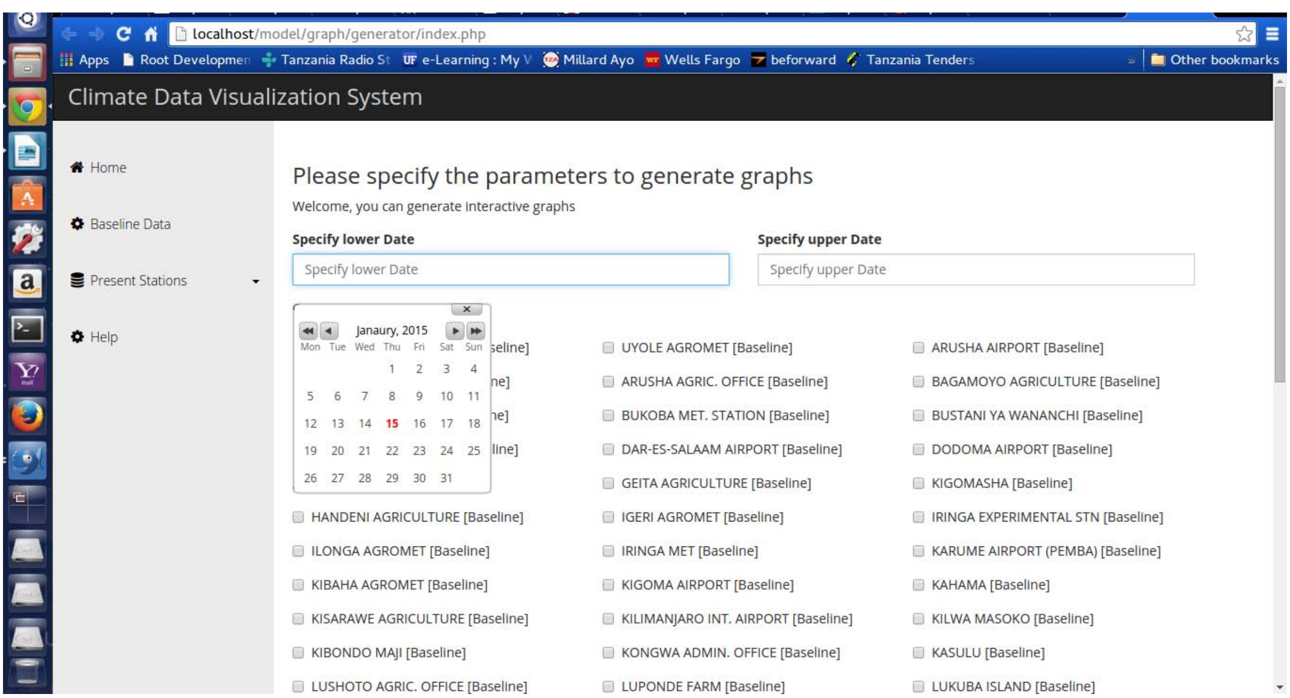

Figure 6: The desired dates and places choose by the user

Global Journal of Computer Science and Technology (H) Volume XVII Issue I Version I Year 2017

