Use of a statistical computational simulation to predict intensity of malaria transmission from data of clinical symptomatic episodes of malaria and climate

L.N. Ozurumba-Dwight (leon_ozurumba@yahoo.com)
University of Ibadan, Nigeria- Department of Zoology (Parasitology) https://orcid.org/0000-0003-3126-2989

A.A. Hassan
University of Ibadan, Nigeria-Department of Zoology (Parasitology)

A.B. Odaibo
University of Ibadan, Nigeria- Department of Zoology (Parasitology)

Okorie Anyaele
University of Ibadan, Nigeria- Department of Zoology (Entomology and Environmental Biology)

A.A. Adeyemo
National Institute of Health NIH, Bethesda USA- NHGRI , Centre for Research In Genomics and Global Health

O.O. Amodu
Institute of Child Health, College of Medicine, University of Ibadan, Nigeria

C T. Happi
Redeemers University, Ede, Osun State, Nigeria- Department of Biological Science

S.I. Oyedeji
Federal University Oye Ekiti, Ekiti State, Nigeria- iotech ology/Animal and Environmental Biology

Research Article

Keywords: Intensity, prediction, r+ve / r-ve, malaria, prevalence, clinical episode, correlation coefficient

Posted Date: June 23rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-1945901/v2

Abstract

Data on the monthly clinical episodes of malaria and prevalence from laboratory diagnosis of patients for malaria infection was obtained from an array of data gathered from malaria parasite tests conducted on patients clinically diagnosed for malaria in health centers within the study area in Akinyele Local Government Area of Ibadan city in Nigeria, for years 1997, 1998, 1999, 2000, 2001 and 2005 (6years) which falls between years 1997-2005. Also, data was gathered for climatic factors (rainfall, relative humidity, temperature and sunshine hours) for all years between years 1997 and 2005 (9years complete) from Geospatial Laboratory of International Institute of Tropical Agriculture IITA in Ibadan, Oyo State, Nigeria.

Thereafter, we engaged statistical methods with computational support from Microsoft Excel version 2007, to generate a climate based- simulation to predict periods of the years for which there were high malaria intensity for malaria. We could not retrieve complete data for prevalence (laboratory positive results for tests) the month for October. So, we proceeded to determine the correlation between clinical episodes and prevalence for the 6 years for which we retrieved data.

The Pearson moment correlation coefficient **"r"**between clinical symptomatic episode and positive outcomes of tests (prevalence of infection) as computed from Microsoft Excel was +**0.986265** This shows a high enough positive correlation, upon which we could use the clinical episodes to compute of simulations to predict periods of high intensity of clinical symptomatic episodes and which can then be related to the intensity for prevalence of malaria.

The statistical computations indicated high intensity of clinical episodes to correlate (correspond) with rise for the climatic factors, and low intensities for lowered levels of most of the climatic factors for years 2002 and 2004, as they both recorded positive ranges of correlation "r" values between clinical episode and climatic factors. This can be used to predict periods of the year with high intensity of clinical episodes of malaria as our simulated prediction.

Then we conducted two test-runs using two observed variants in the climate based-yearly periods of high intensity (those of years 1998 and 2001). The predictions indicated matches for periods of high intensity transmission using statistical tool of Pearson's moment correlation analysis derived relationships and other descriptive statistical attributes. These range of correlative value matches were between the precise values of correlation coefficients of the obtained laboratory data and that of calculated predictive ranges of these values. Since the Pearson correlation between clinical episode and prevalence of malaria was high (close to 1.0), these simulation can assist to predict prevalence of infection obtained from the laboratory diagnosis. From our analysis and predictive simulations we suggest future extraction of additional related data by other scientists to input into this simulation and run more tests with other support statistical tools to further see how it perform. If successful, this simulative prediction of malaria transmission intensity can be built into algorithm involving use of machine learning platforms.

Introduction

Malaria has persisted for a long period of time as one of the leading global health challenges, primarily prevalent in tropical and sub-tropical countries of the world. It is one of the major causes of illness and death in sub-Saharan Africa (World Malaria report, 2019; Dawak et al 2018; World Health Organization, 2022). Parasites of the genus *Plasmodium* place a huge burden on human life as a result of the malaria disease they cause (Ozurumba et al, 2006), especially in the tropics. Globally, approximately 214 million cases of malaria occur annually and 3.2 million people are at risk of infection (Dawak et al 2018). In Nigeria, there is Nigeria high burden of vector borne diseases such as malaria (Okorie et al, 2014). Malaria deaths and cases have been common among people living in tropical climatic countries and malaria is one of the leading causes of deaths in these countries (Chimezie et al, 2020). Malaria incidence in Nigeria is neither

close to eradication nor firm control. Chimezie et al (2020) opined that malaria incidence in Nigeria rose at some period over the last decade to hyper endemic levels while in some countries in the North Africa it got controlled or eradicated; while positing factors like ecological, economic, infrastructural conditions, poor grass-root outreach, and accountability issues as part of the challenges.

The range of malaria diseases is limited by climate to the warmer regions of the globe, and so anthropogenic global warming (and climate change more broadly) now threatens to alter the geographic area for potential malaria transmission, as both *Plasmodium* malaria parasite and Anopheles mosquito vector have highly-temperature dependent lifecycles, while the aquatic immature Anopheles habitats are also strongly dependent on rainfall and local hydrodynamics (Eckenberry et al, 2018).

Malaria is a disease resulting from infection by a Protozoan *Plasmodium* parasites (various species exist), (Ozurumba-Dwight et al, 2020; WHO, 2022) and spread to people through the bites of infected female Anopheles mosquitoes. Malaria is one of the most widespread vector-borne diseases, while about 8% of all cases are recorded in sub-Sahara Africa, reappearing in areas were control efforts are effective (like in the developed western world where climate is cold, and not conducive for the vector to breed well) and re-emerging in areas considered to be considered free from malaria diseases (Ndamuzi et al, 2021). It is still a disease of notable impact in some countries.

It is well known that the morphological growth processes of mosquitoes strongly depend on ambient temperature, water and availability of stagnant water bodies (Odu et al, 2021).

