



**HAL**  
open science

## Forecasting cross-border malaria case number: towards an early warning system to support malaria elimination plans

Thomas Schincariol, Emmanuel Roux, Sylvaine Jégo, Thibault Catry, Florian Girond, Vincent Herbreteau, Yi Moua, Raphael Saldanha, Emilie Mosnier

### ► To cite this version:

Thomas Schincariol, Emmanuel Roux, Sylvaine Jégo, Thibault Catry, Florian Girond, et al.. Forecasting cross-border malaria case number: towards an early warning system to support malaria elimination plans. 7th International conference on Time Series and Forecasting (ITISE 2021), Jul 2021, Gran Canaria, Spain. hal-03381934

**HAL Id: hal-03381934**

**<https://hal.science/hal-03381934v1>**

Submitted on 18 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Forecasting cross-border malaria case number: towards an early warning system to support malaria elimination plans

Thomas Schincariol<sup>1, 2</sup>, Emmanuel Roux<sup>1, 3, 4</sup>, Sylvaine Jégo<sup>1</sup>, Thibault Catry<sup>1, 3</sup>, Florian Girond<sup>1</sup>, Vincent Herbreteau<sup>1</sup>, Yi Moua<sup>1, 3</sup>, Raphael Saldanha<sup>4, 3</sup>, and Émilie Mosnier<sup>5,6</sup>

<sup>1</sup>ESPACE-DEV, IRD, Université de Montpellier, Université de Guyane, Université des Antilles, Université de la Réunion, Montpellier, France

{name.surname}@ird.fr

<sup>2</sup>Université de Technologie de Troyes, Troyes, France

<sup>3</sup>LMI “Sentinela”, Fiocruz, Universidade de Brasília, IRD, Rio de Janeiro, Brazil

<sup>4</sup>Laboratório de Informação em Saúde, ICICT, Fiocruz, Rio de Janeiro, Brazil

<sup>5</sup>Service des Centres Délocalisés de Prévention et de Soins, Cayenne Hospital, Cayenne, French Guiana

<sup>6</sup>SESSTIM, Aix Marseille Université, INSERM, IRD, Marseille, France

**Abstract.** Malaria elimination, one of the Sustainable Development Goals of the United Nation, is challenged by cross-border context specificities. At the French Guiana-Brazil border, a system was developed to harmonized epidemiological data providing by the two countries. This study evaluates the feasibility of using such harmonized data to build a cross-border early warning system. To this end, the study compared ARIMAX and LSTM approaches. Time-lagged meteorological data were introduced to improve the forecasts. LSTM outperformed ARIMAX, with a 10 to 39% decrease of Mean Absolute and Root Means Square Errors, and better predicted low case numbers. Meteorological data improved significantly model predictions, by considering time-lags from 3 to 7 weeks compatible with the knowledge found in the literature. This study demonstrated the feasibility of building a cross-border malaria early warning system, that would significantly contribute to malaria control and elimination.

**Keywords:** ARIMAX, LSTM, Comparative study, Cross-border malaria, Early warning system, Malaria elimination

## 1 Introduction

Malaria is a mosquito-borne parasitic disease mostly present in the tropical and subtropical areas. The global significant decrease in malaria incidence and mortality rates from 2000 to 2015 (minus 37% and 60%, respectively, [1]) led the United Nations to consider the elimination of this disease as one of the sustainable development goals [2], and the World Health Organization (WHO) to

define a Global Technical Strategy aiming at a 90% reduction in global malaria mortality and incidence by 2030 in comparison with 2015 levels [3]. However, some countries that have reached good results in 2015, experienced an upsurge in the following years. It is the case for Brazil, where about 143,000, 129,000 and 194,000 malaria cases were confirmed in 2015, 2016 and 2017, respectively [4]. In French Guiana – a French overseas territory bordering Brazil to the north – the tendency was comparable with a 131% increase in the number of cases (from 258 to 597 cases) from 2016 to 2017 [5], the majority of which being situated in the municipalities bordering Brazil [6]. In fact, territorial specificities associated with the international borders tend to constitute obstacles to disease control and elimination worldwide [7], and malaria at the border between Brazil and French Guiana presents typical features of “cross-border malaria” [8].

Given such a situation, an automatic system for the harmonization and visualization of the epidemiological data provided by the health surveillance systems of the two countries was developed [8]. However, such system does not include forecast capacities. The prediction of the malaria epidemiological situation in the region would significantly facilitate malaria elimination in both countries by helping in planning and carrying out actions i) prior to the malaria case resurgence, and ii) better targeted both in time and space. These actions include: vector control, promotion of the use of impregnated mosquito nets, early resupply of rapid diagnostic tests and medications to health centers, etc.

