European Population Conference 2020

Leveraging Climate, Land Cover, and Health Monitoring to Develop a Malaria Early Warning System for the Amazon

William Pan[1], Ben Zaitchik[2], Mark Janko[3], Carlos Mena[4], Francesco Pizzitutti[1], Christina Recalde[2], Andres Lescano[5]
1 Duke Global Health Institute & Nicholas School of Environment, Duke University, USA
2 Department of Earth & Planetary Sciences, Johns Hopkins University, USA
3 Institute for Health Metrics, University of Washington, USA
4 Institute of Geography and Galapagos Science Center, Universidad San Francisco de Quito, Ecuador
5 Research Unit on Emerging Diseases & Climate Change, Universidad Peruana Cayetano-Heredia, Peru

Abstract
Malaria is a vector-borne disease causing an estimated 219 million infections and 435,000 deaths annually. Since 2011, no other region in the world has experienced a larger increase in malaria cases than the Amazon. Three factors have driven this malaria rise: strong ENSO events (2011-12, 2016); withdrawal of the Global Fund to Fight AIDS, Tuberculosis and Malaria; social unrest in Venezuela; and policies conducive to resource extraction, which increases both vector habitat and occupational migration associated with malaria transmission. These realities make malaria control challenging for health systems. Current surveillance and control programs rely solely on weekly case reports and respond to outbreaks with incomplete data (case reporting has a 1-4 week lag). Further, health response occurs in political districts, regardless of the strong environmental and demographic factors driving malaria risk, and independent of environmental policy. To address these challenges, we initiated the development of a Malaria Early Warning System (MEWS) in collaboration with the Peruvian government and support from NASA. Our MEWS forecasts outbreaks with 95% sensitivity and 75% specificity 12 weeks in advance in eco-regions (i.e., defined by demographic and environmental characteristics such as population, climate, hydrology, and land cover), and provides estimates of malaria incidence in small administrative districts with minimal error. The MEWS also includes agent-based models to simulate intervention response over short- and long-term time horizons. Our MEWS is the first early warning system capable of accurately forecasting health risks on a time scale that permits sufficient planning by the health system.

EXTENDED ABSTRACT

Malaria is a vector borne disease causing an estimated 219 million infections and 435,000 deaths annually. Since 2011, no other region in the world has experienced a larger increase in malaria cases than the Amazon [1]. Increasing malaria incidence began after the 2011-12 Amazon flood and withdrawal of the Global Fund to Fight AIDS, Tuberculosis and Malaria (GF). Cases doubled in Peru and Venezuela in less than two years [3, 4]; cases in Colombia doubled by 2016 (total of 83,227) and increased 5-fold in Ecuador [4]. In 2017, more cases of malaria occurred in Amazon-basin countries than any other year in the past decade (773,503 cases), the highest increases in Venezuela, Ecuador, Peru, Colombia, and Brazil. There is no single underlying cause for this increase. Our research in Loreto, Peru has demonstrated both positive impacts of GF investment and negative impacts of withdrawal ([2], Figure 1), but there were also strong ENSO events in 2011-12 and 2016, policies expanding resource extraction (i.e., logging, mining, etc. [5]) and human migration [3, 6, 7], all correlated with malaria risk.

