Spatio-temporal probabilistic prediction of appearance and duration of malaria outbreak in municipalities of Colombia

J O Rodríguez Velásquez

Insight Research Group, Bogotá, Colombia

E-mail: grupoinsight2025@yahoo.es

Abstract. 10 Colombian municipalities were chosen and the numbers of malaria outbreaks and of its duration for each year between 2003-2007 were analyzed. The probability of number of malaria outbreaks of each year and of its duration were calculated, and both probability values were combined to obtain the appearance probability of the number of outbreaks with a specific duration at epidemiological weeks. Later, some possible annual dynamics of outbreak and non-outbreak were simulated, calculating in time, with the recurrent temporal probability, the possibility that an outbreak appears or not. The whole of possible series of dynamics of an epidemic were calculated, and some were assessed with the cumulative probability and its difference in the context of probability bunch. The outbreaks with duration of one week had the higher occurrence frequency, the probabilistic analysis showed a loaded behavior. The simulations made showed the prediction ability of methodology. Of whole the O-N series possible, only 1431 are the most probable. It was revealed a probabilistic acausal mathematical order apply to week to week to spatio-temporal prediction of outbreaks number and its duration in Malaria epidemic, in the municipalities of Colombia, applicable to any municipality of the universe, useful in public health decisions making.

1. Introduction
Malaria, also called Paludism, is an infectious disease transmitted by arthropods; it affects tropical and subtropical regions around the world [1,2]. In 2016, there was risk of Malaria transmission in 91 areas or countries [3]. WHO estimates that each year occurs 250 million cases of Malaria [1], and almost a million deaths, being Sub-Saharan Africa’s countries the most affected [1,3,5]. In Colombia, for 2010 the mortality and complicated Malaria cases increased in percentages greater than 90% [4], there are municipalities in Colombia with higher API, by this, study malaria dynamics in the municipalities is of great interest.

The malaria epidemic is a problem of public health, which requires investigations that help to improve their surveillance systems [5]. There were statistical, mathematical and neural networks models, which seek to make predictions considering different causal factors, but its results can’t be generalized to other populations [6-13]. In contrast, studies that have looked for analyze and predict the dynamics of epidemics from the theoretical-physics’ causeless perspective, applying physical and mathematical laws and theories [14-17], this have allowed to predict the epidemic behavior through of infected number, this type of methodologies have allowed improve the prediction’s time to 3 weeks [19], overcoming models such as epidemiological channels which needs data of about 5 to 7 previous years for made estimations about dynamic epidemic [18]. The present work is based on probability theory, this allows to quantify the possible occurrence of an event in the future, with results ranging from 0 to 1 [20-23]. The purpose of this work is to develop a spatio-temporal predictive methodology.
Based on probability theory, which allows to establishing predictions of the number of outbreaks and its duration, allowing to simulate possible annual dynamics of outbreak and not-outbreak to establish real-time predictions of the epidemic that being useful in public health decisions making.

2. Definitions

2.1. Outbreak (O) and non-outbreak (N) series
Spatio-temporal sequence, in each municipality, on the dynamics of the epidemic consisting of 53 weeks, each week will have two possible states: Outbreak (O) and non-outbreak (N).

2.2. Recurrent temporal probability
This probability is calculated over time, starting from the occurrence of an outbreak (O) and not-outbreak (N) in a specific series, so that as the weeks number rises, the probability value varies according with apparition of N or O, quantifying how likely they are to find an O in the following weeks.

2.3. Probability bunch
Whole of possible O-N series with a specific quantity of outbreak, i.e., with the passing of the weeks can occur a O or N, generating different types of annual dynamics, all those possible O-N series will called probability bunch.

2.4. Probability cumulative (Pc) difference
Difference between Pc value (Equation 5) found at the end of the series (on week 53), and the value of Pc in the week preceding the first week of the series, in which the recurrent temporal probability is one.