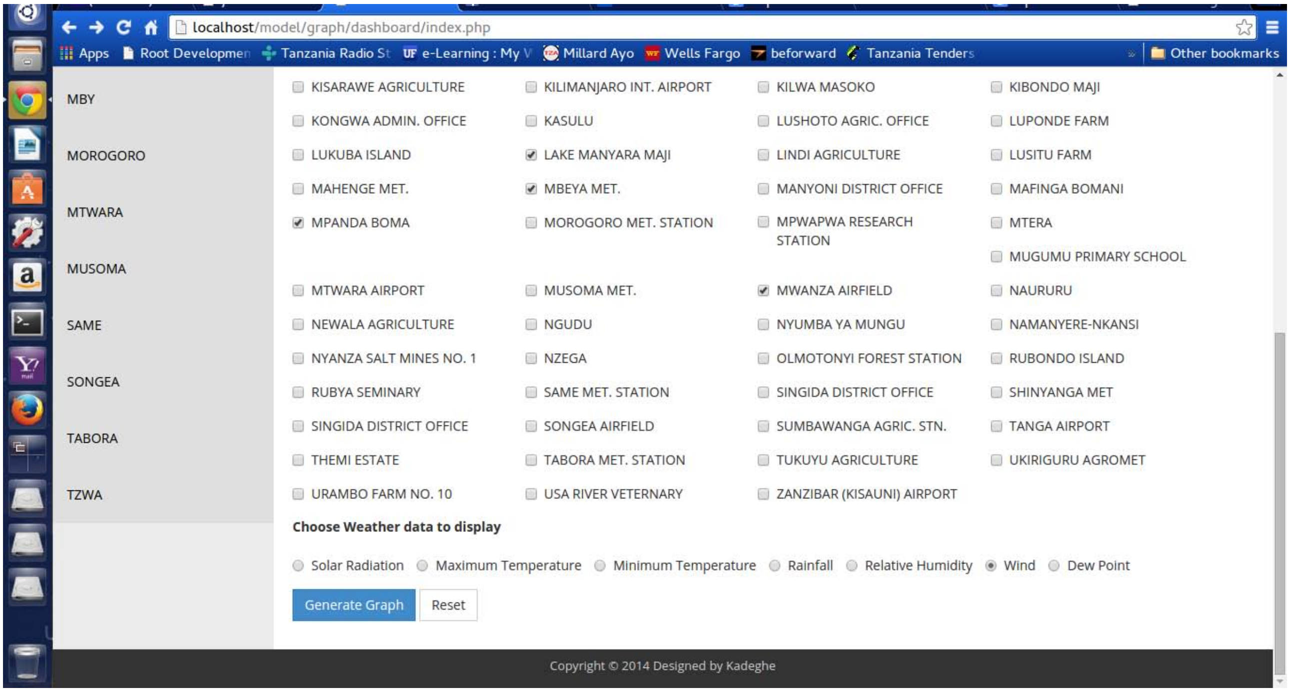

Figure 7: The user specifies the weather data type to be generated

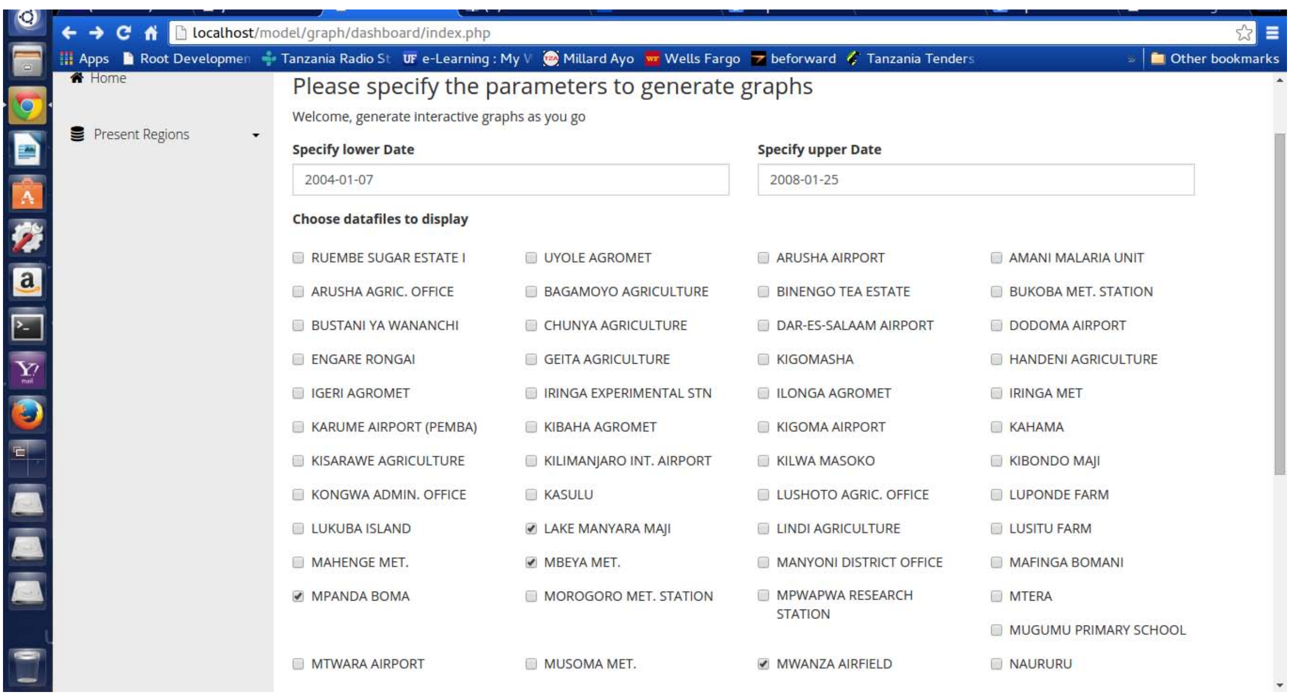

Figure 8: The Main view just before the user presses generate button to generate interactive graphs

The user can now press GENERATE button to instruct the visualization tool to generate the interactive graphs.