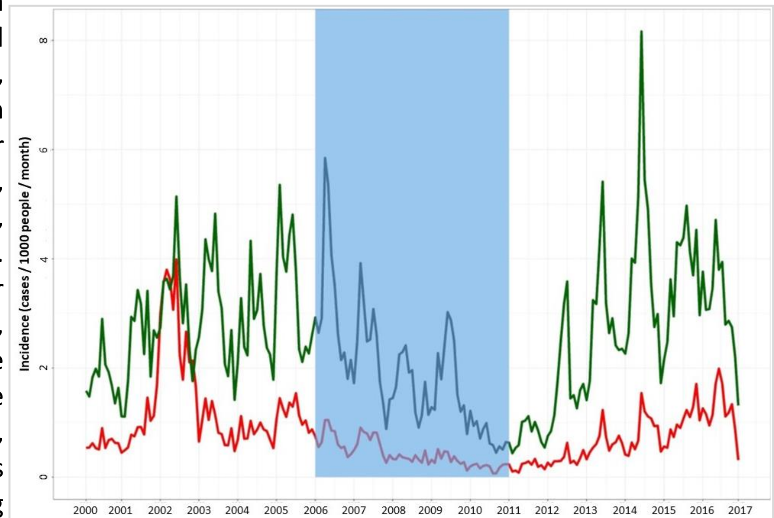

Figure 1. Incidence of ***P. falciparum*** (**red**) and ***P. vivax*** (**green**) in Loreto, Peru, 2000-2017. The shaded blue region is the period during which the Global Fund invested in malaria control in Loreto. Our study shows the GF averted 194,249 malaria cases and withdrawal resulted in an excess of 141,523 malaria cases [2]

Current Surveillance and Response

Current monitoring systems in Peru, Brazil, Colombia, and Ecuador rely on weekly case reports compiled at each health post (clinic) in each administrative health area (e.g., district, canton, municipio). Briefly, when a patient exhibiting malaria symptoms arrives at a clinic, he/she is tested for malaria by microscopy and, if positive, receives treatment. Some cases are verified by rapid diagnostic tests (RDTs). Health post staff report each malaria case to a local Health Center, which is subsequently reported to the regional and national surveillance system. Data transmitted include name, age, sex, national ID, date of diagnosis, malaria type, and residence, but additional data reported varies by country. Each week, ministry personnel tasked with responding to outbreaks compare the reported cases that week to the historical average in that district. If the number of reported cases is >3 standard deviations (SD) above the 6-year geometric mean number of cases from the corresponding week, the ministry declares that an outbreak has occurred (2SD above the mean is considered high risk and 1SD above is medium risk). When outbreaks are officially declared, health ministries (usually) have access to national funds for control, including indoor spraying, bednet distribution, and prophylaxis. However, outbreak declarations are usually based on cases reported 4 weeks prior as reporting delays result in cases reviewed during the concurrent week, on average, are only about 25% complete whereas data from 4 weeks prior are considered 100% up-to-date (according to CDC-Peru). This structural limitation implies that malaria control is reactive rather than proactive, as outbreaks are observed, not predicted. Although the current surveillance system does not provide timely information on malaria risk, **the infrastructure permits the development of probabilistic predictions of outbreaks and surveillance trends, which is a pillar of future malaria control activities** [8].

Malaria and Migration

Human mobility, including permanent residential change and temporary movements (travel, labor, schooling, etc.), is an enigma for human health globally. In the Amazon, population mobility has been associated with a number of adverse effects, including deforestation, urbanization, vector-borne disease risk, and child mortality [9-17]. Rural-rural migration is particularly devastating as increasing population density and policies preventing land ownership induces further settlement into forested areas [12, 13, 18, 19] and has been associated with a phenomenon called *frontier malaria* [20-24]. Labor migration, which is temporary and consists of both long- and short-term migration (i.e., daily to annually), is a strategy to diversify risk through cash income and is regularly practiced by rural families. This type of migration is a major cause of malaria transmission, yet few studies have quantified the proportion of transmission attributable to human movement [25, 26]. For example, a prospective 4-year study by the Amazon ICEMR found that 65% and 85% of incident *P. vivax* and *P. falciparum* cases, respectively, were people who traveled in the past 30 days (Kosek, Pan, Yori, Vinetz, unpublished data), primarily for their occupation. Recent work has tested mobile mapping tools to characterize travel duration, destinations and timing of malaria in the Amazon [25]. And international (labor) migrants have been identified as important sources for malaria reemergence and clustering of cases near borders [27, 28]. Importantly, ABMs developed by our team show that labor migration can sustain hypoendemic transmission [29].

A key component of labor migration is the destination choice, or more specifically, the location of temporary employment. In studies we have conducted in the Peruvian Amazon, the primary forms of occupation migration are logging, mining, fishing, oil/gas extraction or exploration, and construction. The majority of these opportunities occur more than one days travel from a person's home, but in areas connected by river and sometimes by roads. This connectivity often extends beyond the political boundaries of a health system catchment area, but can often be connected ecologically via rivers, forest, and other eco-zones. Connecting space to identify biological niches for animals and plants is a common approach used in environmental science, but not often used in health applications.

