CREATING INTERVENTION MAPS FOR SOIL-TRANSMITTED HELMINTHIASIS: CASE OF KENYA

Internship Report in partial fulfillment of the requirement of
GEOMATICS CERTIFICATE

Presented by :

Gildas BOKONON-GANTA

Supervisors :

For WHO :

- Dr Antonio MONTRESOR
- Ms. Natacha à PORTA

For University of Geneva:

Professor Hy DAO

Genève, January 2015

Content

List of figures

List of tables

Abstract

BACKGROUND: Soil-Transmitted Helminth (STH) infections are caused by four species of nematodes. The life cycle of these parasites includes a human host as site of the adult worms. Through faeces, the worm eggs contaminate the soil and are ingested where there is a lack of hygiene. Simple habits like washing hands can help diminish the contamination. Furthermore, an access to improved sanitation is a significant way to break these parasite cycles. In areas where the STHs prevalence is a matter of public health importance (prevalence estimated to be over 20%), the World Health Organization (WHO) recommends Preventive Chemotherapy (PC), the large scale distribution of anthelminthic drugs to population groups at risk, to reduce the morbidity and the transmission. In this vein, having a clear overview on risk areas is very useful for improving the deworming action. WHO develops for this purpose a mapping manual in two parts: the first for establishing the baseline prevalence in available counties were data are available; the second for predicting the prevalence in all the country, using baseline prevalence from of the first part and predictors variables. This manual was tested with data in some endemics countries. The aim of this study is to test the manual once again and to contribute to its improvement by bringing new ideas. Kenya is the work case.

METHODS AND FINDINGS: For this purpose, a collection of data using two existing data sources and some additional resources helped to build a database with a temporal epidemiological dataset on Kenya from 1966 to 2014. The assessment of the prevalence, following the guideline of WHO, was supplemented with seven predictors for the predictive part. The results issued show two tendencies. Prevalence of STH is high in areas close to water bodies and in the economic pole of Kenya. In the hottest areas with few inhabitants per square kilometre, there is no need of treatment. For more for analysis, the results are compared first with a similar study at the regional level, secondly with a recent national survey. In the first case the comparison shows many matches. In the second case, the difference is perfectly explained by the Lymphatic Filariasis endemicity in the coast and the deworming that occurred in the western part. The results after comparisons proved on one hand the robustness of the prediction. On the other hand, it confirms the usefulness of the manual.

CONCLUSION: The manual gives in details some tools for controlling STHs with less time loss. It has a direct impact on budget in the sense that the policy officers can succeed efficiently in the control of STH as public health problem without facing the difficulties of the national surveys that can be replaced by collection of data on sentinel sites for verifications. Furthermore, the mapping process can be complete with open sources software. Thus it is a tool for all managers interested anywhere, without restriction.

Key words
Epidemiological survey,

Epidemiological database,

Intervention map,

Lymphatic Filariasis,

Prediction model,

Preventive Chemotherapy,

Soil-Transmitted Helminth,

QGIS

Acronym list
GAHI Global Atlas of Helminth Infections

GNTD Global Neglected Tropical Diseases database

JMP Joint Monitoring Programme

LF Lymphatic Filariasis

NTD Neglected Tropical Disease

PC Preventive Chemotherapy

STH Soil-Transmitted Helminth

UNICEF United Nations Children's Fund

WHO World Health Organization

1. Introduction

The public health significance of seventeen Neglected Tropical Diseases (NTDs) is widely known. Soil-Transmitted Helminths (STHs) infections are one among four helminth infections within NTDs. Montresor, Awasthi and Crompton (2003) showed that STHs infections are a major cause of morbidity in children aged 5 – 14 years of many countries worldwide. "The STHs infections are common clinical disorders in man with resultant impairments in physical, intellectual, and cognitive development" (Bethony, et al., 2006). These major STHs (hookworms [*Ancylostoma duodenale* and *Necator Americanus*], roundworm [*Ascaris lumbricoides*] and whipworm [*Trichuris trichiura*]) causing morbidity constitute, considering the long term, a great loss factor of social productivity. Over 800 million children, living where worms are intensively transmitted, are assumed to be infected[1]. However, good health, as people know from their own experience, is a crucial part of wellbeing. Accordingly, improved health contributes to society and economic growth. The World Bank report (1993) shows that improved health reduces production losses by worker illness; permits the use of natural resources that has been totally or nearly inaccessible because of disease; increases the enrolment of children in school and make them better able to learn. Officially, the 54th World Health Assembly meeting[2] urged Member States to take action to control morbidity due to STHs infections by ensuring regular administration of anthelminthic drugs to school-age children living in endemic areas (World Health Assembly Resolution, 2002). However, the appropriate targeting of chemotherapy (WHO, 2012) requires information on the distribution of infection prevalence within countries, to identify high-risk areas that might benefit most from control (Brooker, Rowlands, Savioli, & Bundy, 2000). Furthermore, the previous authors mentionned the need for schistosomiasis and STH control programs in Africa.

Kenya, in the Eastern African region, is one of the endemic countries for STHs infections. It is the country chosen to continue the mapping process already begun in Africa and Asia. To reduce the burden of these helminth infections, different deworming programmes have already been initiated in the country with the goal of reducing morbidity of STHs to levels where it is no longer a public health problem (Ministry of Public Health and Sanitation, 2011). In accordance with this vision, the overall goal of this study is to map the distribution of STHs and to prepare afterwards intervention map in Kenya. For this purpose, the collection of data and the constitution of a database are needed as well as the use of Geographical Information Systems (GIS) to collect and map recent published surveys on STHs epidemiology and its distribution in Kenya.

This report aims to describe the STHs mapping process in Kenya. The first part explains the methodology. It describes the whole process, going from data collection to mapping and analyses. The second part is dedicated to the principal results of the process. The third part explains how we obtained a map for Kenya using heterogeneous data since the analyses have been performed with three sets of data. This was important in order to have a better understanding of the infection distribution. The third part integrates also comments, remarks and suggestions resulting from the analysis process, highlighting the main point of our analyses. The analyses have been performed respecting the framework of the manual developed in two parts by WHO STH team. This manual is titled: Preparing maps for STHs infections.

[1] http://intranet.who.int/sites/evd/documents/ebola-who-new-report-121114.pdf
[2] Geneva, May 2001

2. STHs species and WHO guidelines for their treatment

The STH infection on which this study focuses in Kenya is caused by four species of nematodes. Figure 1, Figure 2 and Figure 3 show respectively the eggs, the male and the female adults of roundworm (*Ascaris lumbricoides*), whipworm (*Trichuris trichiura*) and hookworms (*Ancylostoma duodenale* and *Necator Americanus*).

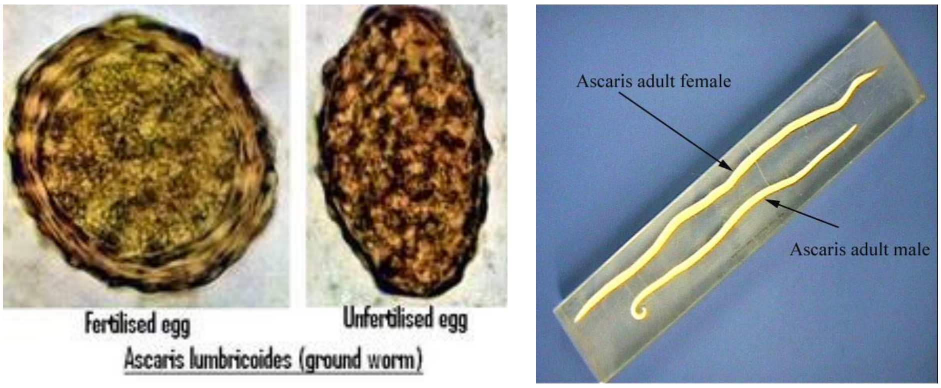


Figure 1: Eggs, adults male and female of *Ascaris lumbricoides* (Wikipedia)

Figure 2: Eggs, adults male and female of *Trichuris trichiura* (Wikipedia)

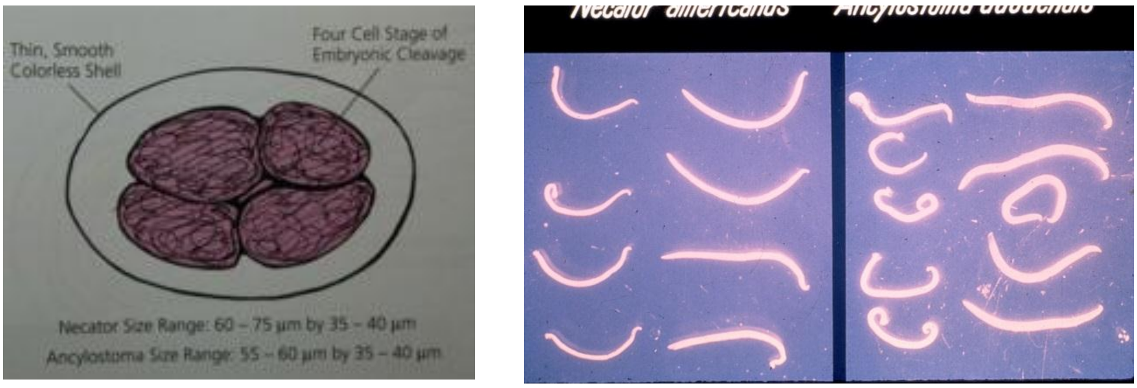


Figure 3: Eggs, adults male and female of *Necator Americanus and Ancylostoma duodenale* (Wikipedia)

The nematode species of our interest have a life cycle that requires a human host. They grow in their host body where, in the long term, they can cause severe health disorders and malfunctions when they are quantitatively significant. The life cycle of these worms is boosted by the open defecation and lack of sanitation which are some of the major characteristics of poor areas in the world. Figure 4 shows the schematic life-cycle of STH. Table 1 shows some characteristics related to each of the four species.

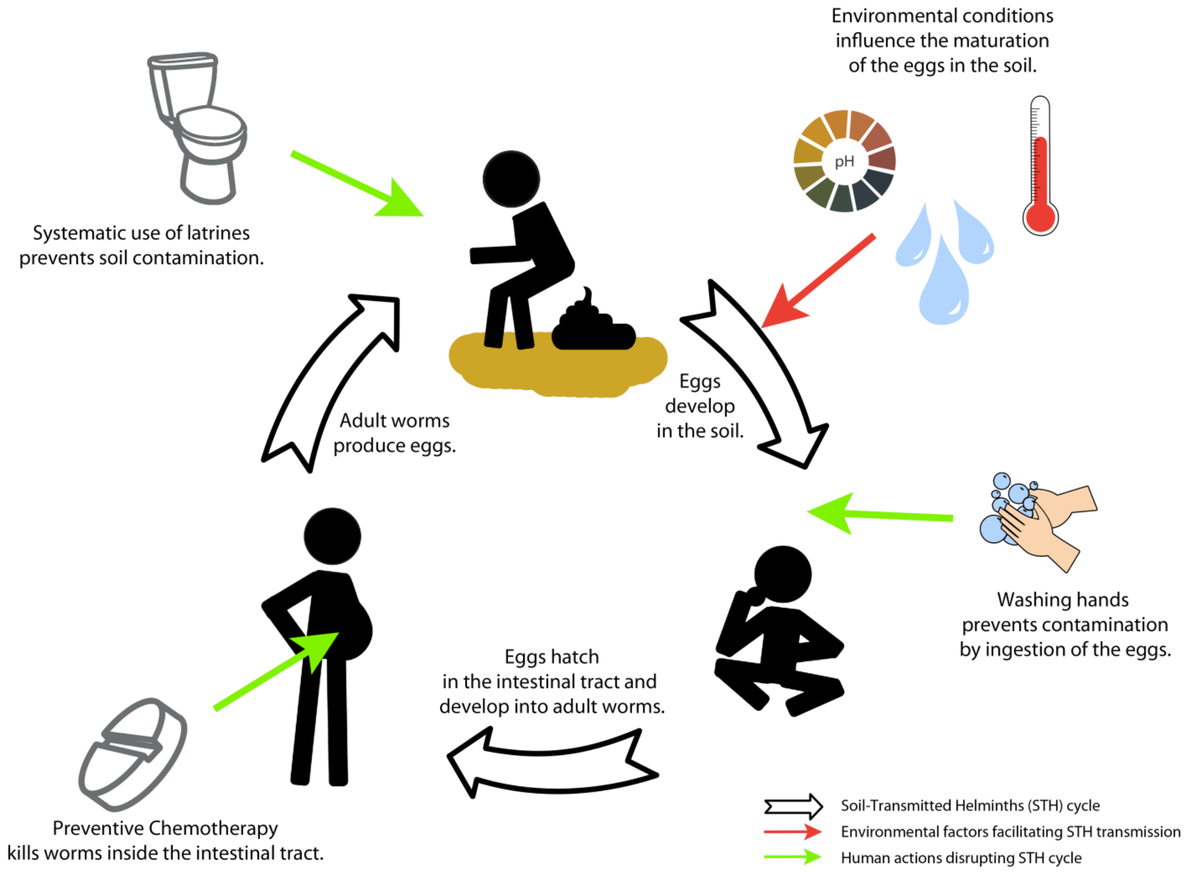


Figure 4: Schematic life-cycle of STHs (In the non-published manual, based on figure 1.3 from WHO, 2011)

Table 1: Some charachteristics of STHs

Species	Mode of infection	infective stage	Site of adult	Geographical distribution
Ascaris lumbricoides	Mouth	Larva 2 in Egg	Stomach and small intestin	Cosmopolitan
Trichuris trichiura	Mouth	Larva 1 in Egg	Caecum	Cosmopolitan
Ancylostoma duodenale	Mouth or Skin	Larva 3	Small intestine	Asia, Africa, Middle East, Pacific
Necator americanus	Skin	Larva 3	Small intestine	Asia, Africa, Middle East, Pacific

WHO recommends, as control intervention, annual treatment in areas where the STHs are a serious concern. In this vision, the organization published maps for STHs infections that present the interpretation of the results from parasitological surveys conducted in some endemic countries. These maps are useful for the manager of the control programme to plan the control intervention. They show three categories of prevalence. Table 2, inspired from (WHO, 2011), summarises the necessary actions in the three cases.

Table 2: Recommended treatment strategy for STH in preventive chemotherapy

PREVALENCE "P" CATEGORY (%)		ACTION TO BE TAKEN
P < 20	**→**	**No Preventive Chemotherapy**
20 ≤ P < 50	**→**	**Preventive chemotherapy once a year**
P ≥ 50	**→**	**Preventive chemotherapy twice a year**

It is important to emphasize that the endemicity of Lymphatic Filariasis (LF) must be taken into account for the treatment of STHs. The drug needed for LF PC has an effect on STHs worms. Therefore, where both LF and STH are endemic, the corresponding PC should be coordinated. Where normally STH PC should be implemented once a year, if LF PC is implemented, no STH PC is needed. Where normally STH PC should be implemented twice a year, if LF PC is implemented, only one STH PC round is needed yearly, six months after LF PC.

3. WHO manual for STHs

Evidence shows that in many countries endemic for STHs, there is a lack of accurate information on the distribution of the infection. In some countries, there is no recent national baseline STH survey. However, developing a Preventive Chemotherapy (PC) program requires the information on the baseline epidemiological situation in the country. There are groups working in the mapping field related to STHs infection distribution. They produce different geographical assessments based on historical data. They have different and personalized framework.

WHO develops a manual "***Preparing maps for Soil-Transmitted Helminth infections***". This manual, split in two parts, aims to standardize the collection and mapping STH data, following the thresholds and PC guidelines of the organization. Part I permits to create a baseline infection distribution map based on historical epidemiological data. Part II of the manual uses environmental and human development indicators to model the prevalence. This predictive map is then used to develop a recommended intervention map. While Part I of the manual can be performed by a country programme manager with solid computer skills, Part II should be executed by an external expert, with the appropriate mapping qualifications.

The mapping process object of this report aims to test the two parts of the manual. On one hand, the test will complete the previous one. Therefore it will provide some more output that can help confirming the first part of the standardized approach for STHs mapping. On the other hand, it will be a new occasion for predicting and improving the model in the second part which relies entirely on the first part. It will be as well source of new reflexions before the validation of the manual.

4. Data used and sources

For the achievement of the expected maps for this study, many datasets have been useful. Table 3 presents all the data and the useful sources for extracting all of them.

Table 3: Data used and the sources

<table>
<tr><th>Data</th><th>Sources</th></tr>
<tr><td>New administrative boundaries in Kenya, Level zero, level one and two</td><td>World Health Organization Headquarters, Polio, Emergencies and Country Collaboration (PEC) cluster, Polio Operations and Research (POL) department, Surveillance, Data and Certification (SAC) team</td></tr>
<tr><td>ESRI Shapefiles</td><td>http://www.arcgis.com/home/item.html?id=5f83ca29e5b849b8b05bc0b281ae27bc</td></tr>
<tr><td>Global Administrative shapefiles</td><td>http://www.gadm.org/download</td></tr>
<tr><td>Historical STHs data extracted from articles: County and Regional level (1966 → 2014)</td><td rowspan="3">http://www.thiswormyworld.org/
http://www.gntd.org/
http://www.who.int/neglected_diseases/preventive_chemotherapy/profiles/en
www.who.int/iris
http://www.ncbi.nlm.nih.gov/pubmed
http://apps.webofknowledge.com
http://www.researchgate.net/
http://scholar.google.ch/
http://www.plos.org/
http://wokinfo.com
http://www.bioline.org.br
WHO Library Stacks
WHO / NTD Archives / Unpublished data received in 2002</td></tr>
<tr><td>STHs Survey data published in 2013</td></tr>
<tr><td>Schistosomiasis Data extracted from articles County level (1966 → 2014)</td></tr>
<tr><td>Lymphatic Filariasis endemicity</td><td>World Health Organization Headquarters / Neglected Tropical Diseases department / Preventive Chemotherapy and Transmission control unit.</td></tr>
<tr><td>Soil acidity</td><td>http://www.worldgrids.org/doku.php</td></tr>
<tr><td>Temperatures</td><td>http://www.worldclim.org/</td></tr>
<tr><td>Rainfall</td><td>http://www.worldclim.org/</td></tr>
<tr><td>Literacy rate</td><td>http://dhsprogram.com / Demographic and health survey 2008-09 report</td></tr>
<tr><td>Poverty rate</td><td>https://www.opendata.go.ke/Poverty/Poverty-Rate-by-District/i5bp-z9aq</td></tr>
<tr><td>Sanitation rate[3]</td><td>https://www.opendata.go.ke/Counties/Kenya-County-Fact-Sheets-Dec-2011/zn6m-25cf</td></tr>
<tr><td>Population</td><td>www.afripop.org</td></tr>
</table>

[3] Access to improved water and improved sanitation (percentage of households)
"Definitions for access to improved water and sanitation are based on the Joint Monitoring Programme (JMP) for Water Supply and Sanitation by the WHO and United Nations Children's Fund (UNICEF). Accordingly, improved water sources include well/borehole, piped and rain-harvested water; while, improved sanitation includes connection to a main sewer, septic tank and cesspool as well as ventilated improved pit (VIP) latrine and covered pit latrine" (Commission on Revenue and Allocation CRA, 2011).