Several computational and statistically-mathematical approaches have been used to study and predict malaria transmission and peak periods. Typical ones include mathematical modeling based on climatic parameters and data of malaria cases (Eckenberry and Gumel 2018; Hoshen and Morse. 2004), negative binomial approach (Makinde et al, 2020), use of field based modeling study on ecological characterization of hourly-host seeking behavior and its associated climatic variables in vector mosquito species (Yin et al, 2019), using mosquito surveillance and weather data from regions or territories – such as that involving numerical simulation of non-autonomous model (Albelrazec and Gumel, 2017), using related time series data (Shi et al, 2020) and use of machine leaning approach of different machine learning classifiers like Logistics regression LC, Random forest RF and K-Nearest Neighbor KNN (Odu et al, 2021) among others.

For instance, a machine learning approach was used to classify this climate variability across the countries of sub-Saharan Africa over a period of twenty-eight years in a study conducted by (Odu et al, 2021). These authors added that the malaria incidence classification model is an early detection mechanism that helps to monitor the spread of malaria, in a unique data driven knowledge discovery system that will assist public health. The result showed non-seasonal changes in three climatic factors (precipitation, temperature and surface radiation) significantly contribute to the outbreak of malaria.

In another study, Mariki et al, (2022) demonstrated machine learning models in diagnosing malaria using patient's symptoms and demographic features by extracting malaria diagnosis datasets in tow regions in Tanzania Morogoro and Kilimanjaro. This study developed a regional specific malaria predictive model based on using machine learning classifiers. One key finding in this study is that malaria transmission depends on climatic conditions that may affect the number and survival of mosquitoes, such as rainfall patterns, temperature and humidity. Furthermore, the study indicated that coughing and joint pain were significant for malaria diagnosis in Morogoro, while dizziness and confusion are important in the diagnosis of malaria in Kilimanjaro; and months that were in the rainy session or just after the rainy session were significant in malaria transmission. Similar study by David et al (2016) conducted in 2011 with data obtained from questionnaires and diagnostics done at the University of Maiduguri Teaching Hospital in

Maiduguri town in Nigeria, using symptoms observed that vomiting had the highest occurrence at 100%. Other symptoms such as fever, headache, joint pain and nausea ranked next to vomiting in occurrence with figures of between 64.7% and 94.1%, while symptoms such as jaundice, loss of consciousness, weight loss, arthralfia, bleeding, pyrexia, cough, backache, and reduced uterine output were the least in occurrence. This study added that peak malaria prevalence was recorded in the month of September and followed by June (15.58% and 14.02% respectively) while the value was least for March followed by November (2.21% and 3.69% respectively).

Malaria control in sub-Saharan Africa is lacking in disease maps which aids in guiding and effective targeting of resources. Invariably, predictive models and the consequent disease maps needs supply of empirical data from health information records, which are also neither adequately kept nor readily furnished to researchers.

Most malaria parasite transmission maps were developed during the optimistic “eradication era” which was eventually shelved for the present “control era”, resulting from an inevitable paradigm shift. This is due to the prevailing socio-economic problems plaguing the African continent and the Third World. The development and utilization of climate based malaria predictive models and risk maps can be useful for pre-emptive interventions against the disease.

Recent epidemiological studies have confirmed that the burden of malaria is still very well prevalent under conditions that offer the risks of infection (Nyasa et al 2021; Bhatt et al 2915; Kimbi et al 2013), some of which could vary with the various prevailing human, vector, environmental and climatic conditions.

In this study, we have used empirically gathered data on malaria incidence from clinics within the study area, to enable us engage statistical computational approach to predict periods of high intensity of malaria transmission. The model was also subsequently test-run for capability to predict intensity of clinical episode of malaria in comparison with the actual figures from the clinical episode of malaria.

Objectives of study are to:

- Use of statistical computational approach to predict malaria transmission pattern in the study area.
- Test-run the model and determine nature of predictive ability based on the simulation performed.

Method

There was a collection of empirical data on yearly malaria incidence from malaria parasite tests conducted on patients presumptively diagnosed for malaria in clinics within the study area in Ibadan, Oyo State, Nigeria; from a licensed Medical diagnostic laboratory (FK International Diagnostic Laboratory) located in Akinyele Local Government Area of the State in 2005. The data covered the period between 6th January, 1997 to 31st December, 2005, while a period in between – of January 2002 to February 2004 was not retrieved- for which we made attempt to predict through our simulative approach engaging Pearson’s correlation coefficient analyses.

The gathered data were estimates of the malaria parasite density (EMPD), an index conventionally deployed in hospitals by medical practitioners for the management of malaria patients. The EMPD values ranged from +, 2+, 3+ and 4+ (for positive outcomes: PO whose total or overall PO is the prevalence of the disease) and – (for –ve or negative outcomes: NO) of the obtained malaria parasite tests (MP) results. It is an approximate method for estimation of parasite density and mostly used for routine laboratory diagnosis of malaria.

To enhance the construction of a climate-based predictive model for malaria transmission and intensity, a nine-year empirical data on climatic factors prevailing in the city of Ibadan was retrieved from the Geospatial Laboratory, at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. IITA is a research institute funded by International

donor countries, agencies and foundations; and devoted to scientific research. Data on rainfall, temperature (minimum and maximum), relative humidity (minimum and maximum) and amount of sunshine were all included in the retrieved data that ranged from the period between the periods of January 1997 to December, 2005 (9 years data). This is notably, data for years 2002 to 2004, which we could not retrieve for malaria prevalence data, but present in the gathered climatic data.

Then, we selected two of the years having variants of the already known transmission pattern (the variants showed divergences of either positively or negatively correlated in relation to relationship between malaria incidence and array of climatic factors investigated; which were those of years 1998 and 2001now selected for the test-run). This was because we had malaria incidence data gathered for these years, and we could check after our test run if our simulation to predict did match the actual values.

The study area and study population:

The subjects who were sampled in the laboratory diagnostic test by Microscopy on blood smear were within the city of Ibadan, precisely in Akinyele Local Government area. Ibadan is a city located in the Tropical rainforest belt in south western Nigeria, with mean total rainfall of approximately 1230mm, mean daily temperature of 26.46^0C and relative humidity of 74.53%. The city enjoys rainfall showers for at least 8 – 9 months annually, a feature of most cities in the Tropical rainforest belt of the globe. There are two distinct seasons of Wet season (between March and October) - through this can slightly overlap in the following month and Dry season (between November and February). This city is located on GPS coordinates of approximately 7^0 22' 36.2496^{0}N and 3^0 56' 23.2296^{0}E (with slight variations depending on the GPS instrument used) (LatLong.net, 2022).

