A GIS based STH predictor

Nam Lethanh* and Antonio Montressor[†] December 2, 2015

Abstract

Soil Transmitted Diseases (STH) are common diseases in developing nations. In order to reduce the spreading of the STH among community, WHO has tried to promote different intervention strategies (also known as control program). Each control program is a combination of type of drug distributed to the community and the frequency of the distribution. It is necessary for policy maker and public health specialist to monitor and appraise the effectiveness of each control program over years.

This GIS based STH predictor was therefore developed to assist policy maker and manager of the program to predict the future prevalence of the STH based on the comparison of effectiveness under different control programs (CPs). The predictor was developed as a Python plug-in in QGIS environment and includes two main predictive models: One when only total prevalence is recorded and other is when the data of intensity of each STH is available. Under each model, users are able to select their prefer CP and the program can generate future prevalence based on user's choice.

This paper aims to provide an gentle manual instruction for general users to use the predictor. It includes installation instruction for QGIS and the developed plug-in as well as the step-by-step instruction on preparing the input shapefile, select and run the model within QGIS environment, and finally display the results vividly on the map using QGIS.

1. Installation

1.1. Quantum GIS (QGIS)

The latest version of QGIS can be downloaded from the official website of QGIS www.qgis.org.

A detailed instruction on installation for different operation systems (Window, Linux,
Mac) is given also on the website www.qgis.org.

^{*}Research Associate, Dr., Institute of Construction and Infrastructure Management., Swiss Federal Institute of Technology (ETH), 8093 Zurich, Switzerland. E-mail: lethanh@ibi.baug.ethz.ch

[†]MD, Ph.D., Department of Neglected Tropical Diseases., World Health Organization (WHO), Avenue Appia 20, 1211 Geneva 27, Switzerland. E-mail: montresora@who.int

1.2. The STH-Predictor plugin

- Download the plugin from Github repository https://github.com/namkyodai/GIS-STHpredictor
- Extract the zip file to a PC's local folder
- Copy/Move the extracted folder to the plugin folder of QGIS

Python plugins are listed together with C++ plugins in QGIS plugin manager. They are searched for in these paths:

- UNIX/Mac: $\sim /.qgis2/python/plugins$ and $(qgis_prefix)/share/qgis/python/plugins$
- Windows: $\sim /.qgis2/python/plugins$ and $(qgis_prefix)/python/plugins$

Home directory (denoted by above \sim) on Windows is usually something like C:\ Documents and Settings \ (user) (on Windows XP or earlier) or C:\ Users\(user)\). Since QGIS is using Python 2.7, subdirectories of these paths have to contain an _init_.py file to be considered Python packages that can be imported as a plugin.

After installation, when QGIS is opened. The icon \blacksquare of STH-predictor plugin can be visible on the panel of QGIS (Figure 0.1).

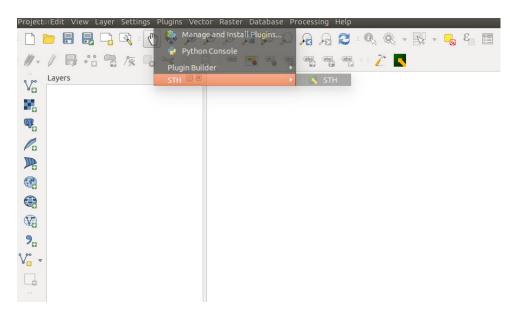


Figure 0.1: Icon of STH-Predictor plugin

2. Process and Application

The main process can be seen in Figure 0.2. This figure shows the highest level of abstraction of the process, which encompasses three main sub-processes. The first sub-process is the task of users to prepare a proper input shapefile to be used with the STH-predictor. The second subprocess is about the plugin itself, which offers an users

friendly interface to run the model. The third subprocess is with visualization and display of results obtained by running the model in previous step.