For instance, if the user chooses four stations to visualize the information, then the system is going to show years only as shown in Figure 8 and Figure 9.

Now, the user may decide to drag between the dates by holding left click and moving the cursor to visualize the information in detail as shown in Figure 10. Figure 11 visualize the chosen date from Figure 10.

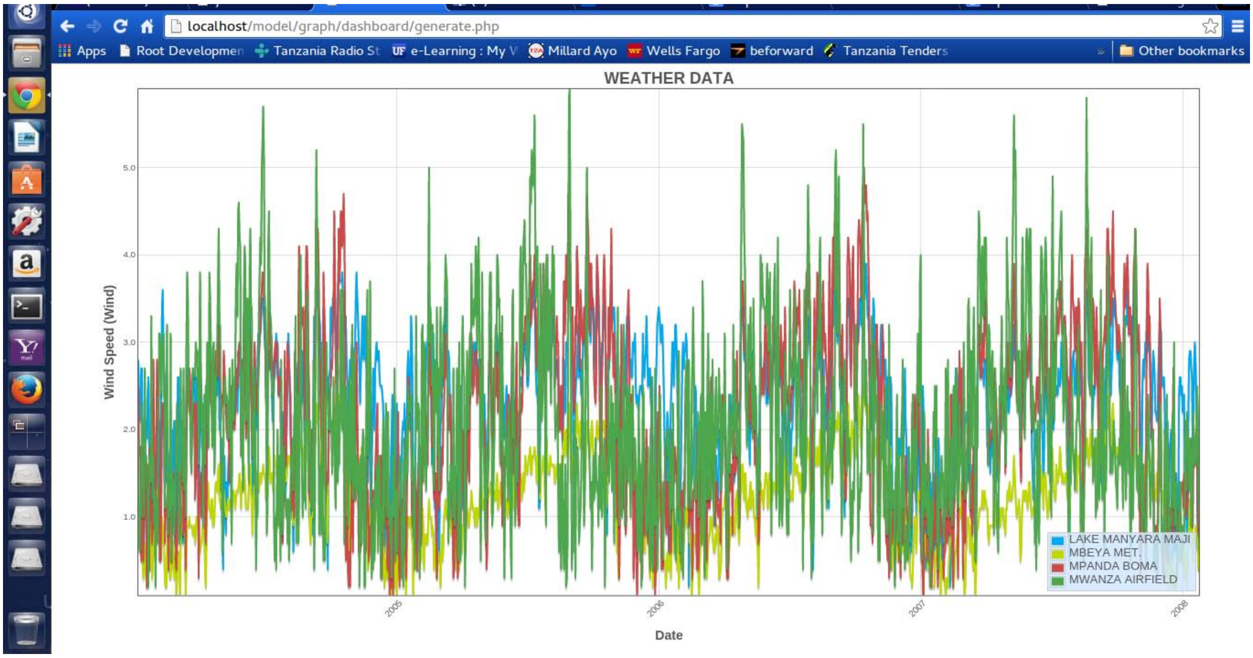

Figure 9: Weather data generator has generated the dataset between 2005-01-04 to 2009-01-23

Figure 11 includes more detailed information that is why the time series now show month with year instead of the years only. If you go further it will show date and then time as shown in Figure 12.