Malaria is a disease that is very sensitive to environmental conditions, which determine the ecosystem suitability for the malaria vectors (mosquitoes of the genus *Anopheles*) and directly impact vectorial density and capacity. Environmental conditions are consequently key predictive factors of the disease transmission. Previous studies have shown that the number of malaria cases could be predicted using statistical methods and environmental data. The most popular predictive models are based on the (Seasonal) Auto Regressive Integrated Moving Average models with Exogenous Input ((S)ARIMAX) formalism (e.g. in [9], [10], [11]). On the other hand, little interest was shown in methods derived from artificial intelligence like Long short-term memory (LSTM) models, whereas such an approach appears particularly promising. In [12], a LSTM model led to an error rate reduction of 84 to 87% in comparison with an ARIMA model. Such an approach was also applied with success to forecast mosquito borne diseases as Chikungunya and Zika [13]. It was used to predict monthly malaria incidence in India [14] [15] and was combined with ARIMAX, Seasonal and Trend decomposition using Loess (STL) and Back-Propagation Artificial Neural Network (BP-ANN) in an ensemble modelling approach, for malaria prediction in China [16].

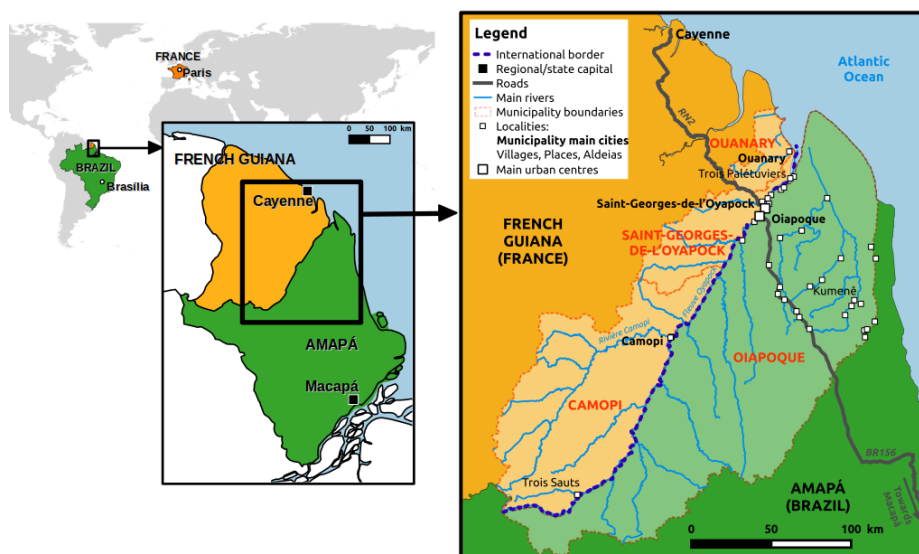
All previously mentioned models consider environmental data (meteorological data provided by in situ meteorological stations, and/or remotely sensed indexes, etc.) as inputs.

In this context, this study aimed at evaluating the feasibility of the prediction of the malaria case number in the French Guiana-Brazil cross-border area, with a view to developing an early warning system. More specifically, the study com-

pared two predictive approaches mentioned earlier: the LSTM and the ARIMAX models.

## 2 Material

### 2.1 Study Area



**Fig. 1.** Map of the study zone, adapted from Mosnier et al., BMC Infectious Diseases (2020) [6]

The cross-border area is defined by the Oiapoque municipality in Brazil and the Ouanary, Saint-Georges-de-l'Oyapock (hereafter referred to as Saint-Georges) and Camopi municipalities in French Guiana (Figure 1). The population is unequally distributed in this area. In the Oiapoque municipality, the 2019 estimation of the Brazilian National Institute of Statistics (IBGE) was 27.270 inhabitants. On the French side, the municipalities of Ouanary, Camopi and Saint-Georges have respectively 201, 1828 and 4220 inhabitants (according to the 2017 estimations of the French National Institute of Statistics and Economic Studies, INSEE). The climate is equatorial with an alternation of rainy (from January to February and from end of April to June) and dry seasons (from July to December and in March).

The main parasite species present in the region is *Plasmodium vivax* [6], which is particularly difficult to eliminate due to the existence of parasites in dormant state in the organism, leading to relapses and contributing to maintain a high parasitic load in the population.