5. Methods

5.1- Country selection

The project started by a decision process regarding the country that can be the territory of the analyses and reflection on STHs, following the framework of the manual developed in WHO for the mapping purpose. Seven endemics countries were listed: Angola, Ethiopia, Kenya, Democratic Republic of the Congo, United Republic of Tanzania, Sudan and Pakistan (Annexe A1).

At this stage we compare the databank of Global Atlas of Helminth Infections (GAHI)[4], Global Neglected Tropical Diseases (GNTD)[5] and World Health Organisation (WHO)[6] resources. The existing maps gave an idea on which country has surveys that cover a large part of the country. The existing data on CSV files on GAHI have been downloaded and the tables within have been carefully checked. The consistency of the data has been compared. The existing bibliographies on GAHI and GNTD regarding each country have been explored. The last countries profiles up to date on WHO website have been checked. After this research step we check-crossed the information. The outcome of this first process showed Pakistan as the less relevant candidate in terms of data availability. No survey was available. Kenya was the most reliable. A long list of epidemiological surveys was available on both STHs and on Schistosomiasis survey data. Considering that schistosomiasis data are also needed in this study, Kenya was the appropriate country for the study (Annexe A6 for some information on Kenya).

5.2- Searching data

The data were collected by searching on the open access existing data bases on the web. Specifically, this step started by the collection of all articles listed in the databases of the GAHI, the GNTD and the WHO intranet resources. The formers are institutes in London and Basel which build maps in the same field. Articles from both sources were supplemented by many open sources platforms such PubMed[7], Web of Sciences[8], research gate[9], Google / Google scholar[10]… Some articles not found on any website were found in WHO Stacks in Library. Some are not at all available. All attempts to obtain them were unsuccessful.

At the end of this step, more than a hundred of papers (articles, reports) have been collected on STHs and schistosomiasis surveys. All the titles have been loaded in the endnote database dedicated to all STH mapping activities in the team (Annexe A2). All the PDFs versions have been assembled in two separate folders. They were assembled in a first folder by Author and year of publication (Figure 23). The author’s folder has been duplicated and the PDFs were labelled following the rule: “*CountryISO2code_Number-given-to-the-article_Author-name_Year-of-publication*”. An example is “*KE0001_Rijpstra_1975*”. These two folders enabled during the data entering process a quick

---

4 http://www.thiswormyworld.org/
5 http://www.gntd.org/
6 http://www.who.int/neglected_diseases/preventive_chemotherapy/profiles/en/
7 http://www.ncbi.nlm.nih.gov/pubmed
8 http://apps.webofknowledge.com
9 http://www.researchgate.net/
10 http://scholar.google.ch/

research by name, by date or by WHO label (Figure 24). All the pdfs have also been printed and assembled in a binder so they can be used without any computer or internet access.

5.3- Filling an access database

This part was the longest step of the process. All the articles obtained in the preview step have been read partially or in detail if needed. The epidemiological information related to STHs have been extracted and entered in a Microsoft Access database prepared by the team for the STH mapping purpose prior to our study. Annexe A4 gives a look on the filling form with all the information required. All the relevant information in all articles, reports and unpublished data have been systematically screened and registered. The crucial information needed are stored thanks to a form (Annexe A4). The challenge on this step was to be concentrated in order to find the right information in papers since some were not comprehensively written. Sometimes it was a matter of lack of information. Sometimes it was a matter of contradiction within the numbers.

5.4- Estimating combined prevalence of any STH

The filled access database has been duplicated. The latest version was treated. The prevalence has been adjusted. The adjustment can be resumed in the way that the estimating combined prevalence of any STH has been calculated from the given prevalence. The calculation is done when the articles didn't mention it. For this purpose, the following formula is implemented in Excel in order to enter prevalence by specie and to have an automatic computing of *total STH prevalence.*

$$STHp = (Ap + Tp + Hp - (Ap * Tp + Ap * Hp + Tp * Hp) + Ap * Tp * Hp)/1.06 \, .$$

STHp: Total STH Prevalence
Ap, Tp, Hp: *A. lumbricoïdes*, *T. trichiuria* and hookworm prevalences respectively.
This formula has been justified by de Silda & Hall (2010).
This operation enables to have therefore a mean prevalence for each record. It can be reminded that the WHO advises to implement one or two rounds of treatment per year in districts where the baseline total STH prevalence is above 20%. This is why this step focusses on evaluating whether the total STH prevalence is more likely to be "high", "moderate" or "low", considering 50% and 20% prevalence as thresholds for those categories.

5.5- The Lymphatic Filariasis data

As we have taken into account the articles related to schistosomiasis, it is recommended to take into account the Lymphatic Filariasis (LF) data since the drug used to treat this NTD have an effect on STHs. It was therefore necessary to obtain from the LF data manager the result of the LF treatment in the endemic areas and to introduce them in the process. The LF programme manager provided a dataset that revealed tree essential information:

- Only the Coast province is endemic for LF in Kenya.

- The implementation units are, in terms of area, lower than the county level.

- The treatment programme has been implemented since 2002 in some locations; it started in 2011 for other locations; some locations still do not benefit from any treatment.

Regarding these information, it was necessary to start a process in order to match each sub location to the corresponding new county. In fact, changes of administrative boundaries occurred at level two in the years preceding. The key to understanding these changes and performing good analyses was to concentrate on substantial resources. For this reason, ESRI shapefile[11], Global Administrative Areas website[12] and Google Earth helped. On one hand the shapes of ESRI and Global Admin have been compared. No discrepancy was observed. On the other hand, we overlaid the administrative boundary level tree and four of Global Administrative areas shapefile with the new official level two obtained from WHO. The operation had helped finding solutions to the belonging of the implementation unit of Lymphatic Filariasis. Finally, the exact place of the new area had been checked with Google Earth and in the Kenyan independent commission on boundaries report (Independant Electoral and Boundaries Commission, 2012). At the end of the process, it was possible to know in which new county LF treatment is implemented or not. Annexe A5 shows the comparison tables.

5.6- Mapping the scores

The two parts of the manual include mapping. Following the guidelines of the manual, at least three maps must be issued. *QGIS version 2.6 "Brighton"* is the software used for the mapping process (Annexe A8). For both parts, the final tables saved as comma-separated values (CSV) were joined to the corresponding shapefiles. The categories have been assessed using the "*Style*" tool in the property dialogue box of the shape. Three categories have been accessed as explained in Table 2.

5.6.1- Working on the first part of the manual

For achieving the map related to the first part of the manual, the main access database has been exported to excel. The Excel table has been cleaned afterwards as recommended. The final Excel table containing all information on the counties prevalence, the category assessment, the sources, the LF endemicity has been checked. This Excel table, saved under comma-separated values (CSV), has been joined to the attribute table of the basic shapefile for issuing the first map.

5.6.2- Working on the second part of the manual

"Zonal statistics" was an additional tool needed for the part two. It was obtained thanks to the corresponding plugin. It has helped for computing automatically from the raster files information the environmental predictors by county (Figure 5).

[11] http://www.arcgis.com/home/item.html?id=5f83ca29e5b849b8b05bc0b281ae27bc
[12] http://www.gadm.org/download

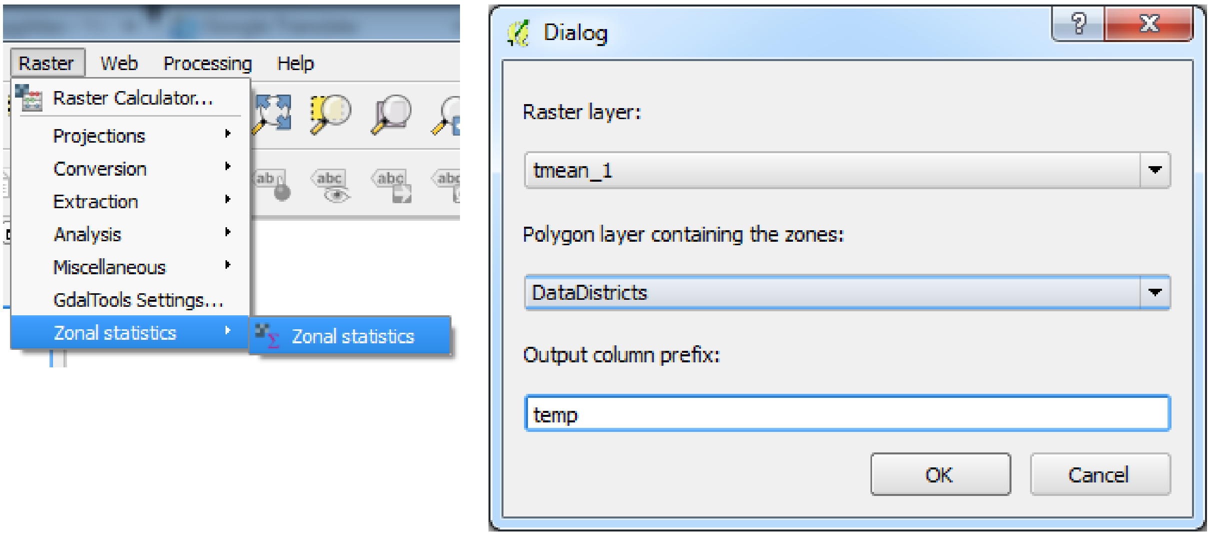


Figure 5: Raster treatment. Computing data from a raster file in a vector file

One result per month was obtained (twenty four columns) for both temperatures and precipitations. The use of the *"Field calculator"* tool followed. It helped to get the annual sum of precipitation. Figure 6 shows an image of the geomatic treatment *("Field calculator">"Addition operator" using the useful "Fields and Values"*).

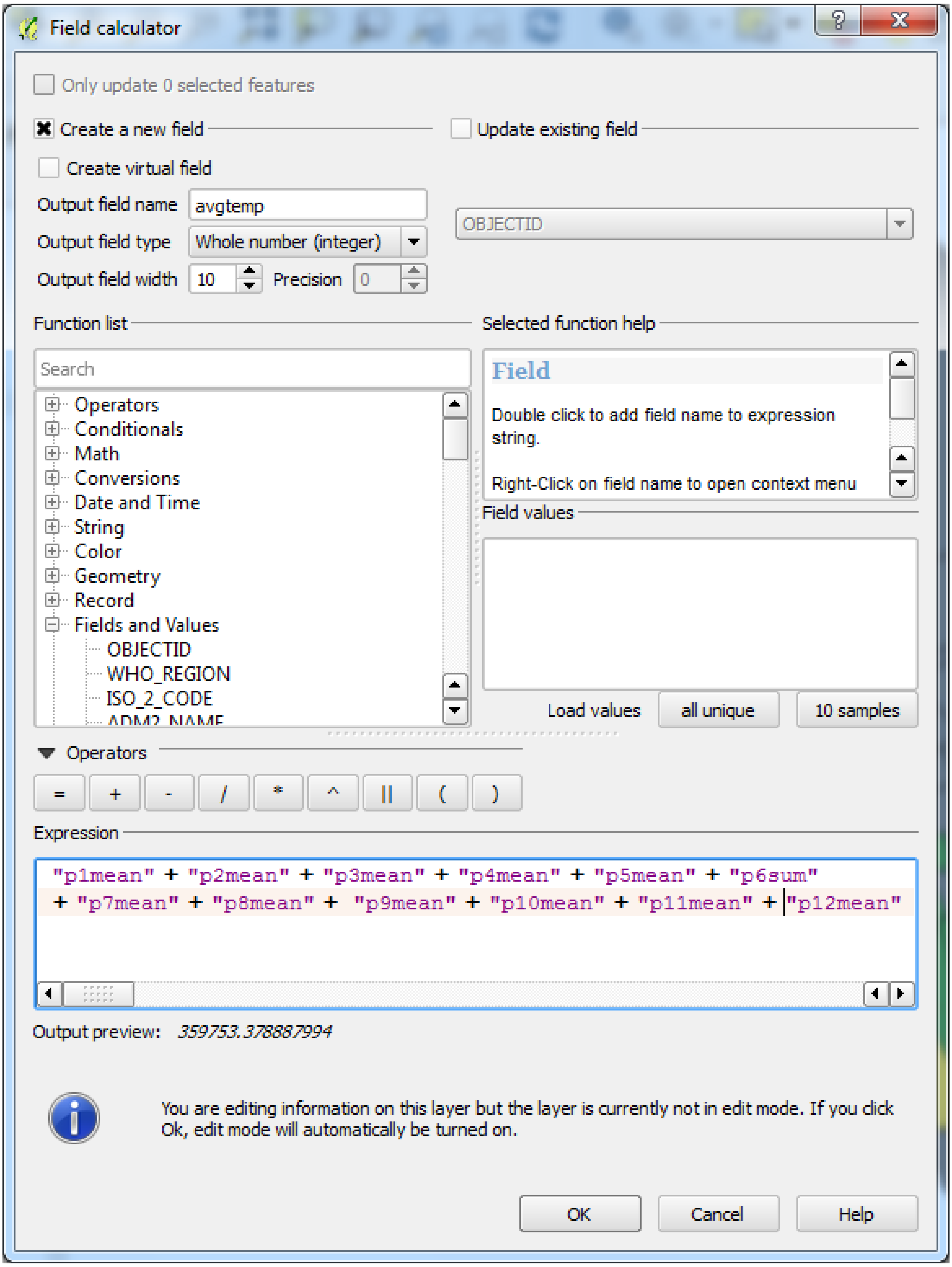


Figure 6: Calculations: Example of the average temperature from monthly means

It helped as well to compute as previously the mean of the monthly values obtained for temperatures with the zonal statistics tool *("Field calculator">"Addition operator" using the useful "Fields and Values"*/12). After all, the new attribute table of the first map was exported to Excel. The Excel table has been cleaned and the information on the predictors allowed preparing a definitive comma-separated values (CSV) file for issuing the environmental predictors maps.

For this part, it was also important to find data required as human predictors. The researches provided many set of data. The data on sanitation were not all consistent. We compared for instance the data on a Non-Governmental Organization[13] to the data in an official report[14]. After comparing the variables and the differences, the official data were finally chosen. The second step was to find the exact rate for the new counties since the report provided data based on old administrative boundaries. At this step, the overlaid of the old and new shapes at the administrative level 2 helped. It was identified the "mother county" from which the news are derived. Table 4 shows some details. We attributed to the new counties the same rate as the old county which generated them. As explained above, the final CSV files allow the visualization of the human predictors in Kenya.
The last step after finding and treating the predictor data was to run the prediction model. *"R"* is the software used. Prior to the study on Kenya, the useful scripts for the prediction have been written and used for other tests. They required as input two tables in the format of "*Text Tab delimited (*.txt)*". Therefore, the tables required for the prediction model have been prepared. They provided all information needed by the script. The script has been run. Two files came out at the end of the process. As output, they were under the same **.txt* format. The table generated as output is the file necessary for mapping the results. It was saved under Excel and CSV after treatment. Annexe A6 gives a brief overview of the script. The final map could afterwards be issued.

Table 4: Example of changes in counties

Old county	Become smaller News counties
BUSIA	Busia
	Teso
NAROK	Narok
	Trans-Mara
Parts of (KIAMBU & MIRANGA)	Thika
HOMA BAY	Homa Bay
	Rachuonyo
	Nyanza
	Suba
KISUMU	Kisumu
	Nyando
KITUI	Kitui
	Mwingi
BUNGOMA	Bungoma
	Mt. Elgon
MURANGA	Muranga
	Maragwa
KAKAMEGA	Kakamega
	Lugari
MIGORI	Migori
	Kuria
BARINGO	Baringo
	Koïbatek
	Bureti
ELGOYO-MARAKWET	Marakwet
	Keiyo
KISII	Kisii (+ a part of Nyamira)
	Gucha
SIAYA	Siaya
	Bondo

13 http://www.majidata.go.ke/county.php?MID=MTE=&SMID=Ng==

14 https://www.opendata.go.ke/Counties/Kenya-County-Fact-Sheets-Dec-2011/zn6m-25cf

6. Results: One manual, two parts, three maps

6.1- Part 1 of the manual: One map from three sets of data

Figure 7, Figure 8 and Figure 9 below show the prevalence of STHs with different sets of data. Figure 10 is the final map for all datasets combined. It shows the prevalence of the infection for all Kenya counties.