A limitation to this study comes from extensiveness of retrieved data.

I had no access to the age and sex of the sampled subjects. This was a restriction on how further extensive I can analyze that data obtained. So, it was not possible to stratify my analysis based on age and sex, to see what the picture looks like when assessed from these angles. However, I did an overview from which other studies can be done from other dimensions and the results and observations future synchronized in future.

Statistical analysis and computations

All gathered data on incidence and climatic factors were entered into Microsoft Excel Statistical Package (version 2007) and used for various descriptive analyses for mean, standard deviation, variance, rank, maxima and minima values. Comparisons between groups were done using Pearson moment correlation coefficient values which were consequently test-run.

Results

Table 1: Monthly Incidence and EMPD values and other deduced statistical parameters

Months	Clinical Symptomatic Episode	EMPD values				Total for Positives	Total for Negatives	Number of years
		+	2+	3+	4+			
January	36	11	13	1	0	25	11	6
Mean ± SD	7.2 ±3.9	2.2±1.2	2.6±2.0	0.2±0.4	0	5.0±2.3	2.2 ±2.2	-
February	41	15	19	1	0	35	6	6
Mean	8.2± 5.8	3.0± 4.1	3.8± 1.7	0.2±0.4	0	7.0±4.7	1.2. ±1.2	-
March	80	34	22	0	0	56	24	6
Mean	13.3± 5.0	5.7±2.4	3.7± 3.5	0	0	9.3± 3.8	4.0± 2.9	-
April	76	37	21	2	1	61	15	6
Mean	12.7± 6.0	6.2± 3.0	3.5±3.6	0.3±0.8	0.2±0.4	10.2±5.4	2.5±1.6	-
May	110	53	24	1	0	78	32	6
Mean	18.3±16.9	8.8±9.7	4.0±3.2	0.2± 0.4	0	13.0±11.5	5.3±5.7	-
June	120	59	27	0	0	86	34	6
Mean	20.0±6.6	9.8±2.3	4.5±2.6	0	0	14.3±3.0	5.7±4.3	-
July	140	72	40	5	1	118	22	6
Mean	23.3±11.2	12.0±7.3	6.7±2.6	0.8±0.7	0.2±0.4	19.7±8.8	3.7±2.9	-
August	159	82	44	2	0	128	31	6
Mean	26.5±14.6	13.7±9.1	7.3±3.5	0.3±0.8	0	21.3±11.0	5.2±4.8	-
September	112	44	37	2	0	83	29	6
Mean	18.7±9.8	7.3±3.8	6.2±2.7	0.3±0.5	0	13.8±5.7	4.8±4.2	-
October	110	Na	Na	Na	Na	Na	Na	6
Mean	-	-	-	-	-	-	-	-
November	129	50	44	2	0	96	33	6
Mean	12.5±8.0	8.3±4.7	7.3±2.54	0.3±0.5	0	16±4.9	5.5±4.7	-
December	63	25	16	3	0	44	19	6
Mean	12.6±7.1	5.0±2.1	3.2±3.1	0.6±0.8	0	8.8±4.3	3.8±5.2	-

TOTAL	1176	482	307	19	2	810	283	-
Mean	98.0	43.8	27.9	1.7	0.2	73.6	23.6	-
SD	38.8	22.4	11.4	1.4	0.4	33.2	9.6	-

*Pearson moment correlation coefficient **"r"** between Clinical symptomatic episode and positive outcomes of tests (Prevalence of infection) as computed from Microsoft Excel was = **+ 0.986265**

This shows a high enough positive correlation, despite the incomplete data for October for the prevalence, which we then left out of entry (Laboratory Test results conducted on the clinically diagnosed patients for October was not found for some years, so we left it out).

Based on this high correlation "r" value, we then used the clinical episodes for our computational simulation to predict the period of the year of high intensity of clinical symptomatic episodes for two (2) years 2002 and 2004 (for which we had no retrieved laboratory data). Then, we used years 1998 and 2001 for which we had retrieved laboratory data, to test run our simulation to see if the simulation matched in terms of predicted period of the year (in terms of the season of the year) for which intensity of clinical symptomatic episodes of malaria was comparatively high.

Table 2: Predictive range of values for Pearson's correlation coefficient relating to the incidence of Malaria and the climatic factors for the years in which data was not obtained.

Using 2000 (r +) for Incidence as (A); 2001 (r-) for Incidence as (B)

Months	2002(A) (r +)	2002 (B) (r -)	2003 (A) (r +)	2003 (B) (r -)	2004 (A) (r +)	2004(B) (r -)
Jan	2	11	2	11	2	11
Feb	1	18	1	18	1	18
Mar	11	19	11	19	11	19
Apr	17	9	17	9	17	9
May	20	12	20	12	20	12
Jun	17	19	17	19	17	19
Jul	19	28	19	28	19	28
Aug	24	32	24	32	24	32
Sep	18	11	18	11	18	11
Oct	12	22	12	22	12	22
Nov	23	22	23	22	23	22
Dec	6	-	6	-	6	-
Mean	14.1667	18.4545	14.1667	18.4545	14.1667	18.4545
SD	7.779	7.3670	7.779	7.3670	7.770	7.3670
N	12	11	12	11	12	11
r1	0.57995	0.4660	0.3139	-0.2209	0.3791	-0.7904
r2	0.3795	0.0196	-0.2156	-0.1876	0.3614	-0.2272
r3	-0.7098	-0.4579	-0.6786	-0.5211	-0.6742	-0.4783
r4	0.8093	0.3006	0.7539	0,5124	0.7928	0.4070
r5	0.5821	0.0898	0.3665	0.5177	0.5947	0.1238
r6	0.4291	-0.3883	0.1275	-0.3582	-0.4162	-0.6649
Mean	0.589	0.220	0.305	-0.062	0.532	0.170
SD	0.152	0.176	0.345	0.340	0.177	0.176
Total	2.345	0.877	1.218	-0.249	2.128	0.679
Rank	P1	N1	P3	N3	P2	N2S

Predictive range correlation coefficient ‘r’ values:
Year 2002: +0.220 to +0.589 (Positive correlation)
Year 2003: -0.062 to +0.305 (Less of negative to more of positive correlation)
Year 2004: +0.176 to +0.177 (Positive correlation)
N: Number of months
P1-P3: Rank for r + column
N1-N3: Rank for r – column

