ACADEMIE UNIVERSITAIRE WALLONIE-EUROPE
UNIVERSITE DE LIEGE
FACULTE DE MEDECINE

Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon

Elisa Solano-Villarreal

FOR ACADEMIC DISSERTATION
Thesis defense: May 22nd., 2024

Members of the jury

President:	**Prof. Philippe Lejeune** Faculty of Gembloux Agro-Bio Tech (GxABT) University of Liege, Liege - Belgium.
Promotor:	**Prof. Marie-Pierre Hayette** Laboratory of Clinical Microbiology, University Hospital of Liège (CHU) University of Liege, Liege - Belgium.
Co-Promotor:	**Prof. Niko Speybroeck** Institute de Recherche Santé et Societé Faculté de Santé Publique Université Catholique de Louvain, Bruxelles - Belgium
Supervisor:	**Ph.D. Angel Rosas** Institute de Recherche Santé et Societé Faculté de Santé Publique. Université catholique de Louvain, Bruxelles – Belgium
External member:	**Prof. Catherine Linard** Department of Geography, University of Namur, Namur - Belgium
International member:	**Prof. Cesare Furlanello** HK3Lab founder, data scientist. Applied Machine Learning. Predictive Models for Health & Environment. Web Valley School. Rovereto (Trento) & Bergamo - Italy

To my mother, my husband, and my lovely daughters
Leticia and Florencia, because you are the center of my life,
my strength and my joy.

Acknowledgments

I would like to express my deepest gratitude to Marie-Pierre Hayette and Niko Speybroeck, for being dedicated, flexible, and encouraging promotors. I won't forget your advice and the way you helped me to solve the problems.

To Angel Rosas-Aguirre for your knowledge and expertise about malaria, which was crucial for improving this research. Also, many thanks for allowing me to share great moments with your fantastic family, Ana, Fabrizio, Sebastian, and lovely Joaquin.

To Catherine Linard, for your kindness in sharing scientific information that was crucial in this thesis and for your final corrections. To Pearcy Morgan for your kind corrections and your friendship. To Cesare Furlanello, for your guidance in different moments of my professional life. To Philippe Lejeune for your guidance as the president of the thesis committee.

To Walter Valdivia for your generosity, for sharing mountains of knowledge in multiple fields, for being my friend, my partner, my inspiration, and especially for holding me in critical moments. To Leticia and Florencia for giving me many reasons to feel fortunate at every sunrise. To my mother, mama Maria, for your endless reason, for your fight during adversity, and, for being my teacher in the wide sense of this word.

To Bruce Millies for your friendship, your English support, and for allowing me to know more about the world through your documents, pictures, and stories. To Erith Muñoz for your generosity in sharing your knowledge on GEE. To Afi Agboli, for listening to my doubts and answering with certainty. To Jean-Claude for your help and friendship. To Ana Cuellar for your support in doing a Ph.D. abroad. To Cesar Cabezas, and Edgardo Nepo for kindly helping to solve questions about Peruvian malaria. To Gabriela Bidegain, Nathalie Maleve, Stephanie Wauquier and Yulissa Vásquez for administrative and logistic support. To Martine Mayer, my landlady. To MAREARTE geek team, the best place to talk about specialized earth science topics. To my extended family for your support and confidence, and for all my friends.

My heartfelt thanks to the University of Liege, the University Catholic Louvain, and the Peruvian University Cayetano Heredia for the support in developing this research and allow

me to grow professionally. To ARES for allowing students around the world to access the science world regardless of national origin, race, gender, or culture.

This study was funded by the Peruvian National Council of Science - CONCYTEC (008-2014-FONDECYT) and the Académie de Recherche et d'Enseignement Supérieur - Commission de la Coopération au Développement of Belgium (ARES-CCD, PRD-Peru 2014–2019).

Table of contents

List of tables

List of figures

List of maps

List of appendix figures

List of appendix tables

List of appendix maps

Abbreviations and acronyms

API	Annual Parasite Index
API	Application Programming interface
AUC	Area Under Curve
BRT	Boosted Regression Trees
CAR	Cumulative Rainfall
CDC	National Center of Epidemiology, Disease Prevention and Control (Peru)
CHA	Community Health Agents
CQ	Chloroquine
COEN	National Emergency Operations Center (Peru)
CONIDA	National Commission for Aerospace Research and Development
DN	Digital numbers
DIRESA	Direction Regional of Health (Peru)
DIGESA	General Directorate of Environmental Health
DMA	Drug Massive Administration
EIRs	Entomological Inoculation rates
EO	Earth observation
ENSO	El Niño-Southern Oscillation
FC	Forest coverage
FL	Forest Loss
FSAT	Focused screening and treatment
GEE	Google Earth Engine
GIS	Geographic Information Science / Geographic Information System
JRC	Joint Research Center
IDE	Interactive Development Environment
IIAP	The Research Institute of the Peruvian Amazon
IGN	National Geographic institute (Peru)
INDECI	National Institute of Civil Defense (Peru)
INEI	National Institute of Statistics and Informatics (Peru)
IRS	Indoor residual spraying
ITN	Insecticide treated net
ITRF	International Terrestrial Reference Frame

IQR	Interquartile range
LA	Latin America
LDCM	Landsat Data Continuity Mission
LST	Land Surface Temperature
LM	Light microscopy
M	Means
Mdn	Medians
MDT	Digital Terrain Model
MIDIS	Ministry of Development and Social Inclusion (Peru)
ML	Machine Learning
MoH	Ministry of Health (Peru)
MODIS	Moderate-Resolution Imaging Spectroradiometer
MOOC	Massive Open Online Course
NASA	The National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NGA	United States National Geospatial-Intelligence Agency
NIH	National Institute of Health (Peru)
NIR	Near Infrared
NMCP	National Malaria Control Program (Peru)
ORAS	Andean Health Organization
PAHO	Pan American Health Organization
PAMAFRO	Project for malaria control in border areas
PCR	Polymerase Chain Reaction
PDP	Partial Dependence Plots
Pf	*Plasmodium falciparum*
PQ	Primaquine
POPD	Population Density
PPE	Personal Protective Equipment
Pv	*Plasmodium vivax*
RC	Relative Contributions
RDT	Rapid Diagnostic Test
RENAES	National registry of health establishments (Peru)
RBM	Roll Back Malaria

RS	Remote Sensing
SD	Standard Deviation
SDR	Euclidean shortest distance to rivers
SDSS	Spatial Decision Support Systems
SIS	Integrated System of Health
SEHINAV	National Hydrography and Navigation Service (Peru)
SELPER	Latin American Society for Remote Sensing and Spatial Information Systems
SNEM	National Malaria Eradication Service (Peru)
SQL	Structured Query Language
SRTM	Shuttle Radar Topography Mission
SWIR	Shortwave Infrared
TED	Technology, Entertainment and Design
TIR	Thermal Infrared
TRMM	Tropical Rainfall Measuring Mission
TPV	Time to major populated villages/towns
UPCH	Peruvian University Cayetano Heredia
UTM	Universal Transverse Mercator
WGS	World Geodesic System
WRS	Worldwide Reference System
ZMP	Zero Malaria Plan

Glossary

Algorithm:	Is a sequence of instructions that should be carried out to transform the input to output. (1)
Coordinate:	An ordered set of data values that specifies a location on the Earth, which can be absolute or relative. The coordinates of a point on the earth's surface are measured in degrees of latitude and longitude. (2)
Geographic Coordinate System:	Reference system that uses latitude and longitude for locating or referencing points on the earth's surface and its representation by zones. The system is based on the measurement of angles in sexagesimal degrees. (2)
Adverse drug reaction:	A response to a medicine that is harmful and unintended and which occurs at doses normally used in humans. (3)
Annual blood examination rate:	The number of people receiving a parasitological test for malaria per unit population per year. (3)
Annual parasite Index:	The number of confirmed new cases from malaria registered in a specific year, expressed per 1,000 individuals under surveillance, for a given country, territory, or geographic area. Annual parasite index (API) refers to high and moderate malaria transmission risk areas. (3)
Anopheles, infected:	Female *Anopheles* mosquitoes with detectable malaria parasites.(3)
Anopheles, infective:	Female *Anopheles* mosquitoes with sporozoites in the salivary glands. (3)
Anopheline density:	Number of female anopheline mosquitoes in relation to the number of specified shelters or hosts (e.g., per room, per trap or per person) or to a given period (e.g., overnight or per hour), specifying the method of collection. Note: This term refers strictly to the population density or abundance of adult female Anopheles *mosquitoes.* Anopheline mosquito density is a highly insensitive measure of malaria transmission. (3)

Anthropophilic:	Description of mosquitoes that show a preference for feeding on humans, even when non-human hosts are available. Note: A relative term requiring quantification to indicate the extent of preference for anthropophily versus zoophily; usually expressed as the human blood index (proportion of mosquitoes that have fed on humans out of total fed). (3)
Artemisinin-based combination therapy:	A combination of an artemisinin derivative with a longer-acting antimalarial drug that has a different mode of action. (3)
Artemisinin-based combination therapy:	A combination of an artemisinin derivative with a longer-acting antimalarial drug that has a different mode of action. (3)
Biting rate:	Average number of mosquito bites received by a host in a unit time, specified according to host and mosquito species (usually measured by human landing collection). (3)
Active case detection:	Detection by health workers of malaria cases at community and household levels, sometimes in population groups that are considered at high risk. Active case detection can consist of screening for fever followed by parasitological examination of all febrile patients or as parasitological examination of the target population without prior screening for fever. Note: Active case detection may be undertaken in response to a confirmed case or cluster of cases, in which a population potentially linked to such cases is screened and tested (referred to as "reactive case detection"), or it may be undertaken in high-risk groups, not prompted by detection of cases (referred to as "proactive case detection"). (3)
Passive case detection:	Detection of malaria cases among patients who, on their own initiative, visit health services for diagnosis and treatment, usually for a febrile illness. (3)
Case detection:	One of the activities of surveillance operations, involving a search for malaria cases in a community. Note: Case detection is a screening process in which the indicator is either the presence of fever or epidemiological attributes such as high-risk situations or groups. Infection detection requires use of a diagnostic test to identify asymptomatic malaria infections.(3)

Case notification:	Compulsory reporting of all malaria cases by medical units and medical practitioners to either the health department or the malaria control program, as prescribed by national laws or regulations. (3)
Case relapsing:	Malaria case attributed to activation of hypnozoites of *P. vivax* or *P. ovale* acquired previously. Note: The latency of a relapsing case can be > 6–12 months. The occurrence of relapsing cases is not an indication of operational failure, but their existence should lead to evaluation of the possibility of ongoing transmission. (3)
Case suspected malaria:	Illness suspected by a health worker to be due to malaria, generally on the basis of the presence of fever with or without other symptoms. (3)
Catchment area:	A geographical area defined and served by a health program or institution, such as a hospital or community health centre, which is delineated on the basis of population distribution, natural boundaries and accessibility by transport. (3)
Cluster:	Aggregation of relatively uncommon events or diseases in space and/or time in numbers that are considered greater than could be expected by chance. (3)
Confirmed case:	Malaria case (or infection) in which the parasite has been detected in a diagnostic test, i.e. microscopy, a rapid diagnostic test or a molecular diagnostic test. Note: On rare occasions, the presence of occult malaria infection in a blood or organ donor is confirmed retrospectively by the demonstration of malaria parasites in the recipient of the blood or organ. (3)
Database:	A system to store information in digital format, organized to enable particular queries. (2)
Diagnosis:	The process of establishing the cause of an illness (for example, a febrile episode), including both clinical assessment and diagnostic testing. (3)
Digital Terrain Model (DTM):	Representation in computer graphics of relief by using x, y, and z coordinates distributed in an organized or random manner. (2)

Drug efficacy:	Capacity of an antimalarial medicine to achieve the therapeutic objective when administered at a recommended dose, which is well tolerated and has minimal toxicity. (3)
Drug resistance:	The ability of a parasite strain to survive and/or multiply despite the absorption of a medicine given in doses equal to or higher than those usually recommended. Note: Drug resistance arises as result of genetic changes (mutations or gene amplification) that confer reduced susceptibility." (3)
Electromagnetic spectrum:	The set of all forms of electromagnetic energy is arranged in ascending or descending order of their respective wavelength, frequency, or energy. (3)
Endemic area:	An area in which there is an ongoing, measurable incidence of malaria infection and mosquito-borne transmission over a succession of years. (3)
Endophagy:	Tendency of mosquitoes to blood-feed indoors. Note: Contrasts with exophagy. (3)
Endophily:	Tendency of mosquitoes to rest indoors. Note: Contrasts with exophily; usually quantified as the proportion resting indoors; used in assessing the effect of indoor residual spraying. (3)
Entomological inoculation rate:	Number of infective bites received per person in a given unit of time, in a human population. Note: This rate is the product of the “human biting rate” (the number of bites per person per day by vector mosquitoes) and the sporozoite rate (proportion of vector mosquitoes that are infective). At low levels of transmission, the estimated entomological inoculation rate may not be reliable. (3)
Erythrocytic cycle:	Portion of the life cycle of the malaria parasite from merozoite invasion of red blood cells to schizont rupture. The duration is approximately 24 h *in P.* knowlesi, 48 h *in P.* falciparum*, P. ovale and P.* vivax and 72 h *in P.* malariae. (3)

