Operational satellite-based temporal modelling of *Aedes* population in Argentina

Manuel Espinosa,1 Eliana Marina Alvarez Di Fino,1,3 Marcelo Abril,1 Mario Lanfri,2 Maria Victoria Periago,1 Carlos Marcelo Scavuzzo2

1Mundo Sano Foundation, Buenos Aires; 2Mario Gulich Institute for Higher Space Studies, Teófilo Tabanera Space Center, Córdoba; 3School of Nutrition, Faculty of Medical Sciences, National University of Córdoba, Argentina

Abstract

*Aedes aegypti* is a vector for Chikungunya, Dengue and Zika viruses in Latin America and is therefore a large public health problem for the region. For this reason, several inter-institutional and multidisciplinary efforts have been made to support vector control actions through the use of geospatial technologies. This study presents the development of an operational system for the application of free access to remotely sensed products capable of assessing the oviposition activity of *Ae. aegypti* in all of Argentina’s northern region with the specific aim to improve the current Argentine National Dengue risk system. Temporal modelling implemented includes remotely sensed variables like the normalized difference vegetation index, the normalized difference water index, day and night land surface temperature and precipitation data available from NASA’s tropical rainfall measuring mission and global precipitation measurement. As a training data set, four years of weekly mosquito oviposition data from four different cities in Argentina were used. A series of satellite-generated variables was built, downloading and resampling the these products both spatially and temporally. From an initial set of 41 variables chosen based on the correlation between these products and the oviposition series, a subset of 11 variables were preserved to develop temporal forecasting models of oviposition using a lineal multivariate method in the four cities. Subsequently, a general model was generated using data from the cities. Finally, in order to obtain a model that could be broadly used, an extrapolation method using the concept of *environmental distance* was developed. Although the system was oriented towards the surveillance of dengue fever, the methodology could also be applied to other relevant vector-borne diseases as well as other geographical regions in Latin America.

Introduction

Mosquitoes are the most important vectors of human infectious diseases in the world, *Aedes aegypti* in particular. It is the main vector of dengue, chikungunya, zika and urban yellow fever (Liang *et al*., 2015). Due to the absence of vaccines for most of these viruses, and the possibility of introduction of other viruses and/or emergence of new diseases (Vasconcelos and Calisher, 2016), vector control is the main tool to control their spread. Transmission is influenced by factors including vector density, virus circulating and human susceptibility (Christophers, 1960). Moreover, the association between the first two factors and climatic variables has been proven, especially with respect to the average daily temperature (Lambrechts *et al*., 2011; Eisen *et al*., 2014; Estallo *et al*., 2015; Mordecai *et al*., 2017). In Argentina, *Ae. aegypti* is the most relevant mosquito in the epidemiologic point of view. This vector is characterized by its presence in urban environments, its preference for breeding in artificial containers (Rueda *et al*., 1990), its egg resistance to desiccation and the feeding behaviour of the female, which bites in multiple occasions during each gonadotrophic cycle (Focks *et al*., 1993a, 1993b).