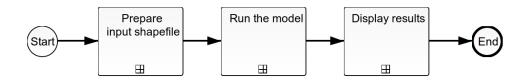


Figure 0.2: Main process

2.1. Prepare input shapefile

This subprocess includes three main tasks (Figure 0.3):

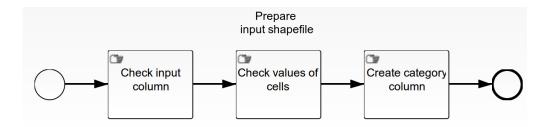


Figure 0.3: Tasks to be prepared input shapefile

- Check input column: in a shapefile to be used with the STH-predictor plugin, there should exists 5 unique columns named "CATEGORY", "STH", "HW", "Tt", and "Al". If there is a missing of these columns, users are requested to create them. To create them, users can use QGIS to open the table of the shapefile and modify them. Other than using QGIS, users can use many commercial or open sources softwares to manipulate the attribute of shapefiles.
- Check values of cells: In each and every cells belong to each columns named "CATEGORY", "STH", "HW", "Tt", and "Al", there should be values. Under column "CATEGORY", cells must be filled in with categorical values "A to F" (e.g. A, B, C, D, E, F). This value is determined based on the value of column "STH" or the sum of three columns "HW", "Tt", and "Al". Values of cells in column "STH" are total prevalence recorded for each region shown in probability ranging from 0 to 1. Similarly, values of cells in columns "HW", "Tt", and "Al" are intensity values of "Hookworms", *Trichuris trichiura*, and *Ascaris lumbricoides*, respectively.
- Create category column: If there "CATEGORY" is missing, users are required to create a column named "CATEGORY" for the selected shapefile. Then assign categorical values from "A to F" based on a rule that 1) category A is when total prevalence is less than 1%; 2) category B is when total prevalence is in between [1,10%); 3) category C is when total prevalence is in between [10,20%); 4) category

D is when total prevalence is in between [20,50%); 5) category E is when total prevalence is in between [50,70%); 6) category F is when total prevalence is greater than or equal to 70%.

In many practical situations, it is often that only total prevalence is available for each region. This means only values of cells of column "STH" are available. And values of cells in columns "HW", "Tt", and "Al" are partially or not available. The STH-predictor plugin still works perfectly. However, users are noted to not select the model for use with these situation. The default model to be run is when all cells of column "STH" receives values.

A shapefile is a file with extension .shp. The .shp format is a widely used type of file supporting geospatial vector, which is used to display information on any GIS software ¹. In the context of prediction for STH, the shapefile is defined with polygon structure. Each polygon represents an administrative management region, in which, the information related to prevalence of the STH is recorded.

For example, in Figure 0.4, each polygon represents a provincial region of India.

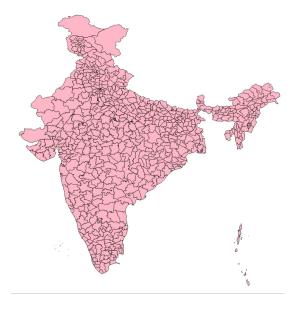


Figure 0.4: Map of provincial regions of India in Polygon structure

In order for the shapefile to be workable, a shapefile itself is linked to additional 3 files, which always come together with the shapefile and they are being created automatically whenever a shapefile is created. Therefore, it is important to note that whenever users want to copy/more a shapefile, they must copy/more all 4 files together and place them exactly in a same working folder. For example, the folder containing input.shp file shown in Figure 0.5 contains also three additional files with extensions .prj, .dbf, and .shx, respectively.

¹https://en.wikipedia.org/wiki/Shapefile

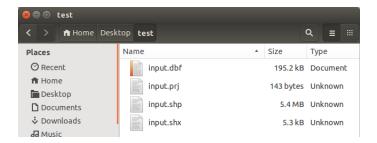


Figure 0.5: Folder contains shapefile and its associate files

Using QGIS, users can add the shapefile into its working environment and display the properties of the selected shapefile. Figure 0.6 shows the structure of a shapefile containing the information of total prevalence of STH and intensities of each STH's species.