Exophagy:	Tendency of mosquitoes to feed outdoors. Note: Contrasts with endophagy; usually quantified as the proportions biting hosts outdoors versus indoors, conveniently assessed by comparative human landing catches outdoors and indoors or by observation of biting rates on non-human hosts outdoors. (3)
Exophily:	Tendency of mosquitoes to rest outdoors Note: Contrasts with endophily; usually quantified as proportions resting outdoors and indoors; used in estimating outdoor transmission risks. (3)
Georeferencing:	The process of assigning a geometric reference to a phenomenon on the Earth's surface; for example, giving coordinates to health services. Georeferencing mechanisms can be generally classified into metric georeferencing and indirect georeferencing. Metric georeferencing, also called continuous georeferencing, is coordinate-based. Indirect georeferencing methods retrieve the metrically georeferenced locations through attribute data. (2,4)
Gonotrophic cycle:	Each complete round of ovarian development in the female mosquito, usually after ingestion of a blood meal, to yield a batch of eggs. Gonotrophic harmony is achieved when every blood meal results in one batch of eggs from the gonotrophic cycle. (3)
GPS (Global Positioning System):	Is a global navigation satellite system that provides location, velocity, and time synchronization. It is based on a global network of satellites that transmit radio signals from medium-earth orbit. There are four constellations that provide this service. However, Global Positioning System (GPS) satellites developed and operated by the United States is most commonly used. (2)
House-spraying:	Application of liquid insecticide formulation to specified (mostly interior) surfaces of buildings. (3)
Hypnozoite:	Persistent liver stage of *P. vivax* and *P. ovale* malaria that remains dormant in host hepatocytes for variable periods, from 3 weeks to 1 year (exceptionally even longer), before activation and development into a pre-erythrocytic schizont, which then causes a blood-stage infection (relapse). (3)

Incubation period:	Period between inoculation of malaria parasites and onset of clinical symptoms. Note: The shortest incubation period in mosquito-borne infections ranges from 7 days for *P. falciparum* to 23 days for *P. malariae*. The long incubation for *P. vivax* and *P. ovale* (from 3 weeks to 1 year and exceptionally many years) is due to the activation of hypnozoites. (3)
Indoor residual spraying:	Operational procedure and strategy for malaria vector control involving spraying interior surfaces of dwellings with a residual insecticide to kill or repel endophilic mosquitoes. (3)
Indoors:	Inside any shelter likely to be used by humans or animals, where mosquitoes may feed or rest. Note: Where indoor-resting mosquitoes can be targeted for indoor residual spraying." (3)
Infectious:	Capable of transmitting infection, a term commonly applied to human hosts. (3)
Infective:	Capable of producing infection, a term commonly applied to parasites (e.g., gametocytes, sporozoites) or to the vector (mosquito). (3)
Infrared:	The portion of the electromagnetic spectrum includes wavelengths between 0.7 µm and 100 µm. (2)
Insecticide resistance:	Property of mosquitoes to survive exposure to a standard dose of insecticide; may be the result of physiological or behavioural adaptation. (3)
Insecticide:	Chemical product (natural or synthetic) that kills insects. Ovicides kill eggs; larvicides (larvacides) kill larvae; pupacides kill pupae; adulticides kill adult mosquitoes. Residual insecticides remain active for an extended period. (3)
Latitude:	The angular distance north or south from the earth's equator measured from 0 through 90 degrees. (2)
Longitude:	The angular distance of a meridian (a great circle passing through the north and south poles) from the prime (0 degree) meridian that passes through Greenwich, England. (2)

Long-lasting insecticidal net:	A factory-treated mosquito net made of material into which insecticide is incorporated or bound around the fibres. The net must retain its effective biological activity for at least 20 WHO standard washes under laboratory conditions and 3 years of recommended use under field conditions. (3)
Malaria control:	Reduction of disease incidence, prevalence, morbidity or mortality to a locally acceptable level as a result of deliberate efforts. Continued interventions are required to sustain control.(3)
Malaria elimination:	Interruption of local transmission (reduction to zero incidence of indigenous cases) of a specified malaria parasite in a defined geographical area as a result of deliberate activities. Continued measures to prevent re-establishment of transmission are required. Note: The certification of malaria elimination in a country will require that local transmission is interrupted for all human malaria parasites. (3)
Malaria eradication:	Permanent reduction to zero of the worldwide incidence of infection caused by human malaria parasites as a result of deliberate activities. Interventions are no longer required once eradication has been achieved. (3)
Malaria incidence:	Number of newly diagnosed malaria cases during a defined period in a specified population. (3)
Malaria infection:	Presence of *Plasmodium* parasites in blood or tissues, confirmed by diagnostic testing. Note: Diagnostic testing could consist of microscopy, rapid diagnostic testing or nucleic acid-based amplification (e.g., polymerase chain reaction assays to detect parasite DNA or RNA). (3)
Malaria mortality rate:	Number of deaths from malaria per unit of the population during a defined period. (3)
Malaria prevalence (parasite prevalence):	Proportion of a specified population with malaria infection at one time. (3)
Malaria reintroduction:	Malaria reintroduction is the occurrence of introduced cases (cases of the first-generation local transmission that are epidemiologically linked to a confirmed imported case) in a country or area where the disease had previously been eliminated.(3)

Malaria risk stratification:	Classification of geographical areas or localities according to factors that determine receptivity and vulnerability to malaria transmission. (3)
Malaria stratification:	Classification of geographical areas or localities according to epidemiological, ecological, social and economic determinants for the purpose of guiding malaria interventions. (3)
Malaria-free:	Describes an area in which there is no continuing local mosquito borne malaria transmission and the risk for acquiring malaria is limited to infection from introduced cases. (3)
Malarious area:	Area in which transmission of malaria is occurring or has occurred during the preceding 3 years. (3)
Map scale:	The relationship (or ratio) between distance on a map and the corresponding distance on the ground. For example, a 1:100,000 scale map. (2)
Near Infrared:	The portion of the electromagnetic spectrum that includes wavelengths between 0.7 μm and 3 μm. (2)
Net insecticide-treated:	Mosquito net that repels, disables or kills mosquitoes that come into contact with the insecticide on the netting material. The two categories of insecticide-treated net are: • conventionally treated net: a mosquito net that has been treated by dipping it into a WHO-recommended insecticide. To ensure its continued insecticidal effect, the net should be re-treated periodically. • long-lasting insecticidal net: a factory-treated mosquito net made of netting material with insecticide incorporated within or bound around the fibres. The net must retain its effective biological activity for at least 20 WHO standard washes under laboratory conditions and 3 years of recommended use under field conditions. Note: Untreated mosquito nets can also provide substantial protection against mosquito bites. (3)
Parallel:	An imaginary circle drawn on the earth that passes through all points that have the same latitude. (2)

Plasmodium:	"Genus of protozoan blood parasites of vertebrates that includes the causal agents of malaria. *P. falciparum, P. malariae, P. ovale* and *P. vivax* cause malaria in humans. Human infection with the monkey malaria parasite *P. knowlesi* and very occasionally with other simian malaria species may occur in tropical forest areas."(3)
Population at risk:	Population living in a geographical area where locally acquired malaria cases have occurred in the past 3 years. (3)
Radar:	A system that emits timed pulses of microwave radiation and determines the direction and distance to an object by measuring the time required for the radiation to reflect back from the object.(2)
Rapid diagnostic test:	Immunochromatographic lateral flow device for rapid detection of malaria parasite antigens. (3)
Raster format:	Spatial representation of the entities through the arrangement of cells or pixels in the form of a numerical matrix of Digital Numbers (DN). (2)
Reintroduction risk:	The risk that endemic malaria will be re-established in a specific area, after its elimination. Note: The risk is typically determined by factors including climate, altitude, vector populations, human susceptibility, socio-economic status, urban or rural and coverage of interventions. (3)
Relapse:	Recurrence of asexual parasitaemia in *P. vivax* or *P. ovale* infections arising from hypnozoites. Note: Relapse occurs when the blood-stage infection has been eliminated but hypnozoites persist in the liver and mature to form hepatic schizonts. After an interval, generally from 3 weeks to 1 year, the hepatic schizonts rupture and liberate merozoites into the bloodstream. (3)
Resolution:	The ability of a sensor system to distinguish detailed information in an object. Examples are spatial, spectral, radiometric, and temporal resolution. (3)
Screening:	Identification of groups at risk that may require further intervention, such as diagnostic testing, treatment or preventive services. (3)

Spectral signature:	A set of wavelengths of electromagnetic radiation emitted or reflected by an object which may be used to identify the object or measure values related to the object. (3)
Spraying, residual:	Spraying the interior walls and ceilings of dwellings with a residual insecticide to kill or repel endophilic mosquito vectors of malaria. (3)
Supervised classification:	The selection of pixels that are representative of specific classes and that will be used by the image processing software as references for the classification of all other pixels in the image.(3)
Surveillance:	Continuous, systematic collection, analysis and interpretation of disease-specific data and use in planning, implementing and evaluating public health practice. Note: Surveillance can be done at different levels of the health care system (e.g. health facilities, the community), with different detection systems (e.g. case-based: active or passive) and sampling strategies (e.g. sentinel sites, surveys). (3)
Thematic maps :	A map that indicates the distribution and behavior of a particular phenomenon, for example, vegetation, geology, rainfall, population distribution, etc. (2)
Topographic maps:	A map that indicates topographical relief using contour lines that represent different values of altitude. (2)
Uncomplicated malaria:	Symptomatic malaria parasitaemia without signs of severity or evidence of vital organ dysfunction. Note: See current WHO definition (Guidelines for the treatment of malaria. Third edition). Malaria-associated disease can be defined more specifically by criteria for the degree of fever (e.g., temperature). (3)
Universal Transverse Mercator Network or UTM system:	A reference system uses a set of strips or meridian zones, defined in a Mercator transverse projection system. Each of these zones has 6º longitude, its origin point is the intersection of the central meridian with the equator. (2)
Unsupervised classification:	The classification of pixels with common characteristics is based on the software analysis of an image without the selection of representative pixels by the user. (2)

Vector competence:	For malaria, the ability of the mosquito to support completion of malaria parasite development after zygote formation and oocyst formation, development and release of sporozoites that migrate to salivary glands, allowing transmission of viable sporozoites when the infective female mosquito feeds again. Note: Human malarias are transmitted exclusively by competent species of *Anopheles* mosquitoes; various plasmodia are transmitted by competent species of mosquitoes of the genera *Aedes, Anopheles* and *Culex* and other haematophagous Diptera.(3)
Vector control:	Measures of any kind against malaria-transmitting mosquitoes, intended to limit their ability to transmit the disease. Note: Ideally, malaria vector control results in a reduction of malaria transmission rates, by reducing the vectorial capacity, to a point at which transmission is interrupted. (3)
Vectorial format:	A representation of a picture in computer graphics by using various geometric shapes such as points, curves, straight lines and polygons. (2)
Vector principal:	The species of *Anopheles* mainly responsible for transmitting malaria in any particular circumstance. Note: Principal vectors may overlap seasonally or alternate in importance. (3)
Vector susceptibility:	The degree to which a mosquito population is susceptible (i.e. not resistant) to insecticides" (3)
Vector:	In malaria, adult females of any mosquito species in which *Plasmodium* undergoes its sexual cycle (whereby the mosquito is the definitive host of the parasite) to the infective sporozoite stage (completion of extrinsic development), ready for transmission when a vertebrate host is bitten Note: Malaria vector species are usually implicated (incriminated) after field collection and dissection indicates that the salivary glands are infected with sporozoites; specific assays can be used to detect and identify circumsporozoite protein, especially where infection rates are low. (3)

Vectorial capacity:	The species of *Anopheles* mainly responsible for transmitting malaria in any particular circumstance. Note: Principal vectors may overlap seasonally or alternate in importance. (3)
Vigilance:	A function of the public health services for preventing reintroduction of malaria. Vigilance consists of close monitoring for any occurrence of malaria in receptive areas and application of the necessary measures to prevent re-establishment of transmission. (3)
WGS84:	Is defined and maintained by the United States National Geospatial-Intelligence Agency (NGA). It is consistent, to about 1cm, with the International Terrestrial Reference Frame (ITRF). It is a global datum, which means that coordinates change over time for objects which are fixed in the ground. (2)
WRS	The Worldwide Reference System (WRS) is a global system that catalogs Landsat data by Path and Row numbers. Landsat satellites 1, 2 and 3 followed WRS-1, and Landsat satellites 4,5,7, 8, and 9 follow WRS-2. (2)

Summary

Malaria remains a significant source of suffering and death. Annually, in Peru, many cases are reported, with 95% of these occurring in the Department of Loreto. Since 2018, the "Zero Malaria Plan" (ZMP) has implemented interventions to eliminate malaria from Peru by 2030. This program promotes a community-based model with three overlapping development phases: (1) Malaria control during the first three years, focusing on eliminating symptomatic infections, and using testing and treating to reduce 70% of the malaria burden. (2) Control during the middle ten years, targeting both asymptomatic and low-parasite-density infections to reduce 99% of the malaria burden. (3) interventions during the entire program to eliminate residual malaria transmission foci, including reintroductions.

There is a need for accurate and timely identification of high malaria transmission areas so that cost-effective malaria prevention, diagnosis, and treatment strategies can be implemented when and where they are needed.

This research is an original study that assesses the risk of co-endemic *Plasmodium vivax* and *Plasmodium falciparum* transmission in the Peruvian Amazon using boosted regression tree (BRT) models based on social and environmental predictors derived from satellite imagery and data to assess and predict high and very high malaria transmission in the department, at the village level. This research also generates technical proposals for malaria control program in the Department of Loreto.

Keywords: Malaria, risk mapping, satellite imagery, boosted regression trees.

Resumen

La malaria sigue siendo un importante problema que genera sufrimiento y muerte en la población. Anualmente, se reporta un gran número de casos en el país, de los cuales el 95% ocurren en el Departamento de Loreto. Desde el 2018, el "Plan Malaria Cero" ha implementado intervenciones para eliminar la malaria del Perú para el 2030. Este programa promueve un modelo basado en la comunidad con tres fases de desarrollo: (1) Control de la malaria durante los primeros tres años, centrándose en eliminando las infecciones sintomáticas y usando pruebas y tratamientos para reducir el 70% de la carga de malaria. (2) Control durante los diez años intermedios, centrándose en infecciones asintomáticas y de baja densidad de parásitos para reducir el 99 % de la carga de paludismo. (3) intervenciones durante todo el programa para eliminar los focos residuales de transmisión de la malaria, incluidas las reintroducciones.