Ovitraps provide useful information on spatial and temporal distributions of *Ae. aegypti*, allowing monitoring of vector activity (Ritchie, 1984) to determine, for example, when and where the most effective vector control activities should take place (Wong *et al*., 2011; WHO, 2016). Important progress has been made on the epidemiology of dengue and other vector-borne and zoonotic diseases, such as Chagas, malaria, leishmaniasis and hantavirus based on the interdisciplinary concept of landscape epidemiology,
which is focused on producing spatial and temporal predictive risk maps based on environmental landscape features together with field data (Porcasi et al., 2005, 2006; Salomón et al., 2006; Arboleda et al., 2009; Estallo et al., 2011, 2012; Polop et al., 2008; Sarfraz et al., 2014; Rocchini et al., 2015; Espinosa et al., 2016a, 2016b; Rotela et al., 2017). These Latin American studies, based on the use of space technology to solve epidemiological problems, take their cue from pioneer studies carried out in United States and Europe (Beck et al., 2000; Hay 2000; Tran et al., 2004; Ostfeld et al., 2005). In 2011, Argentina started to develop an operational dengue project called health early warning system aimed at health authorities and researchers. This tool is basically a dynamic risk map covering all cities in the country (where each city is represented as a point with a determined risk for each year), based on geospatial technology in an inter-disciplinary and inter-institutional context (Porcasi et al., 2012). This platform, developed using Open Source Software in order to generate a free service, is based on four components: entomological, viral, control and environmental. The first three consist of inputs from health agents working in each city and the fourth uses satellite-generated data. This environmental component, in the initial version of the system, had a constant probability of vector presence and a variable module related to the number of viral cycles as a function of the temperature. This fixed probability map (species presence probability map – niche model) is clearly a strong simplification and could be improved. Variables like precipitation and temperature, with local variability, have proven to influence mosquito development, survival, oviposition and vector abundance, therefore disease outbreaks could be better estimated using these variables as a base (Hopp and Foley, 2001; Morin et al., 2013; Estallo et al., 2012, 2016; Waide et al., 2013; Ostfeld et al., 2005, Morin et al., 2015). The sensor moderate resolution imaging spectroradiometer (MODIS) onboard the Earth-observational satellites Terra and Aqua is one of the most suitable for this particular application, due to its temporal, spectral and spatial resolutions provided in sets of preprocessed, freely available products (Tatem et al., 2004) related to vegetation, surface radiation, and land cover (Justice et al., 2002). Specifically, the normalized difference vegetation index (NDVI) and the land surface temperature (LST) products are examples of remotely sensed variables often used in epidemiological applications (Butt et al., 2011; Porcasi et al., 2012). The normalized difference water index (NDWI) is another useful environmental variable obtained from satellites which can monitor changes in the water content of leaves, while the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM) provide precipitation datasets that are of value (Gu et al., 2007). All these environmental data are highly important in the study of mosquito behaviour.

In this framework, the overall objective of this study was to improve the operative application presented by Porcasi et al. (2012) (at the national scale) by incorporating a dynamic forecast for vector population modelling. The work continued using free-access remotely sensed products to assess Ae. aegypti oviposition activity estimations extending the reach to all of the North Argentinean region, a high-risk area for dengue, chikungunya and zika transmission. In addition, a strategy to extrapolate the models based on the concept environmental distance was incorporated.

### Materials and Methods

The methodology is based on a series of previous studies (Estallo et al., 2012, 2015, 2016; German et al., 2018) but oriented towards an operational application. In particular, we included the modelling of four different cities based on a four-year field data time series and with the following environmental remotely sensed variables as predictors: NDVI, NDWI, LST_{day} and LST_{night} and precipitation (TRMM and GPM) as well as two integral LST variables from the previous year expressing temperatures with a threshold of 10°C (see below). Additionally, we incorporated a strategy to extrapolate the models developed in the towns, from which we have field data, to all other towns based on an environmental distance concept: that is, we engaged in building predictive models of Ae. aegypti oviposition activity based on the similarity (from an environmental point of view) between North Argentina and these cities.

