Eur. Phys. J. Special Topics **143**, 125–132 (2007)

DOI: 10.1140/epjst/e2007-00080-4

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

On the study of the dynamical aspects of parasitemia in the blood cycle of malaria

R.M. Zorzenon dos Santos[1], S.T.R. Pinho[2], C.P. Ferreira[3], and P.C.A. da Silva[1]

[1] Departamento de Física, Universidade Federal de Pernambuco, 50670-901, Recife, PE, Brazil
[2] Departamento de Física, Universidade Federal da Bahia, 40210-340, Salvador, BA, Brazil
[3] Departamento de Bioestatística, Universidade Estadual de São Paulo, 18618-000, Botucatu, SP, Brazil

Abstract. Malaria is an important cause of morbidity and mortality worldwide. One striking aspect regarding malaria is the fact that individuals living in endemic areas do not develop immunity against the parasite, falling ill whenever they are exposed to the parasite. The understanding of why immunity is not developed in the usual way against Plasmodium is crucial to the improvement of treatment and prevention. In this work, we study some aspects of the dynamics of the blood cycle of malaria using both modelling and data analysis of observed case-histories described by parasitemia time series. By comparing our simulations with experimental results we have shown that the different behaviour observed among patients may be associated to differences in the efficiency of the immune system to control the infection.

1 Introduction

Over the past 100 years, economic, public-health and medical advances have controlled the worldwide extent of malaria. Nonetheless, population growth and the failure of many of the adopted control and surveillance strategies, have given rise to malaria becoming a significant health problem throughout 90 countries, killing on average two individuals per minute worldwide.

Malaria is caused by four different species of the protozoa Plasmodium, amongst which the most harmful is Plasmodium *falciparum*, responsible for the severest cases, including fatal cases. The human malaria parasite has a complex life cycle that requires both a human host (carrier) and an insect host. In the Anopheles mosquito, the Plasmodium parasite reproduces sexually (by combining sex cells). In people, the parasite reproduces asexually (by cell division), first in liver cells and then, repeatedly, in red blood cells. It is clear that the ability to change forms during its life cycle is the parasite's main mechanism that allows it to escape from the usual control of the immune system (IS).

The bite of the female anopheline mosquitoes inoculates sporozoites into the bloodstream, which then rapidly invade hepatic cells. One sporozoite within the hepatic cell multiplies into 20,000 merozoites, which are released into the blood stream after one or two weeks. Each merozoite that invades a red blood cell (RBC) goes on to produce between 8 and 24 merozoites, which are also then released into the bloodstream after one or two days, thus once again, penetrating other non-infected RBC. This asexual reproduction cycle is hereafter repeated indefinitely. After a number of blood cycles have been completed, a few merozoites start to differentiate into the parasite's sexual form, allowing the human host to contaminate the mosquito, and thus bringing the Plasmodium cycle to a close [1]. The ability to undergo repeated cycles of asexual reproduction means that, in theory, overwhelming infections can result from exposure to a

single sporozoite. The overall behaviour of the parasitemia (parasite counts/μl) is characterised by varying duration periods and a pronounced first maximum followed by a variable number of local maxima at irregular intervals.

In P. *falciparum* malaria, the most dangerous form, infected RBC are sequestered in capillaries of internal organs [2]. By residing in microvascular beds, RBC are protected against immune clearance mechanisms, however adhesive RBC contribute to metabolic disturbances and organs disfunction [3]. For some parasite strains, infected RBC may rosette non-infected RBC [4]. Therefore, the spatial localisation generated by sequestration, plays important role either by protecting the infected RBC from clearance or by agglutinating non-infected ones, mechanisms that favour the dissemination of the infection. There is still much to be understood concerning the interplay between the multiple forms of the parasite, the sequestration of RBC and the development of the defence mechanisms that lead to host's inability of becoming immune to the parasite after multiple exposures. This understanding is not only crucial for improving therapeutic procedures, but also for developing anti-malarial vaccines.