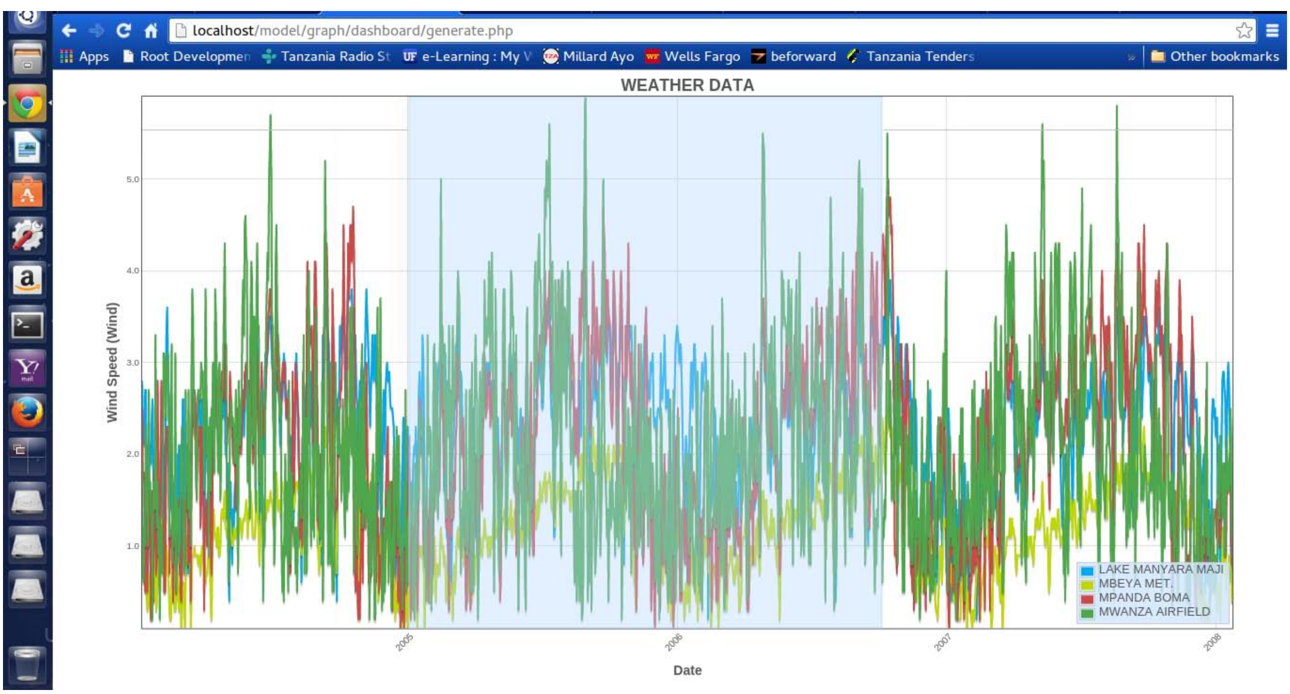

Figure 10: The weather data visualized in detail by dragging between the time-series

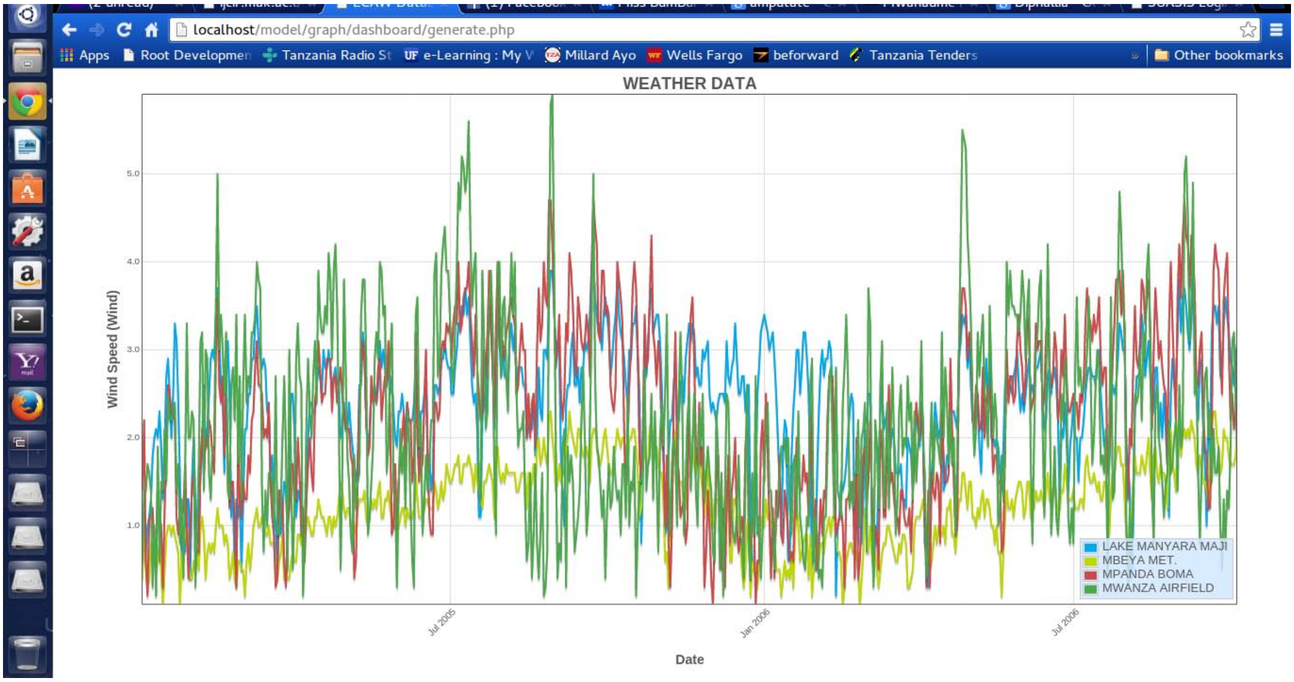

Figure 11: The weather data for the year 2006

Figure 12: The weather data for specific month of April, 2006

Rainfall information is visualized using bar graphs as shown in Figure 13.