### Study area

The four cities used to build the models are presented in Figure 1: i) Tartagal (80,000 inhabitants) is located on the Argentinean north-west (22°32’ S, 63°49' W) and has an altitude of 450 m. The city has a subtropical native forest environment in addition to crops. Its climate is subtropical with an average annual temperature about 23°C (summer average maximum of 39°C and winter average minimum of 9°C). It has an annual precipitation of 1,100 mm, with a dry season from June to October. ii) Clorinda (60,000 inhabitants) is located on the Argentinean north (25°17’S, 57°43’W), 4 km from the Paraguayan border. This region presents a subtropical climate without a dry season; the annual median temperature is about 23°C and the mean annual precipitation 1,300 mm, due to a rainy season stretching from October to May. iii) Puerto Iguazú (42,000 inhabitants) located in the Northeast of the country (25°36’S, 54°34’W), in the phytogeographic region of the Paranaense jungle. The city is situated on the southern margin of the Iguazu River to the North and Paraná (Brazil) to the West. The climate is subtropical, with an average temperature of 22°C and annual rainfall of around 1,900 mm, without a dry season. iv) Pampa del Indio (9,200 inhabitants) is located in the eastern district of the Gran Chaco Ecoregion. It generally has a temperate climate with an average temperature of about 21.8°C. Despite being in a dry region, precipitation is significant, registering an annual average of 973 mm.

### Oviposition data

Ovitraps consisting of black plastic cups of 1,000 cm³ containing 250 ml of water were used. An oviposition substrate in the form of a wooden tongue depressor (15 cm × 2 cm) attracting the female mosquitoes was placed inside each trap. No attractant infusion was used. Oviposition substrates were replaced weekly and transported to the laboratory to count the eggs and calculate the egg index, which is the total number of eggs found over the total number of positive ovitraps (Gomes, 1998).

### Table 1. Periods of monitoring of oviposition activity.

<table>
<thead>
<tr>
<th>Period</th>
<th>City</th>
<th>Number of ovitraps</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-2016</td>
<td>Tartagal</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Clorinda</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Puerto Iguazu</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Pampa del Indio</td>
<td>20</td>
</tr>
</tbody>
</table>
Oviposition was monitored with these ovitraps located in randomly selected houses in each of the urban areas of the four cities (Table 1). In each house, two ovitraps were installed, one inside the house and another outside in a shaded place at ground level in the backyard of each house, following WHO guidelines (WHO, 2015). The monitoring period used in this study was from August 2012 to July 2016, however only the external ovitrap data was used for this study as they would be more directly correlated with the downloaded satellite-generated environmental variables. The oviposition activity, reflected by the total number of eggs obtained from these traps, is presented in Figure 2. As can be seen, it presents the typical annual behaviour of subtropical zones. The city with the greatest oviposition activity was Puerto Iguazú and the one with the least was Pampa del Indio.

Satellite-derived environmental variables

To characterize the temporal evolution of the mosquito habitats, several satellite-derived variables were used. Vegetation conditions (humidity, temperature and precipitation) (Hay et al., 1997; Estallo et al., 2012) were monitored by the NDVI extracted from the MOD13Q1 satellite product (composed over 16 days) with a 250-meter spatial resolution (Gao, 1996; Estallo et al., 2016). The NDWI, which is related with the content of liquid water and humidity in both soil and vegetation, was also used and calculated from the same MODIS product. Gao’s (1996) definition of NDWI was used and calculated using the band values provided by the MOD13Q1 product, which correspond to the middle infrared (MIR) band and near infrared band reflectance.

LST was used since it provides a good remotely sensed approximation of the environmental temperature (Wan, 1999; Peres and DaCamara, 2004; Kalluri et al., 2007). For LST, the MOD11A2 satellite product, which has a 1-kilometer resolution and is an average values of clear-sky LSTs during an 8-day period, was used. This product included one $LST_{\text{day}}$ and one $LST_{\text{night}}$ which in some sense represent the maximum and minimum temperatures of the 24-hour day (Wan et al., 2004). Local precipitation was obtained from the TRMM (Kummerow et al., 1998) until June 2015 at which time it was replaced by GPM (https://earthdata.nasa.gov/trmm-to-gpm) for the remainder of the study.