A recent review of Molineaux and Dietz on malaria blood cycle mathematical modelling (see [5] and references therein) indicates that the number of genuinely different models are much smaller than the number of models found in the literature. The authors also emphasise the necessity of a more rigorous data comparison, and strongly recommend the adoption of discrete-time modelling. In 2001, together with other collaborators [6], they introduced a time-discrete model with a step size of two days to describe the course of the parasitemia in a human host. The model dynamics is driven by two forces: the switching between variants (generated by mutations) of the parasite, and the multiplication of a given variant due to escaping from the IS. This model was able to produce a set of simulated case-histories exhibiting statistically similar behaviour to those of a set of observed case-histories when skipping one of each two observed points. In order to compare simulated and observed cases-histories, they reduced the parameter space by choosing 9 parameters to characterise the parasitemia time-series: (1) initial slope; (2) log(parasitemia+1) at the first local maximum; (3) number of local maxima; (4) slope of the local maxima (including the first maximum); (5) geometric mean (GM) of the intervals between consecutive local maxima; (6) standard deviation (SD) of the logs of the intervals between consecutive local maxima; (7) proportion of positive observations in the first half of the interval between the first and last positive days; (8) proportion of positive observations in the second half of the interval between the first and last positive days; (9) the last positive day.

More recently, spatially structured models have been recognised as a step forward in understanding the dynamics of infection processes, in which the influence of the local dispersal of the pathological agent is relevant for disease persistence in vivo [7,8]. In this work, we consider a discrete mathematical model introduced recently [9] to describe the blood cycle of malaria. The model is based on a cellular automata approach. The spatial localisation considered in the model accounts for the cytoadherence that occurs in the blood vessel during sequestration of infected RBC. The model takes into account the main biological assumptions concerning the blood cycle as well as the role of clearance of the immune system. By employing this approach, we were able to describe the observed daily counts of the parasitemia and not a two-day time step, thus achieving a more realistic comparison between the model results and the observed data. Throughout this work, we have employed this model to simulate a number of case-histories, which are compared to observed case histories, using the same 9 parameters adopted in reference [6].

The observed case histories discussed here were selected from the clustering analysis performed for 35 patients. These were chosen among 334 case histories collected by the USPHS in NIH Laboratories in South Caroline and Georgia in USA, at a time when malaria therapy was used to treat neurosyphilis patients. For each patient the daily parasitemia was measured in order to follow the development of the disease. However most of these parasitemia time series are incomplete, and among the incomplete ones only those with one or two consecutive missing points can be interpolated, otherwise the interpolation would interfere on the actual pattern of the parasitemia time series since the parasite incubation time in the blood cycle is of the order of two days. The 35 patients correspond to those with complete time series, that were not submitted to any treatment or reinfection. By performing a non-supervised clustering analysis

we could separate different time series behaviours. Using the model, we simulated the different case-histories belonging to different clusters obtained from the analysis performed. We finish by discussing the effectiveness of the immune system in killing infected erythrocytes for different patterns, establishing in this way a link between the results of clustering process and efficiency of the immune response.

The paper is organized as follows: in section two, we briefly describe the main aspects of the ensemble of 35 observed parasitemia time series and the respective performed clustering analysis [10,11], searching for different classes of behaviour; in section three, we describe the model and the procedures employed to compare the simulated and observed case-histories; in section four, we discuss our results and present our final remarks.

2 About the data and the clustering analysis

The overall aim of the analysis performed in this section is to detect similar behaviours on the parasitemia time series of the 35 patients mentioned above and attempt to group them according to the characteristics that could bring us information about the action of the immune system. The observed case-histories considered here were collected from adults suffering from neurosyphilis and with no previous history of malaria, who were submitted to malaria therapy with different strains of P. *falciparum* under closed medical supervision. They were inoculated either with sporozoite (in general by mosquito bite) or with infected blood that we refer as spo and blood inoculations, respectively. Microscopic examinations of blood were conducted on an almost daily basis. The detection threshold was taken as 10 parasitized RBC per μl. Extended descriptions and analysis of many aspects of the entire 334 case-histories can be found in [12] and references therein. The 35 patients analysed here were neither subjected to sub-curative nor to curative treatments. Since the parasitemia time series have different sizes, we reduced the parameter space, by choosing few parameters that would characterise all the series, adopting the parameters cited above [6]. Therefore, we could define a distance among patients in this 9-D parameter space.