Table 3: Test-run of the predictive simulation for malaria transmission pattern using the year 1998 (for test-run)

	1998 r+ve evaluation (from year 2000 values)	1998 r--ve evaluation from year 2001 values	
Jan	2	11	
Feb	1	18	
Mar	11	19	
Apr	17	9	
May	20	12	
Jun	17	19	
Jul	19	28	
Aug	24	32	
Sep	18	11	
Oct	12	22	
Nov	23	22	
Dec	6		
Total	170	203	
Mean	14.167	18.455	
SD	7.779	7.367	
N	12	11	

Correlation values			Mean from this Table	Mean From Table 5
R1	0.5532	1.4046	(+)0.9789±0.6020	0.286
R2	0.1658	-0.3273	(+)0.0808±0.3486	0.311
R3	-0.8604	-0.7	(+)0.7802±0.1134	-0.108
R4	0.6963	0.4045	(+)0.5504±0.2063	0.281
R5	0.6998	0.4955	(+)0.5977±0.1445	0.392
R6	-0.1274	-0.6864	(+)0.4069±0.3953	-0.058

Using all 6 r values for the 6 climatic factors			
Mean	0.1879	0.0985	
SD	0.6087	0.8237	

Using only the 4 r values in the transmission defining climatic factors			
Mean	0.5288	0.4943	
SD	0.2514	0.7099	

Table 4: Test-run of the developed bio-mathematical predictive model for climate based malaria transmission pattern using the year 2001 (for test-run).

		2001 r+ve evaluation (from year 2000 values)	2001 r-ve evaluation from year 2001 values	
Jan		2	-	
Feb		1	-	
Mar		11	19	
Apr		17	17	
May		20	53	
Jun		17	23	
Jul		19	39	
Aug		24	51	
Sep		18	38	
Oct		12	25	
Nov		23	31	
Dec		6	24	
Total		170	320	
Mean		14.167	32	
SD		7.779	12.,806	
N		12	10	
Correlation values			Mean from this Table	Mean From Table 5
R1	0.472	0.2546	(+) 0.3633±0.1537	-0.105
R2	0.2308	-0.3455	(-) 0.0574± 0.4075	0.12
R3	-0.6993	-0.5758	(-)0.6376±0.0873	-0.474
R4	-0.1364	-0.4909	(-)0.3137±0.2500	-0.474
R5	-0.085	-0.1515	(-)0.0933±0.0824	-0.453
R6	-0.6505	-0.3091	(-)0.4798±0.2414	-0.517
Using all 6 r values for the 6 climatic factors				
Mean	-0.1364	-0.2697		
SD	0.4683	0.2961		
Using only the 4 r values in the transmission defining climatic factors				
Mean	0.1329	-0.1379		
SD	0.2741	0.3231		

Note: r1 to r6 represents the investigated six climatic factors- rainfall, minimum and maximum temperature, minimum and maximum relative humidity and amount of sunshine respectively.

Table 5: Pearson’s correlation coefficient values (r1 to r6) indicating the relationship between incidence of malaria and the climatic factors under study

Month	Years and values for Clinical symptomatic episodes of malaria					
	1997	1998	1999	2000	2001	2005
Jan	11	3	9	2	11	-
Feb	4	9	9	1	18	-
Mar	5	5	11	15	11	19
Apr	8	4	21	17	9	17
May	2	4	19	20	12	53
Jun	15	13	15	17	19	23
Jul	9	11	34	19	28	39
Aug	12	7	33	24	32	51
Sep	13	9	23	18	11	38
Oct	13	11	27	12	22	25
Nov	6	19	28	23	22	31
Dec	14	4	15	6	-	24
Total	112	105	266	170	203	320
Mean	9.33	8.75	22.17	14.17	18.45	32
SD	4.31	4.69	9.01	7.78	7.37	12.81
Rank	5	6	2	4	3	1

Pearson correlation coefficient “r” values							Mean	Total
r1	0.336	0.286	0.624	0.630	-0.105	0.022	0.299±0.276	1.793
r2	0.326	0.311	0.073	0.555	0.120	-0.519	0.144±0.335	0.866
r3	-0.449	-0.108	-0.823	-0.628	-0.474	-0.604	-0.365±0.423	-2.188
r4	0.398	0.281	0.808	0.781	-0.474	0.613	0.406±0.426	2.434
r5	0.442	0.392	0.699	0.589	-0.453	0.618	0.381±0.387	2.287
r6	-0.187	-0.058	-0.627	-0.441	-0.517	-0.419	-0.375±0.197	-2.250

For all 6 climatic factors (Further computations on Pearson’s moment correlation coefficient “r”)

	1997	1998	1999	2000	2001	2005
Mean	0.145	0.184	0.126	0.248	-0.313	-0.048
SD	0.338	0.193	0.647	0.560	0.237	0.529
Total r	0.867	1.105	0.7542	1.486	-1.876	-0.289

r1: Pearson correlation coefficient value (clinical episode with rainfall)
r2: Pearson correlation coefficient value (clinical episode with minimum temperature)
r3: Pearson correlation coefficient value (clinical episode with maximum temperature)
r4: Pearson correlation coefficient value (clinical episode with minimum relative humidity)
r5: Pearson correlation coefficient value (clinical episode with maximum relative humidity)
r6: Pearson correlation coefficient value (clinical episode with sunshine).

Computational process *1(a-f), leading to the predictive (simulative) analysis (Refer to Tables 2, 3 and 5)

(a). Correlation coefficient “r” analysis of the relationship between clinical episode and the engaged six climatic factors for transmission pattern for all the years of gathered data (presented on Table5).

(b). Estimation of various descriptive values comprising the mean and standard deviation of all “r” values (r1 to r6 in this case) for all the years of gathered data to determine “r” values from the field (presented on Table5).

(c). Selection of the two extremely divergent years form Table 5, the one with the highest positive correlation “r” value (coded r+ve) and the second with the least “r” value (coded r-ve).

(d). Engaging each of the investigated climatic factors for the years of unknown periods of high intensity in a correlation coefficient “r” analysis with the incidence values for each of the months of the years already coded as r+ve and r-ve respectively (presented on Table 2).

(e). Estimation of the various descriptive values comprising the mean and standard deviation values for each of the 2 extremely divergent years (presented on Table 2).