## 2.2 Data

**Malaria Cases** *Individual* malaria cases were provided by the cross-border malaria information system [8] that integrates and harmonizes the epidemiological data provided by the two countries. Data are available since 2003 and 2007 for the Brazilian and French side, respectively. Information on cases includes the *Plasmodium* species, date and locality of notification, residence place, putative place of infection, patients' gender and age (for details on the system and the data, please refer to [8]). Malaria cases on both sides of the border presented several periods associated with significant differences in case number magnitude and spatial distribution, relative importance of *Plasmodium* species, etc. that can be partly explained by significant changes in prevention and control [8]. In this study, the new attacks of *P. vivax* notified during the last period, from January 2014 until December 2019, were considered and weekly aggregated, leading to a 312 data point time-series.

**Meteorological data** Meteorological data were provided by two automatic stations of the French national meteorological service (Meteo-France), situated in Saint-Georges and Camopi localities. Daily measurements are available from 2007 to 2020. Daily minimum, maximum and mean temperature (referred to as TN, TX and TM, respectively) and relative humidity (HN, HX and HM), and accumulated rainfall (RR), were averaged per week to constitute the exogenous variables (or predictors) in the models. For Camopi station, only daily minimum and maximum temperature and rainfall rate were available.

## 3 Predictive Models

### 3.1 General description

**ARIMAX** Auto Regressive Integrated Moving Average models with Exogenous Input (ARIMAX) are defined as follows:

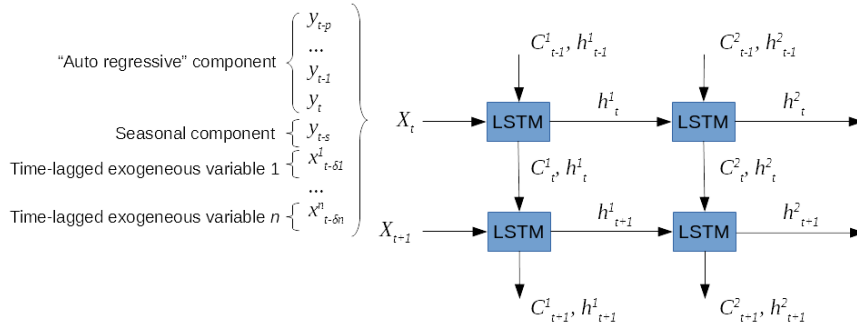
$$y'_t = c + \beta x_t + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \Theta_1 \epsilon_{t-1} + \dots + \Theta_q \epsilon_{t-q} . \quad (1)$$

With:  $y'_t$  the  $d$ -order differencing of the  $y_t$  time-series of malaria case numbers;  $c$  the intercept, a constant;  $\beta x_t$  the exogenous variable part, with  $\beta$  the exogenous variable coefficient and  $x_t$  the value of the exogenous variable  $x$  at timestamp  $t$ ;  $\phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p}$  the autoregressive part, with  $\phi_i$  the coefficient of the  $i^{th}$  autoregressive parameter ( $i \in [1, p]$ );  $\Theta_1 \epsilon_{t-1} + \dots + \Theta_q \epsilon_{t-q}$  the moving average part, with  $\Theta_j$  coefficient of the  $j^{th}$  moving average parameter ( $j \in [1, q]$ ); and  $\epsilon_{t-j}$  the prediction error at the timestamp  $t - j$ .

**LSTM** The LSTM (Long Short-Term Memory) models are part of the Recurrent Neuronal Network (RNN) model family [17]. They can handle time series prediction, with long-term dependencies. They are composed of LSTM *cell*

layers. For each cell, current state ( $C_t$ ) and output (or hidden state) ( $h_t$ ) are updated as a function of the previous cell state and output ( $C_{t-1}$  and  $h_{t-1}$ , respectively) and of the input data vector  $X_t$ , using three filters (referred to as *gates*): a *Forget* gate ( $F_t$ ), an *Input* gate ( $I_t$ ) and a *Output* gate ( $O_t$ ). Current (at  $t$ ) gates, cell state and output are defined as follows [18]:

$$\begin{aligned}
 I_t &= \sigma(w_i[h_{t-1}, X_t] + b_i) \\
 F_t &= \sigma(w_f[h_{t-1}, X_t] + b_f) \\
 O_t &= \sigma(w_o[h_{t-1}, X_t] + b_o) \\
 C_t &= F_t \otimes C_{t-1} + I_t \otimes \tanh(w_c[h_{t-1}, X_t] + b_c) \\
 h_t &= O_t \otimes \tanh(C_t)
 \end{aligned} \tag{2}$$



**Fig. 2.** LSTM model structure

The model structure for the study was composed of two LSTM cell layers, two dropout layers and an output layer (Figure 2). Each dropout layer followed a LSTM layer to avoid overfitting [19]. The training phase was performed using an early stopping function based on the loss function of the test data [20]. The patience of this function was set to 80 epochs. As a consequence, the training did not stop in the staging part of the function while avoiding the overfit. The optimizer used was the Adam version of stochastic gradient descent [21] and the loss function was the mean absolute error.

