

# **Data Warehouse And The Deployment Of Data Mining Process To Make Decision For Leishmaniasis And Development Of Mathematical Models For Phlebotomine Sandflies Seasonality In Marrakech City**

HABIBA MEJHED    SAMIA BOUSSAA    NOUR EL HOUDA MEJHED  
Département Génie Informatique

Laboratoire Micro Informatique Systèmes Embarqués Systèmes sur Puce.  
Ecole Nationale des sciences appliquées, BP. 575, avenue Abdlekrim Khatabi, Guéliz  
Université Cadi Ayyad, Marrakech  
MOROCCO

Laboratoire d'Ecologie et Environnement, Faculté des Sciences Semlalia, Université Cadi  
Ayyad, Marrakech, MOROCCO & Laboratoire de Parasitologie, Faculté de Pharmacie,  
Université Louis Pasteur-Strasbourg I,  
FRANCE

Département Génie Informatique,  
Ecole Nationale des sciences appliquées de Fès.  
Université Sidi Mohammed Ben Abdallah, Fès  
MOROCCO

[mejhed90@gmail.com](mailto:mejhed90@gmail.com)    [samiaboussaa@yahoo.fr](mailto:samiaboussaa@yahoo.fr)    [noramej@hotmail.com](mailto:noramej@hotmail.com)

*Abstract:* - This proposal is based on the implementation and the development of tools to analyze the decision making indicators in the epidemiology for the ministry of the health. First, we introduced new tools of data warehouse and data mining to improve epidemiological knowledge specific to the Leishmaniasis in south of Morocco. We developed an information system to facilitate the decision-making process, access to the information and data storage in the data warehouse. We considered the density of the sandflies population when taking into account the time and climate change. We are interested in patients with suspected Leishmania infection, so we included the transmission of Leishmaniasis to humans. The concept of Data mining is used to select, to explore and to transform our epidemic data to a prediction index. In order to reduce the transmission of leishmaniasis and to control the vector density, we also suggest developing a simple mathematical model with entomological data in Marrakech area. This model predicts the vector density according to local conditions. Simulations have been done on actual data collected in Marrakesh city to confirm the leishmaniasis risk in this area. In this paper we review the type of decision aids which has been successfully implemented and that we have provided to the responsible of the population health to make better decisions.

*Key-Words:* - Epidemiology, Information system, Database, Decision support, Data warehouse, Mathematical model, Morocco.

## **1 Introduction**

In the last decade, the epidemiology applied more and more tools to help to make decision. The aim is to translate epidemic data using the modelling concepts of health information systems for the decision-making. The information is a value-increasing necessary to plan and control the activities of an organism with effectively. It is the raw material that will be transformed by information systems. Often,

the availability of data makes it very difficult, if not impossible, to extrapolate the information that really matter. It is essential to have rapid and complete information needed for the decision-making process: the strategic indicators are extrapolated mainly operational data in a database, through a selection process or synthetic gradually. The widespread use of data analysis techniques has made the information system a strategic element and

policy framework for achieving the business. Therefore, the decision-making systems have emerged in the 80s (decision support system) and offer techniques and means to extract information from a set of memorized data. As a result, the volume of information collected during an epidemiological case study enables the development of new observing systems to analyze and extract some indicators as appropriate clinical decision and public health. The clinical decision support provided epidemiologist technologies necessary to facilitate this difficult task [1], [2], [13].

The data warehouse remains a valuable tool for storage and data accessibility, it is defined as a collection of information that integrates and reorganize the data from a variety of sources and make them available for analysis and assessment to scheduling and decision-making.

If the data warehouse used to store historical data, with the finality analysis, the data mining is defined as a process of exploration and modelling data in order to discover new correlations, to find trends and stable patterns in the data. It proposes a number of tools from different disciplines, in particular, to decision making in epidemiology [3], [4], [5], [6] [7],[25]. Data mining combines between various sciences domains (Databases, Statistics, Artificial Intelligence) to construct models from the data, and under the criteria fixed in advance and make a maximum of knowledge useful to make decision.

In the other hand, a lot of attention has been attracted to the mathematical modelling of epidemic systems [15], [18], [26], [27], [28]. The system is described by compartmental model using the SIR model [18], this model is based on the division of tree compartments: Susceptible, infective and removed individuals. One of the simplest SIR models is the Kermack-McKendrick model [21].

Leishmaniasis represents a complex of diseases with important clinical and epidemiological diversity. Leishmaniasis complex (parasite, vector and reservoir) is highly affected by bioclimatic factors.

In Morocco, leishmaniasis remains a severe public health problem. Many foci were described in rural areas of Ouarzazate [24], Essaouira [22], Azilal [11], Chichaoua [19] and Al Haouz [17], [23] but also in suburban areas, in Taza [20] and Fez [11] Marrakesh is an interesting study site because it lies close to the

focus of cutaneous leishmaniasis in the south of Morocco (Ouarzazate, Chichaoua and Al Haouz), and current studies have classified the area of Marrakesh as being at risk of cutaneous leishmaniasis [8], [9].

As the best choice of a vector-control strategy is dictated by sandfly ecology, we try to simplify this complex of diseases and quantify the climatic factors which can determine the distribution and activity of sandflies vectors in Marrakesh city. According to the WHO [10], the activity of sandfly fauna is affected by many climatic factors as temperature, humidity and wind, besides seasons and according to sandfly species.

This paper is a simple contribution to the fight against leishmaniasis in Morocco. The idea is to propose to epidemiologists an application based on tools of data warehouse and data mining to help them to make decision. In the first time, we conceived and modelled our information system to establish the pattern of the database on which we are going to work. Then, we develop a data warehouse to store and extrapolate data collected in Marrakech city, and we used a data mining tools for Leishmaniasis data analysis to get a better decision-making. In the second time, we present a simple mathematical model predicting the density of vectors according to local climatic conditions. We studied three forms of leishmaniasis according to sandfly vectors collected in Marrakech city by [8], [9]: *Phlebotomus papatasi* proven vector of zoonotic cutaneous leishmaniasis with Rodents as a reservoir; *P. sergenti* proven vector of anthroponotic cutaneous leishmaniasis and *P. longiculis* potential vector of visceral leishmaniasis with canine as reservoir hosts.

## 2 Material and Methods

### A. Sandfly collections

Specimens were collected in Marrakech city (31°36'N, 8°02'W, 471m a.s.l.) between October 2002 and September 2003 as described by [8]. The specimens caught were preserved in 70% ethanol. They were cleared in potash 20% and Marc-André solution, then dehydrated and mounted in Canada balsam [14]. The identification was made by examining the morphology of male genitalia, female spermathecae and pharynges. For *Larroussi*

species, we revised our specimens according to results of Boussaa [16].