In order to search for similar behaviour within the 35 case-histories, we have used the non-supervised Superparamagnetic Clustering Technique (SPC) (see [10,11] and references therein). The SPC technique consists of mapping the ensemble of data or objects (here patients) that we

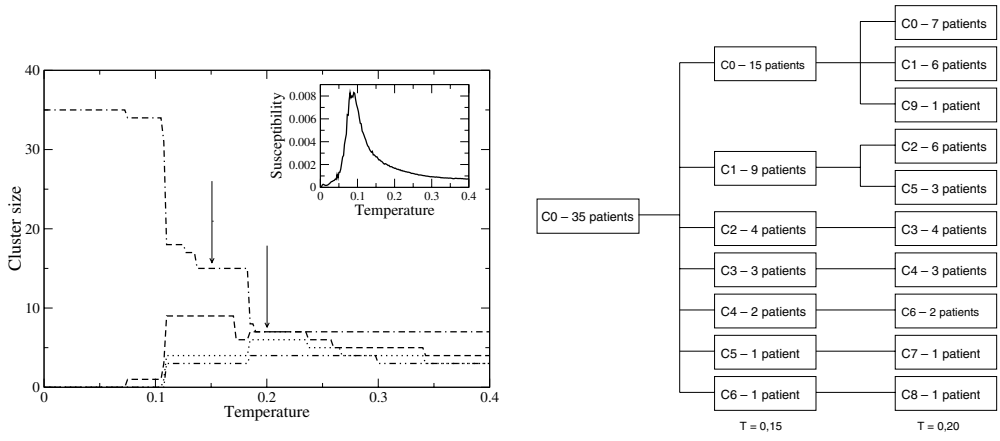

Fig. 1. (a) The cluster size behaviour as a function of temperature (stability parameter) for the four largest clusters: $(\cdot\ -\ -\cdot)$ C0, $(-\ -\ -)$ C1, $(\cdots)$ C2 and $(-\ \cdot\ \cdot\ -\ \cdot)$ C3. The size decreases as the number increases. In the inset, the behaviour of magnetic susceptibility versus temperature identifying the transition region using the SPC code [10]. (b) The dendogram obtained for the classification of 35 case-histories of malaria patients without treatment.

wish to classify into the study of the first-order transition of a ferromagnetic q-state Potts model ($q \geq 10$) between the ferromagnetic and super-paramagnetic phases. The interaction between the spins would depend on the distance between the patients (objects) on the 9-D parameter space and their average number of neighbours. By analysing the behaviour of magnetic susceptibility as a function of the temperature (see inset Figure 1(a)), we arrived at the transition region and obtained the super-paramagnetic clusters from which we could identify the patients that belonged to the different clusters. The temperature is the stability parameter that allows for choosing the best cluster partition (see arrows in Figure 1(a)). After identifying the patients that belonged to the different clusters, we searched for similarities among the elements of a given cluster with respect to the available patient information.

In Figure 1(b), we present the dendogram obtained for two choices of temperatures that correspond to the higher stabilities of the clusters among the possible choices: $T = 0.15$ and 0.20. Starting at $T = 0$ with 35 patients, at $T = 0.15$ they are separated into 7 clusters, the final five remaining the same when the temperature was increased. In our analysis, we have not considered the clusters of one or two patients, and have focussed only on the larger clusters C0-C3. At $T = 0.20$ the first 15 patients of the previous C0 splits into three new clusters (C0, C1 and C9) with 7, 6 and 1 patients respectively, while the former C1 splits into two new ones (C2 and C5) with 6 and 3 patients, respectively. At $T = 0.20$ the previous C2 and C3 are renamed as C3 and C4. While the proportion of positive observations in the first half of the interval between the first and last positive days are almost the same for clusters C0 and C1 at the higher temperature separation, the proportion of the second half and log(parasitemia+1) at the first maximum represent the difference between them. In the case of clusters C2 and C5 (for $T = 0.2$) the proportion of positive observations in the second half of the interval between the first and last positive days are equal for both, but the proportion on the first half is the differentiating parameter. In figures 2 and 3, we show the parasitemia time series of the patients grouped into C2 and C1 ($T = 0.20$) respectively.