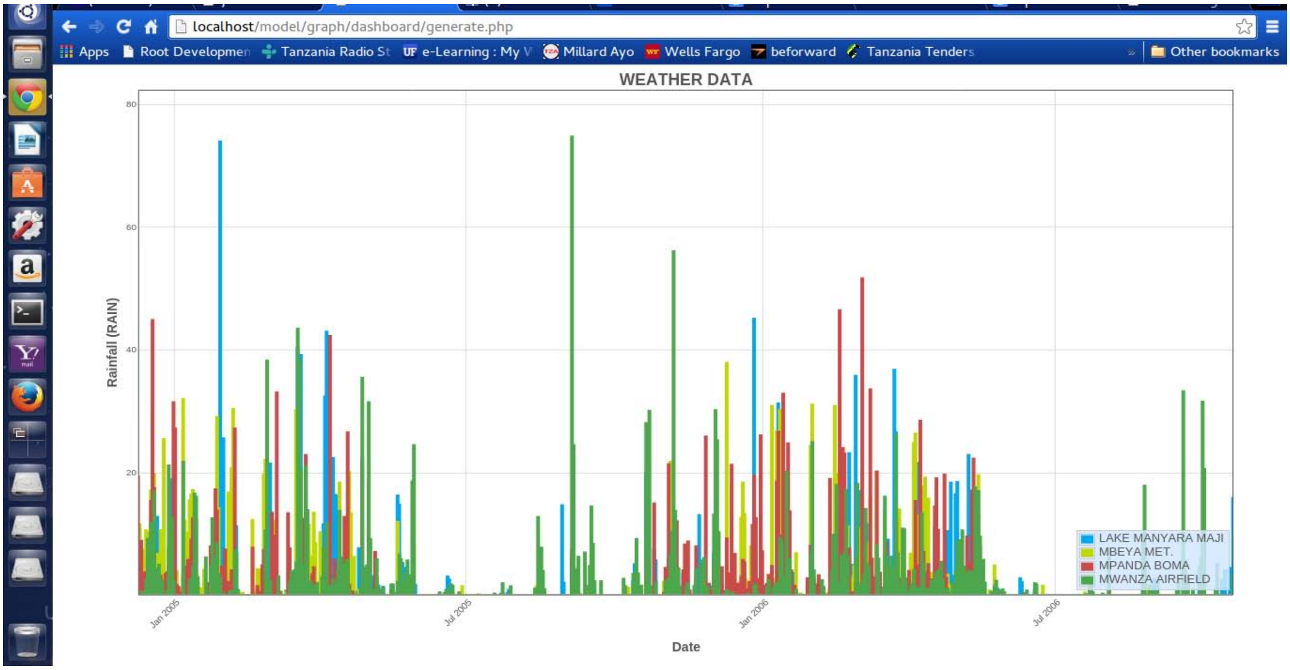

Figure 13: Rainfall data visualization

The rainfall data can be seen in each time, the rainfall was recorded. The detailed data will separate the information so that the differences of rainfall data of the same day can easily be visualized by the user as shown in Figure 14.