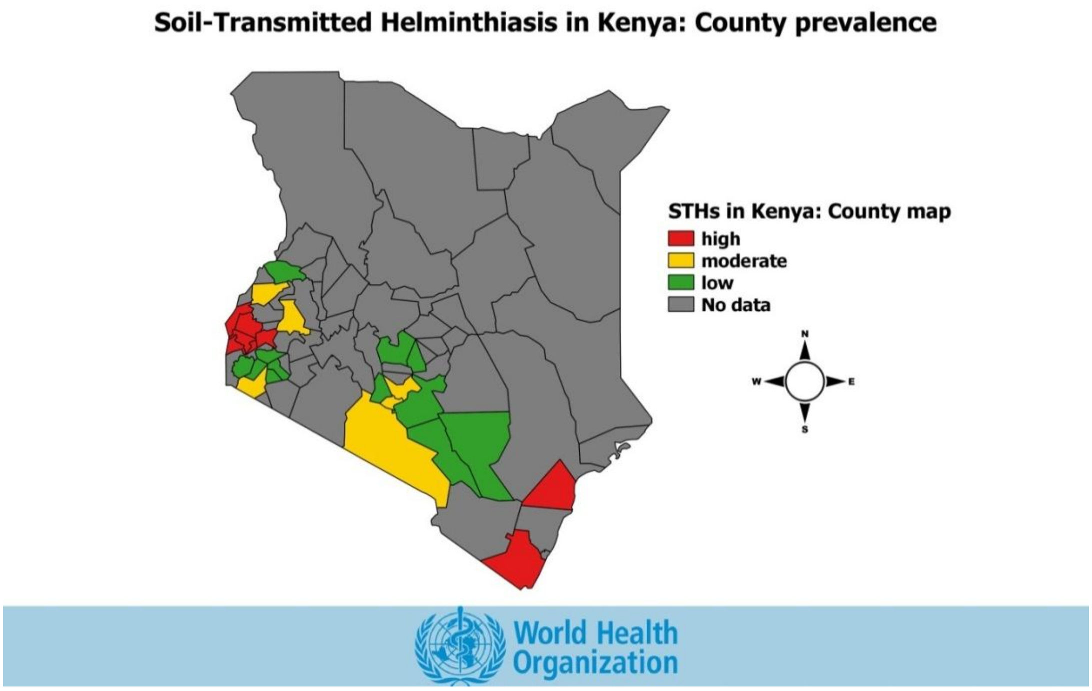


Figure 7: Prevalence in counties (Obtained directly from scientific publications, 1966-2014)

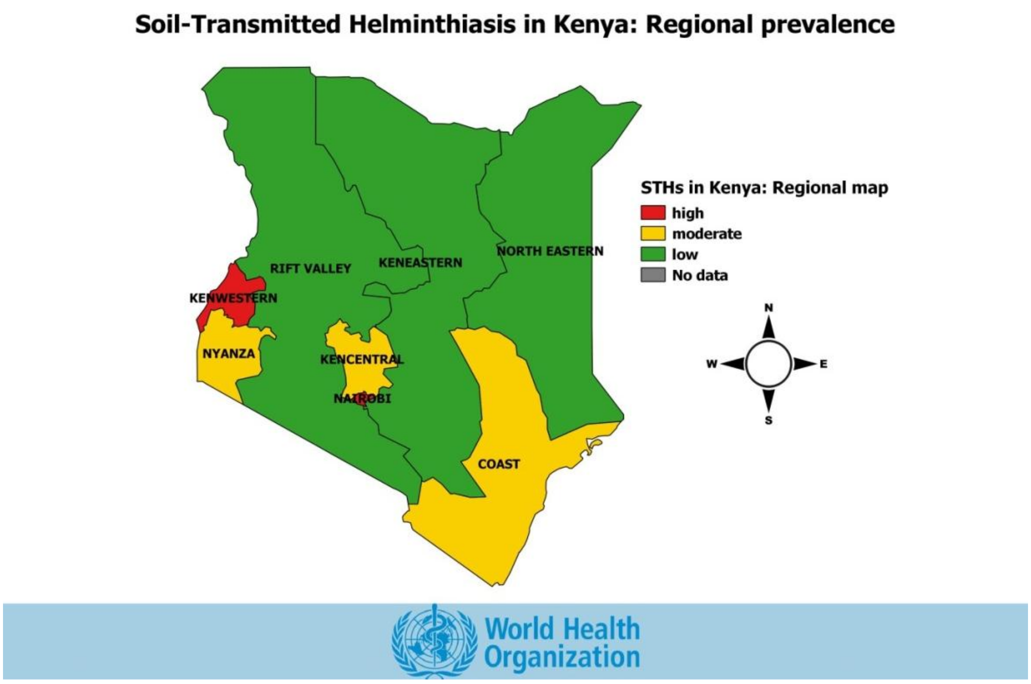


Figure 8: Regional prevalence (Pullan, et al., 2011)

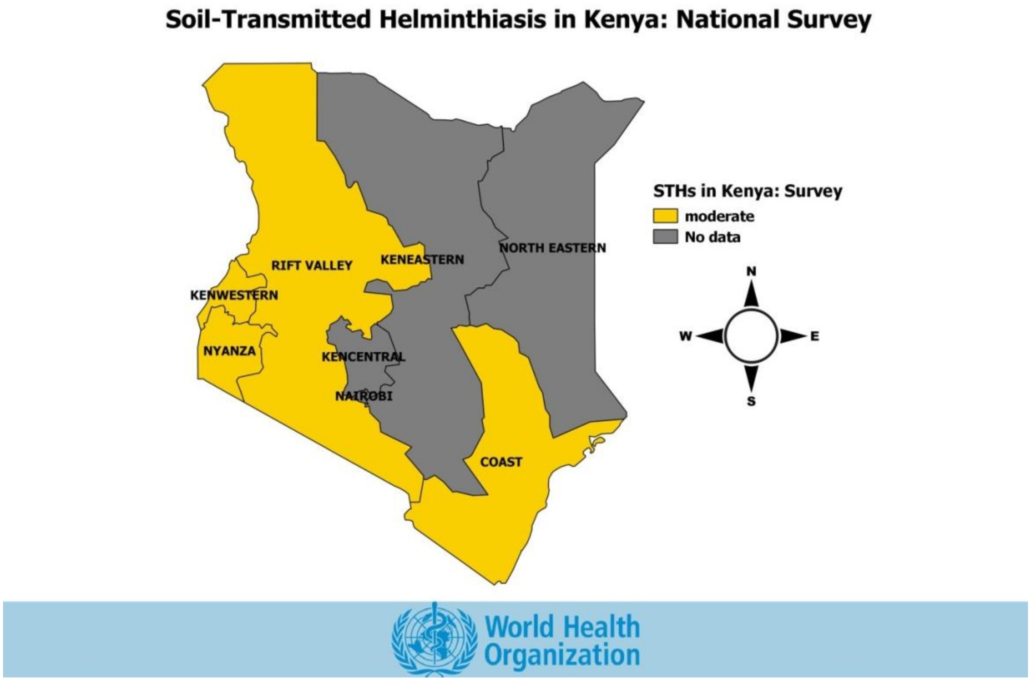


Figure 9: National survey area and prevalence (Mwandawiro, et al., 2013)

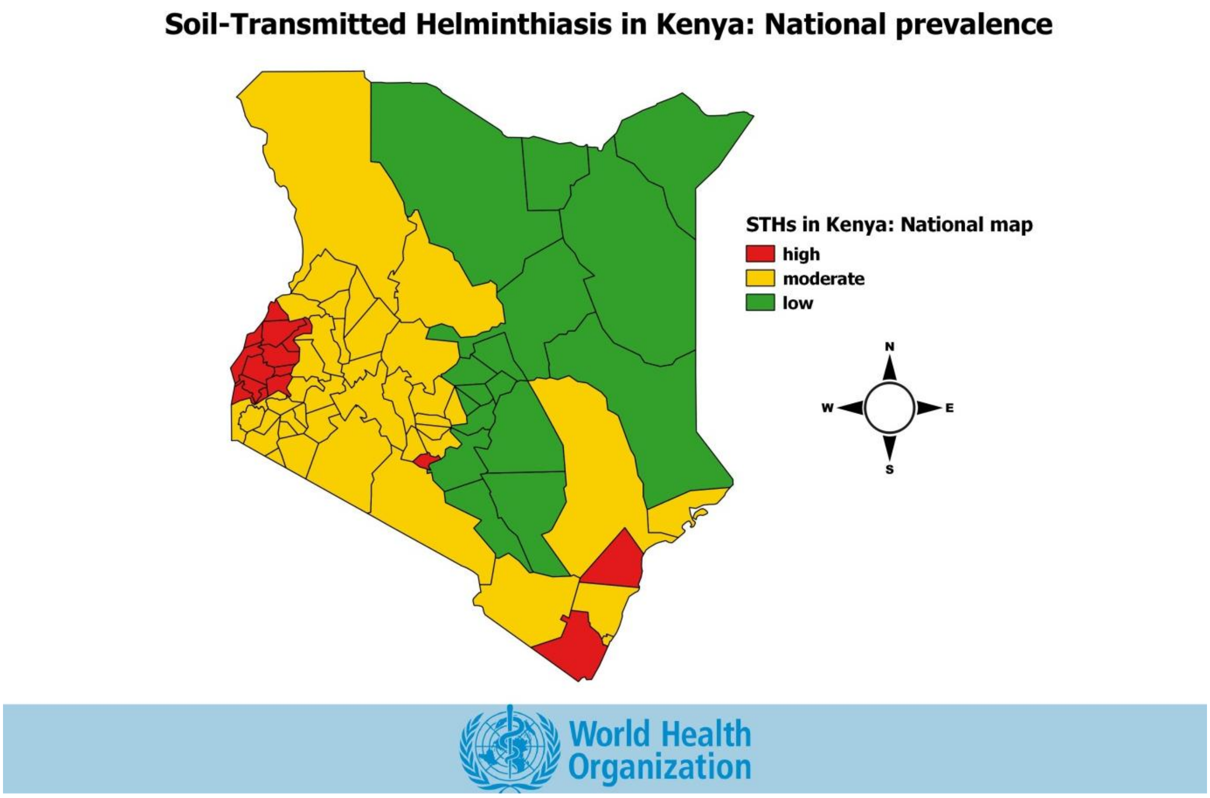


Figure 10: National prevalence (Worst case from all sources, 1966-2014)

6.2- Part 2 of the manual: Maps issued from the predictive data

The following maps are intermediate results. They help visualize all row data obtained and treated for the prediction purposes. They give therefore an overview on the ecological factors in Kenya and on some human factors. Figure 11, Figure 12, Figure 13, Figure 14, Figure 15, Figure 16 and Figure 17 below display the ecological predictors (average temperature, annual rainfall, and soil acidity), the population density and the human development predictors (literacy rate, poverty rate and sanitation rate) respectively.

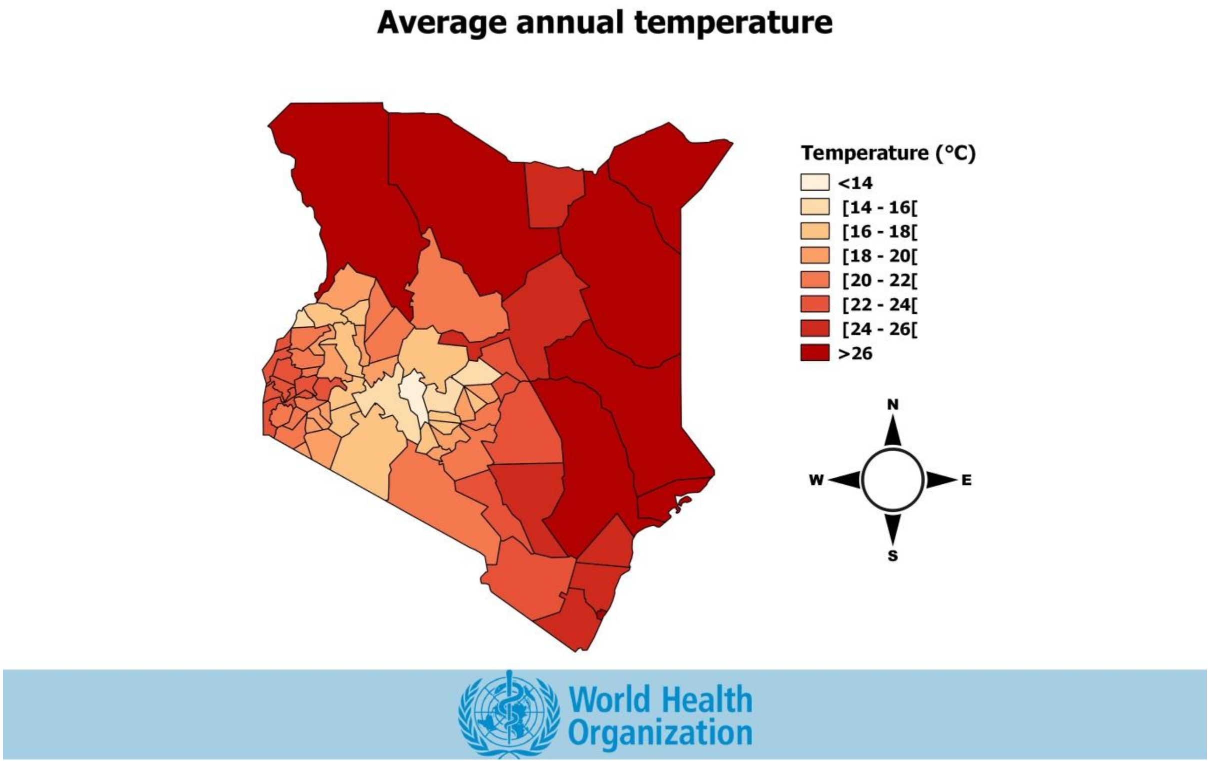


Figure 11: Temperature map (http://www.worlclim.org)

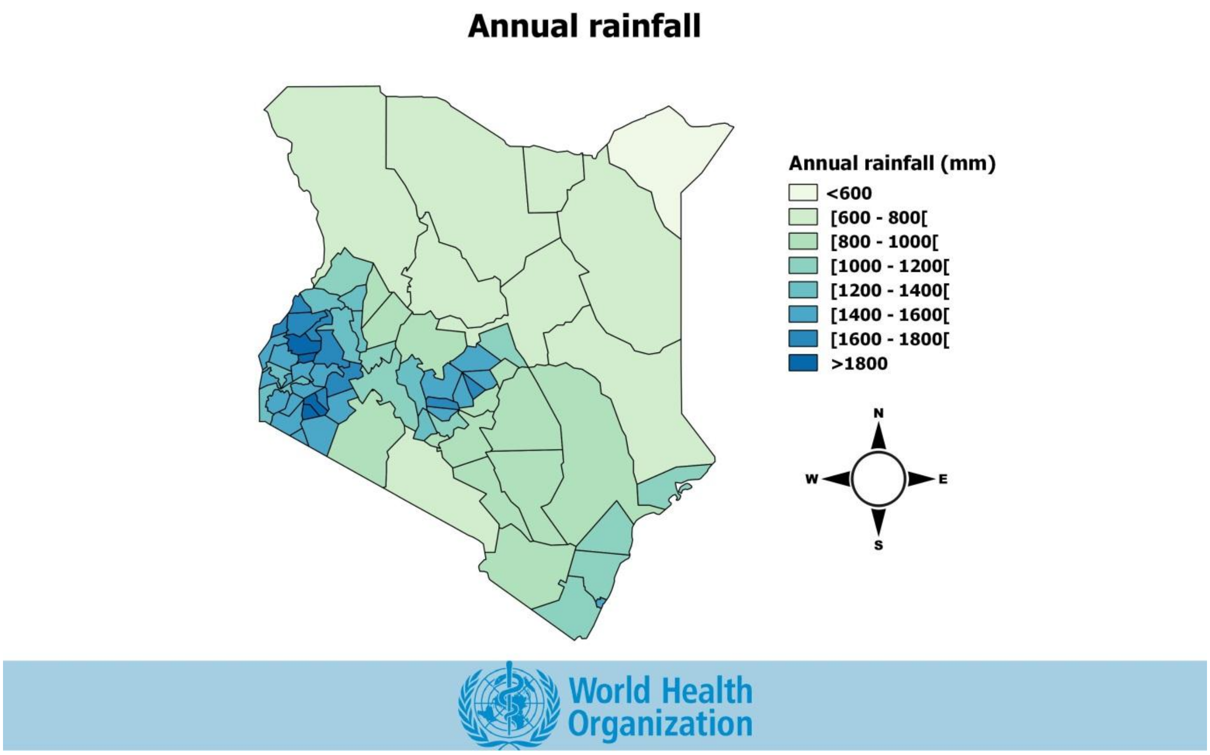


Figure 12: Rainfall map (http://www.worlclim.org)

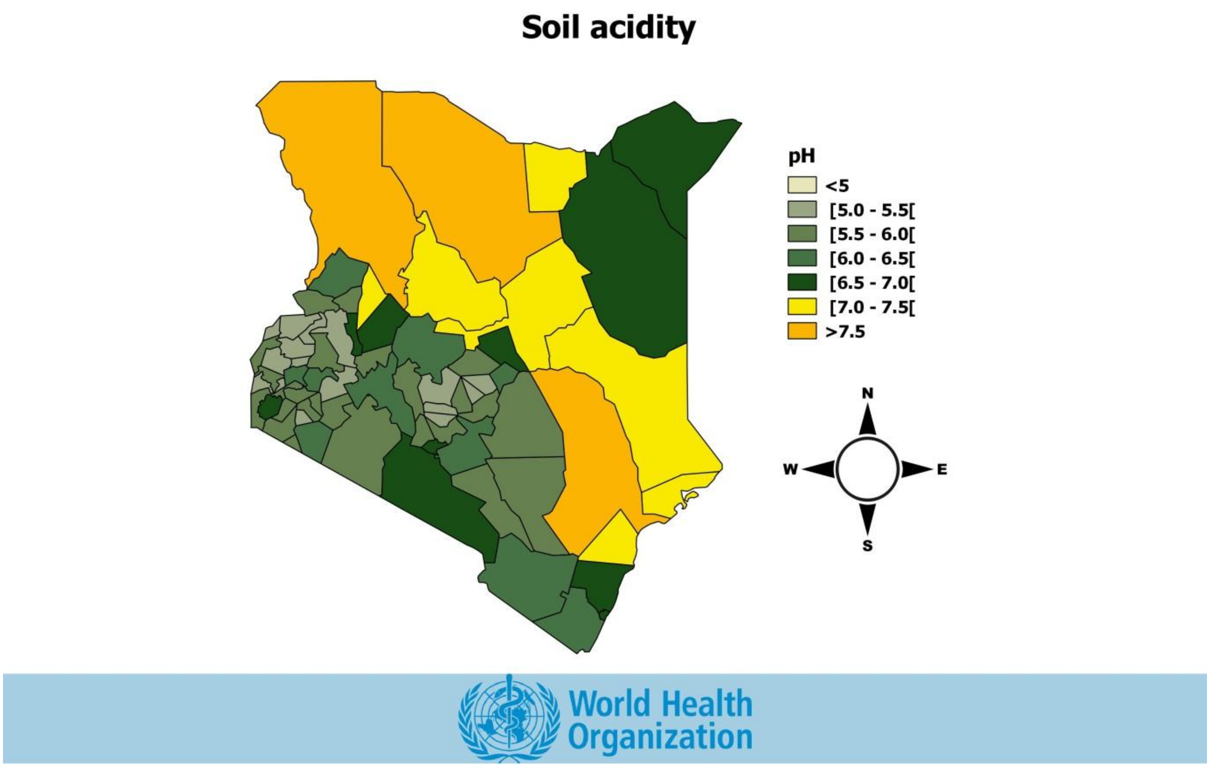


Figure 13: Soil acidity map (http://www.worldgrids.org)