<u>ε</u> <u>ε</u>													
OBJECTID A	_	ADM1_CODE	GUID	_	_	SHAPE_leng	_	GlobalID	CATEGORY	STH	HW	Al	Tt
1885				24-Pargana						67.462298	47.7619210	13.365344000	
1886				24-Pargana						49.681204	42.1028011	4.962860100	
1899		IN0010020				7.55871634				55.985083	48.4654979	2.470838200	
11901	_	IN0010360				6.87382926				33.244060	19.4661322	13.672372000	
1903	Gujarat		{66B61EA1			8.67281016				45.138693	27.7531547	11.583485000	
1904	Gujarat		{8AFABF9D			0.46435566				51.889915	30.9732998	2.769205400	
1905	Maharashtra		{0B38ADFE			10.7803984				57.194810	39.8299416	11.882717000	
11908	Mizoram	IN0010270	{67185827	Aizawal_East	IN0010270	1.87523943	0.11913372	{77876F49	D	31.635261	26.4744542	5.038936600	0.121870600
1909	Mizoram	IN0010270	{BE5B9471	Aizawal_West	IN0010270	2.56092543	0.17889393	{13065B26	D	39.458599	13.5111497	7.418483500	18.5289660
11910	Rajasthan	IN0010320	{E51E7143	Ajmer	IN0010320	7.40055319	0.77661069	{7E5F2D0C	E	59.363276	34.3758167	7.032605800	
11912	Maharashtra	IN0010240	{9E63DB1F	Akola	IN0010240	4.75494084	0.45965240	{0F4523E7	D	29.354969	8.1266369	4.465445500	16.7628870
11913	Kerala	IN0010200	{0F14C5CB	Alappuzha	IN0010200	2.66101861	0.11239792	{408B8BEA	D	25.084621	13.9348365	10.510070000	0.639714300
11914	Uttar_Prad	IN0010360	{4E4C131A	Aligarh	IN0010360	4.19698860	0.35061606	{A22B8429	D	23.787002	0.1386542	6.939641100	16.7087070
11915	Madhya_Pr	IN0010230	{6CE2670B	Alirajpur	IN0010230	3.66557306	0.28555903	{7EF5A9DD	В	9.197406	2.6265871	1.243973700	5.326845500
11916	Uttar_Prad	IN0010360	{B1A26AE6	Allahabad	IN0010360	4.28930537	0.48473293	{DC4C77E7	E	57.828397	43.9257049	3.275735900	10.6269560
11917	Uttarakhand	IN0010370	{E90F6643	Almora	IN0010370	3.30702660	0.30702855	{1CF7F388	E	61.556828	32.4685018	13.365423000	15.7229030
11919	Rajasthan	IN0010320	{BC814439	Alwar	IN0010320	7.58165986	0.75578966	{6889C37C	D	48.292267	24.8632089	11.834968000	11.5940900
11921	Haryana	IN0010140	{63C866D2	Ambala	IN0010140	2.34120975	0.12715462	{938440DD	E	69.227558	38.7846497	14.163807000	16.2791010
11922	Uttar_Prad	IN0010360	{E4E5CF8D	Ambedkar	IN0010360	3.38464190	0.20709476	{ED44A1B8	D	43.983910	32.8793397	2.968001400	8.136568400
11926	Maharashtra	IN0010240	{78F351C7	Amravati	IN0010240	7.84397520	1.02519449	{36FBF7A3	C	18.155002	4.6692422	10.259553000	3.226206900
11927	Gujarat	IN0010130	{028248F9	Amreli	IN0010130	5.63049991	0.64733410	{6DE05514	D	37.962770	30.1649920	0.639947200	7.157830600
11928	Punjab	IN0010310	{07CD25BD	Amritsar	IN0010310	2.72339623	0.25560411	{604F7DC8	C	10.211161	2.8375407	4.070378000	3.303242400
11930	Gujarat	IN0010130	{D2CACA19	Anand	IN0010130	2.92179549	0.24803282	{F34E3091	E	53.869132	40.9128650	2.400939800	10.5553270
11931	Jammu_&	IN0010170	{50910E85	Anantnag	IN0010170	3.35072758	0.23845355	{4EECEFE4	D	26.826023	3.1136900	4.374741300	19.3375920
11932	Andhra Pra	IN0010020	{392B6632	Anantpur	IN0010020	11.8969651	1.65191875	{A6509EA9	E	50.794373	23.1945545	10.970665000	16.6291540
11934	Odisha	IN0010290	{48C58210	Angul	IN0010290	5.01228281	0.56848340	{44EAA00C	С	10.741159	2.8252720	3.675050700	4.240836100
11935	Arunachal	IN0010030	{D259461C	Anjaw	IN0010030	3.46335071	0.56401364	{F2F4D611	E	57.334480	34.6119483	4.204015700	18.5185160
11938		IN0010230				4.01682084				28.582600	9.1624379	14.752329000	

Figure 0.6: Attributes of a shapefile used for prediction with the STH-predictor

As can be seen in the figure, the first 9 columns of the tables is always come with the shapefile itself and the users should not delete or change it. Information of these columns is nowadays defined as a standard. This information is issued by the government of India. In any other country, this information is more or less similar and can be down loadable via various Internet portals.