Existe la necesidad de una identificación precisa y oportuna de las áreas de alta transmisión de la malaria para poder implementar estrategias rentables de prevención, diagnóstico y tratamiento de la malaria cuando y donde se necesiten.

Esta investigación es un estudio original sobre el riesgo de transmisión co-endémica de *Plasmodium vivax* y *Plasmodium falciparum* en la Amazonía peruana, utilizando modelos de árboles de regresión potenciados (BRT) basados en predictores sociales y ambientales, derivados de imágenes satelitales y colección de datos para evaluar y predecir niveles de alto y muy alto riesgo de transmisión de la malaria, a nivel de localidades para toda la región Loreto.

Palabras clave: Paludismo, mapeo de riesgos, imágenes satelitales, árboles de regresión potenciados.

Thesis outline

This thesis isdivided in five chapters:

Chapter 1, *Introduction* – This chapter provides the glossary, rationale, hypothesis and objectives of the thesis.

Chapter 2, *Literature review* - This chapter reviews the relevant literature on three important components: (1) malaria epidemiology and malaria status (2) disease mapping using remote sensing and geographic information systems, and (3) literature related to regression algorithms and boosted regression trees.

Chapter 3, *Methodology* – This chapter describes the study area, the process to achieve the predictors, the use of boosted regression tree models and the risk map elaboration.

Chapter 4, Results – This chapter shows the results of the analysis with BRT. The relative contribution of variables, the partial dependence plots, the model's performance. This chapter also provides population-based malaria risk maps and risk zones.

Chapter 5, *General conclusions, applications and perspectives* – This chapter presents the findings, and describes other applications and avenues for future research.

References, appendix and codes are provided at the end of this thesis.

Chapter 1: Introduction

This chapter provides the positioning of this research, rationale, hypothesis and objective of the thesis.

1.1. Positioning the research

Through the methodology and theoretical aspects, this study relies on several fields of science including, ecology-biogeography, epidemiology, and public health. As ecology-biogeography this research is focused on environmental conditions as essential determinants of distributions of species. As epidemiology, this research studies the distribution and determinants of health-disease states. As a public health investigation, this research generates a proposal for malaria control at the regional scope.

Traditionally, risk mapping is used to characterize the ecological/environmental risk of infection/exposure without active preventive measures. This study goes beyond eco-environmental factors and adds social variables such as population, including, travel time to major populated villages/towns, as an accessibility proxy. In addition, this research develops predictive maps using machine-learning algorithms, over the applied classic risk mapping for description/explanation purposes. To acquire the thesis data uses geospatial technologies that have opened new perspectives of analysis and management of public health programs, particularly in the case of diseases transmitted by vectors.

1.2. Rationale

Despite efforts and investment in control and eradication under Health Strategy for the Prevention and Control of Metaxenic Disease (SNEM-PEM 1957 - 1965), and the Pan-Andean Program for Malaria Control in the border areas (PAMAFRO 2005 - 2010), malaria remains a significant public health issue in Peru (5,6). The Department of Loreto, located in the Amazon rainforest, harbors almost 95% of Peruvian malaria. Historical malaria trends show that malaria increase and reduction are conditioned mainly by environmental factors and by interventions conducted by specialized malaria control programs. After the SNEM-PEM ended in 1965, the number of cases grew, reaching its highest peak in 1998 with 247,229 cases. This rise was linked to the "El Niño" phenomenon and increased resistance to chloroquine, the primary drug against malaria at that time. During the PAMAFRO program, malaria cases in Loreto were reduced by 80%, obtaining the lowest value in 2010 with only 11,504 cases and no cases reported in some districts. Unfortunately, the control measures implemented by the program were not continued by the Ministry of Health (6).

The Peruvian surveillance system is based on the number of reported malaria cases (conventionally confirmed by light microscopy) from a passive case detection strategy. However, asymptomatic or cases with minor symptoms are often not detected by light microscopy and might require more sensitive tests like polymerase chain reaction (PCR) (7). These samples must be sent to specialized laboratories; years before, these were sent from Loreto to Lima (the capital of the country). This conventional surveillance system becomes weak in remote areas.

Anopheles darlingi, the primary malaria vector in Loreto, identified in Iquitos in 1995, is an endophagic, exophagic vector and is highly anthropophilic (8,9). The malaria mosquito has a seasonal behavior, and its population variation is reflected in malaria case rates in Loreto and the country. Its longevity is highly sensitive to environmental conditions including temperature, precipitation, vegetation, deforestation, and natural and man-made water bodies. These variable conditions affect the availability of breeding sites, as well as parasite development itself (10–13). Rains in 2012 probably caused malaria resurgence in Loreto. That year doubled the cases compared to the previous year, and this increasing trend continued. In

2014, reported cases reached 61,000, the highest value in the last decade. In 2015, 61 856 cases in Loreto (14), represented approximately 15% of the total cases reported in the Americas. Of these Loreto cases, 49,745 were caused by *P. vivax* (15).

The main factors responsible for the variation in malaria transmission in the Amazon are environmental characteristics facilitating larvae breeding sites (temperature, precipitation, natural and human-made water bodies) and mosquito resting places (surrounding vegetation, deforestation) for the primary malaria vector of *P. vivax* and *P. falciparum*, *Anopheles darlingi* (10–13). Social factors also influence malaria transmission by increasing exposure (forest-based economic activities) (16,17), by contributing to delayed diagnosis and treatment (poor geographical accessibility to health facilities) (18–20), or by limiting effectiveness of malaria interventions (lower availability and utilization of control resources and preventive measures, and inappropriate treatment-seeking behaviors) (21–23).

In 2019 the Peruvian malaria burden declined, with financial support and intensified interventions under "Zero Malaria Plan" (execution period 2018 – 2030) (24). In the first quarter of 2020, the program saw a reduction of 75% of malaria cases in Loreto. In the last three years, malaria cases have kept it numbers less than in the previous period of the plan. In 2021 were 17,658 cases (80.25% *P. vivax*), in 2020 were 15,520 malaria cases (78.82% *P. vivax*), and in 2019 were 25,871 (74.32 % *P. vivax).*

In 2020-2021 the COVID-19 pandemic hit Peru, which then led the world ranking with 6258 COVID- deaths per million inhabitants. The COVID-19 pandemic exposed the precariousness of Peruvian health system. When the first COVID-19-cases were identified, the government rapidly imposed a lockdown and curfew. These measures interrupted the activities carried out by the malaria program. The pandemic response absorbed community and volunteer health workers. In the first months many malaria campaign activities were discontinued because of the lack of personal protective equipment (PPE). Active case detection, as well as the delivery of treated mosquito nets and malaria drugs were interrupted. People were afraid to visit health care facilities (25).

Another challenge to malaria control is the need to apply tools and interventions at different levels, since Loreto is a hypoendemic malaria area (26,27) and some remote areas have very

low entomological inoculation rates (EIRs) near those reported in Africa. This heterogeneity requires tailored interventions to control malaria in high-risk areas.

Reaching high-risk populations in malaria-eliminating settings becomes an increasingly complex task as parasite reservoirs are increasingly clustered in geographically restricted foci of transmission (6,28,29) and have an undisclosed dynamic. Indigenous populations and mobile dwellers are also often hard to reach. These groups frequently face substantial barriers to accessing health care, including service delivery, and are often missed by standard surveillance systems. These populations act as parasite reservoirs, causing transmission and becoming a threat to malaria elimination (30)(31).

The use of geographic information systems (GIS) (32) provides the ability to delimitate risk areas or find clusters. GIS is designed to store, retrieve, manipulate, analyze, and map geographical data under a reference system to analyze and solve complex problems (33–36). It covers many applications, e.g., engineering, computer science, land use planning, and environmental science. Its application to epidemiology and public health is rapidly increasing (37–47). GIS can access an enormous quantity of data, from documentation, tables, statistical reports, and satellite images, even when it may be difficult to find sufficient information from other sources (48,49).

GIS and remote sensing have become a powerful duo for acquiring data for decision-making based on spatial information. They deliver a large amount of information that helps to focus on high-risk populations in disease prevention and control programs. This complex data can be analyzed with the aid of machine learning (ML). Machine learning (ML) covers a broad range of processes that is difficult to define precisely (20). A machine learns when it changes its structure program or data in response to external information to improve its expected future performance. It is useful when some tasks cannot be well-defined, for example, when seeking relationships and correlations inside vast data sets (data mining) or when the amount of knowledge available about specific tasks might be too large for explicit encoding by humans (20,39).

This thesis, which is the first study to assess the risk of co-endemic *Plasmodium vivax* and *Plasmodium falciparum* transmission in the Peruvian Amazon, combines geographic information science (GIS) with satellite data and boosted regression tree (BRT) modeling to

enable malaria high-risk area detection in the Department of Loreto. This study provides new knowledge and operational tools to support malaria case detection strategies and demonstrates the ability to characterize areas and make predictions that are essential aspects of malaria control programs. Furthermore, these tools may be applicable to other diseases or fields where spatial prioritization or planning is required.

1.3. Objectives

General objective:

To assess, predict, and map the risk of co-endemic *P.vivax* and *P. falciparum* occurrence at the village level in Loreto, using boosted regression tree (BRT) models based on social and environmental predictors derived from satellite imagery for malaria control.

Specific objectives:

- To geocode villages in the Department of Loreto as points and aggregate the population and *P. vivax* and *P. falciparum* incidence by village and year.
- To use remote sensing and geographic information systems (GIS) to derive environmental and social predictors from satellite imagery and freely available data collections.
- To assess the malaria risk by identifying areas of likely high human-vector contact and human parasite carriage using boosted regression tree analysis.
- To predict yearly cross-validated BRT models to discriminate high-risk (annual parasite index API > 10 cases/1,000 people) and very-high-risk for malaria (API > 50 cases/1,000 people).
- To elaborate risk maps to identify areas of high malaria transmission.

1.4. Abstract

This thesis, which is the first study to assess the risk of co-endemic *Plasmodium vivax* and *Plasmodium falciparum* transmission in the Peruvian Amazon, combines geographic information science (GIS) with satellite data and boosted regression tree (BRT) modeling to enable malaria high-risk detection in the Department of Loreto Yearly cross-validated BRT models were created to discriminate high-risk (annual parasite index API > 10 cases/1000 people) and very-high-risk for malaria (API > 50 cases/1000 people) in 2766 georeferenced villages of Loreto department, between 2010–2017. The predictors identified were cumulative annual rainfall, forest coverage, annual forest loss, annual mean land surface temperature, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), shortest distance to rivers, time to populated villages, and population density. BRT models built with predictor data for a given year efficiently discriminated the malaria risk for that year in villages (area under the ROC curve (AUC) > 0.80), and most models also accurately predicted malaria risk in the following year. Cumulative rainfall, population density and time to populated villages were consistently the top three predictors for both P. *vivax and P. falciparum* incidence. Maps created using the BRT models characterize the spatial distribution of the malaria incidence in Loreto and can contribute to malaria-related decision making in the area (15).

Chapter 2: Literature review

This chapter reviews the relevant literature on three important components: (1) malaria epidemiology and malaria status (2) disease mapping using remote sensing and geographic information systems, and (3) literature related to regression algorithms and boosted regression trees.

2.1. Malaria epidemiology

The epidemiology of malaria is the study how health and disease processes affect the population. This includes knowledge about the groups in conditions of illness, the spatial and temporal distribution of the cases and its determinants of health (set of personal, social, economic and environmental factors that determine the health status of individuals or populations).

2.1.1. Malaria parasite

Malaria is a parasitic disease caused by a single-cell parasite of the *Plasmodium* family and transmitted by female *Anopheles* mosquitoes (51). Until the end of nineteenth century, malaria was supposed to be caused by noxious believed substances in the air coming from swamps or humid plains. This was the origin the word *malaria bad-air* (52).

Infective mosquitoes inject *Plasmodium* sporozoites from their salivary glands into the bloodstream during feeding, thereby starting a life cycle (Figure 1). Among all *Plasmodium* species known, five are reportedly able to infect human erythrocytes: *P. vivax*, *P. falciparum*, *P. malarie*, *P. ovale* and *P. knowlesi*. of these, *P. falciparum* and *P. vivax* are the most frequently reported in human infections. *P. vivax* and *P. ovale* are characterized by a hypnozoite stage, wich is a dormant form that can persist in the liver for months to years, causing periodic relapses of peripheral parasitemia and illness (53). *P. falciparum* is associated with the most severe forms of the malaria, and is endemic in

Loreto. For more information about the parasite life cycle and the relation with the vector, see Appendix B.