NASA Early Warning System

We provide a brief description of our malaria early warning system. The system was developed with NASA support (Pan, PI). The primary goal was to develop a MEWS that accurately detects outbreaks 8-12 weeks in advance and that the system "is completed and 'qualified' by our stakeholder (Peru's Ministry of Health, MINSA) through testing and demonstration in the targeted decision-making activity" (NASA refers to this as *Application Readiness Level 8*, ARL8*)*. A minimum 8-week forecast is required to allow adequate response time.

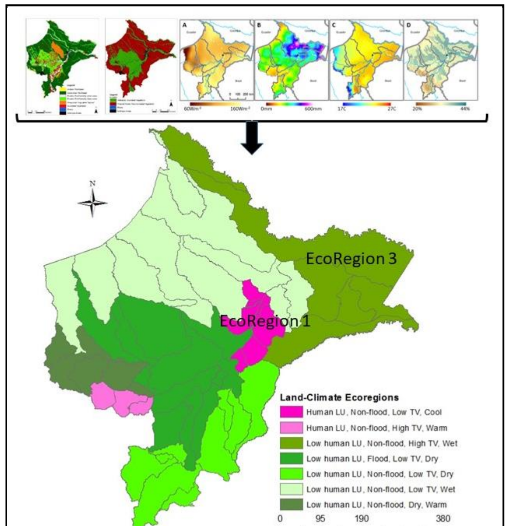

Fig 2. EcoRegions in Loreto, Peru. EcoRegions in Loreto that were identified using Factor Analysis on average monthly values of hydro-meteorological parameters, land cover, and human population size between 2000 and 2010. The small maps above represent NatureServe data, land cover, and 4 LDAS parameters (annual averages).

The MEWS was developed for Loreto, Peru, which consists of 53 administrative districts and a population of almost 900K. The MEWS has four components (detailed in Approach Section): a Land Data Assimilation System (LDAS); EcoRegion forecast model; District-level spatial Bayesian forecast; and Agent-based models. Briefly, the LDAS produces average daily estimates of surface temperature, precipitation, soil moisture, humidity and other hydro-meteorological parameters in 5 KM grid cells that are aggregated to districts. The EcoRegion model produces forecasts 12 weeks in advance using an Unobserved Components Model (UCM) [38-40] in seven predefined areas with similar socio-environmental characteristics (determined via Factor Analysis, Figure 2). The spatial Bayesian model produces complementary forecasts at the district level. The ABM fits scenarios of migration, climate, and interventions to evaluate malaria sensitivity.

Outbreak prediction and location identification occurs in two steps. Since data are not complete in real-time, our forecasts begin 4 weeks prior to the current week (Figure 3). We use the EcoRegion model to detect outbreaks up to 12 weeks in advance, which provides greater accuracy as data are aggregated, minimizing factors such as spatial spillover, vector ecology range, migration (ecoregions tend to have similar occupational labor opportunities), and reporting errors (i.e., residents diagnosed in health posts distinct from location of transmission). If the model detects an outbreak, the spatial Bayesian model is used to determine which district(s) within the EcoRegion will experience an outbreak. The Bayesian model provides a probabilistic interpretation. Ideally, we would like to perform the screening at the district level itself, but diagnostics (sensitivity and specificity) were higher for EcoRegions.

To demonstrate functionality, *P. vivax* forecasts were produced for EcoRegions 1 and 3 (see Figure 2 for locations): EcoRegion 1 surrounds Iquitos (the largest Amazon city in Peru) and is characterized by human land use, low temperature variation, cooler temperatures (compared to other areas of Loreto), and less flood-prone areas; EcoRegion 3 is along the border of Colombia and Brazil, characterized by low human population density, less flood prone, high temperature variation, and high precipitation. The UCM for EcoRegion 1 includes soil temperature and PAMAFRO interventions (ITN distribution, health system strengthening, vector control) as predictors, an autoregressive term, a stochastic trend, and 52-period cycle; EcoRegion 3 UCM included water runoff and PAMAFRO interventions as predictors, an autoregressive term, a stochastic trend, a 3-week dependent lag, and three cycles with periods 3, 26 and 53 weeks. The forecast declares an outbreak if the 95% confidence interval of the forecast overlaps the outbreak threshold (1SD above the 6-year weekly mean). For example, Figure 3 shows the 12 week forecast beginning 4/29/2018, but the 95% CI does not overlap the outbreak threshold for any week.