The last 5 columns is important for users of the STH-predictor. Values of cells under each column must be created/recorded by users. In practical situation, many local district or provincial personnel record data using spreadsheet such as MS Excel application. Such recorded data can be imported directly into shapefile.

If the user's shapefile does not contain such 5 columns, users must create them by using QGIS. Steps to create them are as followed

• Step 1: Load input shapefile: Click icon \(\frac{\frac{1}{2}}{2} \) (add vector layer) on the left panel of the QGIS (or go to Layer -> Add Vector Layer). (Figre 0.7)

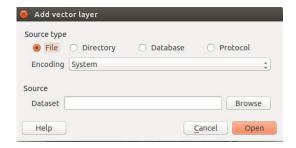


Figure 0.7: Add vector layer (QGIS)

Click button "Browse" and navigate to folder containing the input shapefile, then select the input shapefile ending with .shp format.

One the shapefile is opened, a map will appear in random color. Following figure (Figure 0.8) shows the map of India, with all its regions. Each region is represented by a polygon.

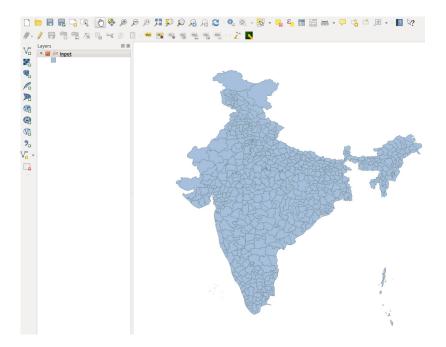


Figure 0.8: Map of india with all its regions

• Step 2: Open attribute table: Click icon [(Open Attribute Table) on QGIS main panel (or go to Layer -> Open Attribute Table).

The table of the shapefile will be appeared and look like the one shown in Figure 0.6. If any column with name "CATEGORY", "STH", "HW", "Tt", or "Al" is missing, users can create them within QGIS table attribute environment.

+ In Figure 0.6, Click icon pencil ∠ located on the top left corner of attribute table. All icons of the attribute table (Figure 0.9) will be visible and users can click on each of them for editing the data.



Figure 0.9: Icons of attribute table

The last three icons shown in Figure 0.9 are used to delete, create, and manipulate columns of the table. Users of the STH-predictor are advised to read some basic concepts and steps in order to manipulate vector data shapefiles from QGIS documentation repository such as the user guide manual $(QGIS2.2)^2$

Following code is given for users to assign categorical values to column "CATE-GORY" when using the last icon of Figure 0.9 to manipulate the data of the table (Figure 0.10).

```
CASE

WHEN STH < 1 THEN 'A'
WHEN STH < 10 THEN 'B'
WHEN STH < 20 THEN 'C'
WHEN STH < 50 THEN 'D'
WHEN STH < 70 THEN 'E'
ELSE 'F'

END
```

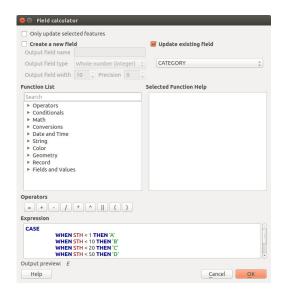


Figure 0.10: Field calculator for the attribute table

• Step 3: Display prevalence on the map: By open the "Properties" of Input layer and Load the color style (as shown in Figure 0.20). A map of prevalence of input will appear as shown in Figure 0.11.

²https://docs.qgis.org/2.2/en/docs/user_manual/index.html

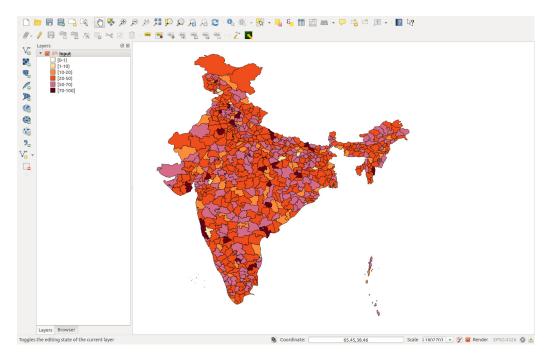


Figure 0.11: Prevalence map of India

The default color style for the STH-predictor is saved in sub-folder named "colorstyles" of the main plugin STH-predictor folder

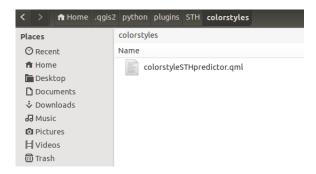


Figure 0.12: Color style file saved in STH-predictor plugin folder

2.2. Run the model

To run the model on a specific shapefile, users are required to perform sequential or parallel tasks shown in the diagram of Figure 0.13.