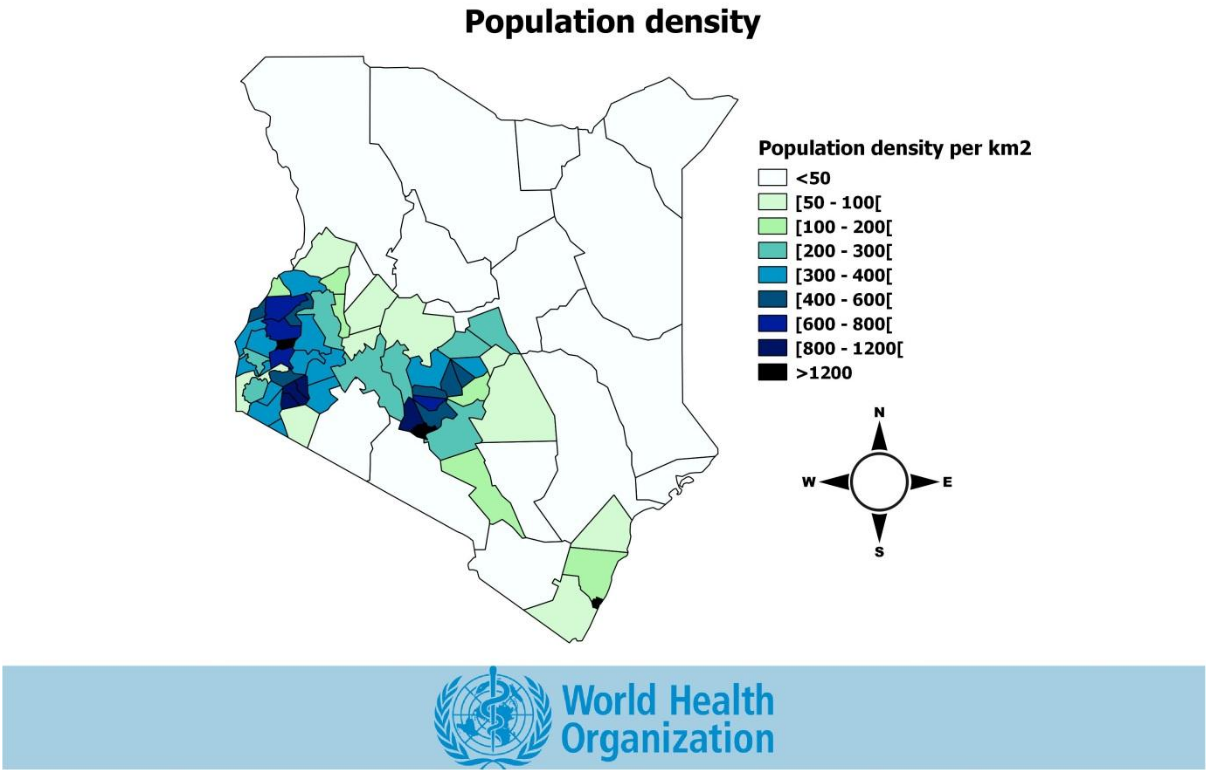


Figure 14: Population density map (www.afripop.org)

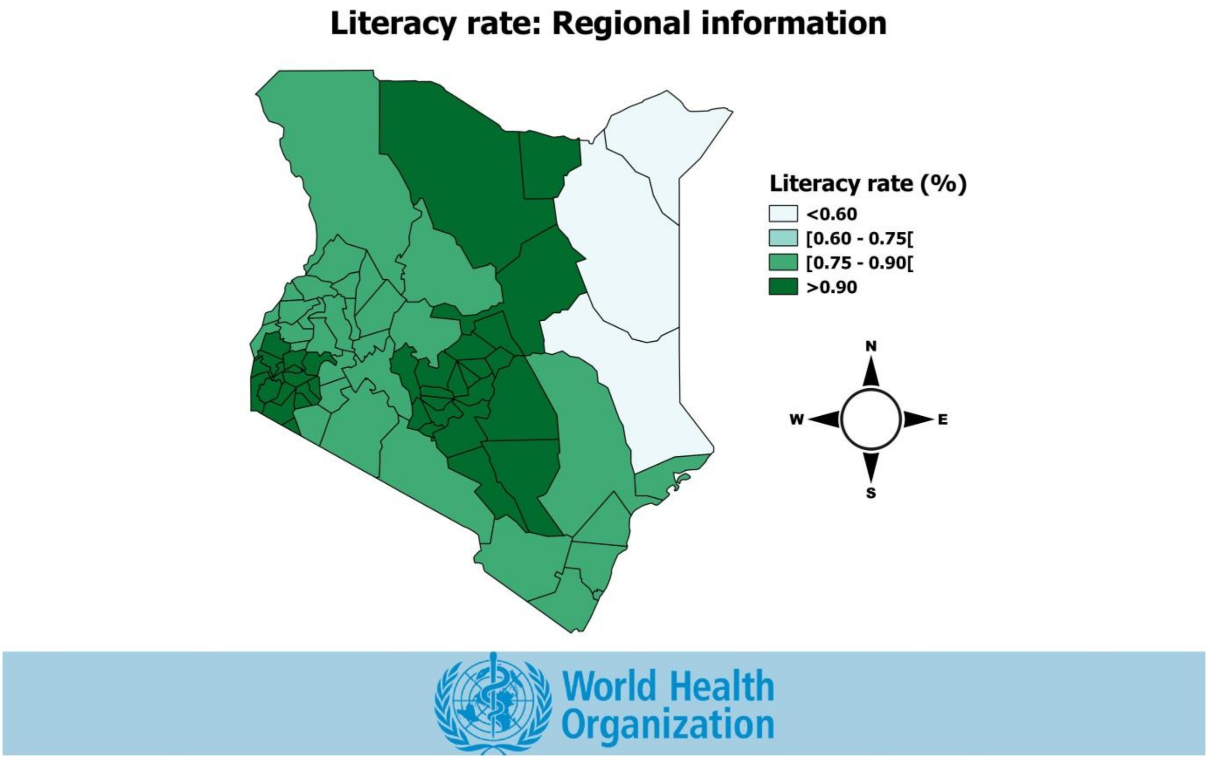
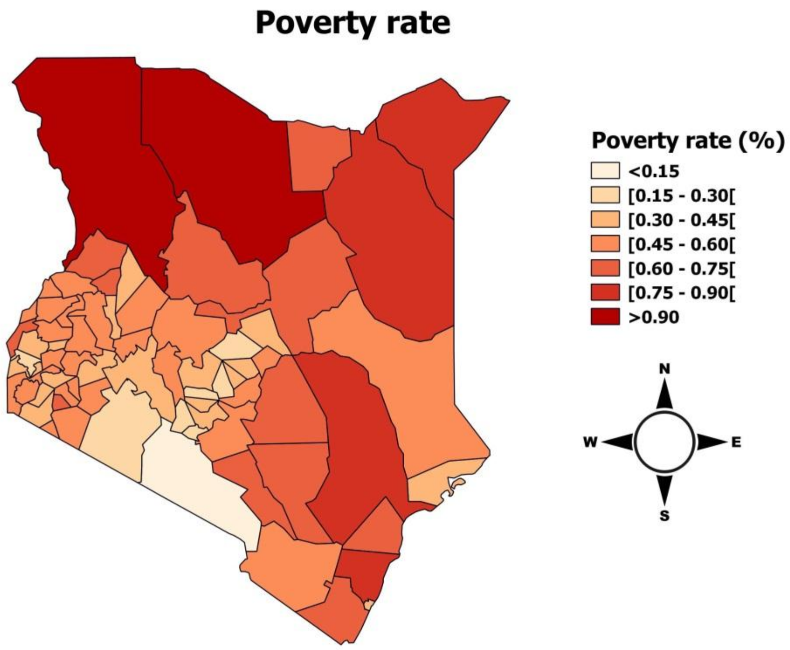
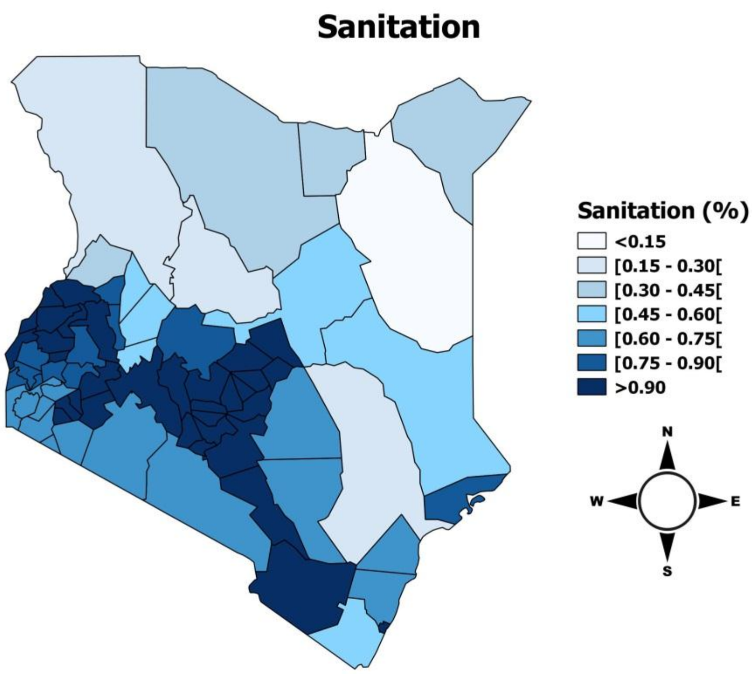


Figure 15: Literacy map (Kenya National Bureau of Statistics KNBS, 2010)



World Health Organization

Figure 16: Poverty map (https://www.opendata.go.ke)



World Health Organization

Figure 17: Improved sanitation map (Commission on Revenue and Allocation CRA, 2011)

6.3- Part 2 of the manual: Two final maps

Figure 18 and Figure 19 show the predicted STHs map and the intervention map. These two maps are after all the process the most useful for a decision making.

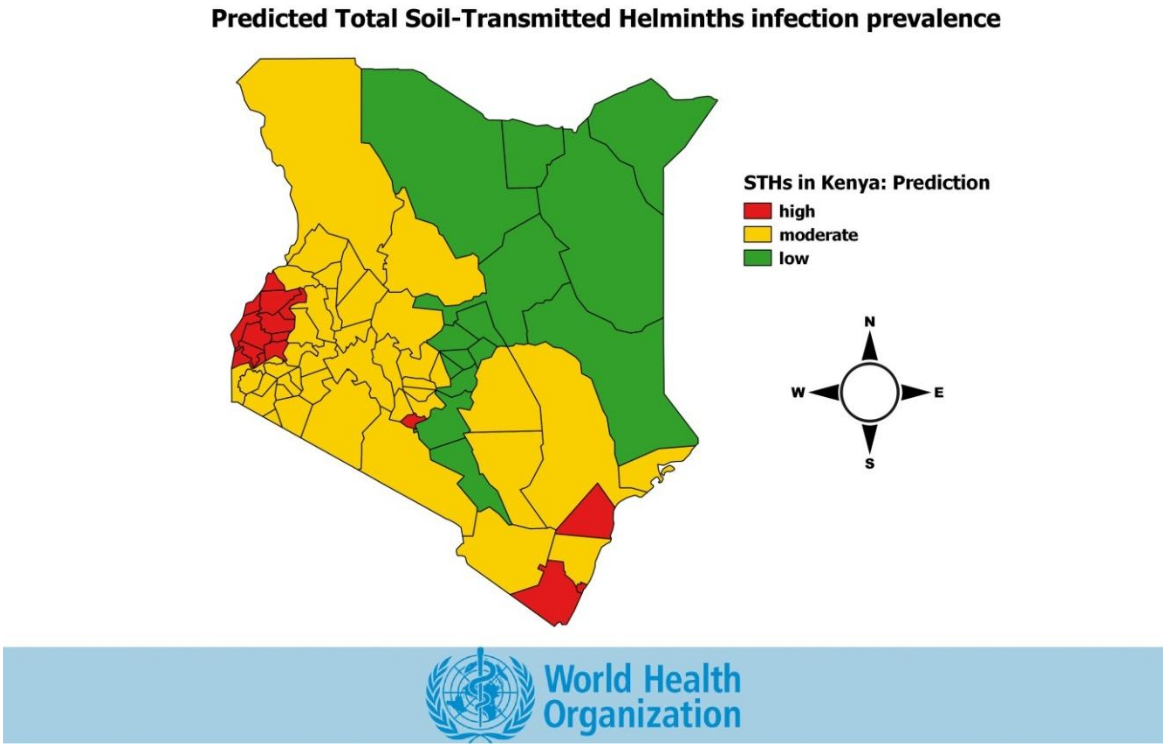


Figure 18: Predicted STHs map (All sources, 1966-2014)

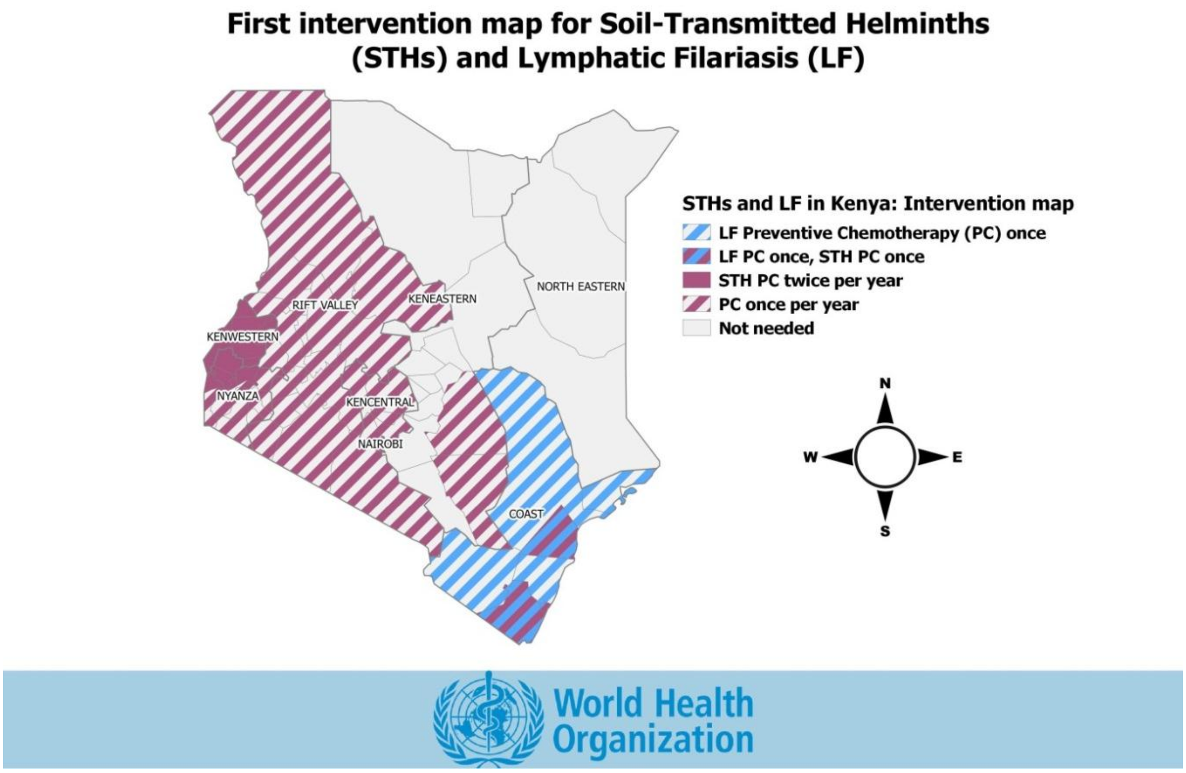


Figure 19: Intervention map (All sources, 1966-2014)

7. Discussion

7.1- One map from heterogeneous data

Before explaining the results, it is important to explain the production process of the map on Figure 10 despite the heterogeneity of data. At the end of the data collection, three categories of data have been obtained: historical data at the county level from 1966 to 2014, historical data at the regional level, the survey data published in 2013.

7.1.1- Historical data at the county level

The first map (Figure 7) shows the prevalence at the county level for all the areas on which we got data. The lack of historical data in all counties is obvious. The result covers twenty six counties out of sixty nine. Many reasons explain this result. For instance, no data was obtained for some counties. Certain titles of articles obtained on the PubMed database were available neither on internet nor in WHO library stacks. Many unpublished data were not obtained as well. Around forty reports listed were issued by the division and Vector Borne Diseases from 1975 to 2009 and some direct contributions by authors. Those reports were available neither on internet, nor in WHO library stacks.

7.1.2- Historical data at the provincial level

To partially address the limitation of the first map, the second map (Figure 8) has been issued at the regional level. The data were extracted from a study which compiled data from 1974 to 2009 (Pullan, et al., 2011). The map shows that there are three risk regions in Kenya: Western Kenya which has the highest risk. Nyanza, Central Kenya, Nairobi and the Coast present a moderate prevalence. I would like to emphasize that Nyanza and Western Kenya are two regions bordering the Victoria Lake. The coast region borders the see. The North Eastern region and Eastern region are riskless according to the data.

7.1.3- Survey data

The third map (Figure 9) shows simply the results of the national survey conducted in 2012 (Mwandawiro, et al., 2013). It should be pointed out the fact that this survey, called “National Survey” doesn’t cover the whole country as shown on the map. The authors explained a specific method of random sampling of 21,508 children from 200 schools in 20 counties from 4 provinces (Mwandawiro, et al., 2013). The results of this survey reveal that the infection is moderate everywhere.

7.1.4- One map from all data

The final map which represents an estimation of the prevalence of STHs at Kenya national level is the result of concatenation of all data,

- avoiding the data overlapping,

- considering the baseline case after removing data after a deworming,

- removing the data after the implementation of the program of LF treatment.

After the previous treatments, such map is obtained by building an Excel table with all dataset per category in columns. The statistic formula "Max (d1, d2, d3)" is used to compute prevalence per County; d1 for counties historical data, d2 for regional historical data and d3 for the survey data.

7.2- About the first results

Figure 10 is the final map at the end of the part 1. It shows the category prevalence of STHs in Kenya. The tendency is moderate or high prevalence of the infection in the counties bordering water bodies. This result confirms the fact that the prevalence of STHs can be high (above 50%) in water bodies area and in the major economics center with poor areas. The coastal counties: Kilifi (67%), Kwale (86%) present heavy prevalence rate while the National Survey's results wrote about a moderate infection: 24%. Lamu, Malindi (41% each) and the other coast counties, even with a moderate prevalence, show a higher difference compared with the national survey. This discrepancy must be pointed out. To explain this, our analyses came back on the LF data which revealed that all these coastal counties are endemic for LF whose PC is implemented in Kenya since 2002. Taking into account the result of the LF data helps to find out the real baseline prevalence of STHs in this region of Kenya. The result is not moderate as indicated by the survey but high. Making an average in the Coast region artificially smoothes internal variations. In the same vein something interesting occurred in the Western on Kenya. While the survey indicated moderate prevalence, our study shows high prevalence. This discrepancy is due to deworming occurred in Nyanza region. The deworming was certainly not taken into account by the survey. The Table 5 below shows the difference after we performed analyses by referring to the manual framework. After considering all the data and specifically the deworming, the difference is obvious.