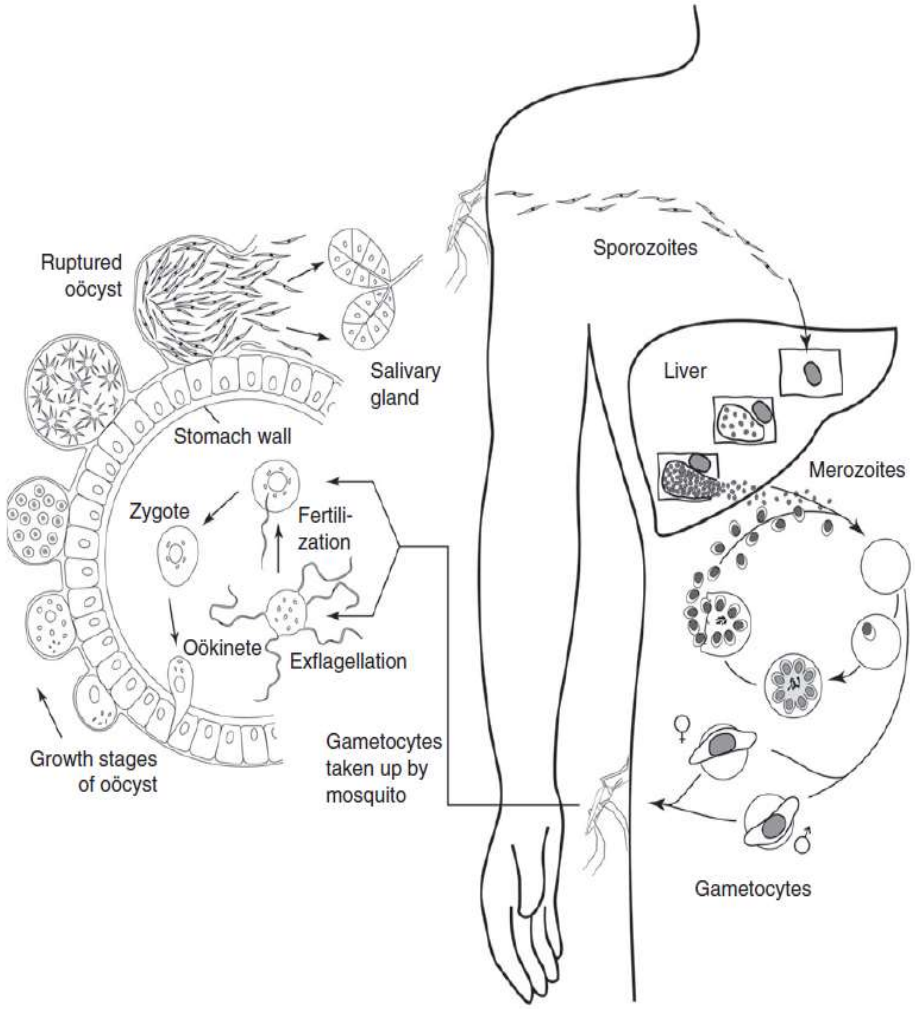


Figure 1: The life cycle of *Plasmodium* (52)

2.1.2. Malaria transmission

Human malaria is transmitted by biting female mosquitoes belonging to the genus *Anopheles*. There are about 400 different *Anopheles* mosquitoes throughout the world and approximately 70 of these are malaria vectors under natural conditions (54). The efficiency of various species as *Plasmodium* vectors varies across regions, so that a particular species of *Anopheles* may be an essential source of transmission in one area but not in another (53,55). Different *Anopheles* can display different behaviors in terms of breeding or larval habitat (e.g., fresh or brackish water; flowing streams, still pools, man-made habitats, shaded or sunny sites), feeding preferences (e.g., time of day for peak biting activity, preference for people over animals, feeding indoors or outside), and resting habits (resting indoors after feeding or leaving the house before resting) (9). These differences in mosquito behavior can affect the epidemiology of malaria and the choice of a malaria control strategy.

Eighteen *Anopheles* species are listed as vectors of malaria in Loreto (56,57). *Anopheles darlingi* is the principal vector, although other species, such as *An. benarrochi,* are sporadically involved in local transmission. *An. darlingi* is the most efficient malaria vector (58) and is known for erratic behavior depending on locality and environment (59). Available data on larval ecology indicates that *An. darlingi* is primarily a riverine species (8,60) and is also found in irrigation canals, rice fields, flooded cane fields, and pastures. Some studies also suggest a positive association between *An. darlingi* with shade and submerged vegetation (12). The largest reported flight range of *An. darlingi* in the Brazilian Amazon is 7.2 km (61), but most mosquitoes fly within 400-500 meters of their eggs hatched (62,63). The association with the deforestation is still debated; *Anopheles* was found associated with areas subject to deforestation in one study (11,13), while the opposite pattern was observed in another entomological study in the Brazilian Amazon (64). Primary and secondary vectors currently recognized in the Region of the Americas are listed in the Appendix B-2.

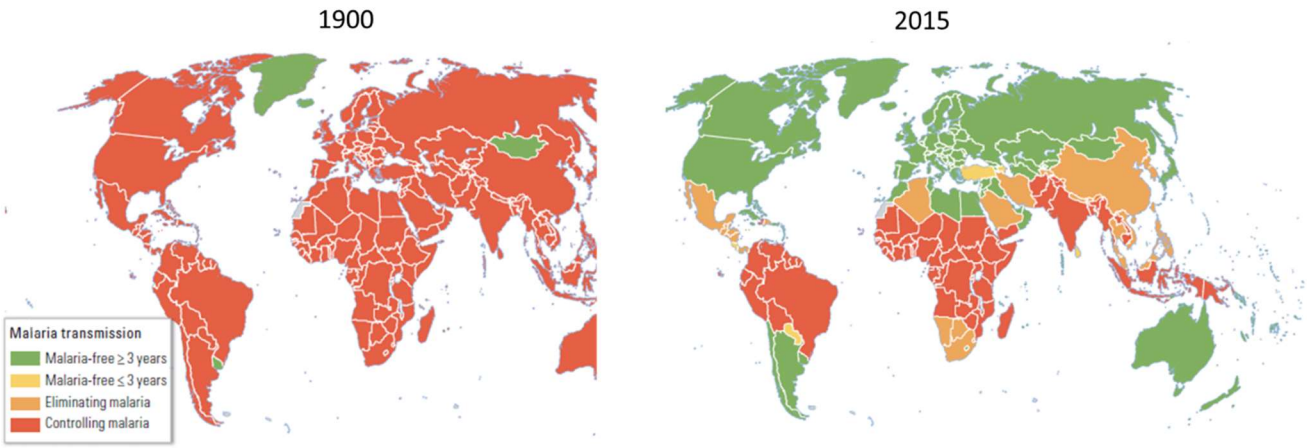
2.1.3. Malaria status in the world

Malaria is widely distributed in tropical and subtropical areas worldwide (Map 1). WHO malaria report in 2020 estimated that there were 241 million malaria cases and 627,000 malaria deaths worldwide. This represents an increase of 14 million more cases, and 69,000 more deaths compared to 2019. Approximately two-thirds of these additional deaths (47,000) were linked to disruptions in the provision of malaria prevention, diagnosis and treatment during the COVID-19 pandemic (65). Sub-Saharan Africa continues to carry the heaviest malaria burden, accounting for about 95% of all malaria cases and 96% of all deaths. About 80% of deaths in the region are among children under 5 years of age.

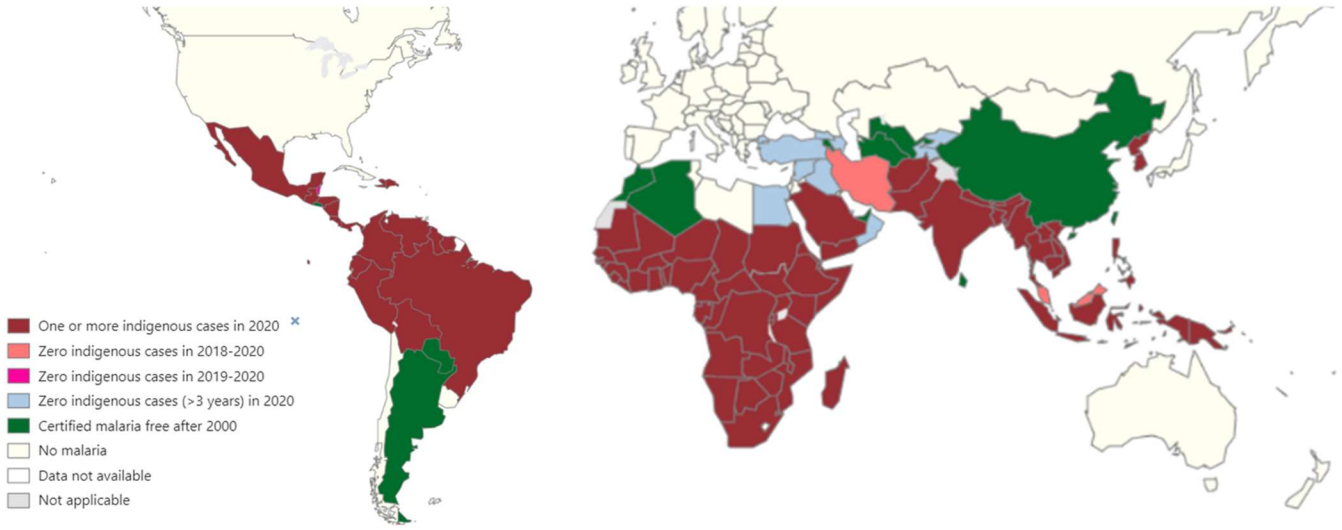
Since 1900, there has been a tremendous advance in reducing global malaria (Table 1 and Map 2) sparked by investment in malaria control and elimination carried out by worldwide malaria programs. Until the mid-nineteenth century, malaria was endemic in tropical and subtropical around the world (7). The availability of chloroquine for treatment and dichloro-diphenyl-trichloroethane (DDT) for vector control allowed malaria elimination in 29 countries by 1949. From 1955 to 1969, WHO conducted the Global Malaria Eradication Program (GMEP) to interrupt transmission in all endemic areas outside Africa (7,66). Between 1992 - 1998, several studies were made on the effectiveness of long-lasting insecticide-treated nets (LLINs), rapid diagnostic tests, and artemisinin-based combination therapies (ACTs) under the Global Malaria Control Strategy and the Roll Back Malaria initiative in 1998 (67). In 2015, WHO approved the *WHO's (2015a) Global Technical Strategy for Malaria 2016–2030* (31) and *Action and Investment to Defeat Malaria 2016–2030* with the goal of launching a new effort to eliminate malaria.

Table 1: Number of countries and territories where malaria was wrradicated, by WHO Region, 1900–2015 (7)

Indicator	Americas and Caribbean	South Asia and East Asia and Pacific	Europe and Central Asia	Middle East and North Africa	Sub-Saharan Africa	Total
Total number of countries	46	39	58	2	45	211
Malaria free						
1900	2	13	3	1	1	20
1900-49	0	0	9	0	0	9
1950-78	23	5	35	4	1	68
1979-90	0	1	2	2	1	6
1991-2015	2	1	9	6	0	18
Total number of malaria-free countries	27	20	58	13	3	121



Map 1: Malaria Transmission Worldwide, 1900, and 2015 (7)



Map 2: Countries with indigenous cases in 2000 and their status by 2020 (65)

The latest data on malaria for 2022 has not yet been published. The global trend from 2000 to 2021 in shown in Figure 2.

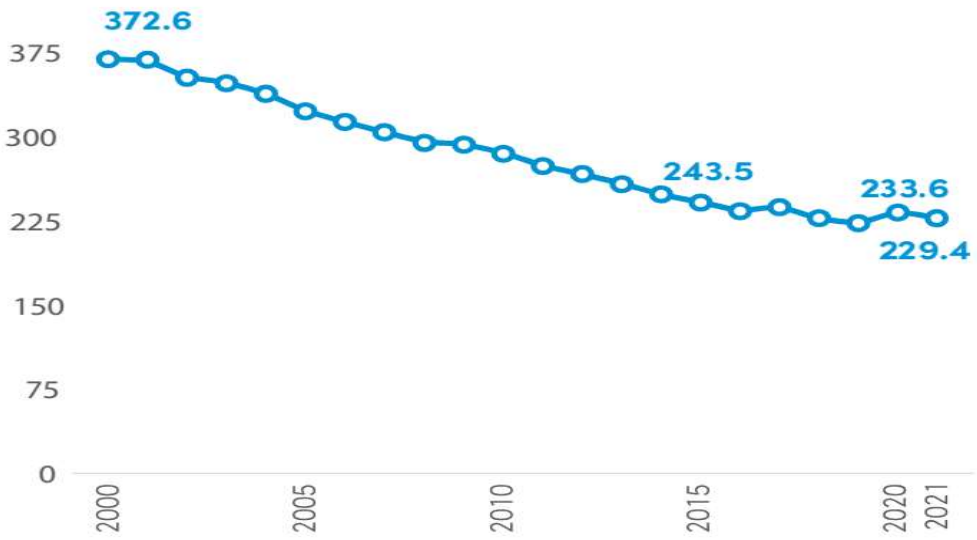
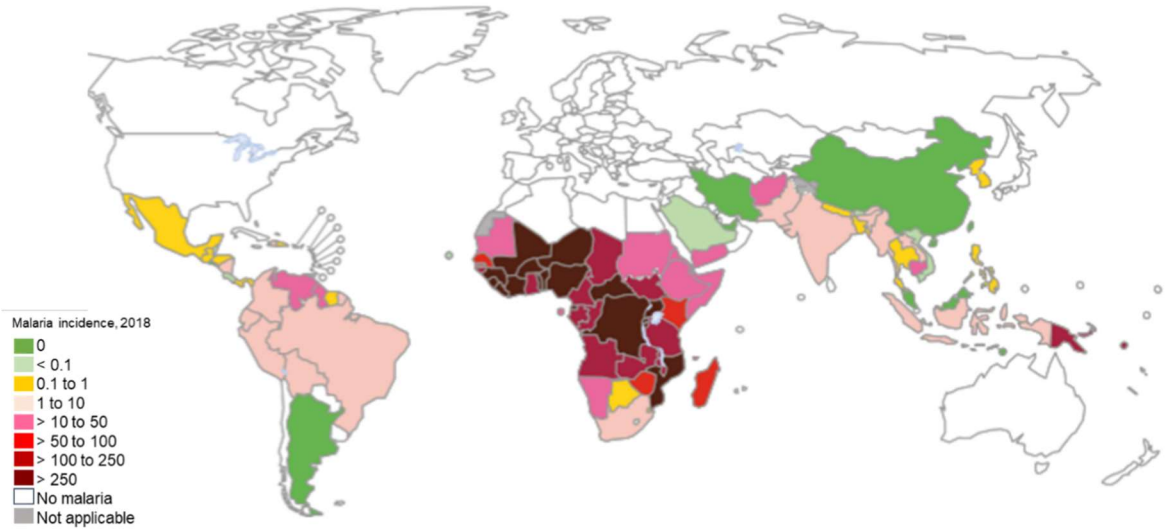


Figure 2: Global trend in malaria incidence (cases per 1,000 population at risk) (65).