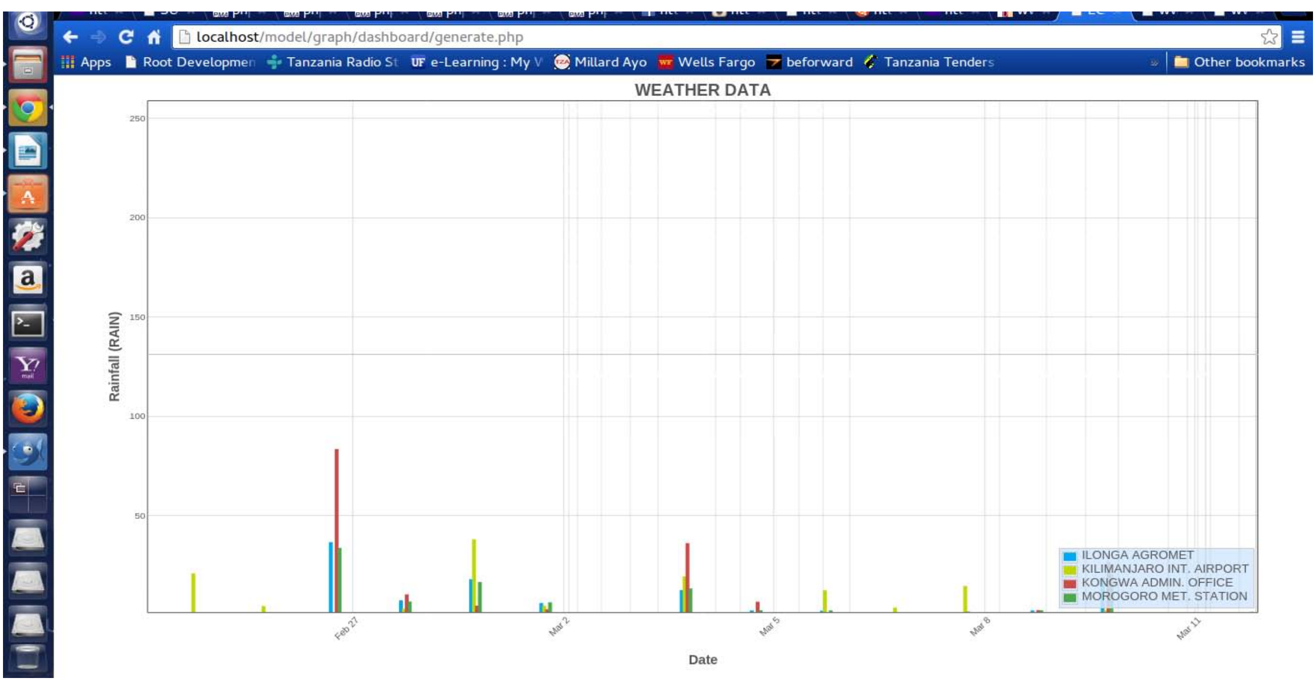

Figure 14: Very detailed rainfall data visualized as bar graph

Also, the system can be used to compare current weather conditions to modeled weather conditions as shown in Figure 15. It means it is possible to compare baseline information to the near-term conditions (2010-2039), Mid-Century (2040-2069) or End-of-Century (2070-2099) using four Representative Concentration Pathways (RCP2.6, RCP4.5, RCP6.0 or RCP8.5) as shown in Figure 16. The left sidebar has present stations which present data for different regions. As the matter of fact, the years are maintained 1980-2009 for compatibility reasons. Near-term means plus 30 years, Mid-century means plus 60 years while End-of_Century means plus 90 years. All the months correspond to each month across years.