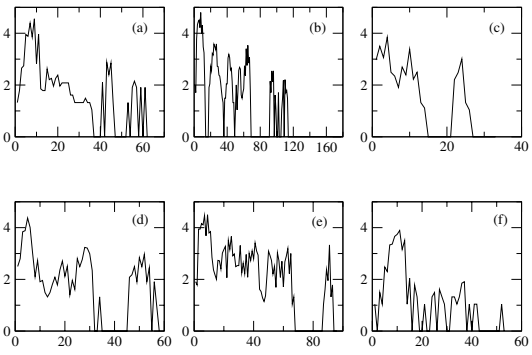

Fig. 2. Log(parasitemia+1) versus time (days) for each of the 6 patients grouped in C2 ($T = 0.2$) of the dendogram shown in Figure 1(b). The original code associated to those patients are: (a) S457, (b) S1212, (c) S910, (d) S775, (e) S895, (f) S1249.

The reason for choosing these clusters is owed to the fact that the similarities between the cluster patients and the behaviour differences between the clusters may be detected by the naked eye.

From the seven patients belonging to C1, six had been inoculated by blood and one by spo. Six out of the seven had been infected with strain Mc Lendon and one with strain Costa. From the six patients belonging to C2, four had been inoculated by blood and 2 by spo via mosquito bites. All of them had been infected with strain Mc Lendon. While in C1 the patients revealed a longer time series, with a monotone downward trend of the local maxima and strong oscillations of the parasitemia at the end of the series, in C2 the time series were shorter (except in patient S1212), exhibiting less oscillations and fewer observations at the second half of the interval between the first and last positive days.

In the following section, we employ a mathematical model to simulate the different case-histories of these two clusters, in order to test if we can extract additional information concerning

the parameters used in the simulations for the different cases, especially those related to the role of the immune system, differences in the parasite reproduction rates, etc.

3 The model

In order to simulate the patterns observed on the parasitemia time series of malaria patients grouped in the clusters shown in Figures 2 and 3, we use a stochastic cellular automata that we have recently introduced [9] to describe the blood cycle of malaria. The CA model captures the essential features of the blood stage of human malaria by taking into account the dynamics of parasite replication, the spread of infection among RBC neighbours at the cytoadherence sites, and the action of the immune system. We consider a two-dimensional lattice ($L \times L$) that will describe the sites of sequestration and adherence of the RBC. To each site we have associated a three-state automaton that represents the non-parasitized, parasitized and dead RBC. We consider Moore neighbourhood to define the rules of local interactions. One time step corresponds to the parallel update of the entire lattice and will correspond to one day.

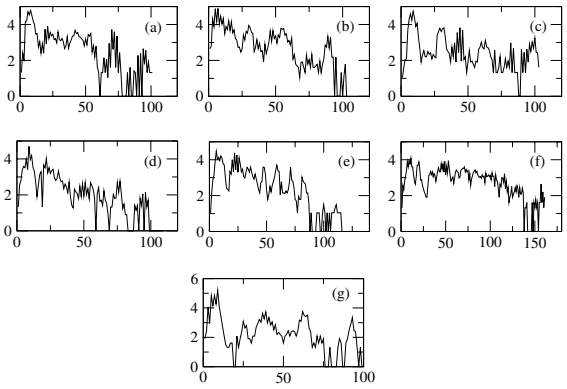

Fig. 3. Log(parasitaemia+1) as a function of time (days) for 7 patients grouped in C1 ($T = 0.2$) of the dendogram shown in Figure 1(b). The original code associated to those patients are: (a) S465, (b) S731, (c) S463, (d) S1102, (e) S1288, (f) S567, (g) S678.

The initial configuration is composed of non-parasitized RBC with a fraction β of parasitized that corresponds to the initial burst of parasitemia generated by the release of merozoites from the liver cells into the bloodstream. After primary exposure there are two phases: the first, varying from $t = 0$ to $t = t_{prim}$, where t_{prim} is the period of time necessary for the immune system to develop specific immune responses. It is only after this period of time that the immune system starts to clear the parasitized cells. The rules of the model will be the same in both stages, except for the action of the immune system on the infected RBC in the second stage. In the model, the non-parasitized RBC either may be infected if there are at least IR parasitized cells in its neighbourhood or may die by regular apoptosis, considering its average life span as 120 days. The parasitized RBC have an incubation period that may vary between 1 and 2 days from cell to cell in the case of P. *falciparum*. In the first stage, during the incubation period, infected RBC may die only by apoptosis, but in the second stage they may either die due to apoptosis or to the action of the specific response with probability p_{is}. A parasitized RBC after incubation time, dies due to the rupture releasing $\mathrm{N}_{mer} = [12, 32]$ parasites in its neighbourhood. The released parasites may infect the neighbours inside a Moore neighbourhood of radio $\mathrm{R} = 2$. Dead cells represent vacancies generated by the death of cells due to the mechanism of apoptosis with probability p_{death}. Vacancies may later be later occupied by healthy RBC, with probability p_{reple} or infected cells circulating in the blood (incoming flux) with probability $p_{reple} * p_{infec}$.