(Map 3), compared to 231 million cases in 2017 and 251 million cases in 2010. Most cases occur in the Africa Region (213 million, 93%), Asia Region Southeast with (8 million, 3.4%) of cases and the Eastern Mediterranean Region with (5 million, 2.1%). Nineteen countries in

sub-Saharan Africa and India accounted for almost 85% of the global burden of malaria. More than half of all malaria cases worldwide were concentrated in six countries: Nigeria (25%), the Democratic Republic of the Congo (12%), Uganda (5%), and Ivory Coast, Mozambique, and Niger (4% each).



Map 3: Map of malaria case incidence rate (cases per 1,000 population at risk) by country, 2018 (7)

2.1.3.1. Malaria control program guidelines

According to "Global technical strategy for malaria 2016 – 2030" (WHO) there is no a unique strategy for malaria control. The document encourages countries to develop a set of strategies based on the intensity and stratification of transmission and on epidemiological, ecological and social characteristics (31). Further, the Global technical strategy for malaria 2016 – 2030 is based on three pillars:

(a) **Ensure universal access to malaria prevention, diagnosis and treatment**, through implies vector control, chemoprevention for pregnant women and children in areas of moderate and high transmission, and for tourists, diagnostic testing and treatment.

(b) **Accelerate efforts towards elimination and achivement of malaria-free status**, through focus programs, enacting legislation, renewed political commitment, deeper regional

collaboration, reducing the number of undetected infections, implementing targeted malaria vector control, preventing re-establishment of local malaria transmission, implementing transmission-blocking chemotherapy, detecting all infections to attain elimination and prevent re-establishment, use of medicines to reduce the parasite reservoir, devising *P. vivax*-specific strategies, and increased using surveillance.

(c) **Transform malaria surveillance into a core intervention**, that aims to detect all malaria infections, whether symptomatic or not, to investigate each individual case of infection, to differentiate imported cases from those acquired locally, and to ensure that each detected case is promptly treated in order to prevent secondary infections (31). Some of these strategies are being implemented in Peru under the Zero Malaria Plan. Malaria programs constitute an important element for the implementation of designed interventions, commonly programs, mobilize people, deliver equipment, strengthen services as health facilities and laboratories, train microscopists and community leaders.

2.1.3.2. Malaria status in the Americas region

Although Sub-Saharan Africa and South and Southeast Asia account the major malaria cases, this disease remain endemic in 19 countries of Latin America and Caribbean. Malaria causes high morbidity in rural areas of this region. Bolivia, Brazil, Colombia, Ecuador, Guyana, French Guiana, Peru, Suriname and Venezuela account for 93% of malaria cases in this area see Figure 3.

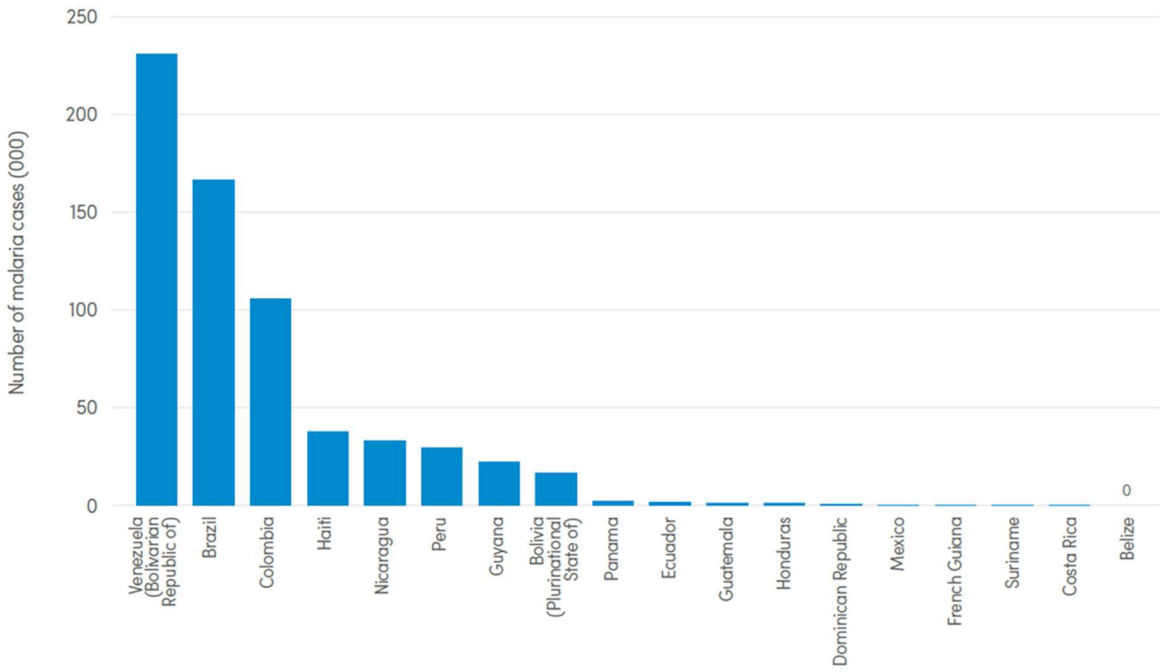


Figure 3: Malaria cases by country in the WHO Region of the Americas area (65)

After sustained work in malaria reduction from 2005 to 2014, the Americas Region experienced an increase in the total number of cases between 2015 and 2019 mainly in Venezuela and the increasing transmission in Brazil, Colombia, Guyana, Nicaragua, and Panama. Outbreaks have also occurred in (Costa Rica, Republic Dominican, and Ecuador) (68,69) . Since 2019 the number of cases and the incidence has decreased in this region. The total account between 2000 and 2020, malaria cases and case incidence reduced by 58% (from 1.5 million to 0.65 million) and 67% (from 14.1 to 4.6 cases per 1,000 population at risk), over the same period, malaria deaths and the mortality rate reduced by 56% (from 909 to 409) see in detail in Table 2 and the incidence trend in (Figure 4)

Table 2: Estimated malaria cases and deaths in the WHO Region of the Americas, 2000–2020
Estimated cases and deaths are shown with 95% upper and lower confidence intervals (65).

Year	Number of cases (000)				Number of deaths		
	Point	**Lower bound**	**Upper bound**	**% *P. vivax***	**Point**	**Lower bound**	**Upper bound**
2000	**1,540**	1,391	1,699	71.6%	**909**	665	1,169
2001	**1,297**	1,169	1,432	67.2%	**832**	597	1,092
2002	**1,183**	1,077	1,298	67.8%	**764**	513	1,022
2003	**1,159**	1,066	1,262	68.6%	**726**	480	984
2004	**1,147**	1,069	1,235	69.6%	**711**	462	985
2005	**1,273**	1,202	1,358	70.3%	**687**	439	960
2006	**1,097**	1,032	1,174	68.4%	**581**	346	843
2007	**989**	908	1,074	70.3%	**503**	293	738
2008	**696**	644	760	71.1%	**470**	224	747
2009	**688**	634	753	70.6%	**463**	230	737
2010	**818**	741	901	70.9%	**502**	247	793
2011	**615**	570	671	68.9%	**464**	205	727
2012	**585**	545	634	68.9%	**430**	211	652
2013	**576**	531	629	64.5%	**470**	232	709
2014	**475**	444	509	69.5%	**348**	193	485
2015	**602**	552	665	70.7%	**414**	227	579
2016	**688**	637	747	67.3%	**529**	264	749
2017	**946**	878	1,031	73.9%	**664**	290	958
2018	**929**	862	1,013	78.2%	**571**	271	815
2019	**894**	826	981	77.4%	**509**	231	738
2020	**653**	604	708	68.3%	**409**	185	579

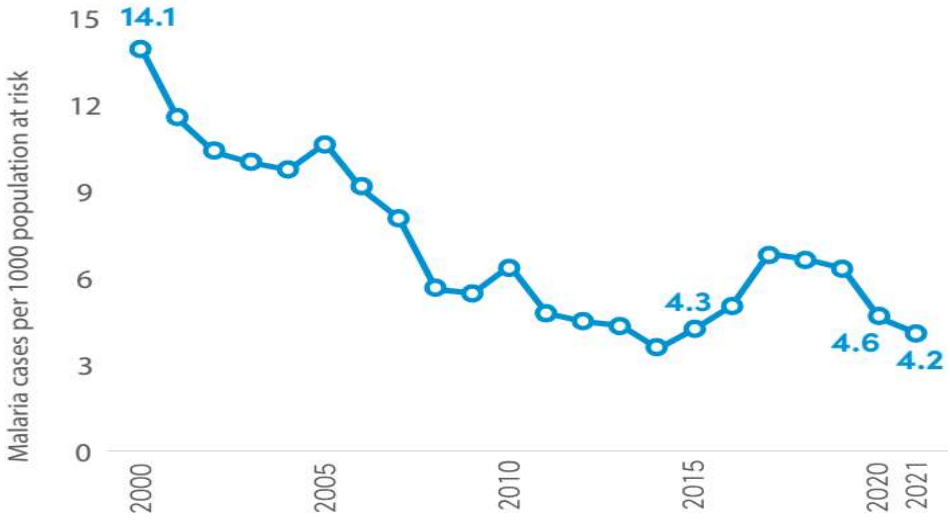
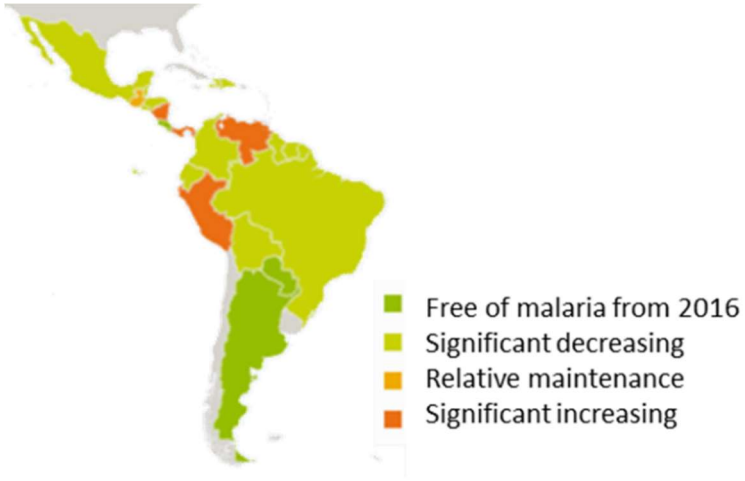


Figure 4: Americas trend in malaria case incidence (cases per 1,000 population at risk).

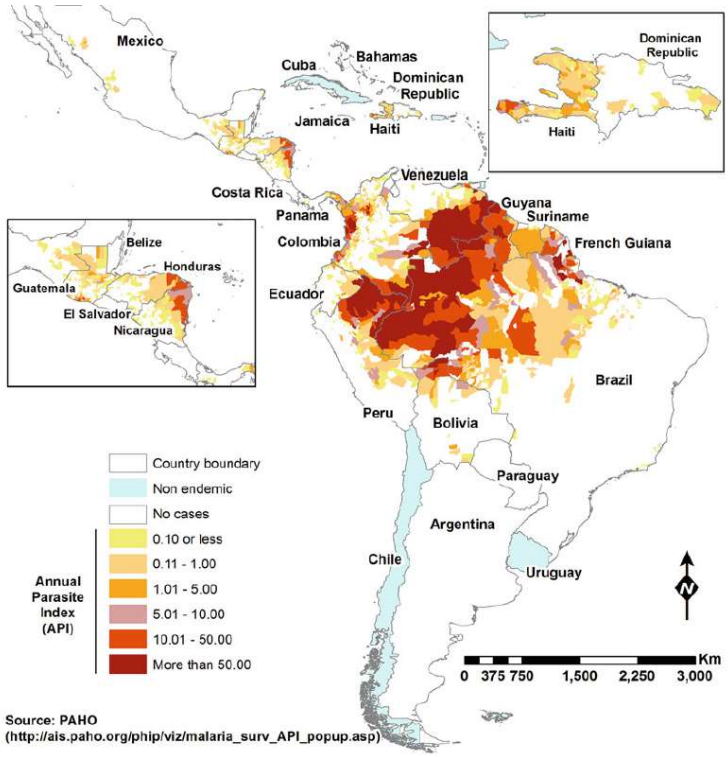
Malaria is endemic in all to malaria except Paraguay and Argentina, which were declared free of malaria in 2018 and 2019, respectively (Map 4).



Map 4: Countries in LA that are endemic to malaria, except Paraguay, and Argentina (7).

In LA Four nations contributed to more than ~80% of the 875,000 malaria cases reported (Venezuela 30%, Brazil 24%, Peru 19% and Colombia 10%) Annual malaria reports by country PAHO 2016 (Map 5).

Malaria annual parasite index (API) refers to high and moderate malaria transmission risk areas: The calculation is the number of confirmed new cases from malaria registered in a specific year, expressed per 1,000 individuals under surveillance, for a given country, territory, or geographic area.



Map 5: Malaria Annual Parasite Index in Latin America and the Caribbean, 2015 (134)

2.1.4. Malaria status in Peru

Peru is located on the west coast of South America between Ecuador, 0° and 10 ° South of latitude, and has an extension of 1,205,216 km2. Its neighbor countries are Ecuador and Colombia in the north, Brazil to the east, and Bolivia and Chile to the southeast. Geopolitically Peru is divided into 24 departments and the Constitutional Province of Callao. According to the most recent (2017) census, Peru has 31 million 237,385 inhabitants (INEI). Females constitute 50.8%and males 49.2%. Peru has great geographic and climatic diversity. Its territory includes dry coastal plains, high Andes mountains, and Amazon valley rainforest. The latter comprises 70% of the national territory (56).