3. Materials and methods
Starting from the Monitoring System and Early Warning (SIVIGILA by its name in Spanish) database of the National Institute of Health in Colombia for the period 2003-2007, was evaluated the number of infected per week during 53 epidemiological registered weeks. Of 820 Colombian municipalities were selected 10, which were divided into four groups according to their number of infected between 2003-2007, these municipalities were chosen since they presented specific dynamics that are prototypes from municipal dynamic in Colombia; to choose the groups, municipalities were ordered according infected number during the years 2003-2007, the groups were formed according their positions; group A positions 1 to 6 (Tierralta, Cáceres, Puerto libertador, Buenaventura, Monte Líbano y Turbo); group B positions 10 y 11 (Necoclí y Taraza); group C position 193 (Tuluá) and group D position 402 (Los Palmitos). A probability space of outbreaks was generated with the 53 epidemiological weeks of the years 2003-2007 (supplementary data), where the number of outbreaks presented in the annual dynamics of each municipality was calculated; thus, establishing the probability that an annual dynamic with different numbers of outbreak, this probability was calculated according to the Equation (1):

$$P_b = \frac{\text{Number of years that presented a certain number of outbreaks}}{\text{Total number of years from probability space}}$$

In this case, the denominator of this expression is 50 years; because for each of 10 municipalities taken, 5 years were studying, which multiplied gives 50 years. Then the outbreaks duration was studied, quantifying the number of outbreaks that lasted one week, two weeks, etc., so generating a frequencies distribution of outbreaks duration for the municipalities and period studied; and the next probability was defined according to the Equation (2),

$$P_f = \frac{\text{Determined Frequency of outbreaks duration}}{\text{Sum of all frequencies duration}}$$
Having these two probability spaces; \( \text{P}_b \) and \( \text{P}_f \), a new space was found combining the probabilities of the two spaces, in this way was found the probability \( \text{P}_{fb} \) of finding a certain number of outbreaks that have a specific duration of weeks (Equation 3), allowing to find the most likely configuration of the number of outbreaks and duration, and in general, find the probability distribution of the whole combinations.

\[
\text{P}_{fb} = \text{P}_f \times \text{P}_b
\]  

(Equation 3)

Subsequently, with the number of outbreaks per year, the number of all possible O-N series was calculated, by calculating the possible combinations that can be given with each of the values of outbreaks per year, taking into account the different dispositions which an outbreak may. The maximum values of outbreaks occurrence are taken from the municipalities data, and O-N series were generated calculating the recurrent temporal probability (see definitions) during 53 epidemiological weeks, evaluating the probability that on particular week "O" or "N" occur considering preceding O-N series to this week (Table 1). Later, for some of possible dynamics simulated, the recurrent temporal probability for each week was found, then probability of complete series was calculated, for this all week probabilities were multiplied (Equation 4):

\[
P_{\text{serie}} = \prod_{i=1}^{53} \text{P}_{\text{week } i}
\]  

(Equation 4)

Finally, with the 10 simulated dynamics, probability bunch (see definitions) were studied through of cumulative probability \( P_c \) and its difference, for this, two dynamics with two outbreak occurrence, and two dynamic with one outbreak occurrence were analyzed, then “cumulative probability” \( P_c \) was calculated with the Equation (5):

\[
P_{cj} = \sum_{n=1}^{j} P_n
\]  

(Equation 5)

Where \( P_{cj} \) is the cumulative probability of the week \( j \), and \( P_n \) is the recurrent temporal probability of the week \( n \).

4. Results

Only municipalities of the A group showed outbreaks. In the 50 years studied, the occurrence frequency of outbreaks distribution oscillates between 1 to 11 (Table 1).