Table 5: comparison of Survey and Baseline prior to deworming in some Nyanza counties

<table>
<tr><th colspan="2">Counties</th><th colspan="2">Prevalence (%)</th><th rowspan="2">Sources</th></tr>
<tr><th>Historical</th><th>Maintained / Become</th><th>Survey
(Mwandawiro, et al., 2013)</th><th>Baseline</th></tr>
<tr><td>Bungoma</td><td>Bungoma</td><td rowspan="4">42</td><td>81</td><td>(Assefa, et al., 2014)</td></tr>
<tr><td>Kisumu</td><td>Kisumu | Nyando | Kitui</td><td>77</td><td>(Verani, et al., 2011)</td></tr>
<tr><td>Migori</td><td>Migori</td><td>42</td><td>(Peterson, et al., 2011)</td></tr>
<tr><td>Siaya</td><td>Siaya | Bondo</td><td>66</td><td>(Mwinzy , et al., 2012)</td></tr>
</table>

The results in the western of Kenya part shows moderate prevalence for the counties in which data were obtained. It attests at least the expectation of moderate risk due to the presence of Victoria Lake. It is even confirmed by the high prevalence rate in the regional estimation of Pullan (2011): 80%. The same observation is made in Nairobi, biggest city of Kenya where the historical data revealed a moderate prevalence of 41%. This result is raised by Pullan 2011, confirming Nairobi as a potential area for STHs (76%). This results matches with the expectations.

Except the high risk zones analysed in the previous lines, all the other zones are moderate area zones. It confirms the fact that, although the survey didn't cover all the country, it is likely to have at least a moderate risk in all the endemic counties which are 66 (Mwandawiro, et al., 2013).

The Eastern part and North-Eastern part of Kenya are areas in which we obtained no data. It's about 16 counties. An affirmation made by Pullan (2011) identifies most of these counties as "biologic limit of transmission" of STHs. Regarding this affirmation, it is likely to consider the lack of data as normal. However, considering that only 4 counties in Kenya are not endemic, the best way to confirm is to make some sample for confirming the limitation of the transmission. Even if the density of these areas is not considerable, an infection can be transmitted through mobility. The risk of transmission in this last case is of course low. Only an in situ verification can confirm. In this area, the prevalence according to Pullan (2011) is very low: 13.3% and 0.002% respectively. This can be theoretically justified by the highest temperature and the small amount of inhabitants confirmed by the intermediate maps. For instance, Isiolo, Marsabit and Wajir are characterized by the highest temperatures and the smallest population size (Annexe A7 for identification of the counties)

7.3- Mapping the scores

Although it is possible to calculate national average prevalence based on the county prevalence (Figure 7), it is clear that such figures can be potentially misleading. In particular, in certain counties, very few prevalence surveys have been undertaken, and these have typically chosen and examined areas of known high transmission. This suggests that such extrapolated national infection prevalences are more often a reflexion of the numbers of surveys conducted and their location, rather than a reliable indication of prevalence. This can illustrate a potential inaccuracy of prevalence estimates based on only a few studies within a country and then extrapolated to the country as a whole, as this belies the geographical heterogeneity within countries. However, the present data have been mapped within defined administrative boundaries because control approaches are implemented at the district level even if it is unclear whether district units are an appropriate unit of comparison, because such study belies the finer-scale spatial heterogeneity of infection. Counties vary in size, in geography and in demography, so a more useful unit of analysis might be based on the number of people in a given area, or based on ecological zones without regard to administrative boundaries. According to this last vision, the second part of the manual takes into account the population size and the ecological data in the area in order to be more accurate. Clearly, there is a trade-off between what is desirable and what is realistic. Despite these limitations, the current data provide a crude estimate of overall disease burden for Kenya.

7.4- Limitations of the methodology

- It was not possible to obtain many unpublished data listed on the GAHI website. All our attempts to get them were unsuccessful.
- The changes occurred in Kenya regarding administrative boundaries are a parameter to consider as limits of analyses. The administrative decoupage of Kenya moves from 47 districts to 69 counties. The shapefiles of the previous districts on Global admin[15] and ESRI[16] websites obtained and compared with shapefile from WHO database confirmed the differences. When extracted and compared, the attribute tables of the shapes reveal even more information. The new decoupage accordingly introduces some new names that were not in the previous district level. Some names in the previous disappeared (Annexe A5). Considering that the epidemiological database we built start in 1966 it's possible that some information may be attributed to another land with the new changes. This therefore should not affect deeply the results in the sense that the geographical points are the same.
- LF is endemic only in one Region of Kenya: Coast region. The data related to the coastal part obtained from the data manager in WHO are available on a lower level than level 2 (Annexe

[15] http://www.gadm.org/ or
[16] http://www.arcgis.com/home/item.html?id=5f83ca29e5b849b8b05bc0b281ae27bc

A5). Furthermore, It started in the same county for instance in 2002 for some regions and in 2011 for other regions. Therefore, an idea was to perform the analyses on level 3 or level 4. But to be able to do that, the corresponding shapefiles are needed. All our attempts to get the shapefiles at this level failed. The analyses could not be done with shapes which do not match the current boundaries. We stayed at the county level for this part. It was a barrier to our will to go more deeply in the analysis.

7.5- Introducing new ideas

The map of Figure 20 is the trigger of a new reflexion. After a team reflexion, the decision is to add the population density data as new predictor considering the evidence that STH needs humans host in area with ecological conditions favourable to their maturation. The new model gives the result of Figure 18, more consistent and realistic. This certitude needs explanation. The result obtained after running the prediction model prepared prior to this study shows a huge discrepancy between the mean normal expectations and the map observations. For example in the North eastern Province, Garissa, Mandera and Wajir are the counties with a very low prevalence of 0,002%, prevalence totally down under the transmission break point (Annexe 7 for identification of counties). The northern and the eastern regions are one of the five main naturals regions in Kenya. They are arid plateaux and *"consists of a vast but largely uninhabited region of desert and semi-desert in the east and the north of the country"* (Hall, et al., 1982). These regions, according to the same authors, because of the complicated and diversified physical environments found in any country, are ecologically different from the rest of the country. Therefore, it was not realistic that the model takes a so great step, making those counties area of high risk (from 0,002% to 50% and more). The last affirmation can be supported by the concrete fact showed by data that densities in this zone are the lowest in the country (Figure 14); temperatures are the highest (Figure 11). Furthermore, the result is totally different when compared with the map of Figure 10. Once again, we mention here that the data obtained in the case of Kenya are not homogeneous as explained above. Considering the fact that the model used the number of population tested and the number of positive at the county level, the model has been run by using the data that have generated the map of Figure 7 since it is a set of data at county level. Most of these data were taken in the four other mains naturals regions. In the northern and eastern parts of the country, no survey was available at the county level. Therefore no information was available in the input table on a number of persons tested in the corresponding counties, what could affect the formula built in the modelling process. The specificities of the northern part stimulate therefore the choice of the population density as predictor. The result obtained after introducing population density in the model (Figure 18) is more consistent and realistic. For more insurance and proof of robustness, we compare this result to the map of Figure 10: National prevalence (Worst case from all sources, 1966-2014)Figure 10. The prevalence categories are similar in all the country except in tree coast counties: Kitui, Mwingi and Mombasa. The differences in these tree cases can be perfectly accepted in the sense that in Kitui (14%) and Mwingi (13%), the model passes from low to moderate (22% and 20% respectively). In Mombasa it runs from moderate (41%) to high. The last result shows the necessity to consider the possible impact of environmental changes in the countries in order to provide the necessary input to the model for expecting more realistic predictions.

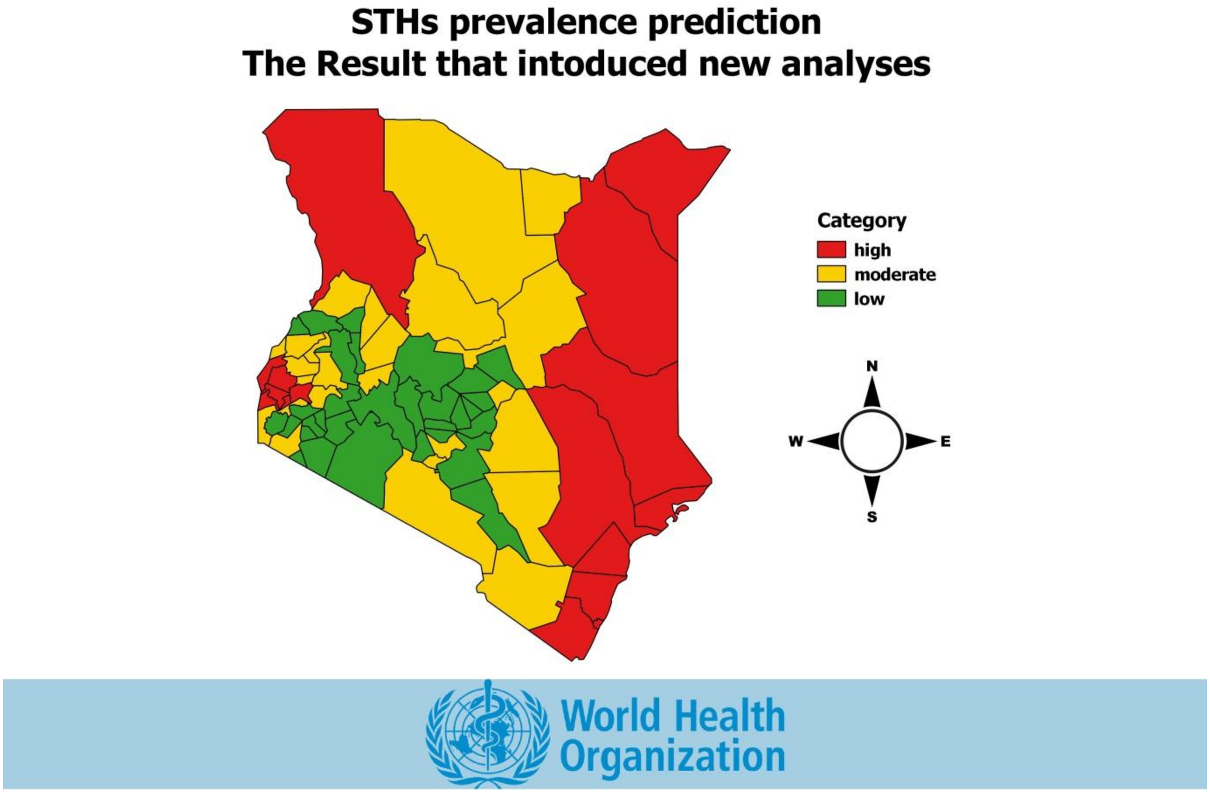


Figure 20: Result of the first prediction

7.6- Using the developed manual as framework

The manual used for reference for this study is undoubtedly the result of a long reflexion. It is clearly felt that it is the reflexion of combining long experiences of actors in the public health field and in the mapping field. The category assessment remains within WHO framework.

In charge of the development of maps on Kenya, I can confirm that with the exception on the limitations mentioned above, it is possible to have a result on the first part of the manual without being an expert in the mapping field. The basic knowledge in the computer field can help if the person in charge is motivated to follow strictly the steps of the manual. It is all the more feasible in the sense that it is proposed in the manual the initiation to the mapping software “QGIS”. However, the second part needs the intervention of a specialist in an appropriate field, who can understand the reflexion behind the prediction model in order to adapt it if needed. This is important because geographic areas can be deeply different on the environmental and on the human point of view.

8. Conclusion

As is the case in many fields, mapping is a tool that can help to manage STHs infection. It is a tool that can provide accurate information on the prevalence. It can guide to manage more efficiently the deworming program in a country. The experience on Kenya complements the maps already made in other endemic countries: Indonesia, India and Ivory-Coast. It reconfirms that the development of such a manual, subject to the availability of reliable data, is considerably useful in several respects. More than a simple manual, it can be a useful tool in the sense that:

- It may be “a less time consuming tool”: countries managers and policy makers are helped to gain in time wasting. A complete national survey can be avoided.

- It may be “a money saver tool”: a part of what might have been a budget in the massive deworming (without an overview on the distribution of the infection, its prevalence in all areas and the approximate quantity of drug needed) or national survey can be saved.

This exercise is in the health field. One reflexion possible in term of improving the reliability and the validity of the maps findings is to frame what is the threshold for considering the quantity of data available as reliable for the mapping and the trustful for a deworming action.

9. Internship in WHO

From August 25th to February 28th, I have spent twenty eight weeks in WHO. On many points, I made some new experiences. Evidence shows that in most perfect world, everything cannot be perfect. So comparing deception points to satisfaction ones, I can honestly affirm that my stay in WHO was a success thanks to many factors as, the capacity to work in a team, the ability to work independently, the work environment, the actors of the work environment…

9.1- Conditions of internship

Prior to my duty, during the brainstorming phase, I followed the first many presentations aiming to give a better understanding on the NTDs, the four helminth infections and the framework of the manual. It was the trigger of my will to learn more on parasitology in order not to map an infection I didn't know about, but rather to understand deeply this infection and its operation.

English as mandatory language: It was the second time I worked at a place where English is mandatory. I forgot a lot since my internship in International Institute of Tropical Agriculture (IITA) in 2008. The opportunity to speak, read and write in English was a great challenge. It has enhanced my self confidence in the sense that I can apply to post even in Anglophone countries in Africa or in the world in the case the post fits with my skills and desire.

The ability to work independently: I could not have a result without being able to set my goals, choose my methods, find ways to issue convincing conclusions from the data. Beyond this last point I emphasize that "One manual, two parts, tree maps" is a simple manual, clearly declined and well detailed. Meanwhile, to be able to follow it requires from the user some abilities in computer manipulation and the confidence in the using of the mapping tools. This work could not be achieved without basics knowledge in data treatment regarding software like Excel, Access, R, GIS. I should mention that prior to the start of the internship, I have never used "QGIS" but the learning of "QGIS" could not be easy if I was not prepared with software like GIMP, ArcGIS and others.

The work environment: I got all the equipment to work decently. Everything was prepared in order that I could feel comfortable and generate the best output possible. This is another point very important that gave motivation and energy. For instance I started with a one screen desk not really performant for this research process. That was comprehensible regarding the need. Meanwhile, as soon as a screen was available, the supervisor proposed me to supplement the material in order to work efficiently.

9.2- Personal feelings

The actors of the work environment and the team work: I feel so good with my colleagues and my supervisors. They made me feel so at ease that I had no pain to communicate freely. Generally I do not talk too much. But because of my sensation of ease, I felt so free that I informed the team about everything: from the little difficulties to the serious troubles during the process. Every time I have submitted a preoccupation, the supervisors showed no hesitation to come to my assistance. That was psychologically an assurance and a motivation to find the resources in order to perform the best analysis I could for the satisfaction of the team. I should mention specifically their efficient presence in finding solutions to all preoccupations that could be for me a deception or an obstacle. I relate here for instance two small stories. During the period of extension of my stay, it happened that I lost the access to my computer. I stayed two weeks without computer. Of course this was the fault of nobody but an automatic system access problem. During this period, I felt uncomfortable to go at work without being able to work. I find an intermediate solution: going to work with my personal computer. But my supervisor proposed me to work from home till the problem could be solved. I did not stay at home but I thought this was the proof that I was in an attentive team that cares about my feelings. It gives me more energy to continue the work with my intermediate solution. Prior to the last story, it happened that my first computer crashed. They did not give up in following with attention the situation till I got a new computer. I cannot relate here many stories but I will conclude saying this: “When I’m in a team where there is attention, small disappointments or misfortunes do not destroy the joy I get from the good vibe around”. If one asks to me: “If given, could you start again from the beginning?”, my answer would be: “Yes if it is the same team”.

Multiculturalism: On a social point of view, twenty eight weeks are a long time for new experiences. I met people from different countries and cultures. It could be strange the first days if one was not aware. However WHO, apart from the work for human health in the world, is as well a place of crossroads, of culture junction where I worked, shared with different people. Thus I discovered many languages; expand my horizons, smile to other realities. This experience gave me more understanding on peoples, their cultures and the world. So speaking, after productivity, fraternity was my second goal.

Bibliography

Assefa, L. M., Crellen, T., Kepha, S., Kihara, J. H., Njenga, S. M., Pullan, R. L., et al. (2014). *Diagnostic accuracy and cost-effectiveness of altenative methods for detection of Soil-Transmitted Helmints in a post-treatment setting in Western kenya.* PLOS Neglected Tropical Diseases.

Bethony, J., Brooker, S., Alboniko, M., Geiger, S., Loukas, A., Diermet, D., et al. (2006). *Soil-transmitted helminth infections: ascariasis, trichuriasis and hookworm.* Lancet 2006 ; 367: 1521–32.

Brooker, S., Rowlands, M., Savioli, L., & Bundy, D. (2000). *Towards an Atlas of Human Helminth Infection in sub-Saharan Africa: The Use of Geographical Information Systems (GIS).* Parasitology Today, vol. 16, no. 7.

Commission on Revenue and Allocation CRA, K. (2011). *Kenya County fact sheets.*

de Silda, N., & Hall, A. (2010). *Using the Prevalence of Individual Species of Intestinal Nematode Worms to Estimate the Combined Prevalence of Any Species.* www.plosntds.org, Volume 4 | Issue 4 | e655.

Hall, A., Latham, M., Crompton, D., Stephenson, L., & Wolgemuth, J. (1982). *Intestinal parasitic infections of men in four regions of rural Kenya.* Transactions of the Royal Society of Tropical Medecine and Hygiene.

Independant Electoral and Boundaries Commission, K. (2012). *Preliminary report on the first review relating to the delimitation of boundaries of constituencies and wards.*

Kenya National Bureau of Statistics KNBS, K. (2010). *Kenya Demographic and Health Survey 2008-2009.*

Ministrty of health and sanitation Kenya. (2011). *National Multi - Year Strategic Plan for control of Neglected Tropical Diseases.*

Ministry of Public Health and Sanitation, K. (2011). *National Multi - Year Strategic Plan for control of Neglected Tropical Diseases.*

Montresor, A., Awasthi, S., & Crompton, D. (2003). *Use of benzimidazoles in children younger than 24 months for.* Acta Tropica 86 (2003) 223-232.

Mwandawiro, C., Birgit, N., Kihara, J. H., Ozier, O., Mukoko, D., Mwanje, M. T., et al. (2013). Monitoring and evaluating the impact of national school-based deworming in Kenya: study design and baseline results. *Parasites & Vectors*.