The country has three malaria transmission zones that have environmental conditions suitable for malaria vectors: the north coast, mountainous Amazon, and the Amazon basin rainforest. Approximately 77% of malaria cases occur in the Amazon rainforest, 18% in the mountainous Amazon, and 5% in the north coast. Both *P. vivax* (80.6%) and *P. falciparum* (19.4 %) are found in the Amazon basin. The Department of Loreto, where this study was conducted, is the largest department in the Peruvian Amazon.

Malaria has been part of Peruvian life since at least the 1500s (70). In 1943, Peru launched an intervention directed toward mosquito larvae to reduce the malaria burden. In 1946, the Ministry of Health (MoH) implemented the first intra-home spraying with insecticide (dieldrin). In 1956, dichloro diphenyl trichloroethane (DDT) was implemented, as in other parts of the world. Between 1957 and 1965, the SNEM-PEM program initiated several interventions and studies which resulted in almost eliminating *P. falciparum.* (71). Between 1960 to 1970 the number of malaria cases was low, with only several dozens of cases reported annually. Subsequently however, antimalaria policies were neglected and investment in control programs decreased. Beginning in 1980 the curve of the number of cases steadily increased, reaching its highest peak in 1999.

The use of DDT was discontinued in 1988 and associated with the reintroduction and dispersion of *An. darlingi*, a very efficient anthropophilic vector, resulted in a 50-fold malaria increase in Loreto over the next decade. In the early 1990s, malaria incidence increased caused by *P. falciparum* in the Amazon, reached around 30% of the cases; linked to An. darlingi presence in the department (72).

In 1998, the number of confirmed cases brutally skyrocketed, from 33,705 in 1991 to 247,229 in 1999 (Figure 5). It was the highest peak in the last 77 years (15), this event was related to the strong (El Niño Southern Oscillation, ENSO) that, is a climatic pattern consisting of the oscillation of the meteorological parameters of the equatorial Pacific. It has two opposite phases that influence malaria cases, one of warming and rain in the eastern Pacific, known as the El Niño phenomenon, and the other phase of cooling, called La Niña. This warming and rain conditions affects the north coast malaria transmission zone creating thousands of puddles that became vector breeding sites (species: *Anopheles*, *Aedes*, and *Culex*), increased in *P. falciparum, P. vivax*, and other infectious diseases such as dengue fever, cholera, and diarrheal diseases.

In this period of malaria peak, the number of malaria cases in the north coast (Tumbes, Piura, Lambayeque, La Libertad and Ancash) were 84,451 versus in the Amazon (Loreto, San Martin, Ucayali y Madre de Dios) were 59,644 (73). These years also *P. falciparum* chloroquine resistance was another condition for the malaria increasing.

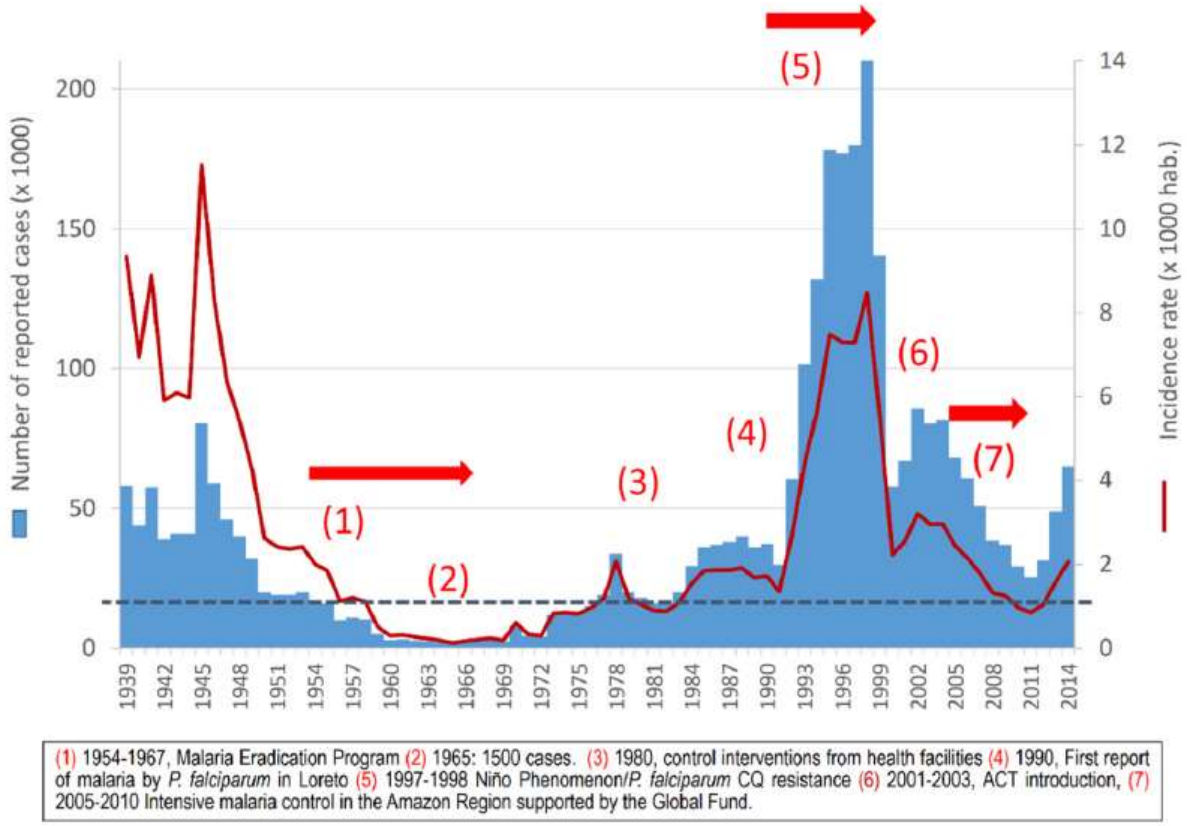


Figure 5: Historical trends of reported malaria incidence in Peru: 1939-2014 - MoH.

From 2005 to 2011, the comprehensive control activities supported by the Global Fund Malaria Project "PAMAFRO" with a community-centered approach, focused in border areas of Peru, Colombia, Ecuador, and Venezuela (see the map in appendix), steadily reduced the incidence of the disease in Loreto. In 2011, the end of the program, only 11,793 malaria cases were reported (6). After PAMAFRO the number of malaria cases rose again due the reduction of financial support allocated to malaria control and intervention activities, combined to important floods that occurred in 2012, in this year 31706 were reported, and in 2013; 43,737 cases. Currently, the country has implemented the Zero Malaria Plan, this program has the goal of eradicate the malaria burden in Peru, focusing on the Loreto department, the period proposed for this development goes from 2018 to 2030. After this intervention the number of malaria cases has been reduced. The Table 3, shows the malaria cases since 2010 and the response of the ZMP plan during the last three years.

Table 3: Peruvian malaria cases, *P. vivax, P. falciparum,* percentage *of P. vivax,* and deaths for years 2010 to 2021, CDC-MoH.

	Malaria cases				
Years	Total	*P. vivax*	*P. falciparum*	*% Pv*	Deaths
2010	29339	27036	2303	92.15	0
2011	23060	20421	2639	88.55	0
2012	31706	27702	4002	87.37	8
2013	48887	40949	7929	83.76	10
2014	64652	54361	10274	84.08	4
2015	63187	50524	12654	79.96	5
2016	56577	41256	15305	72.92	7
2017	54309	41328	12978	76.10	10
2018	44406	35199	9207	79.27	8
2019	25871	19227	4642	74.32	2
2020	15520	12388	3131	79.82	1
2021	17658	14170	3487	80.25	6

In Peru, microscopy is the main laboratory diagnostic method. In 2010 rapid diagnostic tests (RDTs) were introduced as a tool for malaria brigades. These are portable and simple to use and do not require laboratory infrastructure or complicated apparatus, facilitating their use by community agents in remote areas, and obtaining results in short time. RTD has good sensitivity and specificity in detecting the *Plasmodium* histidine rich protein 2 (HRP2) *for P. falciparum.* However for *Plasmodium* lactate dehydrogenase (pLDH) of *P. vivax* it depends on the product, i.e., OptiMAL has high sensitivity (74). ZMP use the SD Bioline

Malaria Ag *P.f/P.v* antigen test that has high values in both parameters, sensitivity and specificity. For *P. falciparum* 99.7% and for *P. vivax* 95.5% the specificity is 99.3%.

According to technical standards for malaria treatment are used in Peru:
For *P. vivax* uncomplicated: three days of Chloroquine (250 mg.) and seven days of Primaquine (7.5 mg.). For *P. falciparum* uncomplicated: Artesunate (AS) 250 mg for three days + Mefloquine (MQ) 205 mg for two days and primaquine (PQ) 15 mg on the first day. However, in case of resistance there are three elective schemes, depending on the level of response for the treatment of *P. falciparum*, based on (a) Sulfadoxine / Pyrimethamine + Artesunate; (b) 2. Mefloquine + Artesunate, and, (c) Quinine + Clindamycin + Primaquine. (75)

2.1.4.1. Malaria and coronavirus

In early 2020 the COVID-19 pandemic reached Peru. The pandemic, and the government's reaction to it, severely affected the provision of, and access to malaria health services. On March 15, 2020, the Peruvian government declared a national emergency due to the COVID-19 outbreak (Decree No. 044-2020-PCM). Restrictions placed on commercial activity. Only essential services, such as commercial establishments for food, pharmaceuticals establishments and health services remained active. Transport by air, sea, river, and land was suspended.

Malaria program strategies to control and eradicate disease include delivering insecticide-treated nets (ITNs), indoor residual spraying (IRS), early diagnosis, treatment, active case finding, and surveillance. The mechanisms to deliver these interventions depends on health workers, community agents and access to services. However, due to COVID-19 restrictions, Peru faced (1) reduction in the number of malaria healthcare workers, (2) overburden of health personnel by COVID-19 related activities, (3) closure of laboratories and the cessation of malaria diagnostic and research activities, (4) interruption of the activities carried out by community health agents and volunteers, and (5) reduction in active case finding due to the unavailability of proper personal protection equipment (PPE). Also, people at home were dominated by fear, especially for attending health facilities.

During the COVID-19 pandemic, malaria was often misdiagnosed, or coinfection overlooked, particularly when clinicians relied primarily on symptoms since malaria and COVID-19 can

have common symptoms i.e., fever, breathing difficulties, headache and tiredness (76). These conditions affected malaria data capture, and the number of cases fell. In 2019, 25,871 malaria cases were reported in the country. In 2020, 15,652 cases, a decrease of 35.65% compared to the previous year. In 2021, 17,658 cases were reported, a reduction of 27.40% compared to 2019 (Figure 6).

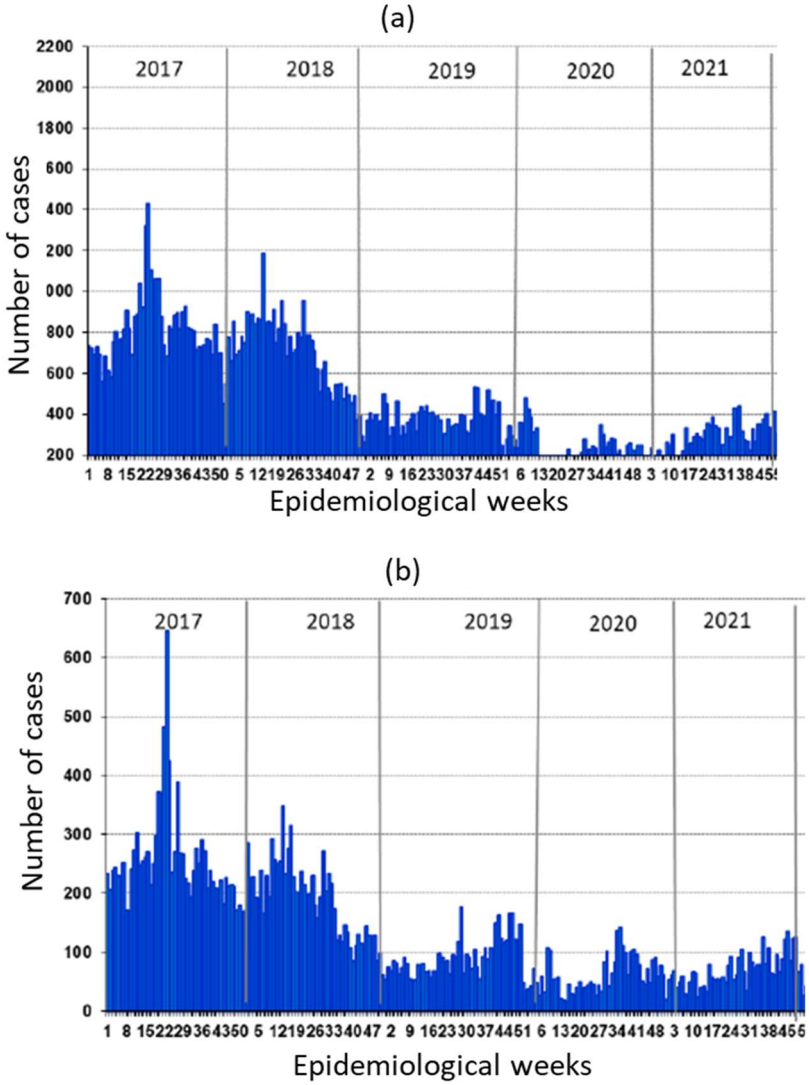


Figure 6: Number of malaria cases in Peru, by epidemiology weeks, from 2016 until 2021 (a) *P. vivax*, (b) *P. falciparum* CDC – MoH – Peru.