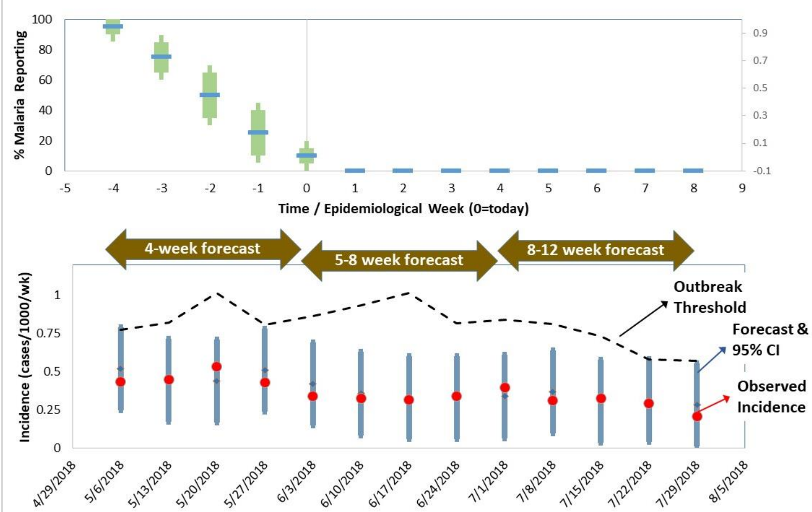

Fig 3. Real-time data reporting, forecast periods, and EcoRegion 1 Forecast between May-July 2018 in Loreto, Peru. The top graph represents the assumed completeness of data from today (day 0) and 4 weeks prior. The bottom graph is the EcoRegion forecast initiated on 4/29/2018: the blue bars are the malaria incidence forecast and 95% confidence interval for EcoRegion 1; the red dot is the observed malaria during that week; and the black dotted line is the 6 year mean plus 1SD used to determine whether an outbreak is expected in the EcoRegion.

Sensitivity (Se), specificity (Se), positive predictive value (PPV) and negative predictive value (NPV) of detecting an outbreak was computed for 1-4, 5-8 and 9-12 weeks from the date of the forecast. 2016 was used as this was the last year with at least two epidemiological weeks were classified as an outbreak. The assessment consisted of producing a 12-week forecast beginning 11/8/2015, repeating every 4-weeks (equivalent to a forecast produced each month). Each 4-week forecast block was compared to outbreak thresholds in that block. If the forecast and observed rates indicated an outbreak, we defined this as a successful detection. If not, it was classified as a false positive or false negative. 9-12

Table 1. 2016 Forecast performance, EcoRegions 1 & 3 (n=13 weeks)

Forecast weeks		TP	FN	FP	TN	Se	Sp	PPV	NPV
Eco-Region 1	1-4	3	0	0	10	100%	100%	100%	100%
	5-8	3	0	1	9	100%	90%	75%	100%
	9-12	3	0	3	7	100%	70%	50%	100%
Eco-Region 3	1-4	1	1	1	10	50%	91%	50%	91%
	5-8	1	1	1	10	50%	91%	50%	91%
	9-12	2	0	3	8	100%	73%	40%	100%

TP=True positive; FN=False Negative; FP=False Positive; TN=True Negative weeks in advance, the models correctly identified all outbreak weeks in each EcoRegion (100% Se); however, Sp was 70% and 73% for EcoRegion 1 & 3, respectively (Table 1).