Mwinzy , P., Montgomery, S., Owaga, C., Mwanje, M., Muok, E., Ayisi, J., et al. (2012). Integrated community-directed intervention for Schistosomiasis and Soil-Transmitted Helminths in Western Kenya - a pilot study. *Parasites & Vectors*.

Peterson, L., Ondiek, M., Oludhe, D., Naul, B., & Vermund, S. (2011). Effecttiveness of a school-based deworming campaign in Rural Kenya. *Journal of Tropical Pediatrics*.

Pullan, R., Gething, W., Smith, J., Mwandawiro, C., Sturrock, J., Gitonga, C., et al. (2011). *Spatial Modelling of Soil-Transmitted Helminth Infections in Kenya: A Disease Control Planning Tool.* www.plosntds.org; February 2011 | Volume 5 | Issue 2 | e958.

Verani, J. R., Abudho, B., Montgomery, S. P., Mwinzy, P. N., Shane, H. L., Butler, S., et al. (2011). Schistosomiasis among young children in Usoma, Kenya. *American Journal of Tropical Medecine and Hygiene*, 787-791.

WHO. (2011). *Helminth control in School-age children: a guide for managers of control programmes, 2nd edition.* Geneva.

WHO. (2012). *Eliminating Soil-Transmitted Helminthiases as a Public Health Problem in Children: Progress Report 2001-2010 and Strategic Plan 2011-2020 .*

World Bank. (1993). *World Developement report: Investing in Health.* Oxford University Press, London and New York.

World Health Assembly Resolution. (2002). *Schistosomias ans soil transmitted helminth.* WHO, WHA54.19.

Annexes

A1- Candidate countries

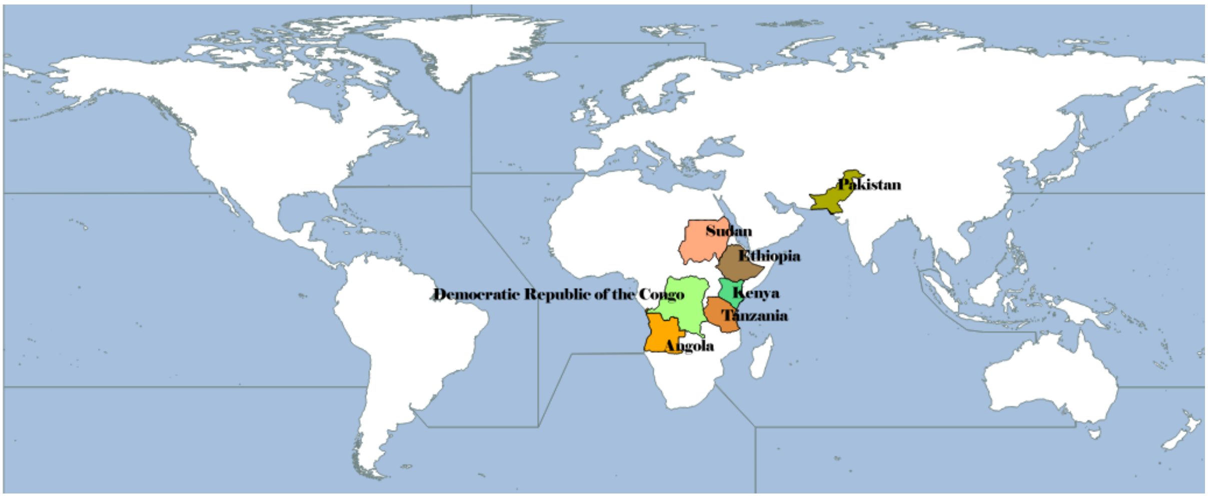


Figure 21: Candidate countries for the STHs mapping process.

A2- Endnote interface

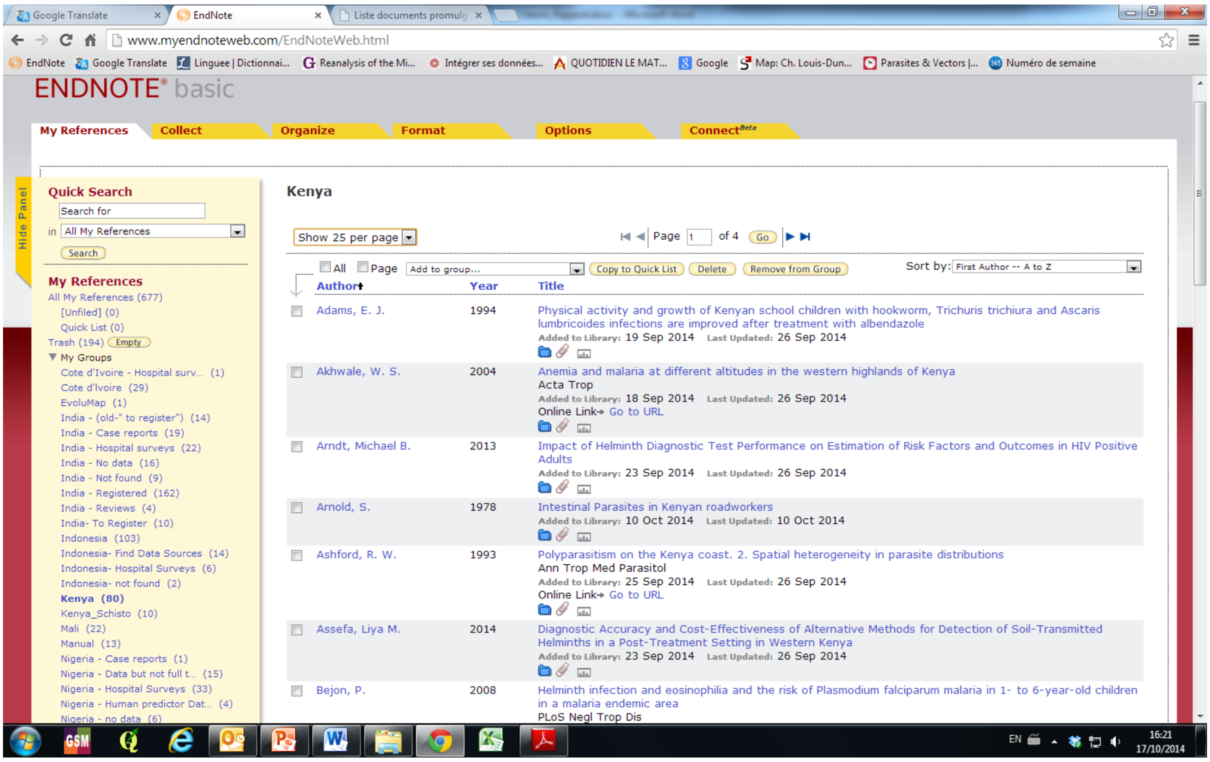


Figure 22: Endnote online data store

A3- Epidemiological survey stored on STH and Schisto.

Name	Date modified	Type	Size
WHOAfrica_2013.pdf	18/09/2014 11:25	Adobe Acrobat D...	2,621 KB
Wekesa_2014.pdf	23/09/2014 11:15	Adobe Acrobat D...	500 KB
Walson_2010.pdf	22/09/2014 13:54	Adobe Acrobat D...	439 KB
Verani_2011.pdf	22/09/2014 11:15	Adobe Acrobat D...	513 KB
Thiong'o_2001.pdf	22/09/2014 14:01	Adobe Acrobat D...	351 KB
Suchdev_2014.pdf	17/09/2014 15:49	Adobe Acrobat D...	475 KB
Sturrock_2010.pdf	17/09/2014 15:47	Adobe Acrobat D...	930 KB
Sturrock_1996.pdf	25/09/2014 11:44	Adobe Acrobat D...	1,112 KB
Stephenson_1993_2.pdf	26/09/2014 12:03	Adobe Acrobat D...	423 KB
Stephenson_1993.pdf	23/09/2014 09:56	Adobe Acrobat D...	1,057 KB
Stephenson_1989_3.pdf	19/09/2014 14:25	Adobe Acrobat D...	1,178 KB

Figure 23: Historical epidemiological survey stored by Author

KE0001_Rijpstra_1975.pdf	25/09/2014 11:40	Adobe Acrobat D...	1,263 KB
KE0002_Stephenson_1979.pdf	25/09/2014 12:05	Adobe Acrobat D...	1,095 KB
KE0003_Stephenson_1980.pdf	23/09/2014 10:39	Adobe Acrobat D...	85 KB
KE0004_Stephenson_1989.pdf	19/09/2014 14:25	Adobe Acrobat D...	1,178 KB
KE0005_Stephenson_1989.pdf	19/09/2014 14:24	Adobe Acrobat D...	1,047 KB
KE0006_Stephenson_1989.pdf	19/09/2014 14:20	Adobe Acrobat D...	1,329 KB
KE0007_Chunge_1991.pdf	25/09/2014 11:22	Adobe Acrobat D...	3,053 KB
KE0008_Booth_1992.pdf	22/09/2014 16:22	Adobe Acrobat D...	529 KB
KE0009_Ashford_1993.pdf	25/09/2014 11:09	Adobe Acrobat D...	6,465 KB
KE0010_Stephenson_1993.pdf	23/09/2014 09:56	Adobe Acrobat D...	1,057 KB
KE0011_Stephenson_1993.pdf	26/09/2014 12:03	Adobe Acrobat D...	423 KB
KE0012_Adams_1994.pdf	26/09/2014 12:42	Adobe Acrobat D...	1,476 KB

Figure 24: Historical epidemiological survey stored by label

A4- Access database formulary

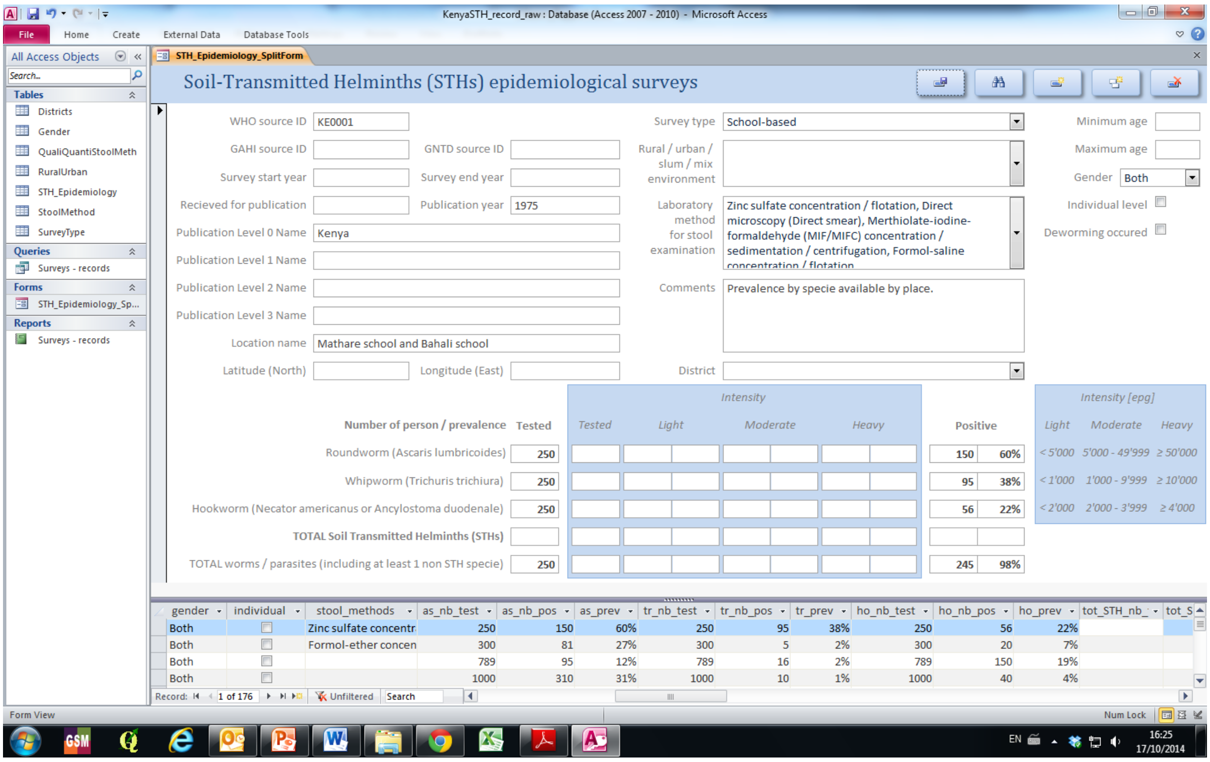


Figure 25: Formulary for data storing

A5- From 47 Districts to 69 counties: better understanding

Table 6: Old districts and new counties

GADMIN			WHO		
NAME_1	VARNAME_2	NAME_2	ADM2_NAME	ADM1_NAME	Matching
Central		Kiambu	KIAMBU	KENCENTRAL	Central
Central		Kirinyaga	KIRINYAGA	KENCENTRAL	Central
Central		Machakos	MARAGWA	KENCENTRAL	Central
Central		Murang'a	MURANGA	KENCENTRAL	Central
Central		Nyandarua	NYANDARUA	KENCENTRAL	Central
Central		Nyeri	NYERI	KENCENTRAL	Central
Coast		Kilifi	THIKA	KENCENTRAL	Central
Coast		Kwale	KILIFI	COAST	Coast
Coast		Lamu	KWALE	COAST	Coast
Coast		Mombasa	LAMU	COAST	Coast
Coast	Taita	Taita Taveta	MALINDI	COAST	Coast
Coast		Tana River	MOMBASA	COAST	Coast
Eastern		Embu	TAITA TAVETA	COAST	Coast
Eastern		Isiolo	TANA RIVER	COAST	Coast
Eastern		Kitui	CENTRAL MERU	KENEASTERN	Eastern
Eastern		Machakos	EMBU	KENEASTERN	Eastern
Eastern		Makueni	ISIOLO	KENEASTERN	Eastern
Eastern		Marsabit	KITUI	KENEASTERN	Eastern
Eastern		Meru	MACHAKOS	KENEASTERN	Eastern
Eastern		Nithi	MAKUENI	KENEASTERN	Eastern
Eastern		Wajir	MARSABIT	KENEASTERN	Eastern
Nairobi		Nairobi	MBEERE	KENEASTERN	Eastern
North-Eastern		Garissa	MOYALE	KENEASTERN	Eastern
North-Eastern		Mandera	MWINGI	KENEASTERN	Eastern
North-Eastern		Wajir	NORTH MERU	KENEASTERN	Eastern
Nyanza		Homa Bay	SOUTH MERU	KENEASTERN	Eastern
Nyanza		Kisii	THARAKA	KENEASTERN	Eastern
Nyanza		Kisumu	GARISSA	NORTH EASTERN	North-Eastern
Nyanza		Migori	MANDERA	NORTH EASTERN	North-Eastern
Nyanza	North Kisii	Nyamira	WAJIR	NORTH EASTERN	North-Eastern
Nyanza		Siaya	BONDO	NYANZA	Nyanza
Rift Valley		Baringo	GUCHA	NYANZA	Nyanza
Rift Valley		Bomet	HOMA BAY	NYANZA	Nyanza
Rift Valley	E. Marakwet	Elgeyo-Marakwet	KISII	NYANZA	Nyanza
Rift Valley		Kericho	KISUMU	NYANZA	Nyanza
Rift Valley		Laikipia	KURIA	NYANZA	Nyanza
Rift Valley		Nakuru	MIGORI	NYANZA	Nyanza
Rift Valley		Nandi	NYAMIRA	NYANZA	Nyanza
Rift Valley		Narok	NYANDO	NYANZA	Nyanza
Rift Valley		Samburu	NYANZA	NYANZA	Nyanza
Rift Valley		Trans-Nzoia	RACHUONYO	NYANZA	Nyanza
Rift Valley		Turkana	SIAYA	NYANZA	Nyanza
Rift Valley		Uasin Gishu	SUBA	NYANZA	Nyanza
Rift Valley		West Pokot	BARINGO	RIFT VALLEY	Rift Valey
Western		Bungoma	BOMET	RIFT VALLEY	Rift Valey
Western		Busia	BURETI	RIFT VALLEY	Rift Valey
Western		Kakamega	KAJIADO	RIFT VALLEY	Rift Valey
Western		Vihiga	KEIYO	RIFT VALLEY	Rift Valey
			KERICHO	RIFT VALLEY	Rift Valey
			KOIBATEK	RIFT VALLEY	Rift Valey
			LAIKIPIA	RIFT VALLEY	Rift Valey
			MARAKWET	RIFT VALLEY	Rift Valey
			NAKURU	RIFT VALLEY	Rift Valey
			NANDI	RIFT VALLEY	Rift Valey
			NAROK	RIFT VALLEY	Rift Valey
			SAMBURU	RIFT VALLEY	Rift Valey
			TRANS NZOIA	RIFT VALLEY	Rift Valey
			TRANS-MARA	RIFT VALLEY	Rift Valey
			TURKANA	RIFT VALLEY	Rift Valey
			UASIN GISHU	RIFT VALLEY	Rift Valey
			WEST POKOT	RIFT VALLEY	Rift Valey
			BUNGOMA	KENWESTERN	Western
			BUSIA	KENWESTERN	Western
			KAKAMEGA	KENWESTERN	Western
			LUGARI	KENWESTERN	Western
			MT. ELGON	KENWESTERN	Western
			TESO	KENWESTERN	Western
			VIHIGA	KENWESTERN	Western
			NAIROBI	NAIROBI	Western