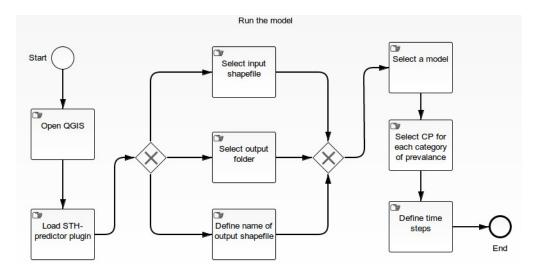


Figure 0.13: Run the model

Open QGIS

If a version of QGIS has been installed in your operation system, users can find the softwares easily in the program panel.

Load STH-predictor plugin

By clicking the icon of the STH-predictor, the STH-predictor will be loaded automatically and its interface will appear on the screen (Figure 0.14).

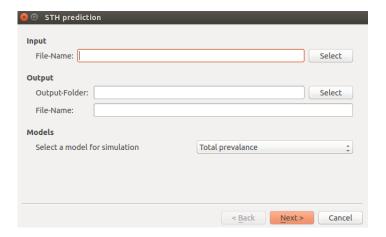


Figure 0.14: Welcome screen

Select input shapefile

By clicking the "select" button in the input section, users can navigate to the folder containing the shapefile file and choose it (Figure 0.15).

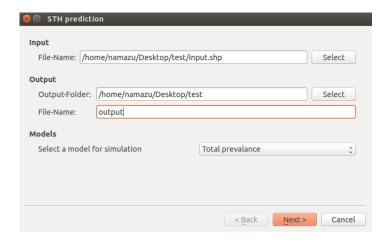


Figure 0.15: Input and output selection

Select output folder

By clicking the "select" button in the output section, users can navigate to the folder, in which, users would like to save the output shapefiles (Figure 0.15).

Define name of output shapefile

name of output's shapefile can be typed inside the box "file-name". Users do not need to specify the extension format of the shapefile (e.g. output.shp), the program itself will create automatically shapefiles and stored them in the selected output folder (Figure 0.15).

Select a model

In this step, users need to decide which models they want to use for prediction. There are two models built in the plug-in. The default model is used when there is only information of total prevalence of the STH. The other model is when information of intensity of Hookworms, *Trichuris trichiura*, and *Ascaris lumbricoides* is available.

Once a model is selected, users can move forward by clicking "Next" button shown in the Figure 0.15.

Select control program (CP) for each category

When the "Next" button shown in the Figure 0.15 is selected, a new interface will appear. This interface gives users choices to select which control program (CP) to be used under 6 categories mentioned earlier.

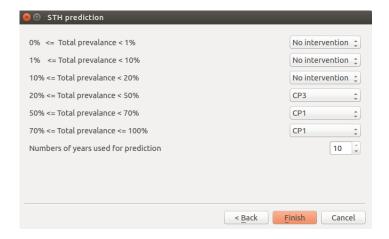


Figure 0.16: Select the control program

The default CP for each category is automatically appeared. This default CP is set based on expert opinion. However, users can define a CP differently from the default one.

To choose another CP other than the default one, users can click and choose CPs from com-box box under each category.

When users move cursor over each com-box box, a tool-tip text will appear and give users the definition of each CP.

The definition of each CP is given below.