The NDWI, NDVI and LST (night and day) for each city were extracted as a mean of two areas of 85 ha (Estallo et al., 2008), one located within the city (the urban area) and the other covering the native vegetation surrounding the city (the rural area). All images were downloaded from the US National Aeronautics and Space Administration (NASA) website (http://e4ftl01.cr.usgs.gov) and imported to GRASS 7.1 software (https://grass.osgeo.org/grass7/) to calculate the mean for each of the two areas for every date. After that, all average values, with the corresponding dates, were exported to produce a Table using R Studio software, where the final temporal series were made and the statistical model built. Taking into account that the periodicity of environmental variables and the

Figure 1. Study area with the four study cities.
sampling captures were different, a linear interpolation was performed in order to obtain values for all the sampling dates (a value for every epidemiological week). Finally, three lag functions were added from the original time series for each variable, corresponding to a one-, two- and three-week lag.

In addition, two integral variables regarding low temperatures from the previous year with a threshold of 10°C were incorporated into the model as it could theoretically play a role for the following summer’s mosquito populations. The first variable, named weeks, was obtained by counting the number of weeks under this threshold in one epidemiological year (August–July). The second, called degrees, gave the number of degrees under the 10°C threshold for each week considered in the weeks variable (Gómez-Braga et al., 2017; German et al., 2018). The procedure to generate the time series in the context of the complete system according to Porcasi et al. (2012) is schematized in Figure 3.

Modelling

As a first step, four models were created, one for each city. From the original 41 environmental satellite-derived variables (German et al., 2018) taking into account the correlation with the oviposition time series and the co-linearity between them, a subset of 11 variables were retained for the modelling: i) NDVI rural lag-1; ii) NDVI urban lag-1; iii) NDWI rural lag-1; iv) NDWI urban lag-1; v) LSTday rural lag-3; vi) LSTday urban lag-3; vii) LSTnight rural lag-1; viii) LSTnight urban lag-2; ix) TRMM rural lag-3; x) Weeks (Number of weeks under the 10°C threshold in rural areas); xi) Degrees (Number of degrees under 10°C threshold in rural areas). Based on previous experience with modelling epidemiological issues with environment remotely sensed variables (Andreo et al., 2009; Rai et al., 2012; Estallo et al., 2016), linear regression modelling was used to develop different multiple linear models. For each city, a first model was developed. Then, the first 10 sampling dates with the most error were removed and a new model created (German et al., 2018). Errors were calculated using the R package (R Core Team, 2014) for plotting model results. Specifically, we checked nonlinear patterns plotting residual errors over fitted values. Then, to assess if residuals were normally distributed, a Q-Q pplot (https://www.rdocumentation.org/topics/qqPlot) was made. Finally, a spread location plot was made to assess if residuals were spread equally along the predictor ranges, and checked for homoscedasticity (i.e., if they all had the same finite variance). The four models, one for each city, were obtained by the following:

\[
Oviposition = \beta_0 + \sum \text{Coef} \times \text{EnvironmentalVariable}
\]  

where \(\beta_0\) represents the value of intercept of the model and Coef, the coefficient for each satellite-derived environmental variable.

Results

Individual models

Table 2 summarizes each model and Figure 4 shows the results of the predicted values of each model and the oviposition activity

![Figure 2. Oviposition activity based on external traps in the four study cities.](image)