Since we have performed a detailed analysis of the parameter space (not dealt with herein), we were able to search for parameter ranges that could simulate case-histories similar to the observed cases we wished to study. Some parameters of the model are kept constant, since they are inspired in biological data, as for instance the incubation period $t_{incub} = [1, 2]$, probability

of cell death p_{death} based on the average life time of 120 days of the erythrocytes, probability of replenishment p_{reple} that we consider 0.95. After choosing the appropriate range for each parameter, we used the genetic algorithm [13] to generate 10.000 randomly chosen sets of parameters that are used to simulate different case-histories inside the pre-determined intervals. To each simulation (or parameter set) we estimated the prediction error, which is an absolute measure of the dissimilarities of each simulation with respect to the observed case. The prediction error e is given by:

$$e^j = \sqrt{\frac{\sum_{i=1}^{N}(y_i^j - x_i)^2}{N}} \tag{3.1}$$

where j stands for the sample j, N is the length of the time series, y_i^j corresponds to the parasitemia at time i of sample j and x_i is the parasitemia at time i of the observed case-history. The smaller e would correspond to the best set. We consequently proceeded by using the best set obtained in this analysis to generate 400 samples. Somewhere in the region of 10–20% of these samples were discarded since the behaviour of the time series did not permit the estimation of one of the nine parameters. Therefore, by using such variables, such samples could not be compared with the observed case. The differences between the samples time series reside in the infection evolution differences due to distinct spatial distributions of the parasitized cells in the initial configuration. For each of these samples we estimated the prediction error. The best simulation would correspond to the smaller value of e.

In Figures 4, at the same plot we show the best simulation and the observed case-history for two of the 7 patients of cluster 1: S567 (a) and S1288 (b). In table 1, we present the values of the 9 parameters adopted to characterise the time series for the observed case S1288

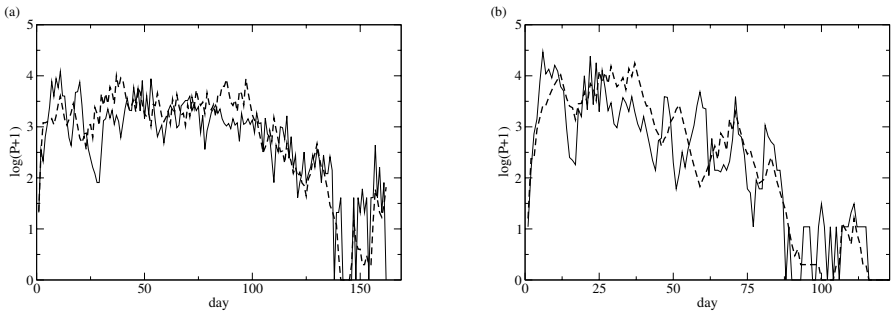

Fig. 4. The continuous line corresponds to the parasitemia counts of the observed case-history and the dashed line to the best simulation. In (a) the best simulation of patient S567 was obtained by adopting the following parameters: $L = 216, t_{prim} = 7$, $Nmer = [12, 19]$, $\beta = 1.910^{-4}$, $p_{is} = 0.416$, $p_{infec} = 5.510^{-4}$, $IR = 6$. For this sample $e = 0.57$. In (b) the best simulation of patient S1288 was obtained for the following parameters: $L = 268$, $t_{prim} = 9$, $Nmer = [12, 20]$, $\beta = 2.310^{-4}$, $p_{is} = 0.415$, $p_{infec} = 1.110^{-4}$, $IR = 6$. For this sample e = 0.63.