### B. Data Analysis

We have a data file.xls containing all the information on the activity of sandflies *P. Papatasi*, *P. Sergenti* and *P. Longicuspis* based on climate change (date, temperature and density). We can present data from Excel files in the following diagrams:

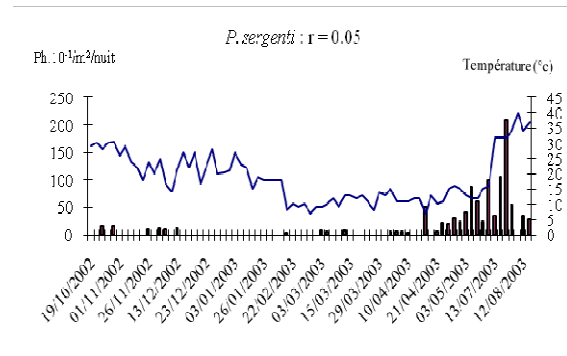
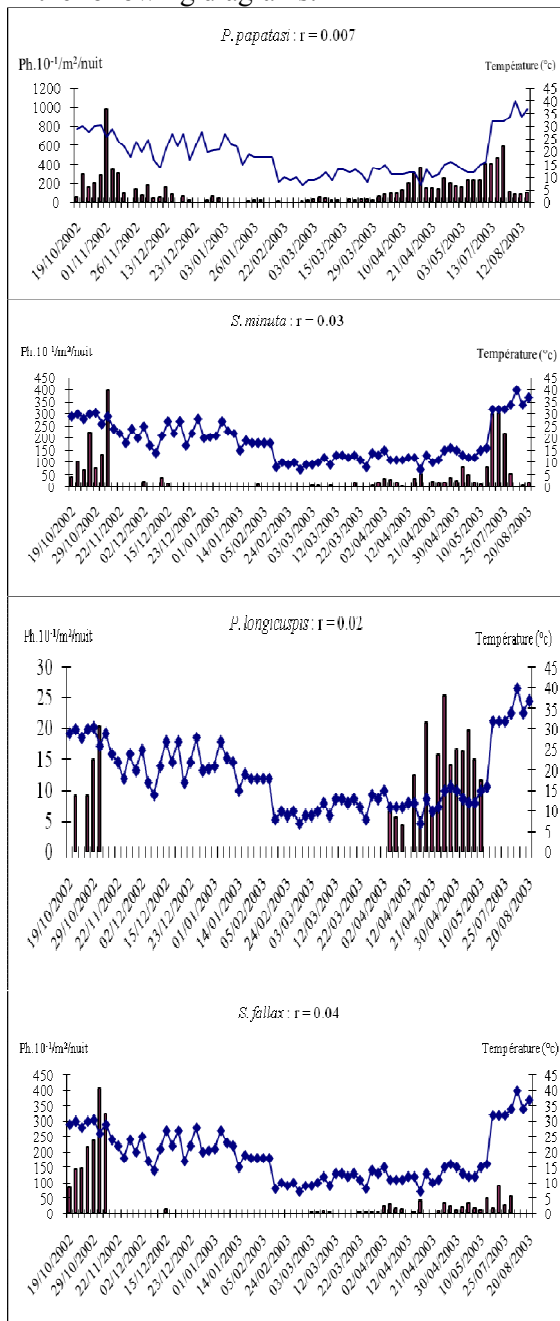


Fig. 1 : The activity of sandfly population in Marrakech area

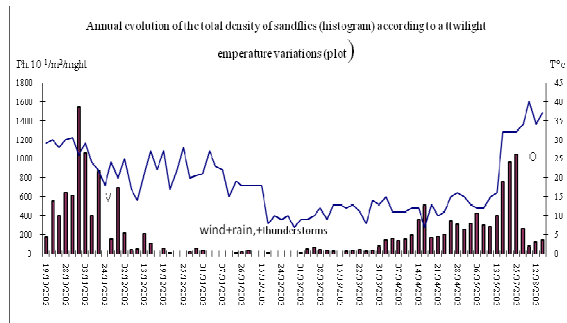


Fig. 2: Annual evolution of the global density of the sandflies vector according to the time and the climatic change

## 3 Development of a data warehouse (DW) Leishmaniasis

In this section we proceed to the conception of our data scheme. Our investigations on the sandflies collections were carried out in Marrakech city. The goal is to design a database, to develop DW and to load this database, and then the relationships in the cube are built automatically to give the answer to question posed in this case.

### 3.1 Background

The creation of a data warehouse involves several steps:

**The conception:** the implementation of a data warehouse usually begins by framing the project, define the needs and goals expressed by policymakers, modeling and designing a data structure. There are two data models, the star pattern, in this model, we must define one (or more) table (s) Fact with one or several measures (values of indicators). Both of them must have multiple dimension tables whose primary keys form the primary key tables done. Warning: The dimension tables are not linked.

Then the model snowflake which is derived from the star schema where the tables are standard size (of the table remains unchanged). With this scheme, each dimension is divided according to his (or her) hierarchy (s).

▪ **The acquisition of data:** The data will be extracted from the sources.

- The static extraction will be performed when the DW must be loaded for the first time and is conceptually a copy of operational data.
- The incremental extraction, is used for the periodic updating of the DW, and captures just the changes in data sources at the last extraction.

The choice of extracting data is based mainly on their quality; selection of data from the database is not a simple task to do.

**Data cleaning:** This phase will improve the quality of duplicate data, inconsistencies between the values logically related, missing data, unexpected use of a field and impossible or wrong value...

**Loading DW:** The loading of data in the DW is the process which permits to load the data cleaned and prepared in the DW.

### 3.2 Needs Specification

We have developed the analysis tools concerning, the population of sandflies, according to the various species listed in Marrakech city and their density, we considered the human population which may become affected after bites of infected sandflies.

These tools allow knowing the following information:

- The density of each species of sandflies listed
- The period at risk for the spread of the disease.
- The rate of infection of humans by infected sandflies.
- The rate per unit time, which a man loses her immunity and becomes susceptible.
- The rate of infected and susceptible in humans

To meet the needs of decision makers, we implemented the DW with respect to the architecture described in the following section.

### 3.3 Data warehouse Architecture

Schematic below shows the architecture of the data warehouse applied to Leishmaniasis data.

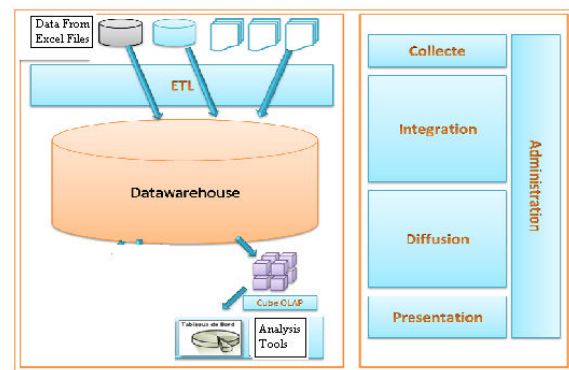


Fig. 3: data warehouse Architecture

To ensure a robust, flexible and portable solution, we adopted a software architecture divided into several parts:

1. Collection of raw data files: the job of this module is periodically connected to all servers, to check the generation of new data files
2. Conversion of the Files: The application of this module is written in Java. This module convert the data files collected by the module above, and filter the information contained therein. It permits to leave only the information that will be used in the future treatments.
3. Loading data into the data warehouse: ETL process allows Extraction, Transformation and Loading of data from various sources (databases, files) into DW. ETL process is the most important module to design a Data warehouse with respecting two constraints: data sources and data types (data quality).
4. Building of multidimensionnel cubes :

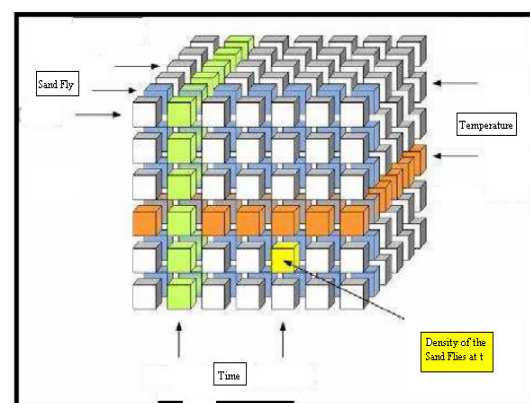


Fig. 4: Example of multidimensional view (Cube) of the Leishmaniasis data.