To make the two model types (ARIMAX and LSTM) comparable, the inputs of the LSTM models are the number of the malaria cases at the  $p$  past weeks (autoregressive component) and  $s$  years (seasonality component) and the time-lagged meteorological data (exogenous variables).

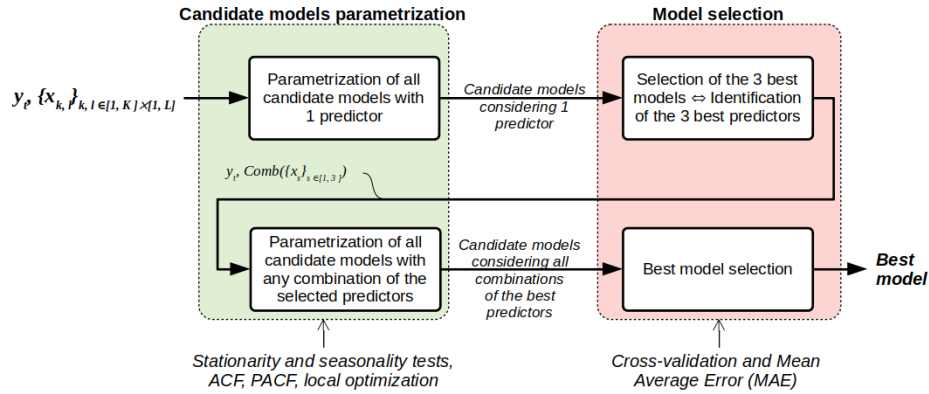
### 3.2 Model parameterization and selection

An iterative parameterization and selection procedure was similarly applied to the ARIMAX and LSTM models. At the first iteration, the best time-lag for

each exogenous variable was determined: 1) a set of candidate models was parameterized, by considering each meteorological station and each time-lagged meteorological variable separately, with time-lags from 0 to 10 weeks. Each candidate model was the one that locally minimized the Akaike Information Criterion (AIC), by considering its possible parameterizations around initial values determined by classical statistical tests and auto-correlation functions (see below for details); 2) Then a cross-correlation procedure was used to select the three candidate models (and their associated time-lagged exogenous variables) with the best 4-week horizon prediction accuracy according to the Mean Average Error (MAE). The MAE was chosen because the comparison is made between models using the same dataset, i.e. in the absence of any scale-dependency problem. At the second iteration, the best combination of previously selected exogenous variables was determined: 1) a set of 9 candidate models was parameterized by considering all possible combinations of the three previously selected exogenous variables; 2) the best final model was the one with a minimum MAE of prediction, using cross-validation.

Models were calibrated with data from 2014 to 2017 and tested with those from 2018 to 2019.

Such a procedure is described in Figure 3. Specificities for each model are details hereafter.



**Fig. 3.** Common overall parametrization and selection procedure for ARIMAX and LSTM models.  $y_t$  is the malaria case time-serie,  $\{x_{k,l}\}_{k,l \in [1,K] \times [1,L]}$  is the set composed of the  $K$  predictors, each of them time-lagged from 1 to  $L$  weeks.  $Comb(\{x_s\}_{s \in [1,3]})$  refers to all the combinations of the selected best predictors.

**ARIMAX** Seasonality and initial model parameters  $p_0$ ,  $d_0$  and  $q_0$  were determined according to statistical tests and auto-correlation function: a seasonality test (Webel-Ollech test) was used to decide or not the inclusion of a seasonal component in the model; a stationarity test (unit-root test) permitted to find the

differencing order  $d$ ;  $p$  and  $q$  parameters were determined by considering the time lag associated with the maximum value of the auto-correlation function (ACF) and partial auto-correlation function (PACF), respectively. These parameters were optimized in a second step by finding a local minimum of the AIC, by testing all the parameter value combinations within the neighborhood ( $\pm 1$ ) of the previously found initial values. Such a procedure were performed iteratively for the ARIMAX models until finding a local minimum of AIC, and only once for the SARIMAX models, due to a prohibitive computing time. Finally, model residuals were tested to check if it corresponds to a white noise.