- **No intervention:** when no drug is distributed. Under this situation, the distribution of prevalence in future will not change.
- CP1 (Administration of anthelminthic (ALB) twice a year): The STH control program is consisting on the administration twice a year of a tablet of albendazole (400 mg) to all school-age children. The drug administration is normally organized using the school infrastructure, also the children of school age that are not enrolled in school are offered drug. School children absent the day of the drug administration are provided the drug when they return to school.
- CP2 (Administration of anthelminthic (MEB) twice a year): The STH control program is consisting on the administration twice a year of a tablet of mebendazole (500 mg) to all school-age children. The drug administration is normally organized using the school infrastructure, also the children of school age that are not enrolled in school are offered drug. School children absent the day of the drug administration are provided the drug when they return to school.
- CP3 (Administration of anthelminthic (ALB) once a year): The STH control program is consisting on the administration once a year of a tablets of albendazole (400 mg) to all school-age children. The drug administration is normally organized using the school infrastructure, also the children of school age that are not enrolled in school are offered drug. School children absent the day of the drug administration are provided the drug when they return to school.
- CP4 (Administration of anthelminthic (MEB) once a year): The STH control programme is consisting on the administration once a year of a tablets

- of mebendazole (500 mg) to all school-age children. The drug administration is normally organized using the school infrastructure, also the children of school age that are not enrolled in school are offered drug. School children absent the day of the drug administration are provided the drug when they return to school.
- CP5 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: administration of drug for LF (albendazole + ivermectin) + ALB at six months interval from the LF intervention): The STH control programme is integrated with the control of Lymphatic Filariasis. The drug administration is based on the administration once a year of albendazole (400 mg) and ivermectin (150 μ g/Kg) to the entire population plus the administration, at approximately 6 months of interval, of albendazole (400 mg) to all school age children.
- CP6 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: administration of drug for LF (albendazole + ivermectin) + MEB at six months interval from the LF intervention): The STH control programme is integrated with the control of Lymphatic Filariasis. The drug administration is based on the administration once a year of albendazole (400 mg) and ivermectin (150 μ g/Kg) to the entire population plus the administration, at approximately 6 months of interval, of mebendazole (500mg) to all school age children.
- CP7 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: (ALB) at six months interval from the administration of drug for LF (albendazole + DEC)): The STH control program is integrated with the control of Lymphatic Filariasis. The intervention is based on the administration once a year of albendazole (400 mg) and dyethilcarbamazine (6 mg/Kg) to the entire population plus the administration of albendazole (500mg) to all school age children.
- CP8 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: (MEB) at six months interval from the administration of drug for LF (albendazole + DEC)): The STH control program is integrated with the control of Lymphatic Filariasis. The intervention is based on the administration once a year of albendazole (400 mg) and dyethilcarbamazine (6 mg/Kg) to the entire population plus the administration of albendazole (500mg) to all school age children.
- CP9 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: administration of drug for LF (albendazole + DEC) only): The STH control program is already covered by the control of Lymphatic Filariasis. The intervention is based on the administration once a year of albendazole (400 mg) and dyethilcarbamazine (6 mg/Kg).
- CP10 (Administration of anthelminthic for STH in the context of lymphatic filariasis (LF) control program: administration of drug for LF (albendazole + IVR) only): The STH control program is already covered by

the control of Lymphatic Filariasis. The intervention is based on the administration once a year of albendazole (400 mg) and ivermectin (150 μ g/Kg).

Define time steps

This is the last step with the modeling section. Users can choose the time intervals (in years) for prediction. The default time intervals is 10 years (Figure 0.16) as it is convinced that a period of 10 years is good enough for prediction and decision making regarding the determination of the optimal set of control programs.

After selecting the time interval, users can click the button "Finish" shown in Figure 0.16 to complete the task.

2.3. Display results

Once users click the button "Finish" shown in Figure 0.16, the program will perform calculation and prediction. Results of this process will be automatically recorded in the output shapefile, which users have chosen earlier.

Both input and output shapefiles are loaded in QGIS environment and attributes of fields in each shapefile can be displayed as maps, with different color for the purposes of visualization and reporting.

The steps in displaying results are depicted in Figure 0.17.

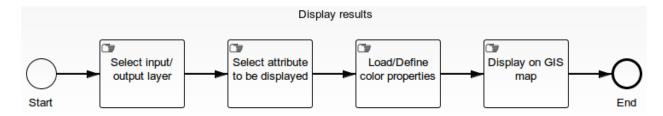


Figure 0.17: Display results

Select input/output layer

Under layer panel of QGIS, users can choose the layer, which is basically the input or output shapefile, to be shown as map in the main window.