(f). Estimation** of the range of obtained individual and grand “r” values (Table 2).

Computational process 2 in the test-run is as follows (Tables 3 and 4).

We engaged steps *1(a-f) steps a-f from Process*1.

(g). Selection of two of the years of variants which are the already known period of the year with high intensity of transmission showing divergence (either positively or negatively correlated with respect to the relation between clinical episode and array of climatic factors investigated) (years 1998 and 2001 were randomly selected for the test-run), since we had malaria incidence data gathered for these years, and we could check after our test run if our simulation to predict did match the actual values.

(h). Comparison of all the computed predictive range of “r” values on each of Tables 3 and 4 with the actual “r” values (on Table 5) earlier obtained from various correlation analysis from the field.

Thus, two test-runs were conducted.

Discussion

Based on a Pearson moment correlation coefficient “r” analysis previously conducted on all the years in relation to the 6 engaged climatic factors, the obtained order of decreasing “r” values was: minimum relative humidity > maximum relative humidity > rainfall > minimum temperature > maximum temperature > sunshine hours (“r” value using both

mean and total “r” values for comparison) (+0.41 and +2.43 > +0.38 and +2.29 > + 0.30 and +1.79 > + 0.14 and +0.87 > -0.37 and -2.19 > -0.38 and -2.25 for the “r” values of each of the respectively listed climatic factors above, obtained from the study area) (Figures 1 and 2; Table 5).

This implies that relative humidity, rainfall, minimum temperature and maximum temperature, were the climatic factors (coded 4ecf) that were observed to be positively correlated (with a positive relationship) with values for clinical episodes of malaria for the period of data gathered. As such, they appeared to be the major defining factors for clues to periods of high intensity of malaria transmission in the study area.

From Table 1, involving part of the predictive model whose data were not retrieved, the order of increasing correlation with the climatic factors is:

2002 (r+ve) < 2004 (r+ve) < 2003 (r+ve) – in terms of all the r+ve correlation comparisons [0.589 ± 0.152 and 2.354 (P1) < 0.532 ± 0.177 and 2.128 (P2) < 0.305 ± 0.345 and 1.218 (P3) respectively for mean and total “r” values], and 2002 (r-ve) < 2004 (r-ve) < 2003 (r-ve) – in terms of the “r-ve” correlation comparison [0.220 ± 0.176 and 0.877 (N1) < 0.170 ± 0.176 and 0.679 (N2) < -0.062 ± 0.340 and -0.249 (N3) respectively for the mean and total “r” values in the “r-ve” analysis]. Thus, the predictive range of correlation coefficient “r” values for years 2002, 2003 and 2004 all enumerated beneath Table 2 indicates only year 2003 to have the highest possibility or likelihood of a negative correlation (predictive range of r = - 0.062 to +0.305) with the six climatic factors engaged.

The degree of fitting-in (match) was more noticed when all engaged climatic factors (6cf) were used for the model than when only the ascertained and previously determined actual climatic factors (4ecf in this study) that influenced the periods of high intensity transmission of the disease in the study area. The wider the pool of climatic factors engaged in the model, the more the level of fitting-in or accuracy. Hence, the actual periods of high intensity transmission defining climatic factors were used as a check within the simulation and to create a broader picture of understanding of the relationship amongst the varying groups of factors (Table 2).

The statistical computations indicated high intensity of clinical episodes to correlate (correspond) with rise for the climatic factors, and low intensities for lowered levels of most of the climatic factors for years 2002 and 2004, as they both recorded positive ranges of correlation “r” values between clinical episode and climatic factors. This was used to predict periods of the year with high intensity of clinical episodes of malaria as our simulated prediction.

The four climatic factors we found to determine the period of high intensity of transmission which have linkages with the provisions or enabling of a humid (wet) condition in the environment (minimum and maximum relative humidity, rainfall and minimum temperature),. They enhance the provision of suitable conditions for the vector to breed and participate in the process of transmission through infection and re-infection of unsusceptible and susceptible humans recorded.

Similar finding from Mariki et al (2022) opined that malaria transmission depends on climatic conditions that may affect the number and survival of mosquitoes, such as rainfall patterns, temperature and humidity. This study further observed that symptoms were important in the diagnosis of malaria in Kilimanjaro province of Tanzania in the months of rainy session or just after the rainy session, during which they were significant in malaria transmission.

In a related finding from the study of Abdelrazec and Gumel (2017), it was reported that observed peak mosquito abundance for temperature and rainfall values recorded for months of July and August in Peel region of Ontario in Canada. These are in the season of summer in Canadian which is called the rainy season in the hot tropical and sub-tropical climates where malaria is a more common disease and endemic here. A related finding from a study by David et

el (2016) revealed that Peak malaria prevalence was recorded in the month of September and followed by June (15.58% and 14.02% respectively) while the value was least for March followed by November (2.21% and 3.69% respectively).

Singh et al (2019) used a dynamic model (VECTR) for malaria transmission that accounts for the influence of population and climatic conditions to investigate malaria transmission dynamics for a highly endemic region in India (State of Odisha) to do numerical simulations for years 2000-2013 and found out that temperature, adult mosquito population and infective biting rates have increased over this period and malaria vector abundance higher during the summer monsoon season. The peak malaria transmission occurred when the monthly mean temperature is in the range of 28-290C and monthly rainfall accumulation in the range of 200-360mm.

According to Stuckey et al (2013), evaluating the effectiveness of malaria control interventions on the basis of their impact on transmission is increasingly important as countries move from malaria control to pre-eradication programs. They added that mathematical modeling can examine relationships between malaria indicators, allowing translation of measured data into measures of transmission, as simulations show nature of statistical correlation, allowing direct comparisons of malaria transmission using data collected across a range of transmission intensities and seasonal patterns.

Test-run of predictive analysis as simulated

The test-run was necessary in order to check the correctness of the predictions on periods of high malaria intensity made for years 2002, 2003 and 2004, and the capability (effectiveness) of the model. It was more of a validation of previous predicted results and from the model. Basically, the procedures in the sequence of operations earlier outlined were followed up to the stage of estimation of the range in the obtained results. Thus, comparison of predicted ranges of “r’ values and the actual field values in the case of the test-run could be conducted unlike in the earlier predictions done for years 2002, 2003 and 2004. This is because there were actual figures from the field to compare with and help validate results for consistency.