Table 1. The behaviour of the 9 parameters adopted to characterise the parasitemia time series, for 325 samples out of 400 trials for patient 1288.

variable	median	deviation	minimum	maximum	patient	best simulation
init. slope	0.276	0.108	0.006	0.649	0.647	0.210
log 1st max.	3.367	0.542	1.322	4.444	4.469	4.019
No max.	6.969	2.021	2.000	12.000	11.000	9.000
slope max.	–0.015	0.014	−0.114	0.025	−0.030	−0.033
GM interv.	15.648	5.886	8.088	52.459	9.910	10.397
SD log interv.	0.246	0.077	0.000	0.478	0.163	0.273
prop. + 1st	0.749	0.241	0.185	1.000	1.000	1.000
prop. + 2nd	0.533	0.293	0.015	1.000	0.828	0.661
last + day	113.249	21.621	23.000	129.000	114.000	110.000

and the best simulated sample. Similar results were obtained for the other patients not shown here, and parameter interval for the best simulations obtained for all patients belonging to this cluster, can be summarised as follows: $L = [268, 337]$, $Nmer = [9, 20]$, $t_{prim} = [7, 9]$, $\beta = [1.010^{-4}, 8.810^{-4}]$, $p_{is} = [0.395, 0.416]$, $p_{infec} = [110^{-4}, 5.410^{-4}]$ and $IR = 6$. We observe that 8 out of 9 parameters are in excellent agreement when we compare the estimates for the best samples and the observed case-histories of C1. In Figures 5 we show the best simulation and the observed case-history for two of the 6 patients of cluster 2: S775 (a) and S895 (b). In table 2, we present the values of the 9 parameters adopted to characterise the time series for the observed case S775 and the best sample obtained from the simulations. In all the simulations performed for C2, when we compare the estimates for the best samples and the observed case-histories 8 out of 9 parameters used to characterise the time series are in excellent agreement. The interval of parameters for the best simulations obtained for patients belonging to this cluster, can be summarised as follows: $L = [307, 393]$, $N_{mer} = [9, 16]$, $t_{prim} = [4, 7]$, $\beta = [1.510^{-4}, 2.6410^{-4}]$, $p_{is} = [0.661, 0.863]$, $p_{infec} = [2.3610^{-4}, 3.0410^{-4}]$ and $IR = [6, 7]$.

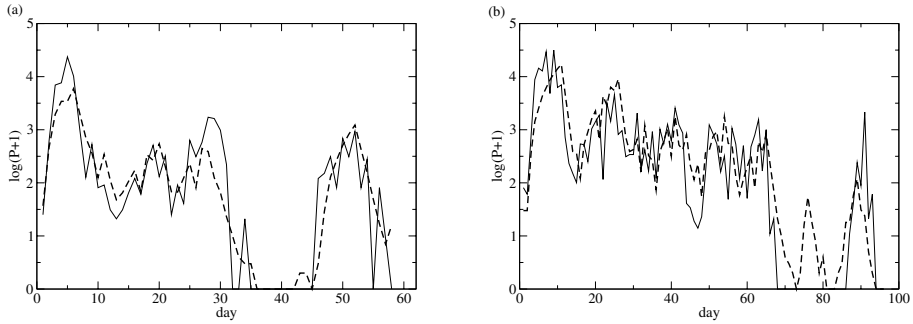

Fig. 5. The continuous line corresponds to the parasitemia counts of the observed case-history and the dashed line to the best simulation. In (a) the best simulation for patient S775 was obtained using the following parameters: $L = 393$, $t_{prim} = 6$, $Nmer = [10, 26]$, $\beta = 2.610^{-4}$, $p_{is} = 0.66$, $p_{infec} = 3.010^{-4}$, $IR = 7$. For this sample $e = 0.58$. In (b) the best simulation of patient 895 was obtained using the following parameters: $L = 309, t_{prim} = 7$, $Nmer = [9, 16]$, $\beta = 1.510^{-4}$, $p_{is} = 0.72$, $p_{infec} = 2.910^{-4}$, $IR = 6$. For this sample $e = 0.66$.