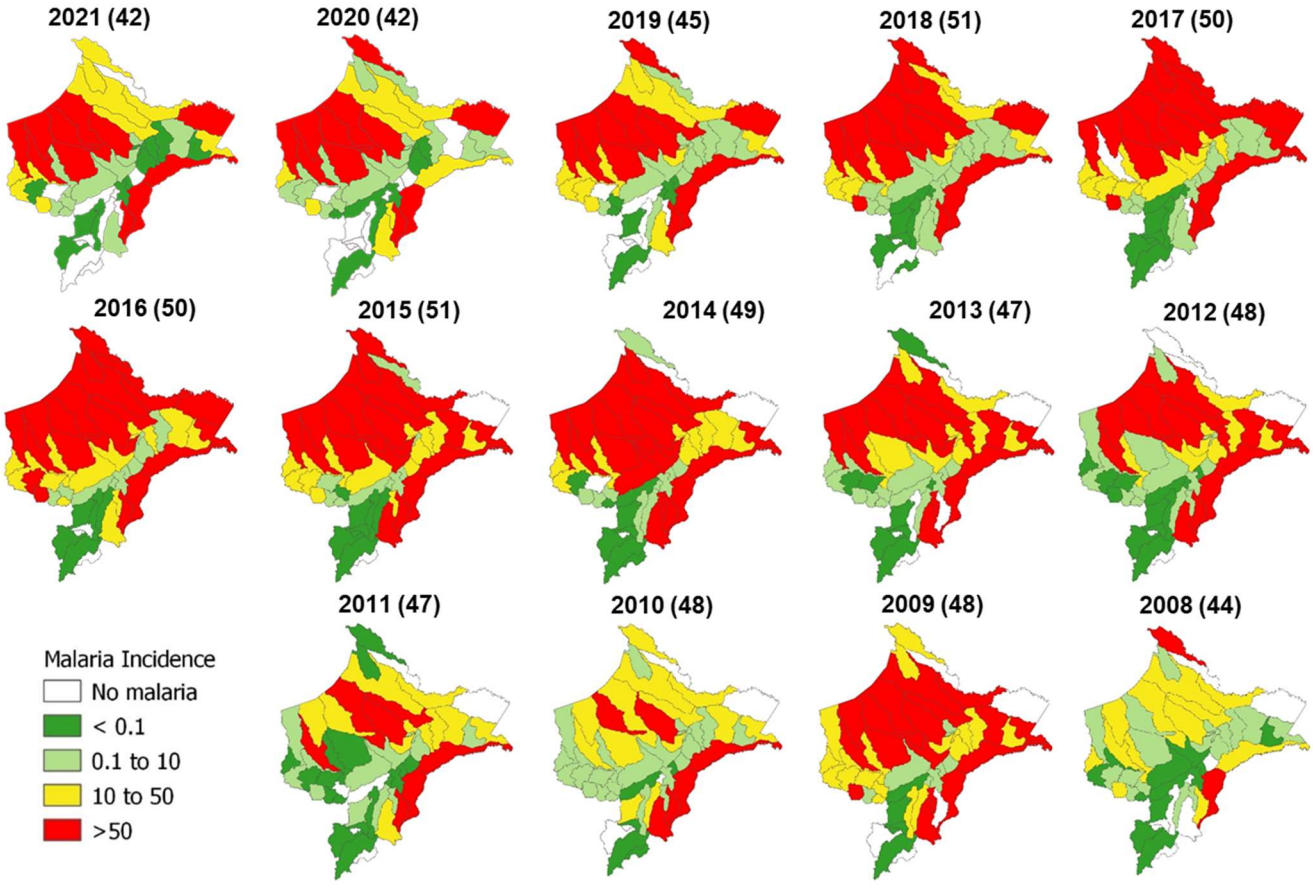
Historically, when malaria control has been neglected, the disease springs back. Therefore, it is a necessary to continue all malaria control activities and research projects, while maintaining necessary restrictions and procedures to address the COVID-19 epidemic.

2.1.4.2. Malaria status in Loreto

In the last three years, malaria cases decreased in Loreto. In (2019) 24,325 malaria cases were reported, 19,607 (80.6%) were by *P. vivax* and 4,716 (19.4%) of *P. falciparum* (*Pv/Pf* 4.15); the cases were concentrated in 47/53 districts (Map 6).

In 2020. 13,398 malaria cases were reported, 10,384 (77.5%) were by *P. vivax* and 3,014 (22.5%) of *P. falciparum* (*Pv/Pf* 4.45); the cases were concentrated in 42/53 districts (Map 6).

In 2021. 15,490 malaria cases were reported, 11,955 (77.2%) were by *P. vivax* and 3,534 (22.8%) of *P. falciparum* (*Pv/Pf* 3.38); the cases were concentrated in 42/53 districts (Map 6).



Map 6: Loreto malaria incidence by districts (53) and number de districts affected by year from 2008 to 2021. Data from National Center for Epidemiology, Disease Prevention and Control CDC - Peru.

In the last 14 years districts with mayor number of malaria cases were: San Juan Bautista, Andoas, Tigre, Pastaza, Napo, Trompeteros, Yavari, Ramon Castilla, Iquitos, Punchana, Mazan, Alto Nanay, Urarinas, Nauta, Balsapuerto, Torres Causana (Figure 7) and districts with high average API were Soplin, Alto nanay, Pastaza, Tigre, Andoas, Yaquerana, Trompeteros, Yavari, Torres causana, Napo, Alto tapiche, Mazan, Urarinas, Putumayo, Morona, were the districts with mayor number of malaria cases in Loreto.

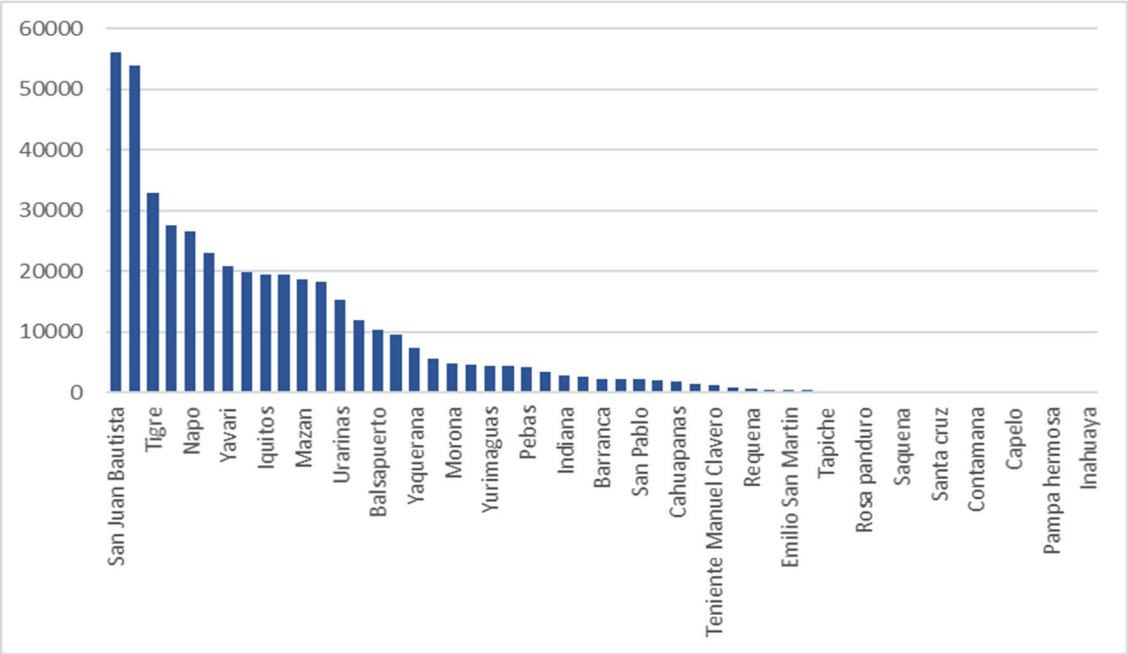


Figure 7: Loreto districts with malaria cases (2008 -2021). CDC - MoH - Peru.

Before Zero Malaria Plan intervention, districts that reported 80% of malaria cases were: Andoas (10 096 -19.06%), San Juan Bautista (6 728 - 12.70%); Tigre (6 113 - 11.54%), Napo (4,203 - 7.93%), Pastaza (3,630 -6.85%), Trompeteros (3,348 - 6.32%), Urarinas (2,621 - 4.95%), Iquitos (2,238 - 4.22%), Punchana (1,508 - 2.85%), Alto Nanay (1,455 - 2.75%), Torres Causana (1,431 - 2.70%).

After ZMP interventions, in 2019, the districts that reported 80% of malaria cases were: Andoas (5,723 - 26.05%), Trompeteros (1,977 - 8.99%), Tigre (1,940 - 8.83%), San Juan Bautista (1,580 - 7.19%), Pastaza (1,327 - 6.04%), Urarinas (842 - 3.83%), Yavari (817 - 3.72%), Iquitos (721 - 3.28%), Alto Nanay (716 - 3.26%), Balsapuerto (637 - 2.9%), Mazan (620 - 2.82%), Napo (596 - 2.71%).

2.1.4.3. Malaria control program

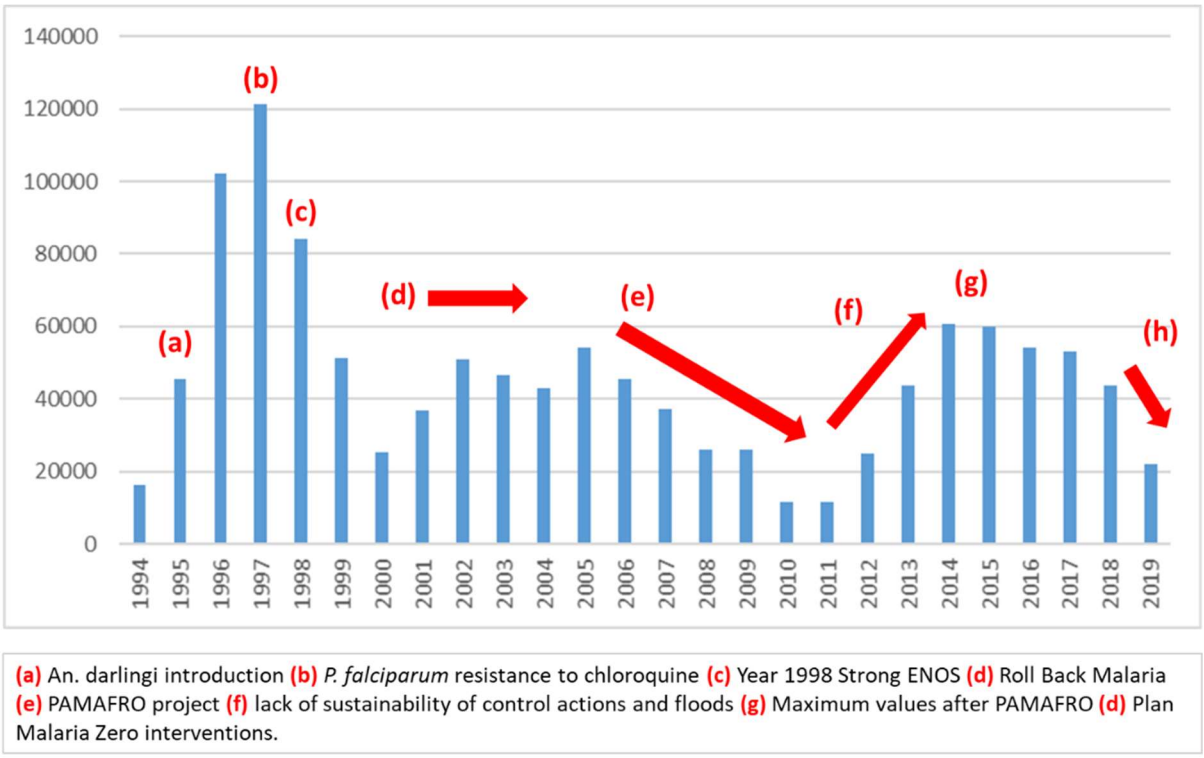
Recent activities against malaria in Peru are developing under the "Zero Malaria Plan" a taskforce between Peruvian Ministry of Health, Regional Health Direction (DIRESA - Loreto) and PAHO / WHO guidelines, is operative from 2017 (77) although intervention activities started by the end of 2018. This program with emphasis in Loreto, promotes a management model of malaria based on community health agents (CHA) with an integral, intercultural and health services approach with the objective of eliminating malaria by the end of 2030. The strategy is based on: (1) To ensure the quality of the diagnosis and treatment of malaria in health facilities at the community level, reducing the inequality of access and ensuring the multisectoral participation of the different levels of government in the control and elimination. (2) To reduce the risk of malaria transmission through the implementation of preventive interventions with community workers. (3) To promote healthy lifestyles with actions to eradicate malaria. (4) To optimize the malaria management through robust and responsive surveillance systems (24).

This program will be developed in three phases, (1) phase I, denominated malaria control, in which it is intended to reduce clinical malaria by at least 70% over the three-year period, and (2) phase II, towards the elimination of malaria, in which it is intended to drastically reduce asymptomatic and sub patent malaria, in which new strategies will be used (i.e. population treatment, with variants of mass treatment of the reservoirs of endemic populations such as FST) and new diagnostic methods, the goal is to reduce malaria in 99%, (3) phase III, of elimination of residual malaria, in which the strategies and methods tested in phase II will be used in scenarios of residual malaria.

In this brief period of interventions before the COVID-19, the program has succeeded in reducing cases to less than half in the country (Figure 8). In Loreto, the number of cases were reduced. However, Tumbes (department in the northwest part of Peru had an outbreak by *P. vivax* with 34 cases reported, of which 22 were indigenous and 12 imported from Venezuela.