Predictive correlation ‘r’ value ranges for 1998 and 2001 are given below:

(I) for 1998 are as follows (Table 3):

(+) 0.0985 to (+) 0.1879 (for all 6cf)

(+) 0.4043 to (+) 0.5288 ± 0.2178 (for only 4ecf)

Actual values of correlation coefficient ‘r’ values from the incidence obtained from field studies:

(+) 0.184 ± 0-193 (for all 6cf)

(+) 0.318 ± 0.044 (for 4ecf).

(II) For 2001 are as follows (Table 4):

(-) 0.2975 to (-) 0.1364 (using all 6cf)

(-) 0.1379 to (+) 0-1329 (using only 4ecf).

Actual values of correlation coefficient ‘r’ values for the year 2001 from the incidence obtained from field data:

(-) 0.313 ± 0-235 (using all 6cf)

(-) 0-221 ± 0-242 (using 4ecf).

Note: The incidence for year 2005 was used for the r+ve and r-ve evaluations of incidence in relationship with climatic factors of year 2001 by Person correlation coefficient analysis.

( ) Brackets have deliberately been for easy understanding, clarity positive and negative correlative values with respect to the two seasons of the year.

The year 2001 incidence results showed least correlation coefficient 'r' value by inspection of results obtained (on Table 5), next to that of year 2005 (minimum 'r' value for all 6 years studied) form the actual values. Hence, it (year 2001) was utilized for the evaluations.

Test run for 1998 (Table 3)

A comparison of the actual correlation coefficient 'r' values (on Table5) and the predictive correlation 'r' values for 1998 of (+0.184 ± 0-.93) on Tables 2 and 3 (using all 6climatic factors - 6cf) is in line with the results from the model.

Firstly, both results from the actual values shown on Table 5 (for 1998) and the predictive one on Table 2, both indicate a positive correlation between malaria incidence and the six climatic factors. This shows exactness in this area of prediction (100%).

Secondly, the actual (+) 0.184 ± 0.193 'r' values (for all 6cf) is well within the predictive (+) 0.0985 to (+) 0.1879 'r' range of values (for all 6cf). The former (actual values) fits in 100% within the latter's predictive range.

A predictive analysis of the years 2002, '03 and '04 (on Table 2) in which data on incidence were not obtained, indicated a most likely tendency towards high intensity of transmission that was related to the seasons, with respect to the climatic factors influencing it and a yearly non-perennial high intensities in transmission for the years 2002 and 2004 with positive 'r' value ranges, while year 2003 predictably had the most likely tendency for a non-season related high intensity transmission (range of 'r' = -0.062 to +0.305) amongst the three years. This formed the basis for the "developed mathematical predictive model", which was also test-run as a check for its predictive outcome of which was alright and in line with actual field 'r' values.

The capacity of climate based simulations to predict periods of high intensity of malaria transmission could be of benefit to control measures in areas of policy formulation, planning and proper execution of such control measures in such endemic and bio-geographical area; with added cost-benefit attributes in addition to possibly reducing incidence (perhaps to insignificant levels in some or most areas), and associated health and economic burdens..

The predictions from the model on Table 3 indicate a positive correlation and association of malaria incidence with climatic factors (season related high intensity of malaria transmission). This is exactly what was previously obtained from the actual results on Table 2.

In view of the foregoing, the predictive simulation for malaria intensity was able to predict periods of high intensity of transmission correctly. In terms of the precision of the exact value (which is a greater or more demanding task, of- the extent of association (correlation) between malaria incidence and climate, the capability is alright. The predictive range of correlation coefficient 'r' values was well within the range of actual 'r' values' obtained from the field studies, which was used to test-run for the exact relationship existing between clinical episode of malaria and climatic factors probed for.

Test run for 2001 (Table 4)

A comparison of the actual correlation coefficient ‘r’ values for 2001 and the predictive one on Tables 2 and 5 of the constructed predictive model indicate the following:

Firstly, the results from the actual values shown on Table 5 (for year 2001) showed a negative correlation between clinical episode and the 6cf combined. This indicates exactness in this area of prediction (100%).

Secondly, the actual (-) 0.313 ± 0.235 ‘r’ values for all 6cf is well within the predictive (-) 0.2697 to (-) 0.1364 ‘r’ value range for all 6cf (from Tables 4 and 5 respectively). When values are taken to one decimal place, it fits in 100% within the predictive range.

These deductions above indicate a negative correlation and association of clinical episode of malaria with the engaged climatic factors (non-seasonal regarding mean correlation with the engaged 6 climatic factors). A cross-check also, showed that this was exactly what had been previously obtained from the actual field data on Table 5 (graphically depicted in Figure 2).

Conclusion

In view of the foregoing analysis and discussions, the simulation method was able to predict the periods of high intensity for malaria transmission. In terms of the precision of the exact value of the extent of association (correlation) between clinical episode of malaria and climate, the actual correlation “r” values were within the predictive ranges of “r” values. From our analysis and predictive simulations we suggest future extraction of additional related data by other scientists to input into this simulation and run more tests with other support statistical tools to further see how it perform. If successful, this simulative prediction of malaria transmission intensity can be built into algorithm involving use of machine learning platforms. We encourage other researchers working on other approaches to predictive simulations for intensity of malaria transmission and other diseases, as we are all in a scientific community where we share ideas and build-up on insights.

Studies that emerge with climate-based predictive simulations to predict periods of high intensity of malaria transmission, if found to maintain predictive performance, could be of benefit to preventive medical and public healthcare, and in planning to support avenues for control. Then seek how to write software program on such predictive simulations to enhance ease of usage, where the performance shows that it is sustained.

Declarations

Conflict of interest: We declare none.

Acknowledgements

Appreciation to Dr Femi Aina and Mr. K. Eniayewu for furnishing Ozurumba-Dwight LN with climatic data from the Geospatial Laboratory of IITA at the Computer Training Unit of International Institute of Tropical Agriculture (IITA) in Ibadan, Nigeria. Appreciation to Mr. Olugbenle the Director of KF international diagnostic laboratory, NISER-Ojoo area, Ibadan, Oyo State, for enabling access in 2006 to records on patients referred to this Laboratory for malaria diagnostic test. Gratitude to Cellular Parasitology Laboratory of Department of Zoology, University of Ibadan, Nigeria; and Department of Zoology at OAU University Ile-Ife, Nigeria. Also, gratitude to Registrar and Dean of Faculty of Science, University of Ibadan, Nigeria, during the period of this study.