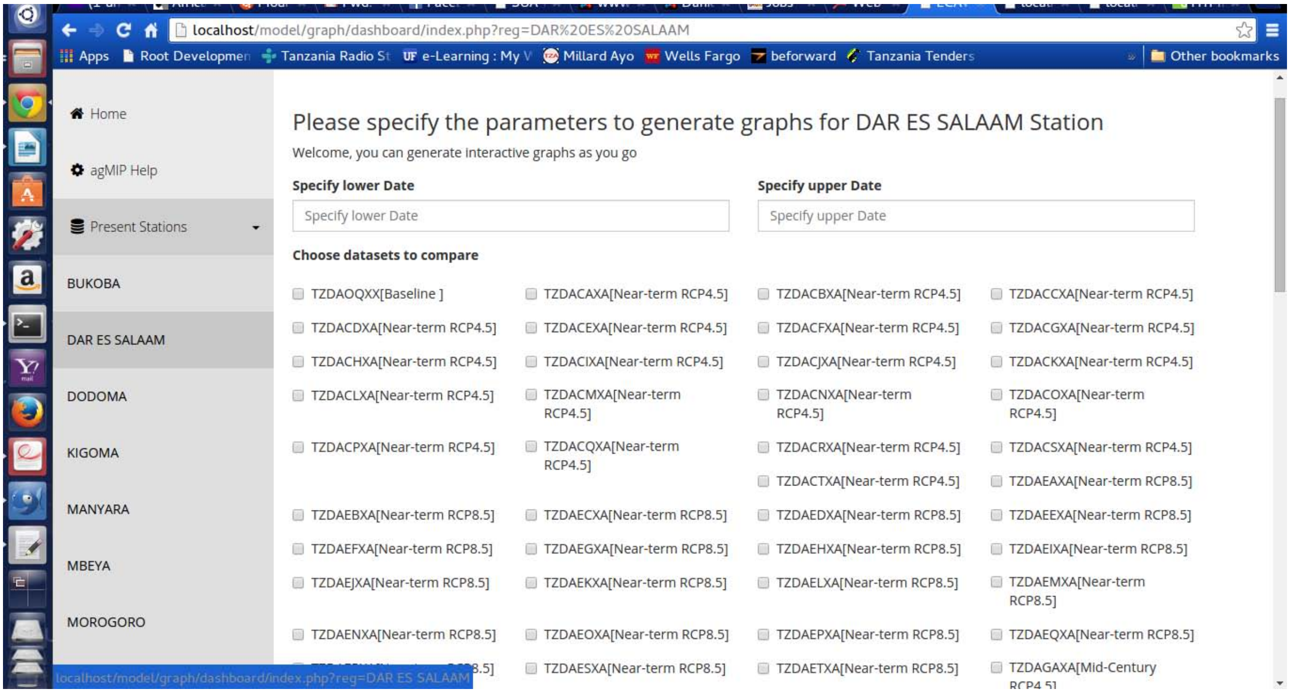

Figure 15: Choosing station to compare the baseline with the predicted weather conditions

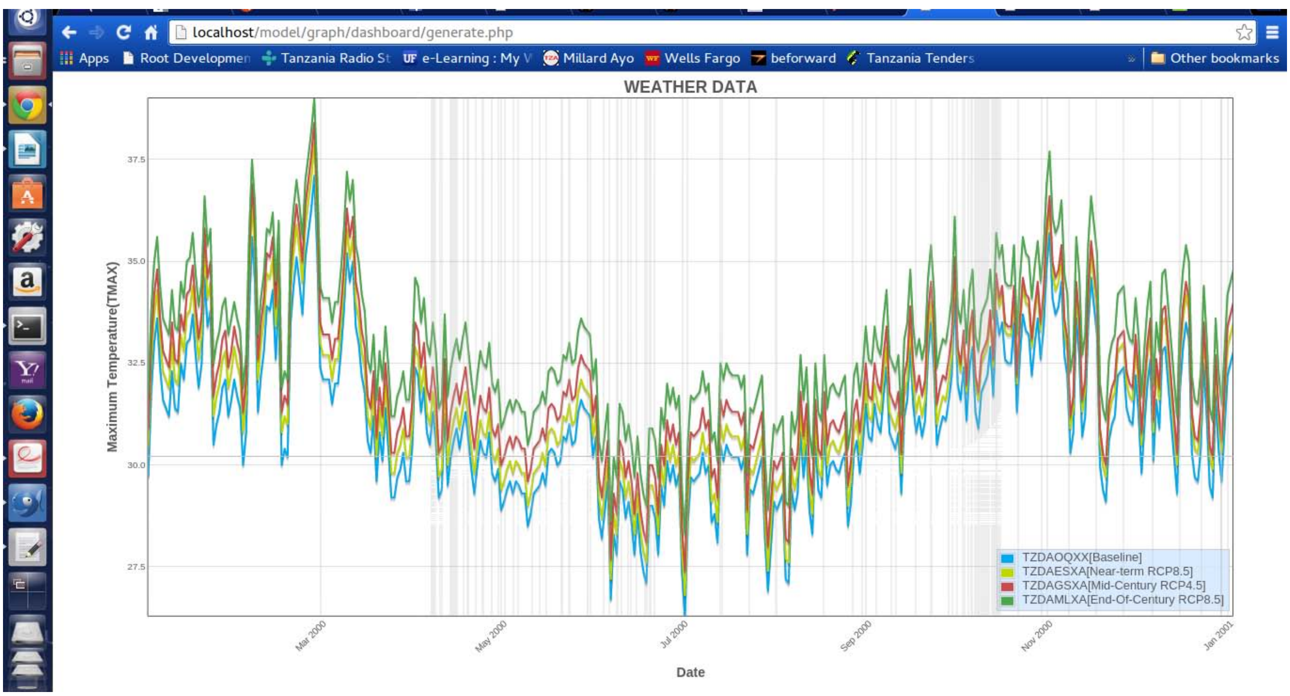

Figure 16: Generated graph for all year 2000, 2030[near-term], 2060[Mid-Century], 2090[End-of-Century]

The results from this paper are better compared to that from Luhunga and Chang'a (2016) who presented a decision supports system for determining effects of climate change. Their system was not interactive and hence, not adapted to rural farmers. On the other hand, ECAWSoft[2] is an open system (Sanga et al., 2016) for open data for the climate and weather visualization. It is in use online. Future[3] data visualization system needs to embed data warehousing, data mining and artificial intelligence algorithms (Woodard, 2016).