<table>
<thead>
<tr>
<th>Performance</th>
<th>Global model</th>
<th>Tartagal model (M-1)</th>
<th>Pampa del Indio model (M-2)</th>
<th>Clorinda model (M-3)</th>
<th>Puerto Iguazú model (M-4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>Estimate P-value</td>
<td>Adjust R2 = 0.429</td>
<td>Estimate P-value</td>
<td>Adjust R2 = 0.691</td>
<td>Estimate P-value</td>
</tr>
<tr>
<td>Intercept</td>
<td>-188.9 0.01881</td>
<td>-167.5 0.02605</td>
<td>-178.7 0.03599</td>
<td>-179.3 0.05969</td>
<td>-1412.8 0.06957</td>
</tr>
<tr>
<td>NDVI, rural lag1</td>
<td>213.3 1.82 e-15</td>
<td>-14.820 0.00131</td>
<td>49.490 0.03599</td>
<td>94.140 0.06957</td>
<td></td>
</tr>
<tr>
<td>NDVI, urban lag1</td>
<td>-207.3 0.00012</td>
<td>-25.620 0.02474</td>
<td>53.420 0.16746</td>
<td>174.0 0.2599</td>
<td></td>
</tr>
<tr>
<td>NDWI, rural lag1</td>
<td>-50.53 0.00190</td>
<td>33.420 0.02474</td>
<td>153.420 0.16746</td>
<td>238.5 0.0722</td>
<td></td>
</tr>
<tr>
<td>NDWI, urban lag1</td>
<td>99.33 0.01687</td>
<td>35.319 0.00131</td>
<td>90.180 0.00131</td>
<td>157.1 0.0970</td>
<td></td>
</tr>
<tr>
<td>Weeks (&lt;10°C)</td>
<td>-8.990 1.92 e-16</td>
<td>-1.187 0.01717</td>
<td>0.287 0.00131</td>
<td>157.1 0.0970</td>
<td></td>
</tr>
<tr>
<td>Degrees (&lt;10°C)</td>
<td>1.459 1.0000754</td>
<td>1.792 0.02617</td>
<td>2.165 0.00303</td>
<td>1.652 0.06381</td>
<td></td>
</tr>
<tr>
<td>LSTday rural lag3</td>
<td>2.462 0.01989</td>
<td>1.266 0.000164</td>
<td>0.1291 0.02506</td>
<td>2.229 0.03006</td>
<td></td>
</tr>
<tr>
<td>LSTday urban lag3</td>
<td>-0.1695 0.86350</td>
<td>1.585 0.03587</td>
<td>-1.027 0.01875</td>
<td>0.53 0.05008</td>
<td></td>
</tr>
<tr>
<td>LSTnight rural lag1</td>
<td>-1.752 0.04635</td>
<td>0.799 0.02723</td>
<td>0.0425 0.32717</td>
<td>1.919 1.986-06</td>
<td></td>
</tr>
<tr>
<td>LSTnight urban lag2</td>
<td>-0.06599 0.94097</td>
<td>-0.517 0.150376</td>
<td>0.000121 0.86690</td>
<td>2.724 2.33e-08</td>
<td></td>
</tr>
<tr>
<td>TRMM, rural lag3</td>
<td>-0.05314 5.55e-13</td>
<td>0.007 0.57108</td>
<td>0.001196 0.803</td>
<td>0.73708 0.0062</td>
<td></td>
</tr>
</tbody>
</table>

NDVI, normalized difference vegetation index; NDWI, normalized difference water index; LST, land surface temperature; TRMM, Tropical Rainfall Measuring Mission.
Figure 3. Processing procedure architecture.

Figure 4. Oviposition activity of each city: fitted values for each model. (A) Tartagal model (M-1); (B) Pampa del Indio model (M-2); (C) Clorinda model (M-3); (D) Puerto Iguazu model (M-4).
in each city. In general, predicted values were similar to the ones sampled and the models followed the annual mosquito behaviour in the four cities.

Then, as shown in Figure 5, the capacity of each model to explain the oviposition activity in the other three cities was assessed: the model of each city was tested with the environmental data of the other three cities and then the result of the models was compared to the real oviposition situation. These model turned out to be very different from each other. Each model conformed to the particular environmental characteristics of each city and therefore not capable to describe the oviposition in the other cities (as they have different environmental characteristics).

**Model generalization efforts**

Because the four models created are very different from each other, in a second step, a global model using the dataset from the four cities was generated in the same manner as before. This global model has greater flexibility and its capability to adapt to the different environmental characteristics of the study area was explored. A summary of the model is detailed in Table 2 and the results presented in Figure 6. As shown, the performance of this global model did not improve but resulted in an even poorer picture than the individual ones. It should be mentioned that the model presented a similar annual decrease and increase in oviposition in Clorinda; however, the predicted values did not generally represent the values sampled. In other words, the model did not show the same annual oviposition behaviour.