**Table 1.** Distribution of the occurrence of outbreaks per year between 2003-2005 in Colombian municipalities. OFR: Occurrence frequency in years, ie, number of years in which a specific number of outbreaks (\# O) occurred, \# O: number of outbreaks and P: Probability, this probability was calculated with Equation 1.

| \# O | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|------|---|---|---|---|---|---|---|
| OFR (years) | 5 | 1 | 11 | 5 | 6 | 1 | 1 |
| P    | 0.1 | 0.02 | 0.22 | 0.1 | 0.12 | 0.02 | 0.02 |

It was also found from the 50 years studied (5 years for each municipality), that the frequencies of outbreaks duration varied between 1 and 48 (Table 2).
Table 2. Outbreak duration distribution for 2003–2007, in Colombian municipalities used in induction.
DO: Outbreak duration in weeks, FR: number of outbreaks found with a specific duration, SUM is the sum of all FR for different number of DO, and \( P_f \) is calculated by Equation 2.

| DO (weeks) | 1  | 2  | 3  | 4  | 6  | SUM |
|------------|----|----|----|----|----|-----|
| FR         | 48 | 10 | 12 | 2  | 1  | 73  |
| \( P_f \)  | 0.6575 | 0.1370 | 0.1644 | 0.0274 | 0.0137 |     |

Probability values associated with the number of outbreaks combination with its duration (Table 3) presented values between 0.1447 and 0.0003, showing that the epidemic dynamics is a non-equiprobable phenomenon, for example the probabilities of outbreaks number that may occur in a year varies between 0.02 and 0.22, this latter value correspond to the event that two annual outbreaks occur. Similarly, the duration of outbreaks probabilities (Table 2) not present an equiprobable behavior.

Table 3. Combination of probability spaces \( P_b \) y \( P_f \) the number of outbreaks (#O) and the duration of these. Number of outbreaks (#O) indicates the number of outbreaks that occur and Duration (D (weeks)) is the duration in weeks of outbreaks, the probability values (P) were calculated with Equation (4).

| D (weeks) | # O | 1   | 2   | 3   | 4   | 5   | 6   |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| 1         |     | 0.0132 | 0.1447 | 0.0658 | 0.0789 | 0.0132 | 0.0132 |
| 2         |     | 0.0027 | 0.0301 | 0.0137 | 0.0164 | 0.0027 | 0.0027 |
| 3         |     | 0.0033 | 0.0362 | 0.0164 | 0.0197 | 0.0033 | 0.0033 |
| 4         |     | 0.0005 | 0.0060 | 0.0027 | 0.0033 | 0.0005 | 0.0005 |
| 5         |     | 0   | 0 | 0 | 0 | 0 | 0 |
| 6         |     | 0.0003 | 0.0030 | 0.0014 | 0.0016 | 0.0003 | 0.0003 |

Combining outbreaks duration with number of outbreaks, can be observed how the probabilities have differences of \( 10, 10^2 \) y \( 10^3 \), this behavior can be found in the whole table, achieving quantifies temporarily the occurrence and duration of outbreaks with non-equiprobable probabilities showing a loaded behavior. 26,168,273 possible O-N series were found (Table 4), of these, in the reality, the most frequent were only 1.431.

Table 4. Total number of possible O-N series according to number of outbreaks (Number O), constituting probability bunches.

| O-N Series | Number O | 1     | 2     | 3     | 4     | 5     | 6     | Total |
|------------|----------|-------|-------|-------|-------|-------|-------|-------|
| Number of Series | 53 | 1.378 | 46.852 | 292.825 | 2.869.685 | 22.957.480 | 26.168.273 |

Ten O-N series were simulated from the evaluated data, these showed the predictive ability of the methodology, this can be seen in both Table 5; at some point in dynamics, this becomes completely deterministic dynamics, for example this happens in the week 8, the probabilities from this point show values of one, showing the impossibility of occurrence of an outbreak in this period.