5. Creation of the graphical interfaces, graph and report: used for operation, querying cubes and creating reports.

### 3.4 Data model

This section deals with the transmission of the Leishmaniasis disease from the sandflies vector to human. There are four actors in this case, sandflies species, climatic change and Time and Human. The data dictionary given in the following table:

data	Description of data
Human	<ul style="list-style-type: none"> <li>➤ Sexe</li> <li>➤ Age</li> <li>➤ statut</li> </ul>
sandflies	<ul style="list-style-type: none"> <li>➤ specie</li> <li>➤ Name</li> </ul>
Temperature	<ul style="list-style-type: none"> <li>➤ Degree</li> </ul>
time	<ul style="list-style-type: none"> <li>➤ wear</li> <li>➤ Month</li> <li>➤ Day</li> </ul>

The data layer architecture of Leishmaniasis is illustrated schematically by:

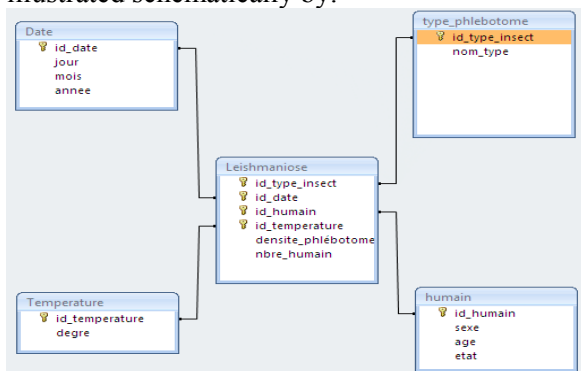


Fig 5: The model data warehouse diagram

The model chosen must comply with the requirements and needs of use, in our case, we opted for a star pattern respecting the nature of the information we have.

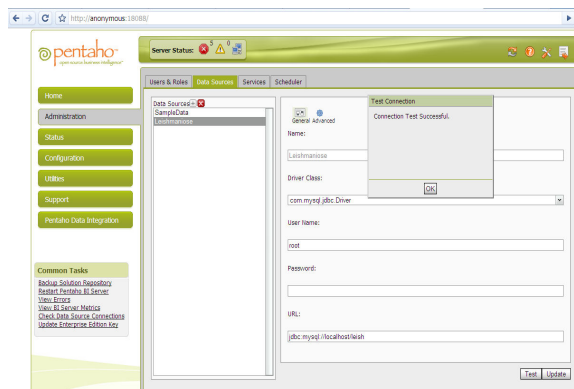
Dimension Tables:

- Date
- Temperature
- Sandflies (*P. Sergenti*, *S. Minuta*, *S. Fallax*, *P. Logicuspis*, *P. Papatasi*)
- Human

Fact Table:

- Leishmaniose

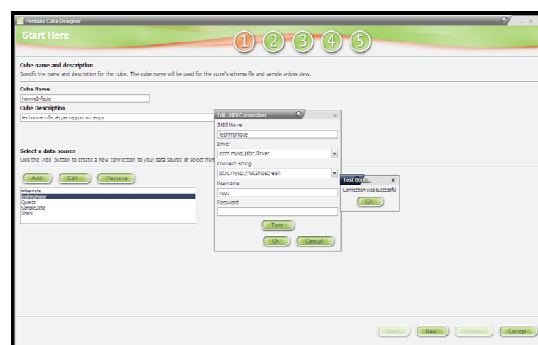
Data is extracted from Excel files using java code. The program consists of two classes, one for extraction and one for export useful data to the database.



### 3.5 Dimensional Cube:

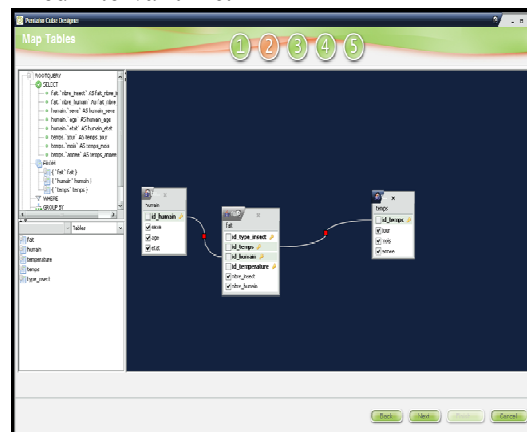
Pentaho integr a CubeDesign tool, it allows to have the cube in XML format and to publish it in the User Pentaho Console. Cube creation through 5 steps as shown in the below:

**Step1:** With the CubeDesigner we establish the connection to database Leishmaniose.



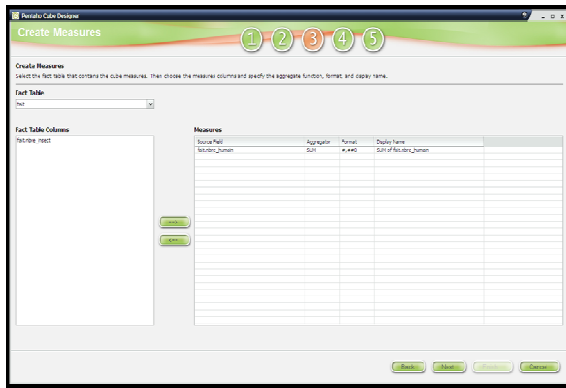
**Step 2 :**

Select tables and views useful to visualise the Cube. For example, to calculate the density of the *P. Sergenti* and the infected human in a fixed interval time.

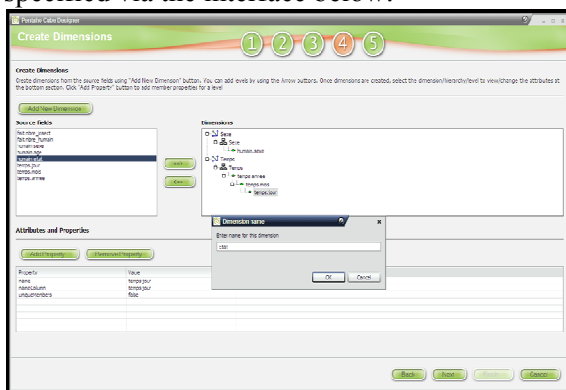


**Step 3:**

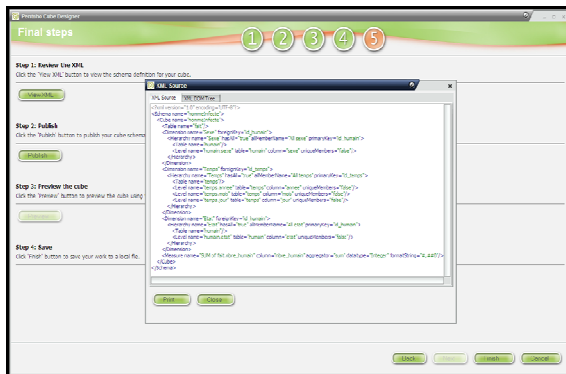
Choose of the measures: Infected human



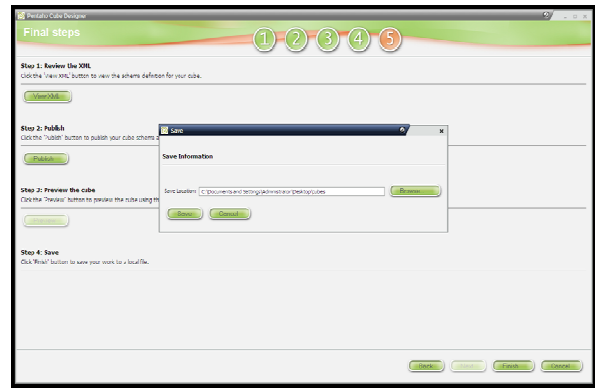
**Step 4 :**  
All the dimensions and their aggregation will be specified via the interface below:



**Step 5 :**  
The interface below allows us to view the file for the model; it's possible to update it as needed:

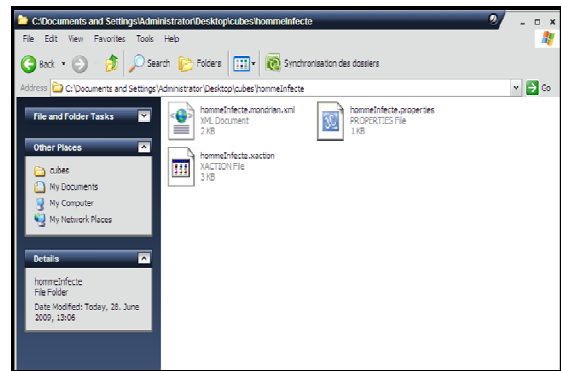


Finally, the pattern is saved:



Tree files are generated :

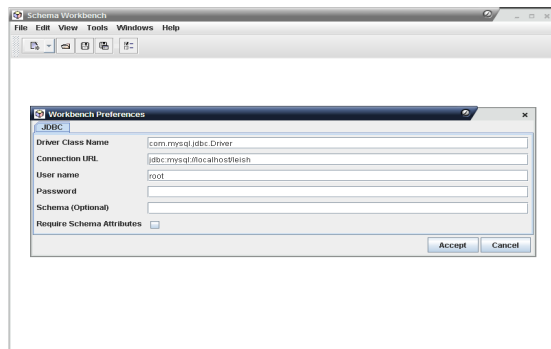
- Xml file to save the pattern produced by CubeDesigner.
- Properties file: for the allocation of the database
- Xaction file: presents a set of all protocols to data access.



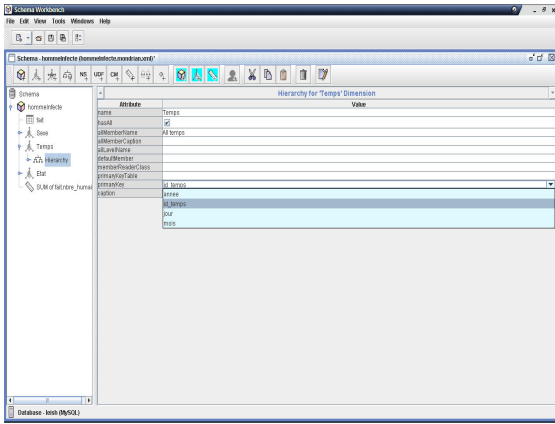
### 3.6 Pattern Publication with Workbench

Workben is a tool to create diagrams, for our case it is just used to refine and publish the pattern designed by CubeDesigner.

**Step 1 :**  
Configuration of the Workbench references to establish connection to database.



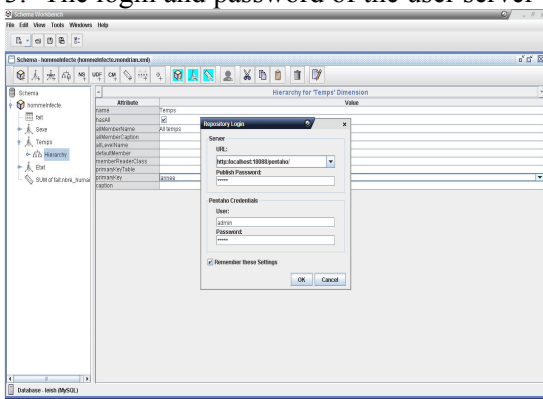
**Step 3 :**  
Pattern refinement.



**Step 4 :**

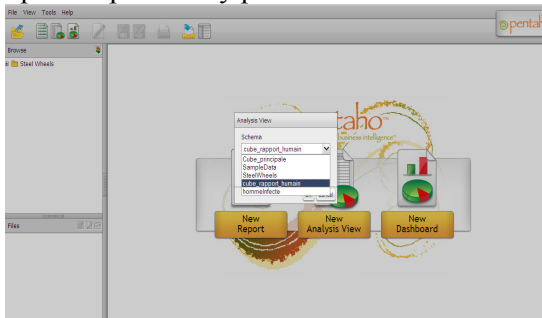
Now, it is important to have:

1. URL for the host user
2. The password for the publisher
3. The login and password of the user server

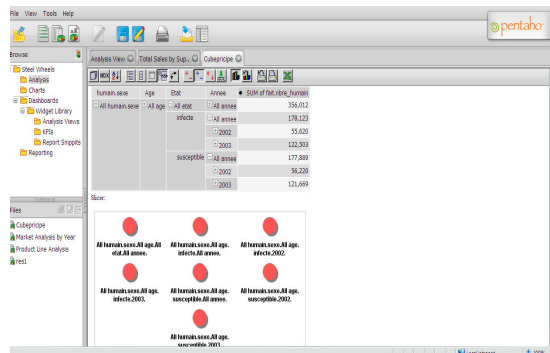
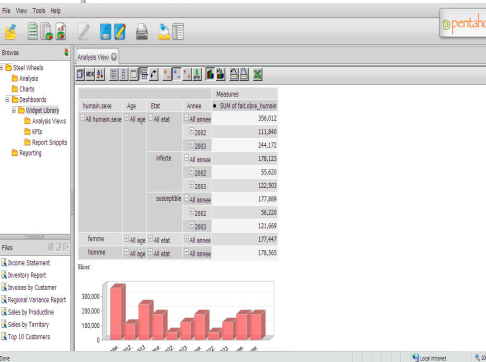
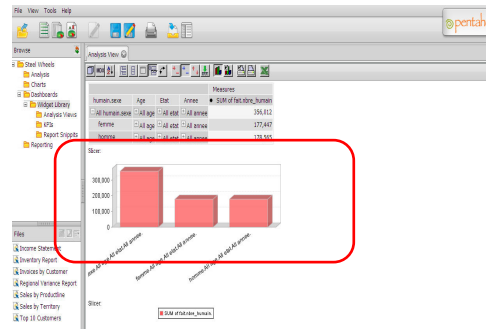
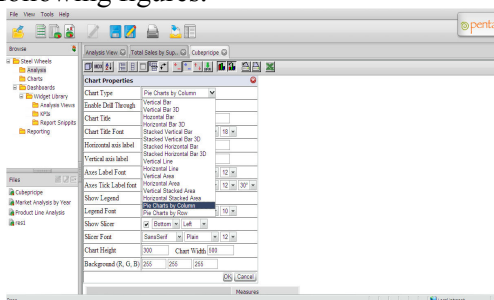


**3.7 Visualization phase**

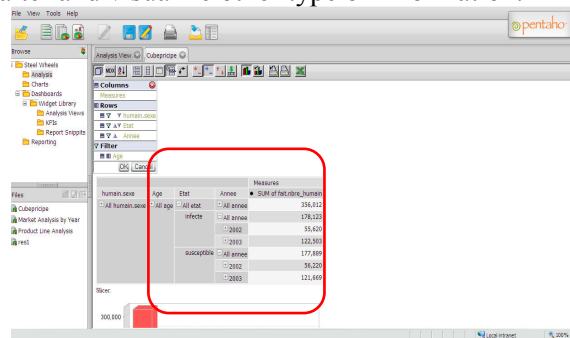
'User Console' tool gives different views of the pattern previously published.