**Further development operational approach**

Some more sophisticated modelling techniques are being evaluated (Scavuzzo et al., 2018), but they are not mature enough for an operational implementation. In the section below some further developments are discussed.

**Environmental distance**

Given that the four models for each city were not interchangeable and that the global model did not fit all four cities, this section details the methodology adopted to establish an approximate model that could be used for any other town in the North of Argentina.

The simplest way to interpolate a known variable with consideration to others is nearest-neighbour interpolation (also known as proximal interpolation). This method determines the closest neighbouring pixel known and assumes the variable value of it rather...
than calculate an average value by some weighting criteria or generating an intermediate value based on complicated rules. For our particular problem, given our known four developed models M-1, M-2, M-3, M-4, this means that for an additional point/town localized at position X, we can set model $M_X$ equal to the known model of the nearest city (nearest neighbour approach), i.e. $M_X = M_j$ where $j$ corresponds to the nearest town with $j = 1, 4$.

An improvement on this approach would be to use an average of the four known models weighted by the inverse of the distance of this new point to each of the cities where we have a model. That is, the model of the nearest city will have a stronger weight and the model that is farther away would have a lower weight, using the following formula:

$$M_{(X)} = \sum \frac{M_j}{L_j}$$  \hspace{1cm} \text{Eq. 2}$$

where $L_j$ represents the normalized distance from the city $j$ to $x$ (the geographic localization of the new town). The problem with the previous idea is that it is more reasonable to think that the vector behaviour in a town at point $x$ would be more comparable to the city that is environmentally more similar and not necessarily the one that is closer. In that sense (in the scheme of nearest neighbour), the model for city $j$ should be used if it were more environmentally similar to the town located at $x$. Closer usually means geographically closer, but we can now think of it in the environmental sense (i.e. more similar environmental variables) which naturally leads to the Normalized Environmental Distance concept.

In order to apply these ideas, which ecologically and biologically sound reasonable, we would need to define the variables involved in the concept environmental similarity and then define an operational environmental distance. For our case we used the 19 WorldClim given by Hijmans et al. (2005) from a large time series and in addition altitude (ALT) and a mean value for NDVI for a 10-year period. Once the variables are defined, we can define a generalized distance based on these environmental variables as follows:

$$\text{Dist}(x_1, x_2) = \sqrt{\sum (v_i - v_i)^2}$$  \hspace{1cm} \text{Eq. 3}$$

where $v_i$ is each of 19 bioclimatic variables plus ALT and NDVI. It should be clear that, to compute the sum of equation 3, the variables

![Figure 6. Fitted values of the Global model over oviposition activity in each city. (A) Tartagal model (M-1); (B) Pampa del Indio model (M-2); (C) Clorinda model (M-3); (D) Puerto Iguazú model (M-4).](image-url)
must be previously standardized and dimensionless (it is possible to do that, simply dividing the original value of each environmental variable by their maximum value). Subsequently, we can estimate the environmental distance from each city with location \( x \) and our four modelled cities \( j \), and recalculate the extrapolation method by Eq. 2 but now using the environmental distance. Operatively, to calculate the distances, we need to define a buffer region of 20km around each \( j \) city to characterize the variables of these cities (as a mean of the pixels into the buffer) and use the probabilities class membership of ENVI software (http://www.sigsa.info/productos/envi) to calculate the normalized environmental distance from each pixel to each of the four modelled cities. The result of the normalized environmental distance for each city is presented in Figure 7. It’s important to note that the normalized environmental distance tells us how similar a city is compared to the other three. Pixels with values close to 1 means that it is more similar to the city compared. As an example of how this methodology works, the normalized distance of Tucumán, Santiago del Estero, Corrientes and Salta are described on Table 3. With these distances calculated, an oviposition estimation on each new town can be calculated, for a given week, using M1 to M4, weighted by \( 1/L_j \) but using the environmental variables \( V_i \) in all the cases of this new town for this specific week. The results of oviposition for the week 2014-02-06 in the four new cities are described in Figure 8. It should be clear that this does not represent a validation of the proposed method. It can only be validated when additional and independent field data are available. So, here we can only say that the proposed environmental interpolation method is novel, easily implementable and ecologically and mathematically consistent.