**LSTM** Each candidate model parameterization was realized in a two step procedure comparable with the one used for the ARIMAX models: 1) determination of initial values of  $p$  and  $s$  according to autocorrelation; 2) identification of a local minimum of the MAE (Mean average error) by exploring the neighboring values of the input pairs ( $p - / + 1, s - / + 1$ ).

The previous model parameterization and selection procedures were applied to predict weekly new *P. vivax* cases with a prediction horizon of 4 weeks and by considering separately three data sets: cases notified for the entire cross-border area, and cases notified by the two national surveillance systems considered separately.

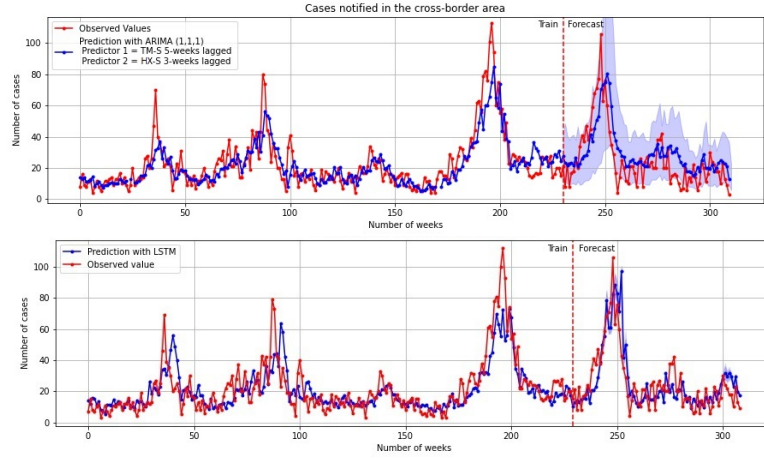
## 4 Results

For the same data set, the two models did not necessarily take into account the same predictors as input (Table 1). Regardless the data set and the model, the selected predictors were the mean temperature (TM, selected 4 times out of 6), the precipitation (RR, 3/6 times), maximum temperature (TX, 3/6 times) and mean and maximum relative humidity (HN and HX, respectively, 1 times each). Meteorological data were time-lagged from 3 to 7 weeks.

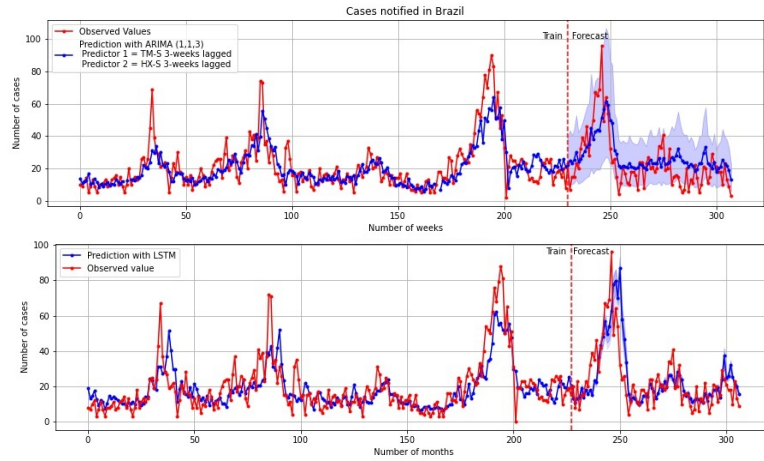
By comparing the models, LSTM model always gave better results except for the prediction RMSE of Brazilian data (with only a 1% increase of the RMSE and with the MAE being 12% lower for LSTM in this case). Improvement in MAE and RMSE range from 10% to 39% (median: 20%). It is worth noting that the higher improvements in the prediction performance was reached with the French Guiana time-series, which presented the lowest numbers of cases per week.

Both ARIMAX and LSTM gave globally satisfactory results in terms of case number dynamics, as shown Figure 4. Both models predicted the 2018 peak (around week number 245 on the figures). However, LSTM better predicted the peak heights and dates of occurrence, the predicted cases increasing more slowly and to a lesser extent with ARIMAX compared to observed cases and LSTM results. Figure 4 also show that LSTM models clearly better predicted low case numbers for the year 2019 (from week 260 to week 312 on the figures) and that prediction uncertainty was significantly lower for LSTM models in all situations.