References

Abdelrazec A. and Gumel AB. (2017). Mathematical assessment of the role of temperature and rainfall on mosquito population dynamics. J.Math.Biol. 2017; 74(6): 1351-95.

Bhatt SJ, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U. et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Public Medical Center. 2015: 565 (7572): 207-211.

Chimezie RO (2020). Makaria hyperendemicity: The burden and obstacles to eradication in Nigeria. Journal of Biosciences and Medicine. 8 (11): Doi: 10.4236/jbm.2020.811015.

David J, Anaso EC, Ozurumba LN (2016). Prevalence of Malaria in a Study Population in Maiduguri, Nigeria, Symptomatic Manifestations and Public Health Implications. Greener Journal of Biological Sciences, 6(5):103-111, http://doi.org/10.15580/GJBS.2016.5.092416145

Dawak S, Al-Mekhlafi M, Ithoi I, Ibrahim J, Atroosh WM, Abdulsalam AM, et al (2018). Is Nigeria winning the battle against malaria? Prevalence risk factors and KAP assessment among Hausa communities in Kano State. Malaria Journal 15: 351. Doi: 10.1186/s12936-016-1394-3.

Eckenberry SE, Gumel AB. (2018). Mathematical modeling of climate and malaria transmission dynamics: a historical review. J.Math.Biol. 2018; 77(4): 857-933.

Hoshen MB and Morse AP. (2004). A weather-driven model of malaria transmission. Malar J. 2004, 3(1): 32.

Kimbi HK, Nana Y, Sumbele IN, Anchang-Kimbi JK, Lum E and Tonga C. (2013). Environmental factors and preventive methods against malaria parasite prevalence in rural Bomaka and Urban Molyko, Southwest Cameroun. Journal of Bacteriology and Parasitology. 4: 1. Doi: 10.4172/2155-9597.1000162.

LatLong.net (2022). Ibadan, Nigeria. www.latlong.net. Accessed 24th May, 2022.

Makinde OS, Abiodun GL, Ojo OT. Modeling of malaria incidence in Akure, Nigeria: negative binomial approach. GeoJournal 2020.https://doi.org/1007/S10708-019-10134-x

Mariki M E, Mkoba E and Mduma N. (2022). Combining clinical symptoms and patient features for malaria diagnosis: machine learning approach. Artificial intelligence. Doi: 10.1080/08839514.20022.203182.

Ndamuzi E, Gahungu P (2021). Mathematical modeling of malaria transmission dynamics: Case of Burundi, Journal of Applied Mathematics and Physics. Doi: 10.4236/jamp.2021.910156.

Nyasa RB, Fotabe EL, Ndip RN (2021). Trends in malaria prevalence and risk factors associated with the disease in Nkongho-mbeng: a typical rural setting in the equatorial rainforest of South west region of Cameroun. PLoS one. 16(5): e0251380. Doi: 10:1371/journak.pone.0252380.

Odu et al, (2021). Prediction of malaria incidence using climate variability and machine learning. Informatics in Medicine Unlocked., 22: https://doi.org/j.imu.2020.100508.

Okorie PN, Popoola KOK, Olayemi MA, Ibrahim KT, Ademowo GO. Species composition and temporal distribution of mosquito populations in Ibadan, Southwestern Nigeria. J Entomol Zool Stud 2014; 2(4): 164-169.

Ozurumba, L.N.; Anumudu, C.I. and Nwagwu, M. (2006). Seasonal variations in antibody response to a *Plasmodium falciparum* recombinant circumsporozoite antigen in two villages in South- Western Nigeria. Afr. J. Biomed. Sc. Jan; Vol. 9 (1): 37 – 43.

Ozurumba-Dwight L.N.; Schumacker, C.Jjnr; Ogbonna C.S.; Enwere O.O.; Sahal N.R., David J., Akintoye N.A., Gadzama U.N, Odu C.E., Nwoke B.E.B., Ukoh V., Oguoma V.M., Balarabe M.L. (2020). Elucidative Intersects and Divergences Involving Cancer and Malaria. Journal of Oncology Translational Research. Extended Abstract. 6, 2. 2020.

Singh PR, Bal PK, Kumar V et al. (2019). Numerical modeling of the dynamics of malaria transmission in highly endemic region of India. Sci Rep 9, 11903(2019). Http://doi.org/10.1038/s41598-019-47212-6.

Stuckey EM, Smith TA and Chitnis N (2013). Estimating malaria transmission through mathematical models. Treads in Parasitology 29(10): P477-482. Doi: http://doi.org/10.1016/j.pt.2013.08.001.

Tompkins AM and Guiseppe FD (2015). Predictability of Malaria in Africa Using ECMWF Monthly and seasonal climate forecasts. Journal of Applied Meteorology and Climatology. 54 (54): Http://doi.org/10.1175/jamc-d-14-0156 (2015).

World Health Organization (WHO) (2022). Malaria Key Facts. Accessed 25th May 2022 from www.who.int/news-room/facts-sheets/details/malaria.

World Malaria report (2019). Analysis on malaria's impact worldwide.p:1-3. Reliefweb. OCHA Accessed 25th May 2022 from www.refiefwb.int/report.

Yin Q, Li L, Guo X, Wu R, Shi B, Wang Y. et al (2019). A field-based modeling study on ecological characterization of hourly host-seeking behavior and associated climatic variables in Aedes albopictus. Parasit Vectors 2019; 12:1-14.