**Table 3. Normalized Environmental Distance to the four models.**

<table>
<thead>
<tr>
<th></th>
<th>Puerto Iguazú</th>
<th>Clorinda</th>
<th>Pampa del Indio</th>
<th>Tartagal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tucumán</td>
<td>0.197</td>
<td>0.011</td>
<td>0.388</td>
<td>0.402</td>
</tr>
<tr>
<td>Santiago del Estero</td>
<td>0.098</td>
<td>0.009</td>
<td>0.576</td>
<td>0.315</td>
</tr>
<tr>
<td>Corrientes</td>
<td>0.491</td>
<td>0.466</td>
<td>0.039</td>
<td>0.002</td>
</tr>
<tr>
<td>Salta</td>
<td>0.112</td>
<td>0.005</td>
<td>0.133</td>
<td>0.749</td>
</tr>
</tbody>
</table>

![Figure 7. Inverse of the Normalized Environmental Distance map (or environmental similarity map).](image-url)
Discussion and Conclusions

This study represents an upgrade of the National Operational Risk System (Porcasi et al., 2012) that is being used in Argentina to allow health decision makers to perform oriented control actions, using a complete set of satellite-derived variables. The methodology we describe herein is based on completely free software and free and globally available space data in order to assure its replication in different endemic regions on the world.

Since the pioneer work of Rogers (1991), Hay (2000), Beck et al. (2000), Gorla (2002), Tran et al. (2004) and Ostfeld et al. (2005), much has been done on the use of remote sensing in epidemiology, especially with regard to vector-borne diseases (Scavuzzo et al., 2018). Despite this, there are very few examples where geospatial technology is used operationally in the national health systems. There is some more evidence on the use of GIS tools for this purpose but very few environmental products derived from remote sensors are used (Rotela et al., 2017; Porcasi et al., 2012). As Yang et al. (2012) stated, the delay in the response and the use of expensive software are the principal barriers for an operational GIS-based early warning system in epidemiology. The use of open-source technologies and free operational and globally available satellite data are therefore the key to guarantee timeliness and affordability.

A conceptual difference between academic study and operative development is that the former starts with the formulation of a scientific hypothesis, while the latter begins with the generation of user requirements that should be translated into system requirements. The contribution of the model presented herein could generate improvements on the national system already in place (Porcasi et al., 2012) by fulfilling the following requirements: i) estimate the risk stratification by applying a multifactor approach at the national level for all the localities; ii) facilitate decision making; iii) update environmental conditions based on remotely sensed data; iv) be expandable in order to cover others regional countries without major changes through the use of a modular concept.

The modular design of the informatics platform allows us to improve the environmental component by introducing a dynamic forecasting of the vector for each town based on oviposition temporal evolution. The system architecture implemented also permits its application of risk mapping to others vector-borne diseases. In China, Yang et al. (2012) developed a similar platform for monitoring schistosomiasis transmission. This contribution represents an improvement of the national system model that has been running since 2011.

This upgrade is related only with the national scale of the system, since decision makers at the national level need a general framework specifically to define the allocation of the economic resources that need to be distributed between the different localities. In general, the algorithm developed drastically improves the temporal resolution of the risk prediction. While the original platform was updated twice a year, this new version is calculated weekly by combining landscape epidemiology concepts and previous data on vector-borne diseases, in this case dengue, supported by research experience in Latin America and new evidence (Porcasi et al., 2005; Porcasi et al., 2006; Salomón et al., 2006; Rotela et al., 2007; Rotela et al., 2010; Khormi et al., 2011).