IV. CONCLUSION

This paper has presented a technique that can easily be loaded using a flash disc and manipulates data for easy visualization and interaction. The future

[2] http://41.73.194.138/model/graph/generator/index.php

work will involve integration with the web mapping system that will allow the users to visualize the information just in upper layer of the maps dynamically so as to make a feel and understanding of spatial information.

ECAWsoft has used simple but somehow complicated scripting to give state of art interaction of the dynamic graphs for large dataset of climate information (i.e. big data analysis)

References Références Referencias

1. Heo, M., & Hirtle, S. C. (2001). An empirical comparison of visualization tools to assist information retrieval on the Web. *Journal of the American Society for Information Science and Technology*, *52*(8), 666-675.
2. Luhunga, P., & Chang'a, L. (2016). Assessment of the impacts of climate change on maize production in the Wami Ruvu basin of Tanzania. *Journal of Water and Climate Change*, jwc2016055, DOI: 10.2166/wcc.2016.0.
3. Woodard, J. (2016). Big data and Ag-Analytics: An open source, open data platform for agricultural & environmental finance, insurance, and risk, Agricultural Finance Review, Vol. 76 Iss: 1, pp.15 – 26.
4. Brodlie, K. (1997, June). Visualization over the world wide web. In *Scientific Visualization Conference (dagstuhl'97)* (pp. 23-23). IEEE.
5. Murray, S. (2013). *Interactive data visualization for the Web*. O'Reilly Media, Inc.
6. Buja, A., McDonald, J. A., Michalak, J., & Stuetzle, W. (1991, October). Interactive data visualization using focusing and linking. In *Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on* (pp. 156-163). IEEE.
7. Huang, B. (2003). Web-based dynamic and interactive environmental visualization. *Computers, Environment and Urban Systems*, *27*(6), 623-636.
8. Lu, S., Li, R. M., Tjhi, W. C., Lee, K. K., Wang, L., Li, X., & Ma, D. (2011, November). A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data. In *Cloud Computing Technology and Science (Cloud Com), 2011 IEEE Third International Conference on* (pp. 618-622). IEEE.
9. Ladstädter, F., Steiner, A. K., Lackner, B. C., Pirscher, B., Kirchengast, G., Kehrer, J., ... & Doleisch, H. (2010). Exploration of climate data using interactive visualization. *Journal of Atmospheric and Oceanic Technology*, *27*(4), 667-679.
10. Tominski, C., Donges, J. F., & Nocke, T. (2011, July). Information visualization in climate research. In *Information Visualisation (IV), 2011 15th International Conference on* (pp. 298-305). IEEE.
11. Sanga, C., Sumari, N., & Tumbo, S. D. (2013). On the Development of Climate Data Visualization tool for Interpretation of Empirical Results from Climate Model: Does it add Value to Different Stakeholders?. In *6th UbuntuNet Alliance Annual Conference* (pp. 223-235).
12. Msongaleli, B. M., Rwehumbiza, F., Tumbo, S. D., & Kihupi, N. (2015). Impacts of climate variability and change on rainfed sorghum and maize: implications for food security policy in Tanzania. *Journal of Agricultural Science*, *7*(5), 124.
13. Mourice, S. K., Mbungu, W., & Tumbo, S. D. (2017). Quantification of Climate Change and Variability Impacts on Maize Production at Farm Level in the Wami River Sub-Basin, Tanzania. In *Quantification of Climate Variability, Adaptation and Mitigation for Agricultural Sustainability* (pp. 323-351). Springer International Publishing.
14. Wambura, F. J., Ndomba, P. M., Kongo, V., & Tumbo, S. D. (2015). Uncertainty of runoff projecttions under changing climate in Wami River subbasin. *Journal of Hydrology: Regional Studies*, *4*, 333-348.
15. Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S. and Asseng, S. (2013). The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agricultural and Forest Meteorology*, *170*, pp.166-182.
16. Sanga, C. A., Masamaki, J. P., Fue, K. G., Mlozi, M. R., & Tumbo, S. D. (2016). Experimenting open agricultural extension service in Tanzania: A case of Kilosa Open Data Initiative (KODI), Journal of Scientific and Engineering Research, 3 (6), pp. 116-124.

[3] https://www.ag-analytics.org/AgRiskManagement/Home

This page is intentionally left blank