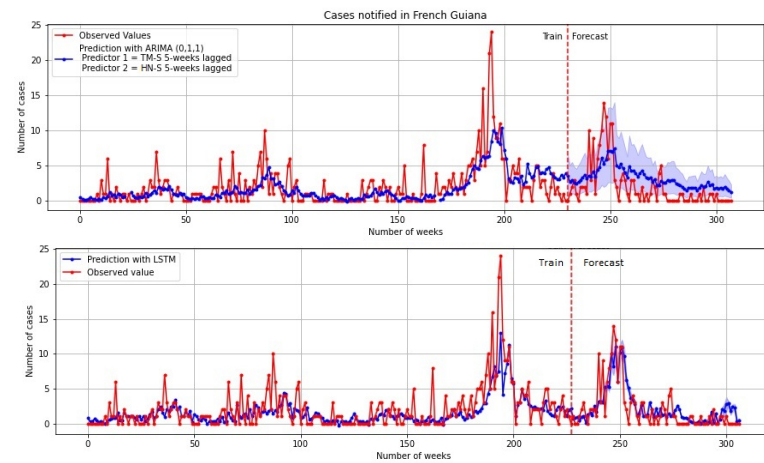




(a)



(b)



(c)

**Fig. 4.** Predictions (4-week prediction horizon) for the training and forecasting periods, for weekly *P. vivax* cases notified in: (a) the cross-border area; (b) the Brazilian surveillance system only; (c) the French Guiana surveillance system only. For each subfigure, ARIMAX and LSTM results are positioned at the top and bottom, respectively. The light blue zone corresponds to the 95% confidence interval.

	ARIMAX ( $p, d, q$ )			LSTM ( $p, s$ )		
Malaria case source	Cross-border area	Brazil	French Guiana	Cross-border area	Brazil	French Guiana
<b>BY WEEK</b>						
<b>Parameter values</b>	(1,1,1)	(1,1,3)	(0,1,1)	(5,2)	(5,2)	(6,3)
<b>Predictor (time lag)</b>	TM <sub>SG</sub> (3); HX <sub>SG</sub> (5)	TM <sub>SG</sub> (3); HX <sub>C</sub> (3)	TM <sub>SG</sub> (5); HN <sub>SG</sub> (5)	RR <sub>SG</sub> (4); TX <sub>C</sub> (7)	RR <sub>SG</sub> (4); TX <sub>C</sub> (7)	RR <sub>SG</sub> (3); TM <sub>C</sub> (5)
<b>MAE</b>	10.5	9.3	2.3	8.4 (-20%)	8.2 (-12%)	1.4 (-39%)
<b>RMSE</b>	12.9	11.9	2.7	11.6 (-10%)	12.0 (+1%)	2.2 (-21%)

**Table 1.** ARIMAX and LSTM model structures and performances at a 4-week prediction horizon estimated by cross-validation. SG and C indexes for meteorological variables refer to Saint-Georges and Camopi stations, respectively. MAE is for Mean Absolute Error and RMSE for Root Mean Square Error.

The computer used for this study was equipped with an AMD Ryzen 4600H processor and 16 Go RAM. The optimization of the ARIMAX and LSTM models lasted 16 m 47 s and 1 h 42 m 53 s, respectively. One prediction with ARIMAX and LSTM lasted 24 s and 3 m 4 s, respectively.

## 5 Discussion

By comparing the most popular predictive models, ARIMAX and LSTM, results showed that both models globally gave satisfactory results in terms of prediction errors. However, LSTM outperformed ARIMAX in many ways: overall prediction errors, malaria peaks temporality prediction, prediction of low case numbers. The overperformance of LSTM was also found in a comparable study [16]. This makes LSTM a very good candidate model for malaria case forecasting and, thus, early warning generation. The fact that LSTM approach particularly improved prediction for low case numbers is of primary interest within the malaria elimination perspective, implying that malaria case number will globally tend to decrease. However, despite such a globally favorable epidemiological situation, it is worth stressing the interest of continuing to monitor and try to forecast malaria cases in this area, particularly because disease resurgences, which anticipation is crucial in the perspective of elimination, are possible locally [22].

On the other hand, the study was performed on the 2014-2019 period, which can be too short to definitely conclude on models performances on longer term. However, as more data is made available monthly, it is fair to expect that the proposed model will deliver better quality results with time. Moreover, we can expect that the introduction of exogenous variables known to be good predictors of malaria since they are directly involved in the eco-epidemiology of the disease, participated to make the proposed model reliable and robust.

In this study, the fact that the temperature (TM and TX) and the rainfall (RR) were frequently selected in the models shows their actual predictive potential. In

fact, mean temperature and rainfall have already been identified as significantly predictive variables of the malaria vector density in the region [23], [24].