Table 2. The behaviour of the 9 parameters adopted to characterise the parasitemia time series, for 344 samples out of 400 trials for patient S775.

variable	average	deviation	minimum	maximum	patient	best simulation
init. slope	0.261	0.126	−0.012	0.656	0.703	0.396
log 1st max.	3.606	0.729	1.518	5.004	4.373	3.775
No max	4.570	1.097	2.000	8.000	4.000	4.000
slope max.	−0.031	0.035	−0.201	0.047	−0.023	−0.011
GM interv.	11.510	4.540	5.429	39.000	14.461	13.481
SD log interv.	0.146	0.071	0.000	0.399	0.174	0.226
prop. + 1st	0.879	0.150	0.345	1.000	1.000	1.000
prop. + 2nd	0.769	0.209	0.069	1.000	0.500	0.483
last + day	55.529	3.218	35.000	57.000	55.000	57.000

4 Conclusions

In this work we have shown that the simulations of the recently introduced cellular automata model [9] reproduces the parasitemia behaviour of the observed malaria patients belonging to C1 and C2 very well. As in CA models for other diseases [8], such successful results reveal the importance of local interactions for the dynamics of infectious diseases within the patient.

From the cluster analysis we observe that the initial slope, the number of local maxima, the slope of the local maxima and the proportion of positive observations in the second half of the interval between the first and last positive days, are the differentiating parameters of clusters C1 and C2. While in C1, the patients have longer time series with a monotone downward trend of the local maxima, in C2 the time series are shorter (except for patient S1212), they exhibit a small number of oscillations and few observations in the second half of the interval between the first and last positive days. Crossing this information with the results obtained from the simulations performed with the CA model for both clusters concerning the variation of the parameters for both sets, we notice that $L, Nmer, \beta, p_{infec}$ and IR are of the same order of magnitude for both clusters. However, there are differences for t_{prim} that are larger in C1 than in C2, and also for p_{is} that is 50% below the efficiency for C1 and above for C2. The results for the patients belonging to clusters C1 and C2 suggest that the long time and monotone decay series are linked to a lower IS efficiency in killing parasitized RBC and a longer delay in generating specific responses. The short time series on the other hand, are linked to a greater IS efficiency, and quicker delays in generating specific responses. Our conclusion is corroborated by a more complete analysis (that will be published elsewhere) using 79 patients case histories without treatment and reinfection, including in this case the parasitemia time series that could be interpolated [9].

This work was partially supported by CNPq and CAPES (Brazilian Federal Agencies) and by FACEPE (Pernambuco State Agency) under the grant PRONEX/FACEPE EDT 0012-05.03/04, FAPESP (São Paulo State Agency) under the grant 05/51671-6 and FAPESB (Bahia State Agency) under the grant 047/2004. We thank Prof. Luzia Aparecida Trinca (IBB-UNESP) for many helpful discussion during this work. We also thank E. Domany for making the SPC software package available to us.

References

1. L.J. Bruce-Chwatt, *Essential Malariology* (John Wiley & Sons, New York, 1985)
2. D. Waklin, *Immunity to Patrasites: How parasitic infections are controlled*, 2nd edn. (Cambridge University Press, Cambridge, NY, 1998)
3. D. Warrel, Parasitology **94**, S53–S76 (1987)
4. D.J. Roberts, A.G. Craig, A.R. Berendt, R. Pinches, G. Nash, K. Marsh, C.L. Newbold, Nature **357**, 689–692 (1992)
5. L. Molineaux, K. Dietz, Parassitologia **41**, 221–231 (1999)
6. L. Molineaux, H.H. Diebner, M. Eichner, W.E. Collins, G.M. Jeffery, K. Dietz, Parasitology **122**, 379–391 (2001)
7. G.A. Funk, V.A.A. Jansen, S. Bonhoeffer, T. Killingback, J. Theo. Bio. **233**, 221–236 (2005)
8. R.M. Zorzenon dos Santos, S.G. Coutinho, Phys. Rev. Lett. **87**, 161102 (2001)
9. R.M. Zorzenon dos Santos, S.T.R. Pinho, C.P. Ferreira, T.B. Gurgel, L. da Silva, preprint (2006)
10. M. Blatt, S. Wiseman, E. Domany, Phys. Rev. Lett. **76**, 3251–3254 (1996)
11. M. Blatt, S. Wiseman, E. Domany, Neural Comp. **9**, 1805–1842 (1997)
12. W.E. Collins, G.M. Jeffery, Am. J. Trop. Med. Hyg. **61**, 4–19; 20–35; 36–43; 44–48 (1999)
13. M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, 1996)