It's possible to have several modes to visualize data from database. As illustrated in the following figures.



The health professional may use OLAP concept to filter and visualize other type of information.



**4 Data Mining: Application to the Leishmaniasis**

Given the seriousness of leishmaniasis in Morocco, it was essential to deploy easy methods to reduce its exploitable spread if not eradicate it completely. Our proposal aims to exploit tools of data mining process on this infectious disease. By definition, the data

mining attempts to extract knowledge from vast volumes of data.

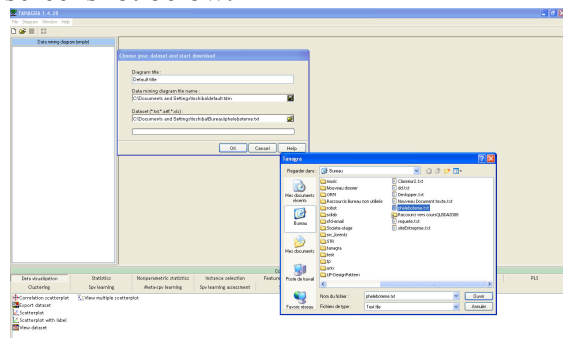
The wealth of information transmitted on vectors of disease allows us to apply these tools to identify methods and anticipate behaviour, therefore, make a good decision.

#### 4.1 K-Means and deployment

The technologies that are on the market, offers comprehensive platforms and integrated data analysis to meet all requests of indicators developed in the industry. We are able to accede to any type of data stored in our data base, to implement operations to analyze the data and present results in a need predefined by the user.

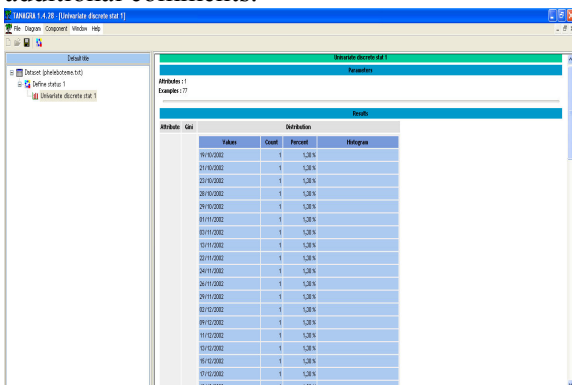
The software that we have developed our application offers a wide range of approaches ranging from methods of descriptive statistical analysis to predictive modeling methods.

The first step is to create a new diagram and import the data as shown in the screenshot below.



#### 4.2 Descriptive statistics

We can do descriptive statistics to variables. We calculate the frequency histograms on all columns to count the number of active and additional comments.



#### 4.3 Method principal component analysis (ACP)

They are three categories of data mining algorithms: supervised methods, unsupervised

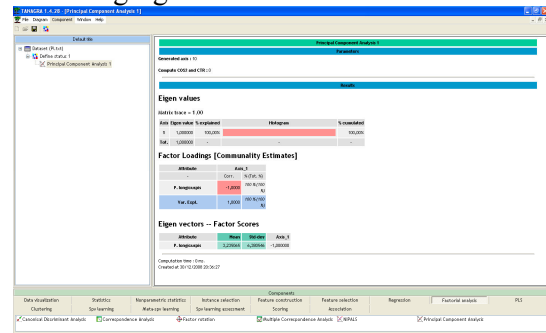
methods and methods of data reduction. Each category is based on a number of techniques.

In this section, we chose the third type using the method of principal component analysis (ACP).

Given a set of observations described by variables exclusively digital ( $x_1, x_2, \dots, x_p$ ), the APC aims to describe the same data set with new variables in reduced numbers. These new variables will be linear combinations of original variables. Principal component analysis can therefore be seen as a technique to reduce dimensionality.

#### 4.4 Visualization of our data

To implement the ACP method, we can see, for example, date and temperature data concerning *P. longicuspis*. After we define an analysis of the variables studied. The result is given in the following figure:



To better assess the relative positions to sandflies in the first factorial design, we add the component display. We put abscissa variable representing the first axis, calculated using the ACP, and ordered the second axis. We get the point cloud :

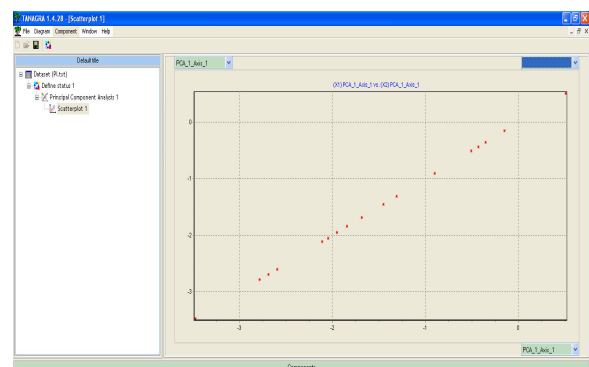


Fig. 6: Point cloud data on (date, temperature, for *P. longicuspis*)

Among the variables are, we want to check the effect of the variable date that can distort our results. We will color the points according to this variable, we select as a variable



component as is shown in this illustrative visualization

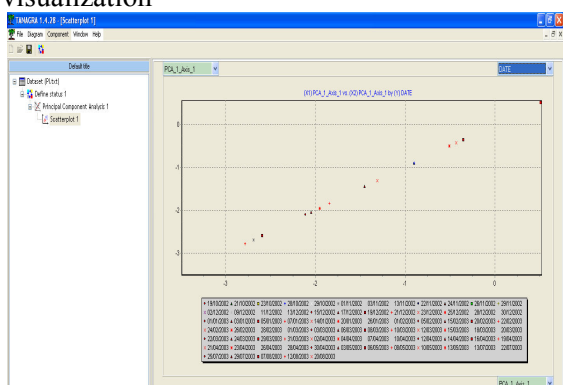


Fig. 7: The use of date\_format as illustrative variable

Now we apply the ACP method on all data:

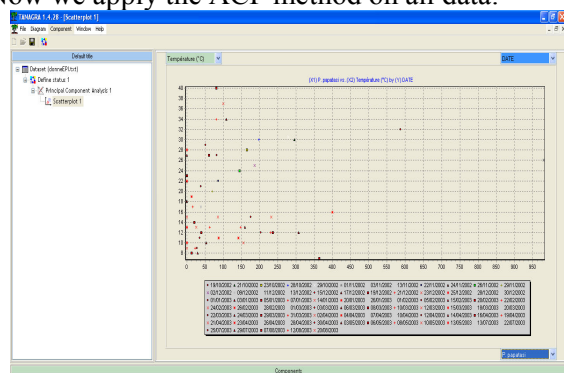


Fig. 8: Implementation of the ACP method on sandflies data

The most difficult decision in any project is to determine what method should be implemented. In our case, we prefer use the open source tools:

- ETL « Talend Open Studio » Talend Open Studio offer: A complete range of components, Traces and statistics of treatment in real time, the addition of specific code and the integration on the decisional open source.
- MySQL to create the database and data warehouse for the persistent storage area.
- Pentaho BI Suite: cover many areas of Business intelligence through various software (owned by Pentaho or integrated into).
- Pentaho Analysis (Mondrian + JPivot), Cube Designe, schema Workbench. For analysis OLAP
- Open source data mining software Tanagra.