Satellite-derived environmental variables were used to build temporal models to predict the oviposition activity in four cities in northern Argentina. It was shown that these variables were good predictors of the conditions in the environment (temperature, humidity and precipitation) which influenced oviposition activity. Also, different time-lagged variables were important and justified their inclusion because for some of the variables, the highest correlation was found even with three time lags (LSTday). Moreover, the inclusion of lagged variables is positive from the point of view of forecast. In addition, integral variables were calculated to be used in the model so as to include the low temperature conditions.
for long periods previous to the oviposition sampling week. The individual models M-1 ($R^2 = 0.69$), M-2 ($R^2 = 0.63$), M-3 ($R^2 = 0.58$) and M-4 ($R^2 = 0.68$), are an improvement of previous studies (Estallo et al., 2008; Estallo et al., 2012; Estallo et al., 2016), because the long-term data used are more adequate for obtaining a more general model and not so specific of the local conditions. Another advantage is the use of a more complete set of remotely sensed variables which are all operatively accessible. This last advantage is very important and in line with our institutional objective, which is to generate forecast modelling for every location in the northern region of Argentina, in that way it is an upgrade of Argentina’s National Risk System (Porcasi et al., 2012).

Furthermore, the methodology presented herein could be used as a base for the automation of the procedure for creating the satellite derived variables and modelling results. This is why all the information processing steps of were based on open-source programming languages (Grass and R) in order to be able to reproduce them elsewhere in a common format and in a short time period. Clearly, an automatic computational procedure is essential to produce a national operational framework for the future, which requires working with large amounts of data (big data).

In order to reach the objective of the study, an innovative procedure was used to generalize the four individual models so that the final product could be widely used. This procedure is based on a biological approach and it consists of an interpolation based on the environmental distance or environmental similitude calculated from historic climatic and vegetation conditions. This concept refers to how similar (from an environmental point of view) one place is with respect to another, regardless of the geographical distance between the two places (Hirzel and Arlettaz, 2003; Díaz Villanueva et al., 2018). In a way, this approach resembles the concept of ecomorphology, that is, the relationship between the ecological role of an individual and its morphological adaptations (Leisler and Winkler, 1985). This can also be seen from a convergent evolution perspective, where ecologically similar but geographically distant communities develop similar morphological features, which appear independently in species of different lineages (Wainwright and Reilly, 1994; Fernandez et al., 2012). Both of these concepts point to the importance of the environment regardless of distance. Therefore, the approach proposed in this paper, supposes the application of an ecological concept to a mathematical model. There is no previous use of this concept for this purpose in the literature but we believe it is valid since the platform serves to model a biological phenomenon based basically on epidemiology ideas.

It would be important, in a second stage, to evaluate the effectiveness of the method developed herein, as well as its limitations. Since this platform is based on the interpolation methodology, conceptually solid, it would not require undergoing a validation process. However, the performance of this method would be improved by incorporating new city models, since new ecosystems could be defined, such as distance classes. Therefore, it is necessary to evaluate how many models will be needed to achieve an optimal oviposition activity prediction of the vector.

Finally, it is also important to emphasize that the main public health strategies for the prevention of dengue, chikungunya and zika lies in Ae. aegypti population control. Therefore, having a flexible system that provides timely information on the dynamics of the population of the vector would not only favour the planning of preventive actions, but also achieve greater efficiency and effectiveness of them. Thus, the upgrade presented in this paper is a simple, novel and accessible way to predict the dynamics of the vector in a weekly period. As a next step, we are planning the incorporation of data from more cities in the region, in order to improve the accuracy of the system.

References


Estallo EL, Benitez EM, Lamfri MA, Scavuzzo CM, Almirón WR,