Evaluating the probability for each series \( P_{sim} \), (Equation 4), it is observed that the values of this probability for the series having the same number of outbreaks are equal (Table 5). The values of \( P_c \) were between 0.925 and 51.59 for the ten O–N series simulated; to the calculate the \( P_c \) difference (see definitions), this can quantify differences between the dynamics with the same outbreaks number but different O–N series. These values show how quickly the dynamics becomes deterministic, i.e., how fast the dynamics can be determined by one specific series of probability bunch.
**Table 5.** Example of possible dynamics change with one outbreak. WN: (week number). O (in bold): Outbreak, N: Non-outbreak, FRAC: Expression of the numbers used on fraction to calculate the probability recurrent temporal (PROB). PC: Probability cumulative.

| WN | O-N | FRAC  | PROB | Pc  | WN | O-N | FRAC  | PROB | Pc  |
|----|-----|-------|------|-----|----|-----|-------|------|-----|
| 1  | O   | 1/53  | 0.019| 28  | N  | 26/26| 1     | 25.94|     |
| 2  | N   | 51/52 | 0.981| 1   | 29 | N   | 25/25 | 1    | 26.94|
| 3  | N   | 50/51 | 0.98 | 1.98| 30 | N   | 24/24 | 1    | 27.94|
| 4  | N   | 40/50 | 0.98 | 2.96| 31 | N   | 23/23 | 1    | 28.94|
| 5  | N   | 48/49 | 0.98 | 3.94| 32 | N   | 22/22 | 1    | 29.94|
| 6  | N   | 47/48 | 0.979| 4.94| 33 | N   | 21/21 | 1    | 30.94|
| 7  | N   | 46/47 | 0.979| 5.94| 34 | N   | 20/20 | 1    | 31.94|
| 8  | O   | 2/46  | 0.043| 6.94| 35 | N   | 19/19 | 1    | 32.94|
| 9  | N   | 45/45 | 1    | 7.94| 36 | N   | 18/18 | 1    | 33.94|
| 10 | N   | 44/44 | 1    | 8.94| 37 | N   | 17/17 | 1    | 34.94|
| 11 | N   | 43/43 | 1    | 9.94| 38 | N   | 16/16 | 1    | 35.94|
| 12 | N   | 42/42 | 1    | 10.94| 39 | N   | 15/15 | 1    | 36.94|
| 13 | N   | 41/41 | 1    | 11.94| 40 | N   | 14/14 | 1    | 37.94|
| 14 | N   | 40/40 | 1    | 12.94| 41 | N   | 13/13 | 1    | 38.94|
| 15 | N   | 39/39 | 1    | 13.94| 42 | N   | 12/12 | 1    | 39.94|
| 16 | N   | 38/38 | 1    | 14.94| 43 | N   | 11/11 | 1    | 40.94|
| 17 | N   | 37/37 | 1    | 15.94| 44 | N   | 10/10 | 1    | 41.94|
| 18 | N   | 36/36 | 1    | 16.94| 45 | N   | 9/9   | 1    | 42.94|
| 19 | N   | 35/35 | 1    | 17.94| 46 | N   | 8/8   | 1    | 43.94|
| 20 | N   | 34/34 | 1    | 18.94| 47 | N   | 7/7   | 1    | 44.94|
| 21 | N   | 33/33 | 1    | 19.94| 48 | N   | 6/6   | 1    | 45.94|
| 22 | N   | 32/32 | 1    | 20.94| 49 | N   | 5/5   | 1    | 46.94|
| 23 | N   | 31/31 | 1    | 21.94| 50 | N   | 4/4   | 1    | 47.94|
| 24 | N   | 30/30 | 1    | 22.94| 51 | N   | 3/3   | 1    | 48.94|
| 25 | N   | 29/29 | 1    | 23.94| 52 | N   | 2/2   | 1    | 49.94|
| 26 | N   | 28/28 | 1    | 24.94| 53 | N   | 1/1   | 1    | 50.94|