## 5. Development of Mathematical models

The mathematical models developed in this section were performed according to model of Ross and the extended model (SIR) developed by [21].

Numeric simulation was carried out by Scilab software ([www.scilab.org](http://www.scilab.org)). It is a program for providing a numerical computing environment for scientific applications. Its syntax and its opportunities are similar to Matlab software.

### 5.1 Model of Anthroponotic Cutaneous Leishmaniasis

Based on the mathematical model of Ross [21], we considered:

Susceptible (S1 human class and S2 sandfly class): population capable of contracting the disease and becoming infective.

Infective (I1 Human class and I2 sandfly class): they are capable to transfer the disease to susceptible.

The transfer of anthroponotic cutaneous leishmaniasis from a class to another with respect to SIR model is illustrated by the scheme in Fig. 9.

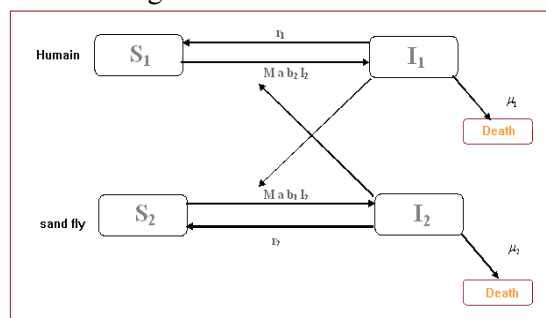


Fig. 9: Schematic representation of the transfer of anthroponotic cutaneous leishmaniasis via SIR model

- a: number of biting per sandfly at time t.
- b<sub>1</sub>: rate of biting per not infectious sandfly which can give infection for vector.
- b<sub>2</sub>: rate of infectious biting which can give infection for human.
- M: sandfly density per human.
- r<sub>1</sub>: rate of human recovery.
- r<sub>2</sub>: probability of infectious sandfly becomes uninfected (to rid of parasite).
- μ<sub>1</sub>: rate of mortality of humans.
- μ<sub>2</sub>: rate of mortality of sandflies.

we get the following equations:

$$\begin{cases} \frac{dI_1}{dt} = \frac{ab_2}{H} I_2 (H - I_1) - (r_1 + \mu_1) I_1 \\ \frac{dI_2}{dt} = \frac{ab_1}{H} I_1 (M - I_2) - (r_2 + \mu_2) I_2 \end{cases} \quad (S)$$

To simplify our equation, we put:

$$i_1 = \frac{I_1}{H} \quad i_2 = \frac{I_2}{V} \quad m = \frac{V}{H}$$

With V is population of sandflies and H is population of humans

Hence:

$$(S) \Leftrightarrow \begin{cases} \frac{di_1}{dt} = ab_2 \frac{V}{H} i_2 (1 - i_1) - (r_1 + \mu_1) i_1 \\ \frac{di_2}{dt} = ab_1 i_1 (1 - i_2) - (r_2 + \mu_2) i_2 \end{cases}$$

For leishmaniasis, we have:

$$\begin{cases} \mu_1 \ll r_1 \\ r_2 \ll \mu_2 \end{cases}$$

Set:  $r = r_1$  and  $\mu = \mu_2$ , the previous equation is of the form:

$$(S) \Leftrightarrow \begin{cases} \frac{di_1}{dt} = mab_2 i_2 (1 - i_1) - r i_1 \\ \frac{di_2}{dt} = ab_1 i_1 (1 - i_2) - \mu i_2 \end{cases}$$

with :  $\frac{di_1}{dt} = \dot{x}$      $\frac{di_2}{dt} = \dot{y}$      $\beta_1 = mab_2$   
 $\beta_2 = ab_1$

The system becomes:

$$(S) \Leftrightarrow \begin{cases} \dot{x} = -\beta_1 xy + \beta_1 y - rx \\ \dot{y} = -\beta_2 xy + \beta_2 x - \mu y \end{cases}$$

We have the equilibrium system if the infectious were eliminated, therefore:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Or

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

But the second case means that we don't have any infection. In this phase, we can use the

model of Ross then you obtain the factor  $R_0 = \frac{\beta_1 \beta_2}{\mu r} = \frac{ma^2 b_1 b_2}{\mu r}$ .

### 5.2 Model of Zoonotic Cutaneous Leishmaniasis and Visceral Leishmaniasis

We try to extend the model developed previously by considering the compartment of reservoir (Canine or rodent). We get the following representation (Fig. 10).

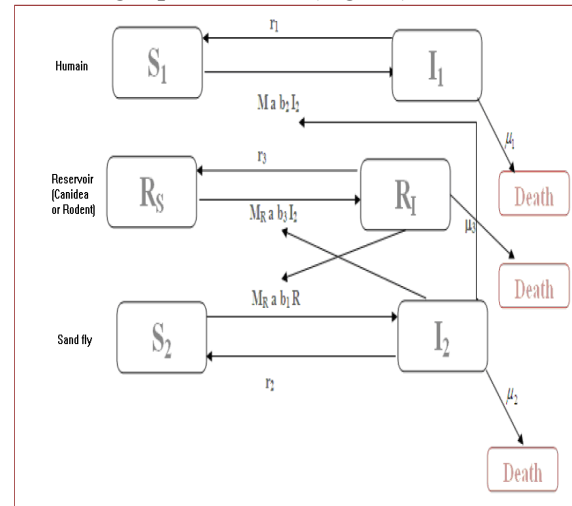


Fig. 10: Schematic representation of zoonotic cutaneous leishmaniasis and visceral leishmaniasis with SIR model

- a : number of biting per sandfly at time t.
- b<sub>1</sub>: rate of biting per not infectious sandfly which can give infection for vector.
- b<sub>2</sub>: rate of infectious biting which can give infection for human.
- b<sub>3</sub>: rate of biting which can give infection for reservoir. V: sandfly population.
- H : human population.
- R: reservoir population (Canidae or Rodent).
- M=V/H: sandfly density per humans.
- MR=V/R: sandfly density per reservoir (Canidae or Rodent).
- r<sub>1</sub> : rate of human recovery.
- r<sub>2</sub> : Probability of infectious sandfly becomes uninfected (to rid of parasite).
- r<sub>3</sub> : Probability of reservoir becomes uninfected.
- μ<sub>1</sub> : Rate of mortality of humans.
- μ<sub>2</sub> : Rate of mortality of sandflies.
- μ<sub>3</sub> : Rate of mortality of reservoir.

According to the same reasoning, we get the following mathematical model:

$$\left\{ \begin{array}{l} \frac{dI_1}{dt} = \frac{ab_2}{H} I_2(H - I_1) - (r_1 + \mu_1)I_1 \\ \frac{dI_2}{dt} = \frac{ab_1}{R} R_1(V - I_2) - (r_2 + \mu_2)I_2 \\ \frac{dI_s}{dt} = \frac{ab_3}{R} I_2(R - R_1) - (r_3 + \mu_3)R_1 \end{array} \right.$$

We apply the same method for get the factor

$$R_0 \text{ following: } R_0 = \frac{Va^2b_1b_3}{R\mu r}$$

The vector seasonality can be explained by the variation of factor a (number of biting per sandfly) at time t.