Select attribute to be displayed

Move the cursor to the layer (e.g. input) and select it, **Right Click** on the selected layer and then select "Properties" (As shown in Figure 0.18)

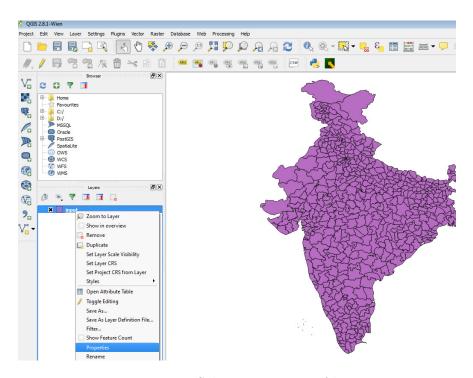


Figure 0.18: Select properties of layer

The properties box will appears as shown in Figure 0.19.

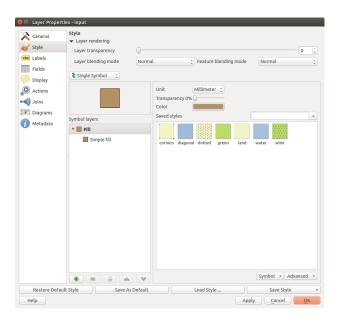


Figure 0.19: Properties box

Users can select "graduated" text and select the attribute to be displayed. The attribute here is the name of the column.

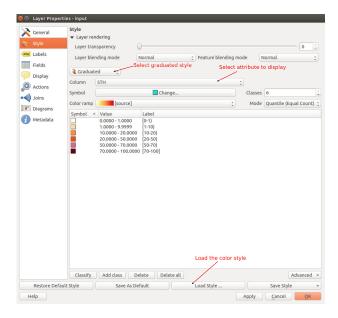


Figure 0.20: Select graduate style and choose the attribute

Load/Define color properties

Users can define a range of preferred colors for the map. For convenience of users, a default range of colors is predefined and it can be loaded (Figure 0.20).

The folder contains the default color style is shown in Figure 0.12

Display results on QGIS

Once the style is selected, the column name is defined, and the color style is loaded. Users can click the button "Apply" or "OK", the results will instantly display with different color on the map.

Following figures show the prevalence at the origin (Figure 0.21) and after one year (Figure 0.22)

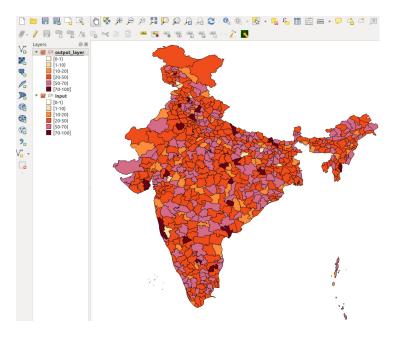


Figure 0.21: Prevalence map at the initial time

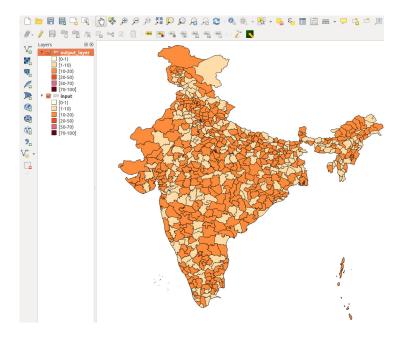


Figure 0.22: Prevalence map at after one year

As can be seen from the colored maps, there will be changes in total prevalence for entire India once the drug distribution programs (control programs) being implemented for each region.

3. Conclusion

This STH-predictor was developed as a Python Plugin for QGIS. It works with shapefile with polygon structure. The modeling path of it was with Markov model [1]. There are two models built for the plugin. The default one is when there is only total prevalence of STH recorded. This case is phenomena in actual practice. The other model is used only when data on intensity level of Hookworms, *Trichuris trichiura*, and *Ascaris lumbricoides* is available.

4. Acknowledgement

This work was financially supported by the Ministry of Health of the Republic of Korea.

Bibliography

[1] Montresor, A.; Gabrielli, A. F.; Yajima, A.; Lethanh, N.; Biggs, B.-A.; Casey, G. J.; Tinh, T. T.; Engels, D. & Savioli, L. "Markov model to forecast the change in prevalence of soil-transmitted helminths during a control programme: a case study in Vietnam". Transactions of The Royal Society of Tropical Medicine and Hygiene, 2013, 107, 313-318