Time lags of 3 to 7 weeks are consistent with the lags found in the literature for the region [23]. In [24], the authors considered that the observed vector density could be related to malaria cases notified between 2 and 4 weeks after, and to meteorological conditions observed from 2 to 4 weeks before, which corresponds to a lag of 4 to 8 weeks between meteorological conditions and notified cases.

The study made use of meteorological data provided by two in situ stations, whereas i) the study zone is relatively large and presents different environmental contexts related to climate and land cover and land use, and ii) the malaria transmission depends on many other environmental factors. Several studies made use of remotely sensed environmental indicators, like the Normalized Difference Vegetation Index (NDVI), to predict malaria cases [11] [16]. In the future, such a complementary predictor should be included in the models, by making use of high resolution satellite images like those provided by the Sentinel-2 mission (European Spatial Agency, Copernicus program). Rainfall estimation from satellites [26] could also be used to cover all the study area. Furthermore, integration of additional information on malaria control interventions on both sides of the border and human mobility (particularly related to illegal gold-mining), should significantly improve the forecasts as these phenomena highly contribute to malaria dynamics in the region. Indeed, in [25] was found that more than 30% of the Brazilian malaria cases studied by the authors were imported from French Guiana and linked to illegal gold-mining. However, human mobility is particularly difficult to monitor in a systematic, regular manner (as required in a early-warning system). Consequently, proxies should be found and evaluated before their integration in models. In fact, the actual consideration of forecasts in the decision making by the authorities in charge of the malaria disease control, will depend not only on model performances, but also on their capacity to explicitly apprehend the highly multifactorial nature of the disease.

In this study, forecasts concern the cases in the entire cross-border area and those notified separately by the two national surveillance systems. It is not strictly speaking a spatialization of the prediction, as both surveillance systems include cases which the residence places and/or putative places of infection can be situated on the other side of the border. Predicting malaria case numbers at the locality level would considerably improve the capacity to implement targeted control actions that become necessary in the perspective of elimination. However, at such a spatial scale, case scarcity can severely limit the possibility to build reliable models and a previous locality clustering would be necessary, as in [22]. To be consistent with the inclusion of environmental exogenous variables, case spatialization should consider putative place of infection (and not patients' residence place), but such an information is still rarely informed on the French Guiana side and the confidence in it, on both sides of the border, is limited [8].

## 6 Conclusion

The study implemented and compared ARIMAX and LSTM modelling approaches to predict the number of malaria cases at the French Guiana-Brazil border. It concluded on the superiority of the prediction accuracy of the LSTM approach, both in terms of peak amplitude and temporality, and especially for low numbers of cases. The feasibility of a reliable prediction of the number of cases with a 4-week horizon, and therefore of the building of an early warning system of the disease, have been demonstrated. Such a system would be a highly valuable tool to assist health actors in planning control actions, particularly in terms of risk prevention campaigns among populations and logistical organisation of local health services, and would thus contribute to efforts to eliminate the disease.

**Acknowledgments.** This work was funded by the CNES/TOSCA program (ANISSETTE project), the European Union/FEDER-Guyane/PCIA program (PRO-GYSAT project) and the Labex DRIIHM/OHM-Oyapock.

## References

1. Cibulskis, R.E., Alonso, P., Aponte, J. et al. Malaria: Global progress 2000 – 2015 and future challenges. *Infect Dis Poverty* 5, 61 (2016)
2. United Nations. Sustainable Development Goals. URL: <https://sdgs.un.org/goals> [accessed 2020-10-07] (2015)
3. World Health Organization. Global Technical Strategy for Malaria 2016-2030. Geneva, Switzerland. URL: [http://apps.who.int/iris/bitstream/10665/176712/1/9789241564991\\_eng.pdf?ua=1&ua=1](http://apps.who.int/iris/bitstream/10665/176712/1/9789241564991_eng.pdf?ua=1&ua=1) [accessed 2020-08-11] (2015)
4. Pan American Health Organization. World Health Organization. PLISA Health Information Platform for the Americas. URL: <https://www.paho.org/data/index.php/en/mnu-topics/indicadores-malaria-en.html> (2020)
5. Santé Publique France/CIRE-Guyane. Surveillance du paludisme - Bulletin périodique : nov. 2017 à janv. 2018. Le point épidémiologique 01/2018:1–3 (2018)
6. Mosnier, E., Roux, E., Cropet, C. et al. Prevalence of Plasmodium spp. in the Amazonian Border Context (French Guiana–Brazil): Associated Factors and Spatial Distribution. *Am J Trop Med and Hyg*, 102(1), 130-141.(2020)
7. Wangdi, K., Gatton, M.L., Kelly, G.C. et al. Cross-border malaria: A major obstacle for malaria elimination. *Adv. Parasitol.* 89:79–107 (2015)
8. Saldanha, R., Mosnier, É, Barcellos, C. et al. Contributing to Elimination of Cross-Border Malaria Through a Standardized Solution for Case Surveillance, Data Sharing, and Data Interpretation: Development of a Cross-Border Monitoring System. *JMIR Public Health Surveill* 6(3):e15409 (2020)
9. Hanf, M., Adenis, A., Nacher, M. et al. The role of El Niño southern oscillation (ENSO) on variations of monthly Plasmodium falciparum malaria cases at the cayenne general hospital, 1996-2009, French Guiana. *Malar J* 10, 100 (2011).
10. Anwar, M.Y., Lewnard, J.A., Parikh, S. et al. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malar J* 15, 566 (2016)