Table 7: Overview on new counties

	GADMIN			WHO		
1	NAME_1	VARNAME_2	NAME_2	ADM2_NAME	ADM1_NAME	Ajout pour comparaison
2	Rift Valley		Baringo	BARINGO	RIFT VALLEY	Rift Valey
3	Rift Valley		Bomet	BOMET	RIFT VALLEY	Rift Valey
4	Western		Bungoma	BONDO	NYANZA	Nyanza
5	Western		Busia	BUNGOMA	KENWESTERN	Western
6	Rift Valley	E. Marakwet	Elgeyo-Marakwet	BURETI	RIFT VALLEY	Rift Valey
7	Eastern		Embu	BUSIA	KENWESTERN	Western
8	North-Eastern		Garissa	CENTRAL MERU	KENEASTERN	Eastern
9	Nyanza		Homa Bay	EMBU	KENEASTERN	Eastern
10	Eastern		Isiolo	GARISSA	NORTH EASTERN	North-Eastern
11	Western		Kakamega	GUCHA	NYANZA	Nyanza
12	Rift Valley		Kericho	HOMA BAY	NYANZA	Nyanza
13	Central		Kiambu	ISIOLO	KENEASTERN	Eastern
14	Coast		Kilifi	KAJIADO	RIFT VALLEY	Rift Valey
15	Central		Kirinyaga	KAKAMEGA	KENWESTERN	Western
16	Nyanza		Kisii	KEIYO	RIFT VALLEY	Rift Valey
17	Nyanza		Kisumu	KERICHO	RIFT VALLEY	Rift Valey
18	Eastern		Kitui	KIAMBU	KENCENTRAL	Central
19	Coast		Kwale	KILIFI	COAST	Coast
20	Rift Valley		Laikipia	KIRINYAGA	KENCENTRAL	Central
21	Coast		Lamu	KISII	NYANZA	Nyanza
22	Central		Machakos	KISUMU	NYANZA	Nyanza
23	Eastern		Machakos	KITUI	KENEASTERN	Eastern
24	Eastern		Makueni	KOIBATEK	RIFT VALLEY	Rift Valey
25	North-Eastern		Mandera	KURIA	NYANZA	Nyanza
26	Eastern		Marsabit	KWALE	COAST	Coast
27	Eastern		Meru	LAIKIPIA	RIFT VALLEY	Rift Valey
28	Nyanza		Migori	LAMU	COAST	Coast
29	Coast		Mombasa	LUGARI	KENWESTERN	Western
30	Central		Murang'a	MACHAKOS	KENEASTERN	Eastern
31	Nairobi		Nairobi	MAKUENI	KENEASTERN	Eastern
32	Rift Valley		Nakuru	MALINDI	COAST	Coast
33	Rift Valley		Nandi	MANDERA	NORTH EASTERN	North-Eastern
34	Rift Valley		Narok	MARAGWA	KENCENTRAL	Central
35	Eastern		Nithi	MARAKWET	RIFT VALLEY	Rift Valey
36	Nyanza	North Kisii	Nyamira	MARSABIT	KENEASTERN	Eastern
37	Central		Nyandarua	MBEERE	KENEASTERN	Eastern
38	Central		Nyeri	MIGORI	NYANZA	Nyanza
39	Rift Valley		Samburu	MOMBASA	COAST	Coast
40	Nyanza		Siaya	MOYALE	KENEASTERN	Eastern
41	Coast	Taita	Taita Taveta	MT. ELGON	KENWESTERN	Western
42	Coast		Tana River	MURANGA	KENCENTRAL	Central
43	Rift Valley		Trans-Nzoia	MWINGI	KENEASTERN	Eastern
44	Rift Valley		Turkana	NAIROBI	NAIROBI	Western
45	Rift Valley		Uasin Gishu	NAKURU	RIFT VALLEY	Rift Valey
46	Western		Vihiga	NANDI	RIFT VALLEY	Rift Valey
47	Eastern		Wajir	NAROK	RIFT VALLEY	Rift Valey
48	North-Eastern		Wajir	NORTH MERU	KENEASTERN	Eastern
49	Rift Valley		West Pokot	NYAMIRA	NYANZA	Nyanza
50				NYANDARUA	KENCENTRAL	Central
51				NYANDO	NYANZA	Nyanza
52				NYANZA	NYANZA	Nyanza
53				NYERI	KENCENTRAL	Central
54				RACHUONYO	NYANZA	Nyanza
55				SAMBURU	RIFT VALLEY	Rift Valey
56				SIAYA	NYANZA	Nyanza
57				SOUTH MERU	KENEASTERN	Eastern
58				SUBA	NYANZA	Nyanza
59				TAITA TAVETA	COAST	Coast
60				TANA RIVER	COAST	Coast
61				TESO	KENWESTERN	Western
62				THARAKA	KENEASTERN	Eastern
63				THIKA	KENCENTRAL	Central
64				TRANS NZOIA	RIFT VALLEY	Rift Valey
65				TRANS-MARA	RIFT VALLEY	Rift Valey
66				TURKANA	RIFT VALLEY	Rift Valey
67				UASIN GISHU	RIFT VALLEY	Rift Valey
68				VIHIGA	KENWESTERN	Western
69				WAJIR	NORTH EASTERN	North-Eastern
70				WEST POKOT	RIFT VALLEY	Rift Valey

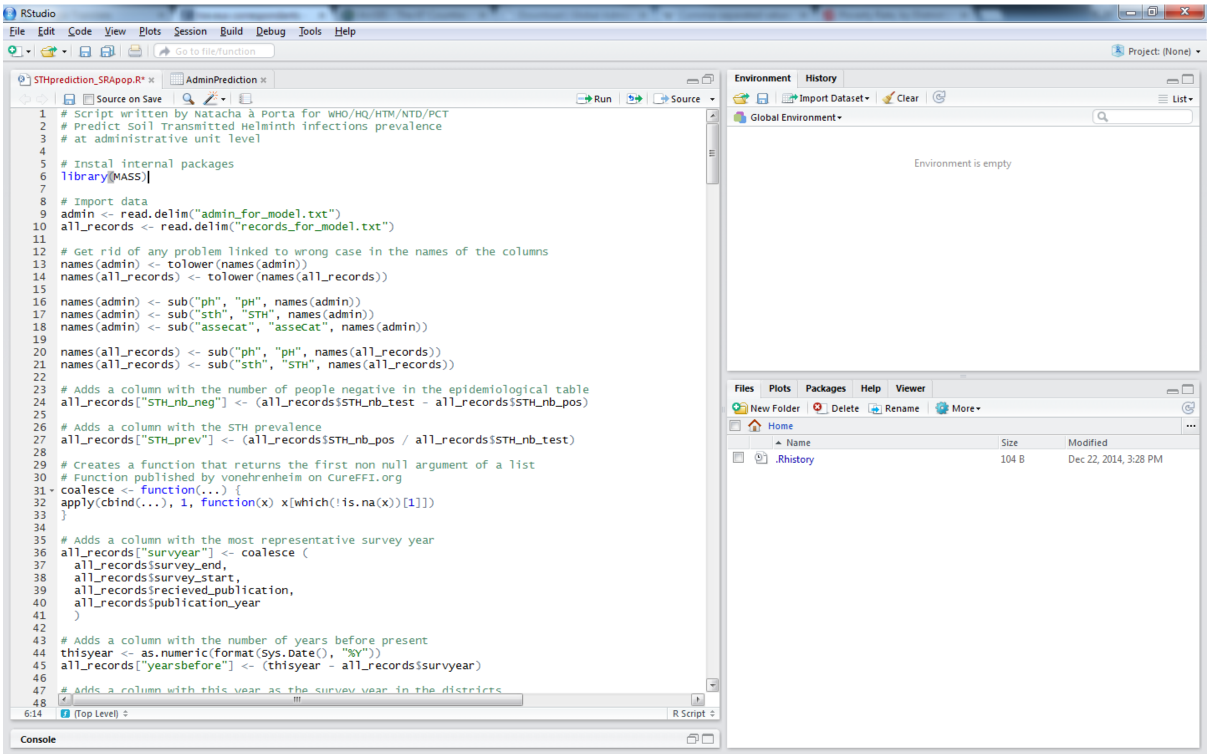
Table 8: Coast differences for understanding LF data

	GADMIN2			WHO2	
	NAME_1	VARNAME_2	NAME_2	ADM2_NAME	ADM1_NAME
1	Coast		Kilifi	KILIFI	COAST
2	Coast		Kwale	KWALE	COAST
3	Coast		Lamu	LAMU	COAST
4	Coast		Mombasa	MALINDI	COAST
5	Coast	Taita	Taita Taveta	MOMBASA	COAST
6	Coast		Tana River	TAITA TAVETA	COAST
7				THIKA	COAST

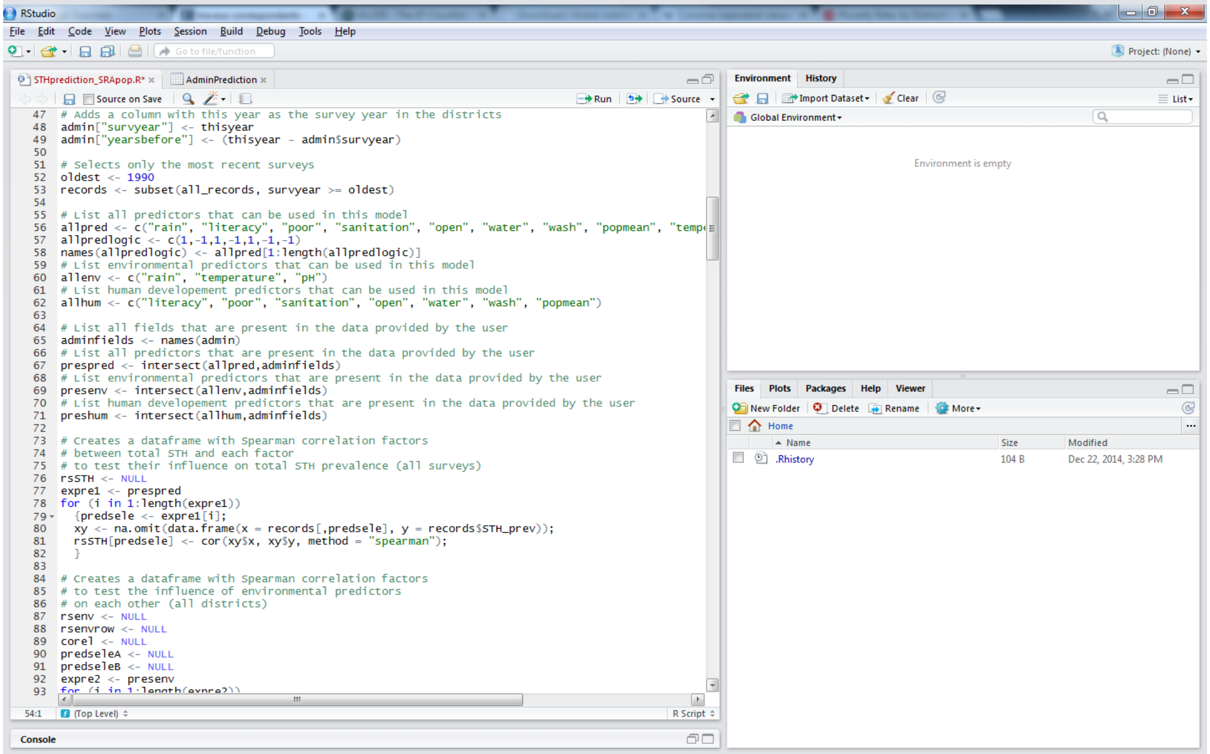
Table 9: Coast areas for understanding LF data

GADMIN3 /Coast			LF result	
Coast	Kilifi	Bahari	Ganze	Kilifi
Coast	Kilifi	Ganze	Kaloleni	Kilifi
Coast	Kilifi	Kalolenli	Kilifi	Kilifi
Coast	Kilifi	Malindi	Magarini	Kilifi
			Malindi	Kilifi
			Rabai	Kilifi
Coast	Kwale	Kinango	Kwale	Kwale
Coast	Kwale	Kubo	Kinango	Kwale
Coast	Kwale	Matuga	Msambweni	Kwale
Coast	Kwale	Msambweni		
Coast	Lamu	Amu	Lamu East	Lamu
Coast	Lamu	Faza	Lamu West	Lamu
Coast	Lamu	Kiunga		
Coast	Lamu	Mpeketoni		
Coast	Lamu	Witu		
Coast	Mombasa	Changamwe	Changamwe	Mombasa
Coast	Mombasa	Island	Kisauni	Mombasa
Coast	Mombasa	Kisauni	Likoni	Mombasa
Coast	Mombasa	Likoni	Mvita	Mombasa
Coast	Taita Taveta	Mwatate	Mwatate	Taita Taveta
Coast	Taita Taveta	Taveta	Taita	Taita Taveta
Coast	Taita Taveta	Tsavo National Park (E&W)	Taveta	Taita Taveta
Coast	Taita Taveta	Voi	Voi	Taita Taveta
Coast	Taita Taveta	Wundanyi		
Coast	Tana River	Bura	Tana Delta	Tana River
Coast	Tana River	Galole	Tana North	Tana River
Coast	Tana River	Garsen	Tana River	Tana River
Coast	Tana River	Madogo		

A6- Brief overview of R script used for obtaining the prediction data



```r
# Script written by Natacha à Porta for WHO/HQ/HTM/NTD/PCT
# Predict Soil Transmitted Helminth infections prevalence
# at administrative unit level

# Instal internal packages
library(MASS)

# Import data
admin <- read.delim("admin_for_model.txt")
all_records <- read.delim("records_for_model.txt")

# Get rid of any problem linked to wrong case in the names of the columns
names(admin) <- tolower(names(admin))
names(all_records) <- tolower(names(all_records))

names(admin) <- sub("ph", "pH", names(admin))
names(admin) <- sub("sth", "STH", names(admin))
names(admin) <- sub("assecat", "asseCat", names(admin))

names(all_records) <- sub("ph", "pH", names(all_records))
names(all_records) <- sub("sth", "STH", names(all_records))

# Adds a column with the number of people negative in the epidemiological table
all_records["STH_nb_neg"] <- (all_records$STH_nb_test - all_records$STH_nb_pos)

# Adds a column with the STH prevalence
all_records["STH_prev"] <- (all_records$STH_nb_pos / all_records$STH_nb_test)

# Creates a function that returns the first non null argument of a list
# Function published by vonehrenheim on CureFFI.org
coalesce <- function(...) {
apply(cbind(...), 1, function(x) x[which(!is.na(x))[1]])
}

# Adds a column with the most representative survey year
all_records["survyear"] <- coalesce (
  all_records$survey_end,
  all_records$survey_start,
  all_records$recieved_publication,
  all_records$publication_year
  )

# Adds a column with the number of years before present
thisyear <- as.numeric(format(Sys.Date(), "%Y"))
all_records["yearsbefore"] <- (thisyear - all_records$survyear)
```



```r
# Adds a column with this year as the survey year in the districts
admin["survyear"] <- thisyear
admin["yearsbefore"] <- (thisyear - admin$survyear)

# Selects only the most recent surveys
oldest <- 1990
records <- subset(all_records, survyear >= oldest)

# List all predictors that can be used in this model
allpred <- c("rain", "literacy", "poor", "sanitation", "open", "water", "wash", "popmean", "tempe
allpredlogic <- c(1,-1,1,-1,1,-1,-1)
names(allpredlogic) <- allpred[1:length(allpredlogic)]
# List environmental predictors that can be used in this model
allenv <- c("rain", "temperature", "pH")
# List human developement predictors that can be used in this model
allhum <- c("literacy", "poor", "sanitation", "open", "water", "wash", "popmean")

# List all fields that are present in the data provided by the user
adminfields <- names(admin)
# List all predictors that are present in the data provided by the user
prespred <- intersect(allpred,adminfields)
# List environmental predictors that are present in the data provided by the user
presenv <- intersect(allenv,adminfields)
# List human developement predictors that are present in the data provided by the user
preshum <- intersect(allhum,adminfields)

# Creates a dataframe with Spearman correlation factors
# between total STH and each factor
# to test their influence on total STH prevalence (all surveys)
rsSTH <- NULL
expre1 <- prespred
for (i in 1:length(expre1))
  {predsele <- expre1[i];
  xy <- na.omit(data.frame(x = records[,predsele], y = records$STH_prev));
  rsSTH[predsele] <- cor(xy$x, xy$y, method = "spearman");
  }

# Creates a dataframe with Spearman correlation factors
# to test the influence of environmental predictors
# on each other (all districts)
rsenv <- NULL
rsenvrow <- NULL
corel <- NULL
predseleA <- NULL
predseleB <- NULL
expre2 <- presenv
```

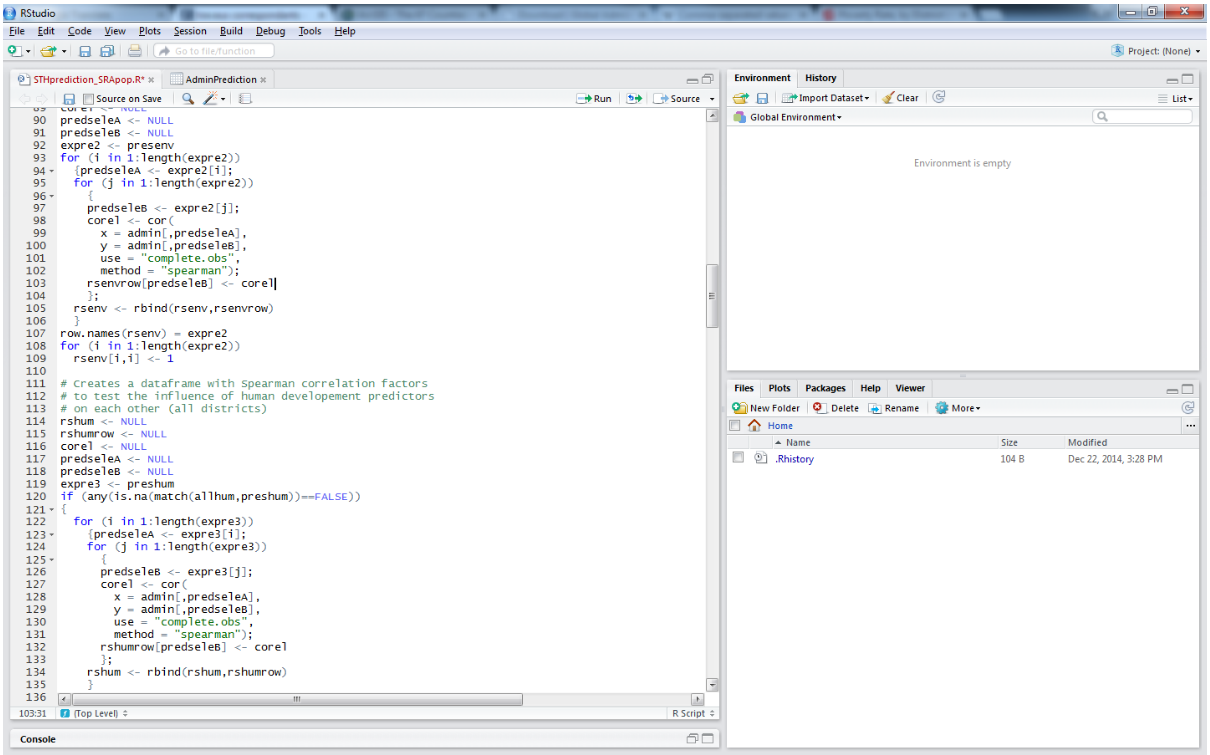


Figure 26: Overview on the R script

A6- About Kenya

Kenya is an East African country on the equator line, well-known for the famous Kilimanjaro Mountain. Coastal country in the south eastern part, Kenya is bordered by Somalia in the East, Ethiopia and Sudan in the North, Uganda in West and Tanzania in the South. The country benefits as well in the western part from the ecosystem services of Victoria Lake one of the most well-known Lake in Africa.