According to factor  $R_0$ , we have:

if  $R_0 \leq 1$  : leishmaniasis disappear

if  $R_0 > 1$  leishmaniasis increases

### 6. Simulation of factor $R_0$

We present the results of numeric simulation for zoonotic cutaneous leishmaniasis (Fig. 11), visceral leishmaniasis (Fig. 12) and anthroponotic cutaneous leishmaniasis (Fig. 13).

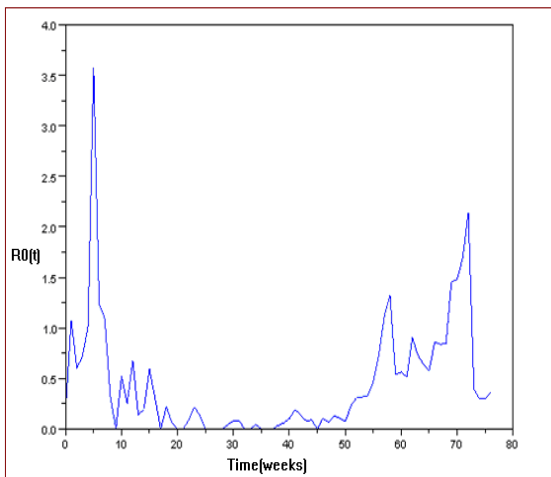


Fig. 11: The behaviour of  $R_0(t)$  Zoonotic cutaneous leishmaniasis (*Phlebotomus papatasi*)

With  $t=[0:76]$ ;  $R=5000$ ;  $MR=V/R$ ;  $a=30$ ;  $b_1=0.74$  ;  $b_2=0.74$ ;  $\mu=3$ ;  $r=0.3$ ;

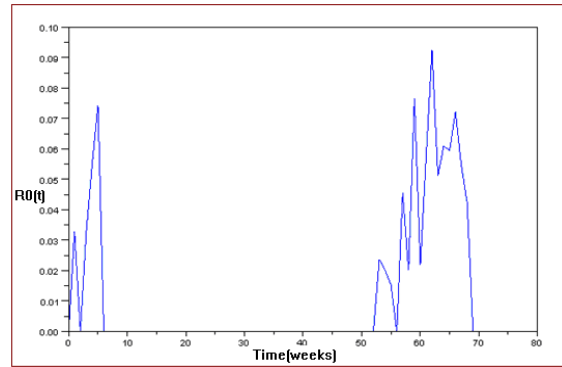


Fig. 12: The behaviour of  $R_0(t)$  Visceral leishmaniasis (*Phlebotomus longicuspis*)

With  $t=[0:76]$ ;  $R=5000$ ;  $MR = V/R$ ;  $a=30$ ;  $b_1=0.74$  ;  $b_2=0.74$ ;  $\mu=3$ ;  $r=0.3$ ;

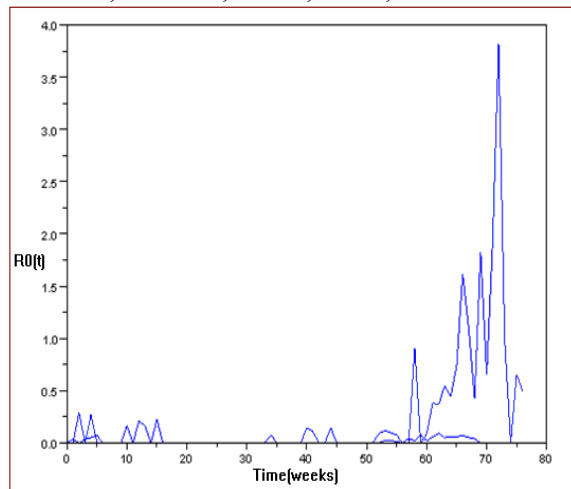


Fig. 13: The behaviour of  $R_0(t)$  for anthroponotic cutaneous leishmaniasis (*Phlebotomus sergenti*)

With  $t=[0:76]$ ;  $h=1000$ ;  $m=V/h$ ;  $a=30$ ;  $b_1=0.74$  ;  $b_2=0.74$ ;  $\mu=3$ ;  $r=0.3$ ;

We developed program that offers the possibility to calculate the factor  $R_0(t)$  in real time and to ensure rapid data entry via a simple graphic interface.

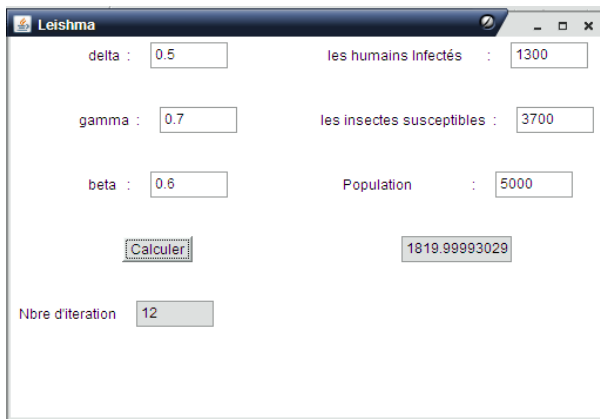


Fig. 14: Leishmaniasis Graphic Interface

## Discussion

Our investigations were conducted in Akioud, an urban district in Marrakech city, during one-year-study. This site was selected, considering the presence, of all the sandfly species inventoried in the urban area of Marrakech [9]. According to the correlation between the weekly density of the three vectors (*P. papatasi*, *P. sergenti* and *P. longicuspis*) and the factor  $R_0(t)$ , we can prevent the risk of leishmaniasis in this area.

- For *P. papatasi* population,  $R_0(t)$  factor is superior to 1 during two periods of the year: November and May-June-July, which correspond to the periods of risk of zoonotic cutaneous leishmaniasis caused by *L. major* in this area.
- For *P. sergenti* population,  $R_0(t)$  is superior to 1 during the period of July–August and inferior to 1 in the rest of the year. So, this period corresponds to the phase of risk of anthroponotic cutaneous leishmaniasis caused by *L. tropica* in this area. We observe that  $R_0(t)$  reaches its peak during the period of August when the temperature is very high.
- For *P. longicuspis*, the results have shown two periods of risk of visceral leishmaniasis in this area: October and May-June.

[8] classified Marrakech area as being at risk of cutaneous leishmaniasis because of the high density of *P. papatasi* throughout the year, its position close to the cutaneous leishmaniasis foci in the arid region [24] and the omnipresence of *Meriones shawi*, main *L. major* reservoir host in Morocco. In this paper,

we conform, by mathematical models, the conclusion of [8] and we exploit these data to assist in the decision on the issue of the fight against leishmaniasis in Morocco.

Relying on the principles of epidemiology modelling, we developed a mathematical model for each type of leishmaniasis. These models can be used to analyze the spread of leishmaniasis in any area, particularly in Marrakech city. Indeed, these are spatio-temporal models which permit to follow and control the evolution of leishmaniasis in terms of time and region, and to find the appropriate threshold of the population of sandflies for stopping the multiplication of the disease. Our model suggests that it is not necessary to completely eliminate the vectors but the outbreak could be stopped if the vector populations were reduced during periods of risk. In Chichaoua (70 Km from Marrakech city), focus of anthroponotic cutaneous leishmaniasis, [15] suggest that the epidemic could be stopped if the vector population were reduced by a factor  $(R_0)^2 = 3.76$ .