11. Girond, F., Randrianasolo, L., Randriamampionona, L. et al. Analysing trends and forecasting malaria epidemics in Madagascar using a sentinel surveillance network: a web-based application. *Malar J* 16, 72 (2017)
12. Siami-Namini, S., Tavakoli N., Siami Namin A., A Comparison of ARIMA and LSTM in Forecasting Time Series. 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 1394-1401 (2018)
13. Coelho, F.C., de Holanda, N.L., Coimbra, B., Transfer learning applied to the forecast of mosquito-borne diseases. medRxiv Preprint. DOI: 10.1101/2020.02.03.20020164 (2020)
14. Santosh, T., Ramesh, D., Reddy, D. LSTM based prediction of malaria abundances using big data. *Computers in Biology and Medicine*, 124, 103859. (2020)
15. Verma, A.K., Kuppili, V. Data-oriented Neural Time Series with Long Short-Term Memories (LSTM) for Malaria Incidence Prediction in Goa, India. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE:1–6 (2019)
16. Wang, M., Wang, H., Wang, J. et al. A novel model for malaria prediction based on ensemble algorithms. *PloS one*, 14(12), e0226910. (2019)
17. Hochreiter, S., Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8), 1735-1780. (1997)
18. Salehinejad, H., Sankar, S., Barfett, J. et al. Recent advances in recurrent neural networks. *Computing Research Repository (CoRR)*. URL: <http://arxiv.org/abs/1801.01078> (2017)
19. Srivastava, N., Hinton, G., Krizhevsky, A. et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.(2014)
20. Caruana, R., Lawrence, S., Giles, C.L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Proceedings of the Neural Information Processing Systems (NIPS) conference 2000*, Denver, CO, USA. *Advances in Neural Information Processing Systems 13*. Leen, T.K., Dietterich, T.G., Tresp, V. (eds.):402–408. (2001)
21. Dozat, T. Incorporating nesterov momentum into adam. URL: [http://cs229.stanford.edu/proj2015/054\\_report.pdf](http://cs229.stanford.edu/proj2015/054_report.pdf) [accessed 2020-10-09] (2015)
22. Mosnier, E., Dusfour, I., Lacour, et al. Resurgence risk for malaria, and the characterization of a recent outbreak in an Amazonian border area between French Guiana and Brazil. *BMC Infectious Diseases*, 20(373), 1-14. (2020)
23. Adde, A., Roux, E., Mangeas, M. et al. Dynamical Mapping of *Anopheles darlingi* Densities in a Residual Malaria Transmission Area of French Guiana by Using Remote Sensing and Meteorological Data. *PLoS ONE* 11:e0164685 (2016)
24. Girod, R., Roux, E., Berger, F. et al. Unravelling the relationships between *Anopheles darlingi* (Diptera: Culicidae) densities, environmental factors and malaria incidence: Understanding the variable patterns of malarial transmission in French Guiana (South America). *Ann. Trop. Med. Parasitol.* 105:107–122 (2011)
25. Gomes, M. D. S. M., Menezes, R. A. D. O., Vieira, J. L. F., et al. Malária na fronteira do Brasil com a Guiana Francesa: a influência dos determinantes sociais e ambientais da saúde na permanência da doença. *Saúde e Sociedade*, 29, e181046.(2020)
26. Ringard, J., Becker, M., Seyler, F., Linguet, L. Temporal and Spatial Assessment of Four Satellite Rainfall Estimates over French Guiana and North Brazil. *Remote Sensing*, 7(12),16441–16459. (2015)