5. Discussion

This is the first work in which it develops a recurrent temporal probabilistic prediction of the number of outbreaks and its duration in epidemiological weeks of malaria outbreaks in 10 prototypic municipalities of Colombia, allowing to evaluate in real-time the dynamics of the epidemic, quantifying when time is most probable the onset of an outbreak with differences in order of magnitude between 10 and 10³. The predictions show that the methodology is useful as a predictive tool in epidemiological surveillance systems and early warning of outbreaks; being easily implemented allowing to develop an automatic and fast monitoring of the dynamics in real time. The underreporting of infected cases from physical and mathematical perspective can predict theoretically, also with the series number of infected probabilities it would deduce whether or not there underreporting by not submitting measures within the predicted range and also can be calculated at any epidemiological week from outbreak/no outbreak series (Table 1). Given the generality of the methodology, being developed from an induction of specific dynamic prototypes may apply to any municipality of the universe.

With this methodology all the possible dynamics of occurrence of outbreaks for any municipality were found, showing that the phenomenon of epidemics is finite and can only occur in certain ways. This can also be observed with the O–N series, because on a given week of the series, the probability value becomes 1, and thus the phenomenon becomes fully deterministic, thus an outbreak no longer be an abrupt change, but it will only be one possible event in a series ON. When the number of possible observer series obtainable in each bunch (Table 4) shows that the most likely dynamics that are 1 or 2 outbreaks can occur only 1.431 forms very small number, which shows the simplicity which can address the phenomenon from this methodology. The probabilistic sample space of the outbreaks number with
its duration allows to find probabilities that show the non equiprobable character of the phenomenon establishing differences on the order of 10, $10^2$ and $10^3$ between events, allowing the development of specific predictions for each dynamic.

In this work were found fundamental underlying orders in the dynamic of the epidemic from non equiprobable probabilistic laws; analogous manner to the behavior of loaded dice sequence outbreak and non-outbreak in time. Predictions in physics are characterized by the search for a value (in classical physics) or one or two values (in quantum mechanics), in this work the predictions are based on dynamic self-organized patterns over time leading to several dynamic predictions. As things stand chance appears when looking to force a phenomena as the dynamics of the epidemic to follow static patterns as classical or quantum physics [26]. In this work the predictions of the number of outbreaks would be predicted by bunch, which go modifying as time goes on but did not infinitely vary but rather are self-regulated by probabilistic laws as can be seen with the cumulative probability difference.

6. Conclusions
The present methodology allowed to make probabilistic predictions of the number of malaria outbreaks and its duration in epidemiological weeks in 10 prototypic municipalities of Colombia.

It was possible to quantify the most probable moment of an outbreak, with differences in order of magnitude between 10 and $10^3$, being useful as a predictive tool in epidemiological surveillance systems and early warning of outbreaks, easy to implement in automatic monitoring systems in real time.

The type of universalist reasoning from which the methodology was developed, allows us to affirm that it can be applied to the prediction of the outbreak dynamics of any municipality in the universe.

Acknowledgements
The present result is part of the products of the INV 2288 project financed by the Universidad Cooperativa de Colombia. Special thanks to Dr. Eva Prada, Director of the Bogota Headquarters, to Drs. Leonardo Galindo, National Director of Research, Andres Mena, Head of Research Bogota Headquarters, and to Dr. María José Villalobos Castro, Dean of the Faculty of Dentistry.

We also like to thank to Signed Prieto Bohórquez, for her dedication, optimism and the heart she put into for this work to be completed.

Dedication
To our children

References

[1] World Health Organization (WHO) 2016 10 facts on malaria Consulted on: http://www.who.int/features/factfiles/malaria/en/

[2] Shuman E 2010 Global climate change and infectious diseases The New England Journal of Medicine 362(2) 1061-1063

[3] World Health Organization (WHO) 2010 World malaria report 2010 (Genova: World Health Organization)

[4] Vigilancia y Análisis del Riesgo en Salud Pública 2010 Comportamiento epidemiológico malaria en Colombia año 2010 (Colombia: Instituto Nacional de Salud)