Our district-level MEWS uses a Bayesian time series framework with a 1st order autoregressive process, and includes cumulative rainfall from the previous 3 months, mean temperature, windspeed, relative humidity, soil moisture and temperature, runoff, season, and PAMAFRO interventions. We include pairwise interactions between seasons and all other covariates to allow environmental and intervention effects to vary by season. We also stratify by district, allowing environmental and intervention effects to further vary spatially. We evaluated overall performance of the district level MEWS by producing 8-week forecasts from 2007-2019 for each epidemiological week, computing Se and Sp for detecting outbreaks of *P. vivax* (Table 2) and the root mean square prediction error to assess model fit, which provides confidence that we are accurately forecasting the number of cases. The MEWS performs considerably better in EcoRegion 1 than 3, with most sensitivities exceeding 0.8 and specificities exceeding 0.70. EcoRegion 3, which borders Colombia and Brazil, never achieves sensitivities above 0.60 (Table 2). We hypothesize that poor performance is related to population movement along the Peru-Colombia-Brazil border. Regardless, there is good overall model fit with low prediction error for Fernado Lores district (Figure 4), which had the lowest sensitivity of all districts, and Ramon Castilla, which had the highest specificity in EcoRegion 3.

Table 2. Sensitivity & Specificity of 8-week district forecasts, 2007-2019

District	Se	Sp
Ecoregion 1		
Iquitos	88%	84%
Fernando Lores	51%	84%
Punchana	89%	74%
Belen	79%	70%
San Juan Bautista	97%	67%
Jenaro Herrera	94%	98%
EcoRegion 3		
Ramon Castilla	57%	79%
Pebas	54%	68%
Yavari	55%	63%
San Pablo	60%	76%

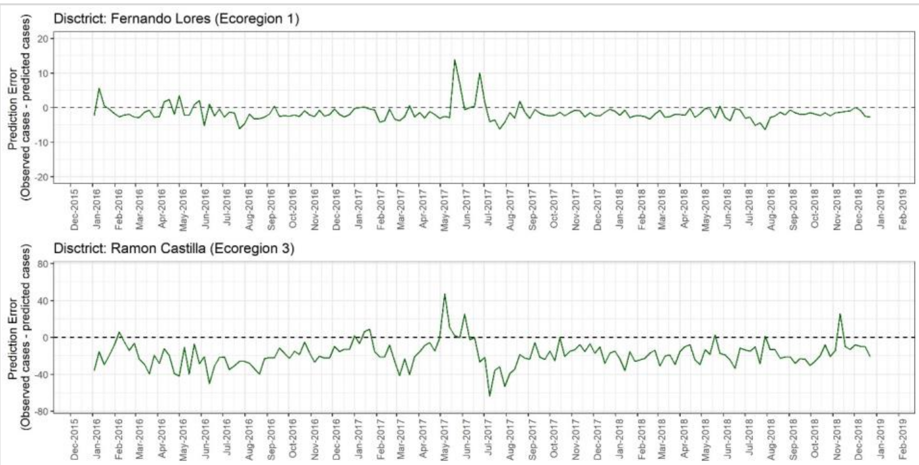

Figure 4. Root-mean square prediction error, Fernando Lores and Ramon Castilla districts, 2016-2019. Both districts have low sensitivity to outbreak detection; however, the RMSE is relatively low, particularly in Fernando Lores. As noted in the text, Ramon Castilla is along the border with Brazil and the model tends to overestimate cases

Implications for Disease Control

Malaria has reemerged in the Amazon and is a major source of morbidity, lost wages, and continued strain on health resources. Malaria's reemergence is coincident with overlapping political, economic, demographic, and environmental factors, including withdrawal of international support for control (exceptions being government-sponsored research), major ENSO events coupled with land use change, neoliberal resource extraction policies, and human migration/mobility (i.e., both internal labor migration and international migration, such as from Venezuela). Creation of an accurate and cost-effective EWS for malaria has been identified as a key component of the regional malaria elimination plan. In 2016, member states of the Pan American Health Organization approved the Action Plan for Malaria Elimination 2016-2020, a pledge to continue reducing malaria through 2020. On World Malaria Day 2019, PAHO announced "Municipalities for Zero Malaria", an initiative to eliminate malaria in areas with the highest burden. The ability to accurately forecast malaria is critical to achieving this goal; however, no forecasting capability exists in any national malaria surveillance system and Amazon-basin countries have an immediate need for such a tool.