Figures

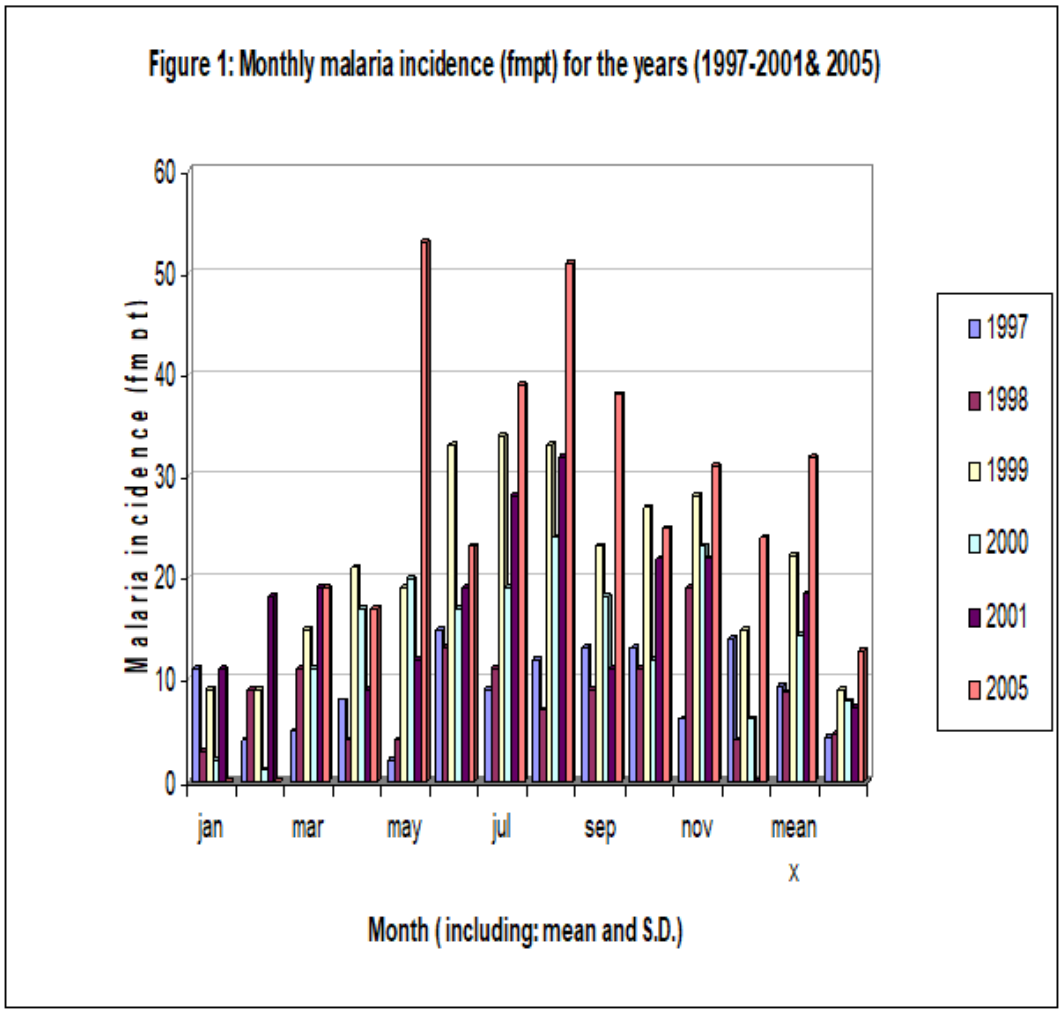

Figure 1

Monthly malaria incidence for the years 1997-2001 & 2005

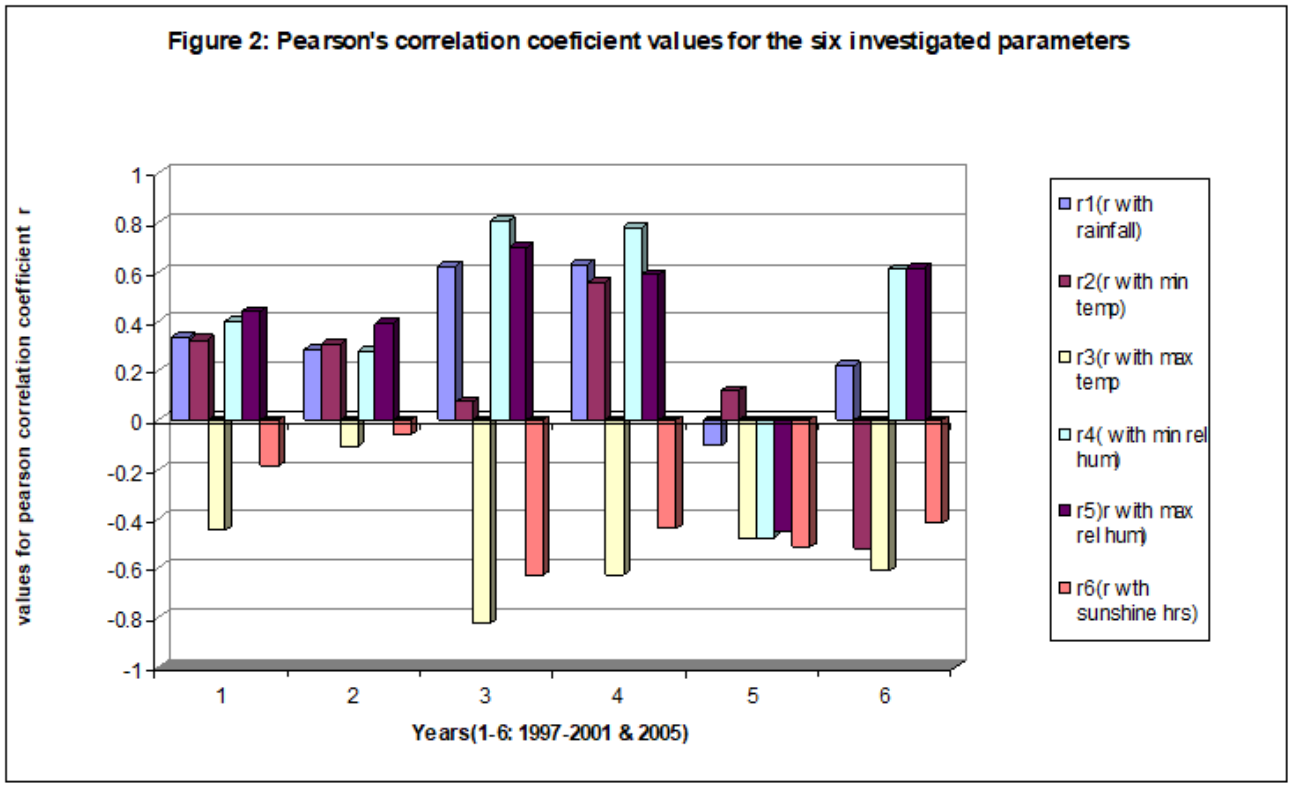

Figure 2

Pearson's correlation coefficient values of incidence with the six investigated climatic parameters

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RESULTSresearchsqcorrectn.SIMULATION.doc