[5] World Health Organization (WHO) 2018 Malaria Consulted on: http://www.who.int/en/news-room/factsheets/detail/malaria

[6] Bouma MJ, van der Kaay HJ 1994 Epidemic malaria in India and El Niño southern oscillation Lancet 344 1638–1639

[7] Bouma MJ, Dye C 1997 Cycles of malaria associated with El Niño in Venezuela J. Am. Med. Assoc. 278 1772–1774

[8] Poveda G, Rojas W 1996 Impacto del fenómeno de El Niño sobre la intensificación de la malaria en Colombia Memorias del XII Seminario Nacional de Hidráulica e Hidrología (Santa Fe de Bogotá: Sociedad Colombiana de Ingenieros)

[9] Poveda G, Rojas W 1997 Evidencias de la asociación entre brotes epidémicos de malaria en Colombia y el fenómeno El Niño — Oscilación del Sur Rev. Acad. Colomb. Cien. 21(81) 421–429
[10] Patz JA, Epstein PR, Burke TA, Balbus JM 1996 Global climate change and emerging infectious diseases J. Am. Med. Assoc. 275 217–223
[11] Martens WJM, Jetten TH, Focks DA 1997 Sensitivity of malaria, schistosomiasis and dengue to global warming Clim Change. 35 145–156
[12] Unninayar SS, Sprigg W 1995 Climate and the emergence and spread of infectious diseases Eos. 76(47) 478
[13] Epstein PR, Chikwenhere GP 1994 Biodiversity questions Science 265 1510–1511
[14] Rodríguez J, Correa C 2009 Predicción Temporal de la Epidemia de Dengue en Colombia: Dinámica Probabilista de la Epidemia Rev. Salud Pública 11(3) 443-453
[15] Rodríguez J 2009 Dinámica probabilista temporal de la epidemia de malaria en Colombia Rev. Med. 17(2) 214-221
[16] Rodríguez J, Vitery S, Puerta G, Muñoz D, Rojas I, Pinilla L, Mora J, Salamanca D, Perdomo N 2011 Dinámica probabilista temporal de la epidemia de dengue en Colombia Rev. Cubana Hig. Epidemiol. 49(1) 74-83
[17] Rodríguez J, Prieto S 2010 Dinámica de la epidemia de Malaria. Predicciones de su Trayectoria Rev. Med. 18(2) 12-20
[18] Bortman M 1999 Elaboración de corredores o canales endémicos mediante planillas de cálculo Rev Panam. Salud Pública 5(1) 1–8
[19] Rodríguez J 2010 Método para la predicción de la dinámica temporal de la malaria en los municipios de Colombia Rev. Panam. Salud Pública 27(3) 211-218
[20] Laplace P 1995 Ensayo filosófico sobre las probabilidades (Barcelona: Altaya)
[21] Mood A, Graybill F, Boes D 1974 Introduction to the theory of statistics 3a Ed (Singapore: McGraw-Hill)
[22] Feynman RP, Leighton RB, Sands M 1964 The Feynman lectures on physics vol 1 (Wilmington: Addison-Wesley) pp 6-1,6-16
[23] Gnedenko BV 1978 The theory of probability 4a Edition (Moscow: Mir Publishers)
[24] Rodríguez J 2010 Entropía proporcional de los sistemas dinámicos cardiacos. Predicciones físicas y matemáticas de la dinámica cardíaca de aplicación clínica Rev. Col. Cardiol. 17(3) 115-129
[25] Rodríguez J, Prieto S, Melo M, Domínguez D, Correa C, Soracipa Y, Mejía M, Valero L, Salamanca A 2014 Predicción del número de linfocitos T CD4 en sangre periférica a partir de teoría de conjuntos y probabilidad en pacientes con VIH/SIDA Inmunología 33 113-120
[26] Feynman RP, Leighton RB, Sands M 1964 The Feynman lectures on physics vol 3 (Wilmington: Addison-Wesley) pp 1–11