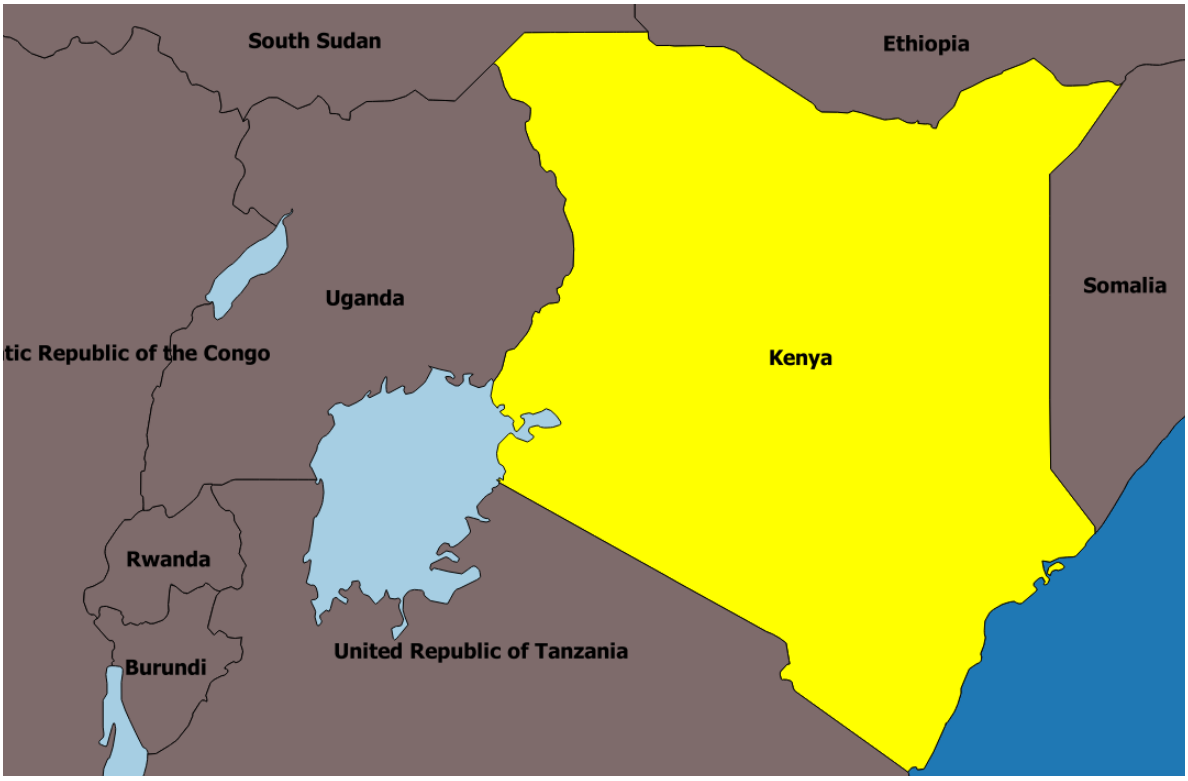


Figure 27: Kenya, East Africa

The population, according to the "*2009 Kenya Population and Housing Census Report published in August 2010*" was confirmed to be 38,610,097 people (Kenya National Bureau of Statistics KNBS, 2010). After the official census took place in the country, estimates are released on a regular basis and in 2011, it was claimed that those numbers had risen to 41,070,934. The estimate for 2014 is 45,941,977[17].

Others socio economics information on the country are summarized in the table below.

[17] http://worldpopulationreview.com/countries/kenya-population/

Table 10: Keys socio-economic information on Kenya

	Information	Observation
Surface area	580,367 square kilometres	47th largest country in the world
Capital city	Nairobi	
Gross domestic product per capita[18]	994.31 USD	In 2013
Gross domestic product[19]	44.1 billion USD	In 2013
Life expectancy[20]	61.08 years	In 2012
Principal Ethnical groups[21]	Kikuyu, Luhya, Luo, Kalenjin, Kamba	22%, 14%, 13% , 12% and 11% respectively

Kenya is unequally shared in eight provinces as visualized on figure 6.

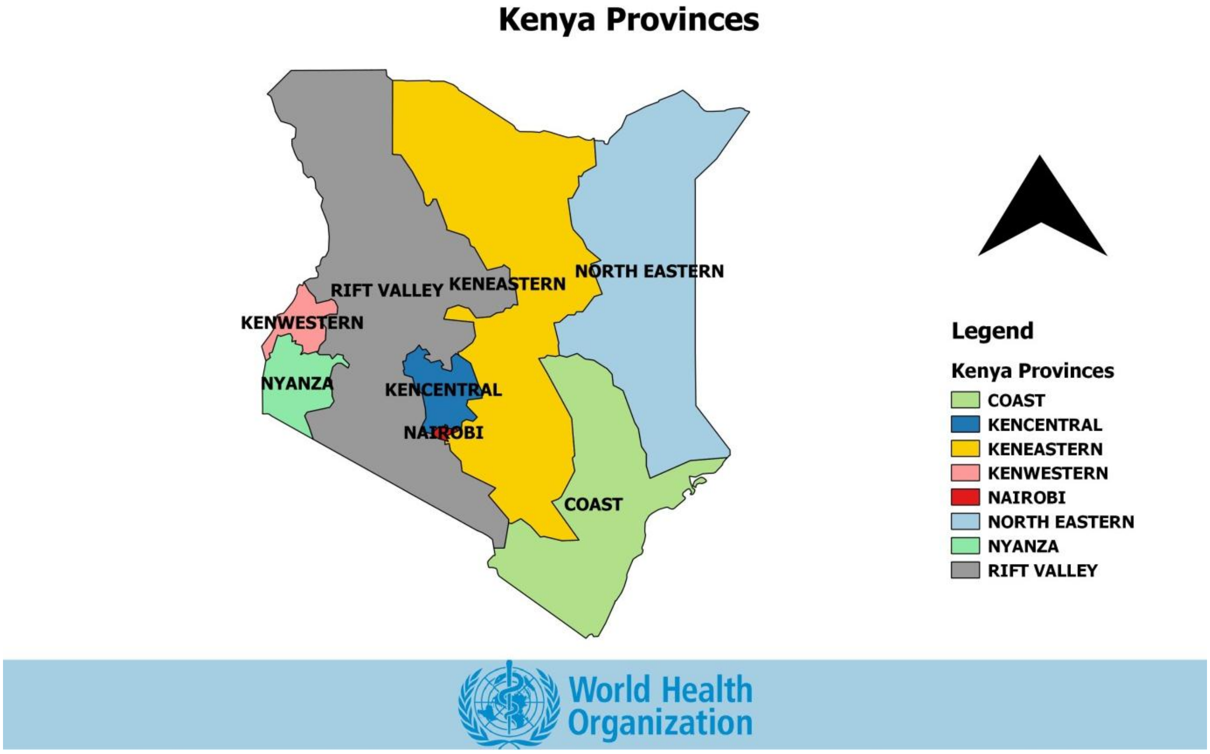


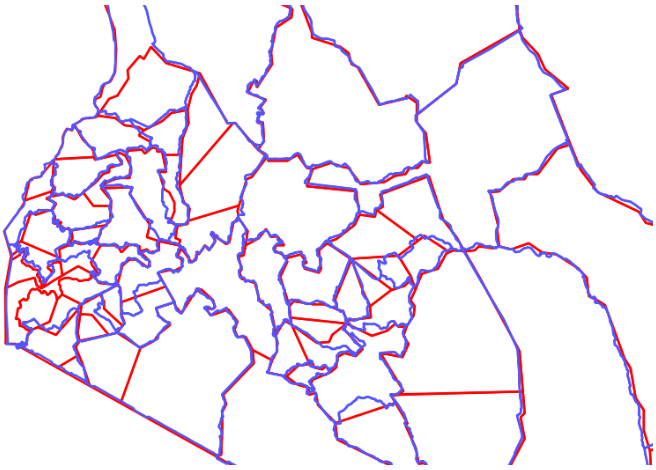
Figure 28: Kenya Provinces

It is important in this report to mention that recently, a change occurred in the administrative boundaries of the country. Therefore, the 47 previous districts within the six regions of Kenya were transformed into 69 counties. More details on those changes can be read in the Independant Electoral and Boundaries Commission report (Independant Electoral and Boundaries Commission,

[18] https://www.google.ch/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=total%20population%20of%20kenya
[19] idem
[20] idem
[21] http://worldpopulationreview.com/countries/kenya-population/ (Independant Electoral and Boundaries Commission, 2012)

2012). Figure 7 shows the overlay of the old and new administrative boundaries at level 2. These changes in the mapping process confronted us with a real challenge in the treatment of the historical data (Annexe A5 / tables 7, 8 and 9).

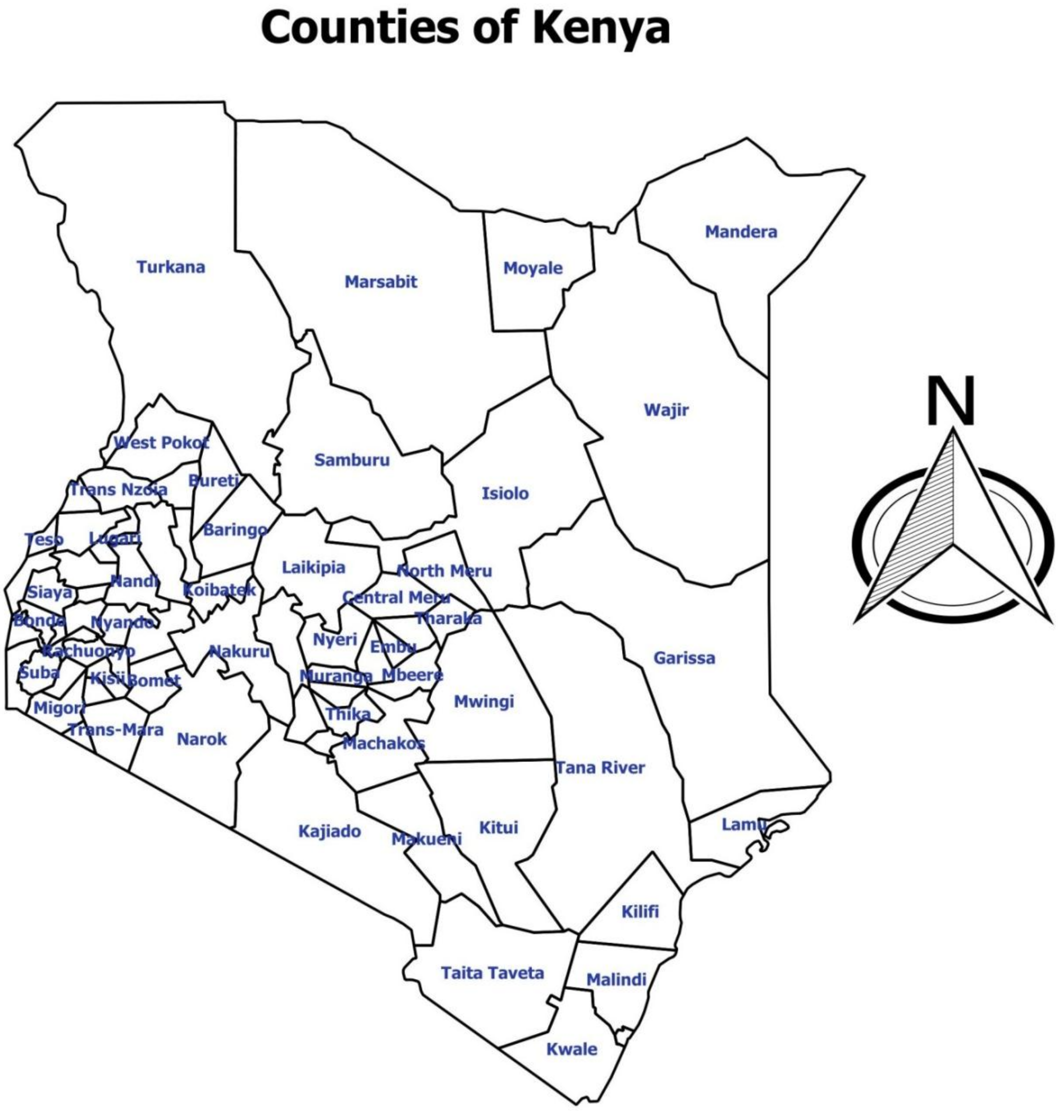
Old boundaries (blue) and new boundaries (red) overlaid.



Old boundaries (blue) and new boundaries (red) overlaid: Zoom on the Western part of Kenya

Figure 29: Kenya old districts and new counties overlaid

A7- Counties in Kenya



Counties of Kenya: Zoom South

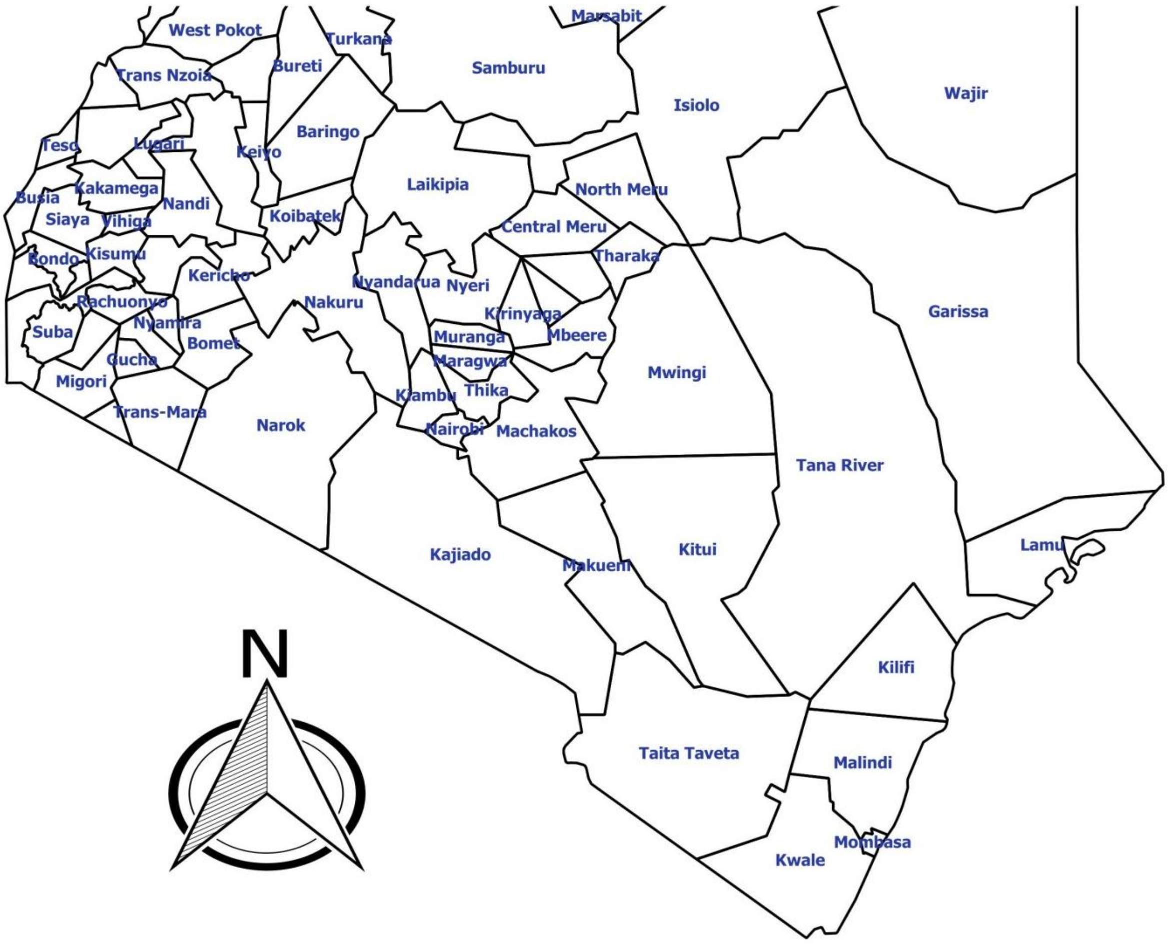


Figure 30: Identification of counties

A8- QGIS

QGIS[22] (previously known as "Quantum GIS") is a cross-platform free and open-source desktop geographic information system (GIS) application that provides data viewing, editing, and analysis capabilities.

Similar to other software GIS systems, QGIS allows users to create maps with many layers using different map projections. Maps can be assembled in different formats and for different uses. QGIS allows maps to be composed of raster or vector layers. Typical for this kind of software the vector data is stored as point, line, or polygon-feature. Different kinds of raster images are supported and the software can perform georeferencing of images.

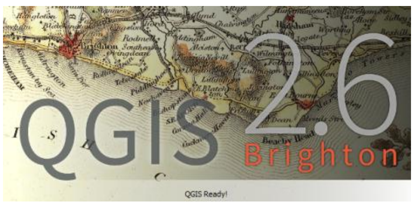
QGIS provides integration with other open source GIS packages, including PostGIS, GRASS, and MapServer to give users extensive functionality. Plugins, written in Python or C++, extend the capabilities of QGIS. There are plugins to geocode using the Google Geocoding API, perform geoprocessing (fTools) similar to the standard tools found in ArcGIS, interface with PostgreSQL/PostGIS, SpatiaLite and MySQL databases.

Written in C++, QGIS makes extensive use of the Qt library. In addition to Qt, required dependencies of QGIS include GEOS and SQLite. GDAL, GRASS GIS, PostGIS, and PostgreSQL are also recommended, as they provide access to additional data formats.

QGIS runs on multiple operating systems including Mac OS X, Linux, UNIX, and Microsoft Windows. For Mac users, the advantage of QGIS over GRASS GIS is that it does not require the X11 windowing system in order to run, and the interface is much cleaner and faster. QGIS can also be used as a graphical user interface to GRASS. QGIS has a small file size compared to commercial GIS's and requires less RAM and processing power; hence it can be used on older hardware or running simultaneously with other applications where CPU power may be limited.

QGIS is maintained by an active group of volunteer developers who regularly release updates and bug fixes. As of 2012 developers have translated QGIS into 48 languages and the application is used internationally in academic and professional environments.

QGIS 2.6 Brighton is the last version of QGIS after QGIS 2.4 Valmiera.



22 http://en.wikipedia.org/wiki/QGIS