## 4 Conclusion

To fight against the spread of the cutaneous and visceral leishmaniasis and to address the need of the responsible for population health to set the policies that determine the nature of health care provided, and most importantly, for fast evaluation of the severity of the multiplication of the disease, we proposed to develop a model of data warehouse to storage appropriate data of the sandflies seasonality and the information about the susceptible human (patient with suspected Leishmania infection). Then, structural key and functional parameters will be measured to ensure the advanced treatment. By employing data mining process we have the ability to extract knowledge and to generate summaries for better health decision-making. Additionally, we developed a mathematical model that combines deterministic SIR model [18] with the existence of the vectors that transfer leishmaniasis. Through this complete and operational platform, it will be easy to control the sandflies seasonality and the rate of the transmission disease.

References:

- [1] Informatique et santé Collection dirigée par P. Degoulet et M. Fieschi, Paris, Springer-Verlag France A. Albert, F.H. Roger-France, P. Degoulet et M. Fieschi Volume 10, 1998
- [2] Gilbert Saporta, *DATA MINING ou FOUILLE DE DONNÉES*, RST « Epidémiologie » Data Mining, 2004
- [3] Daniel T. Larose (adaptation française T. Vallaud): *Des données à la connaissance : Une introduction au data-mining* (1Cédérom), Vuibert, 2005.
- [4] René Lefébure et Gilles Venturi, *Data Mining : Gestion de la relation client, personnalisations de site web*, Eyrolles, mars 2001.
- [5] Pascal Poncelet, Florent Maseglia and Maguelonne Teisseire (Editors). *Data Mining Patterns: New Methods and Applications*, Information Science Reference, ISBN: 978-1599041629, October 2007.
- [6] Stéphane Tufféry, *Data Mining et Statistique Décisionnelle*, Technip, nouvelle édition revue et enrichie, juin 2007.
- [7] Egmont-Petersen, M., de Ridder, D., Handels, H., *Image processing with neural networks -a review*. Pattern Recognition 35: pp. 2279–2301. doi: 10.1016/S0031-3203(01)00178-9. 2002.
- [8] Boussaa, S., Guernaoui, S., Pesson, B., Boumezzough, A., *Seasonal fluctuations of phlebotomine sand fly populations (Diptera: Psychodidae) in the urban area of Marrakech, Morocco*. Acta Trop. 95, pp. 86–91. 2005.
- [9] Boussaa, S., Pesson, B., Boumezzough, A., *Phlebotomine sandflies (Diptera: Psychodidae) of Marrakech city, Morocco*. Ann. Trop. Med. Parasitol. 101, pp. 715–724. 2007.
- [10] WHO. *Lutte contre les leishmanioses*. Série de Rapports Techniques. 2005.
- [11] Rhajaoui, M., Fellah, H., Pratlong, F., Dedet, J.P., Lyagoubi, M., *Leishmaniasis due to Leishmania tropica MON-102 in a new Moroccan focus*. Trans. R. Soc. Trop. Med. Hyg. 98, pp. 299–301. 2004.
- [12] Rioux et al., *Leishmania. Taxonomie et Phylogénèse*. J. A. pp. 365–395. 1986
- [13] TEH All, *Development of a Data warehouse for Lymphoma Cancer Diagnosis and Treatment Decision Support*. Proceedings of the 10th WSEAS International Conference on MATHEMATICS and COMPUTERS in BIOLOGY and CHEMISTRY, Pp. 15–24. ISSN: 1790-5125, ISBN: 978-960-474-062-8. 2009.
- [14] Abonnenc, E., *Les phlébotomes de la région éthiopienne (Diptera: Phlebotomidae)*. Mémoire de l'ORSTOM. 55, pp. 1–289. 1972.
- [15] Bacaër, N., Guernaoui, S., *The epidemic threshold of a simple seasonal model of cutaneous leishmaniasis*. J. Math. Biol. 53, pp. 421–436, 2006.
- [16] Boussaa, S., Boumezzough, A., Remy, P. E., Glasser, N., Pesson, B., *Morphological and isoenzymatic differentiation of Phlebotomus perniciosus and Phlebotomus longicuspis (Diptera: Psychodidae) in Southern Morocco*. Acta Trop. 106, pp. 184–189. 2008.
- [17] Boussaa, S., Pesson, B., Boumezzough, A., *Faunistic study of the sandflies (Diptera: Psychodidae) in an emerging focus of cutaneous leishmaniasis in Al Haouz province, Morocco*. Ann. Trop. Med. Parasitol. 103, pp. 73–83. 2009
- [18] Chaves, L.F., Hernandez, M.J., *Mathematical modelling of American Cutaneous Leishmaniasis: incidental hosts and threshold conditions for infection persistence*. Acta Trop. 92, pp. 245–25. 2004.
- [19] Guernaoui, S., Boumezzough, A., Pesson, B., Pichon, G., *Entomological investigations in Chichaoua: an emerging epidemic focus of cutaneous leishmaniasis in Morocco*. J. Med. Entomol. 42, pp. 697–701. 2005.
- [20] Guessous-Idrissi, N., Chiheb, S., Hamdani, A., Riyad, M., Bichichi, M., Hamdani, S., Krimech, A., *Cutaneous leishmaniasis: an emerging epidemic focus of Leishmania tropica in north Morocco*. Trans. R. Soc. Trop. Med. Hyg. 91, pp. 660–663. 1997.
- [21] Kermack, W.O., McKendrick, A.G., *A Contribution to the Mathematical Theory of Epidemics*. Proc. Roy. Soc. Lond. 115, pp. 700–721, 1927.
- [22] Pratlong, F., Rioux, J.A., Dereure, J., Mahjour, J., Gallego, M., Guilvard, E., Lanotte, G., Périères, J., Martini, A., Saddiki, A., *Leishmania tropica au Maroc*. IV. Diversité isozymique intrafocale. Ann. Parasitol. Hum. Comp. 66, pp. 100–104. 1991.

- [23] Ramaoui, K., Guernaoui, S., Boumezzough, A., *Entomological and epidemiological study of a new focus of cutaneous leishmaniasis in Morocco*. Parasitol. Res. 103, pp. 859-863. 2008.
- [24] Rioux, J.A., et all, *Les leishmanioses cutanées du bassin méditerranéen occidental: de l'identification enzymatique à l'analyse éco-épidémiologique, l'exemple de trois 'foyers', tunisien, marocain et français*. Montpellier, France: Institut Méditerranéen d'Etudes Épidémiologiques et Ecologiques. pp. 365-395. 1986.
- [25] Hang Xiaojiao, Xiubin Zhang, *Comparison Studies on Classification for Remote Sensing Image Based on Data Mining Method*, WSEAS TRANSACTIONS on COMPUTERS. Volume 7, ISSN: 1109-2750, pp. 552-558, 2008.
- [26] Capasso V., *Mathematical structures of Epidemic Systems*. Lecture Notes in Biomathematics, Springer-Verlag, Berlin, 1993.
- [27] Milhail V. and Octarvian Pastravanu, *Exploiting The Componentwise Absolute Stability Of Endemic Epidemic Systems Via SR Models*. CEAI, Vol. 4, pp. 23-28. 2002.
- [28] Silvia Martorano Raimundo, *Drug Resistant Impact on Tuberculosis Transmission*, Proceedings of the 9th WSEAS International Conference on Mathematics & Computers In Biology & Chemistry, Bucharest, Romania , ISBN ~ ISSN:1790-5125 , 978-960-6766-75-6, Pages 176-181, 2008.