This program addresses all learned by its predecessors' programs, the previous was PAMAFRO with a community approach (2005 - 2011) emerged as an initiative of the Andean Area Health Ministers meeting in Bolivia 2002, with the perspective of initiating the social integration of the Andean countries. The project was financed by Global Fund to fight AIDS,

tuberculosis and malaria was, for a total of $ 26 million in 5 years of execution. The scope of action of the PAMAFRO Project includes various areas, located at the borders from Colombia, Ecuador, Peru and Venezuela, which are characterized by sociocultural and ethnic diversity, by conditions of poverty and socio-environmental problems that facilitate the malaria transmission see PAMAFRO scope in Appendix B.



(a) An. darlingi introduction **(b)** *P. falciparum* resistance to chloroquine **(c)** Year 1998 Strong ENOS **(d)** Roll Back Malaria **(e)** PAMAFRO project **(f)** lack of sustainability of control actions and floods **(g)** Maximum values after PAMAFRO **(d)** Plan Malaria Zero interventions.

Figure 8: Loreto trend of reported malaria incidence (1994-2019). CDC - MoH.

2.1.4.4. Malaria case detection

The Peruvian health system has two sectors, the public and the private. The public sector is divided into the subsidized and the direct contributory. This system is segmented and fragmented. About 20% of the population does not have any health coverage. Social security covers only 25% of the people, especially the public and private employees; 5% have private health insurance, and 50% have state healthcare.

Loreto has a low supply of health professionals and services (Table 4), and the access to health services is low because of large distances and sparse populations (75). The provinces of Alto Amazonas, Ucayali, Maynas, and Requena are most connected to health services, while Ramon Castilla, Datem del Maranon and Loreto have less connection to supplies.

Table 4: Supply and health workers in Peru and Loreto, year 2014 and 2015. (78)

Health Supply Indicator		**Years**	**Peru**	**Loreto**	**ratio Supply to Demand *(Peru)**	**ratio Supply to Demand * (Loreto)**
Health workers	Doctors	2015	38065	533	1.219	0.543
	Nurses		39979	796	1.280	0.811
	Obstetricians		14445	425	0.462	0.433
	Dentists		5754	144	0.184	0.147
Health facilities	Hospital	2014	622	15	0.020	0.015
	Health Centers		2519	98	0.081	0.100
	Health post		7719	373	0.247	0.380

*Reference populations (adjusted census 2017 population censed + omitted people).
Loreto pop.: 981,897. Peruvian pop.: 31’237,385

The ministry of Health MoH, organizes the Peruvian malaria surveillance system. Detected malaria cases are reported in the NOTI web system. As we mentioned, the national surveillance system is based on the number of reported malaria cases (conventionally, confirmed by light microscopy) from a passive case detection strategy.

There are three ways in which malaria surveillance system detects and capture the information to the NOTI web.

1. Febrile patient, diagnostic captured for the community health agent (CHA). When a febrile patient is identified, the CHA takes a thick smear sample and a rapid diagnostic test (RDT). If

the RDT is positive, start treatment. If the test is negative, wait for the thick blood smear result, which has been previously sent to a health service with a laboratory, through the contact with the closest primary health worker.

2. Febrile patient in the health establishment without laboratory: the diagnostic is captured by the primary health worker (PHW). The PHW performs the rapid diagnostic test (RDT) and takes the thick blood smear. If the RDT is positive, the treatment is started.

3. Febrile patient in the health establishment with Laboratory. When a febrile patient comes to the health center, this person is first evaluated in the Triage environment. In this place, his vital signs are taken, his medical file is opened (or his existing medical file is searched if the person has one). If a thick blood test is positive, treatment starts.

In each scenario, the use of the epidemiological file by health workers is mandatory (Figure 9), in physical or digital format. This format captures the specific and complementary information of the confirmed malaria cases by thick blood smear. The responsibility of the fulfillment of this format is from the closest health service to the malaria case, also called for surveillance purposes the notifying service or notifying point.

In accordance with “Global technical strategy for malaria 2016 – 2030” (WHO), the Peruvian surveillance system is part of the intervention that means each detected case is promptly treated in order to prevent secondary infections.

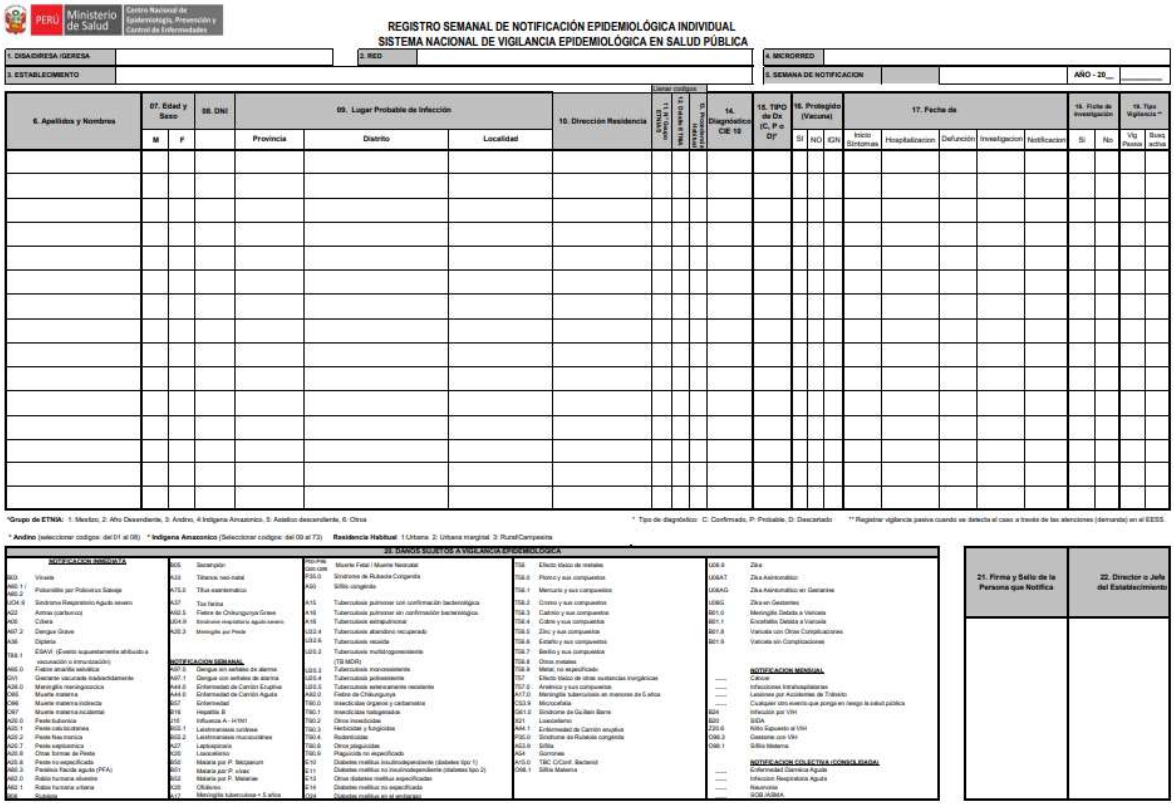


Figure 9: Weekly record of individual epidemiological notification. CDC – MoH.

The notification files are encoded into the NOTI web system in the data entry points in health services with internet and computer facilities, following a weekly calendar. These documents arrive to the entry points by:

- Physical shipment of the format to the data entry points
- Dictated by radio messages.
- Dictated by cell phone from areas with telephone coverage in case the notifying point does not have telephone service.
- Dictation via satellite phone ("Gilat")
- Sending messages or photos of the epidemiological notification by instant messaging service in places where it is possible to obtain an internet connection.

The data registered in the NOTI web system is administered by the CDC-MoH, NHI.

2.2. Geospatial tools

Geospatial tools are the set of elements and techniques that allow the analysis of reality from the knowledge of the geographical space provided for geospatial data. Geospatial systems express its results on various types of maps, showing the shape, size, distribution and other aspects of the phenomena analyzed. This thesis covers both important geo technics i.e., RS and GIS used in this research.

2.2.1. Remote Sensing

Remote sensing is the science and art of obtaining information about an object, area, or phenomenon by a device that is not in contact with the object (79) (RS) enables scientists to study the biotic and abiotic components of the earth surface (80). The relationship between the sensor and the land coverage is expressed as the emissivity of electromagnetic radiation. That is the ratio between the surface emissivity and the perfect black body emissivity at the same temperature. In contrast, that reflectivity is the ratio of the incident flow and the flow reflected by a surface (69). Depending on the portion of the electromagnetic spectrum, it will predominantly be reflection (near-infrared - 0.7 to 1.3 μm) or emission (thermal infrared - 8 to 14 μm). In the middle infrared - 1.3 to 8 μm the signal is mixed. These values are used to capture earth information or to interpret the information. (81).

The history of remote sensing goes back to the first aerial observation platforms: hot air balloons and the invention of photography in the XIX century. The year 1957 is often considered as the entry of remote sensing into the modern era, launched of Sputnik, the first artificial satellite set in orbit around the Earth by the Soviet Union.

Earth Observation (EO) is based on two modes: (1) Passive or optic sensors (which use the sun radiation), or active sensors, also called radar, that generates its energy (81,82). The first optical satellite explicitly dedicated to remote sensing was the American LANDSAT 1, operational between 1972 and early 1978. The last of the series, the LANDSAT 8 Landsat Data Continuity Mission (LDCM), was launched into space on February 11th, 2013. In the field of radar, one of the main milestones for civil applications was the launch of the European Remote Sensing Satellite 1 (ERS 1) in 1991. Currently, there are many sensors orbiting the Earth in various resolutions and revisit period (referring to the periodicity with which a sensor captures images

of the same area). Optical and radar sensors allow the best access to spatial information or earth observation (EO) and expand their applications' range of possibilities (83).

The elements of the data acquisition process are in Figure 10:

(a) Sources of energy,

(b) propagation of energy through the atmosphere,

(c) energy interactions with earth surface features,

(d) retransmission of energy through the atmosphere,

(e) airborne and/or spaceborne sensors,

(f) resulting in the generation of sensor data in pictorial and/or digital form,

(g) interpretation and analysis, involves examining the data using various viewing and interpretation devices. With the aid of the reference data, analyst extracts information about the type, extent, location, and condition of the various resources over sensor data were collected.

(h) this information is compiled, generally, in the form of hardcopy maps and tables or as computer files that can be merged with other "layers" of information in a geographic information system (GIS).

(i) Finally, the information is presented to users who apply it to their decision making process. (79)

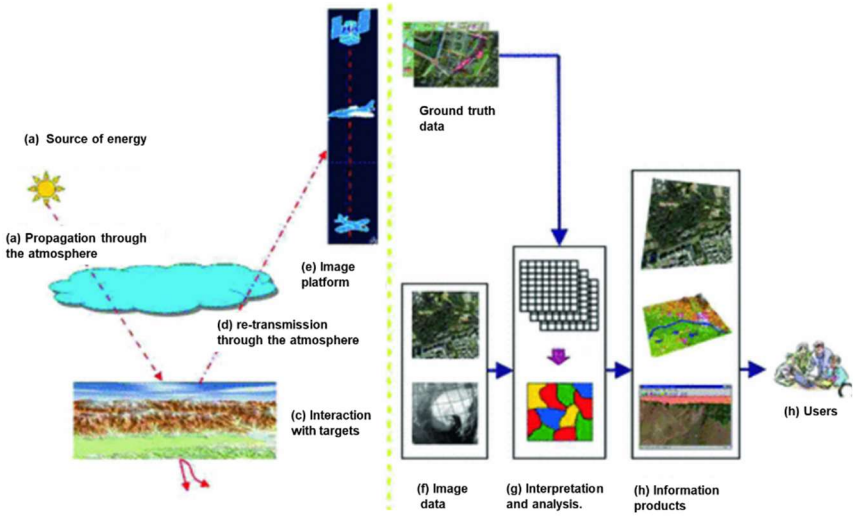


Figure 10: The flow of electromagnetic remote sensing of earth resources. (79)

2.2.1.2. Optic and radar sensors

Optical sensors: is a sensor that requires external power to operate. Cameras and multispectral scanners, for example, are passive sensors (81,82). The visible light is only a limited range of the electromagnetic spectrum, and most satellite equipment is also sensitive to other portions of the spectrum. For example, it is possible to discriminate surface materials based on electromagnetic emissions that the human eye cannot detect (Adams and Huyck, 2006, Sherbinin, 2002). Optical remote sensing has offered data for over four decades, with a few systems dominating anthropogenic land use and cover change LULC analyses due to the length of consistent datasets or the ease of availability (e.g., Landsat since 1972, the Landsat Thematic Mapper since 1983, Satellite Pour l'Observation de la Terre (SPOT) since the mid-1980s and the Moderate Resolution Imaging Spectroradiometer (MODIS) since 1999) (Jensen 2014).

Radar sensor: generates its electric signal and does not require a power source. Several countries secretly developed radar technology before and during World War II (Donnay, 2001, Uwe, 2010). The term RADAR was coined in 1940 by the United States Navy as an acronym for Radio detection and range. This object detection system emits radio waves to determine objects' range, altitude, direction, or speed. The signal emitted travels through the atmosphere, reflects on the surface of the Earth or against an object in its path, and returns to the sensor, which measures signal travel time, strength, and wavelength response. Synthetic Aperture Radar (SAR), for example, uses long-wave signal, which can penetrate clouds or bad weather conditions (84). Standard surveillance radars are typically found in airports or on ships. They make it possible to detect the presence of passive objects (called targets) employing echoes which they send back in response to the emission of an electromagnetic pulse (84) Another example of radar sensors, i.e. (Light Detection and Ranging LIDAR and Synthetic aperture sonar) uses a pulsed laser beam. Distance to object is determined by measuring the delay time between the pulse emission and its detection through the reflected signal.

Landsat and MODIS are the most widely used optical sensors in the Americas. In 2014, the National Aerospace Research and Development Commission CONIDA launched the first Peruvian optical satellite, Perusat-1, which can capture high multispectral image resolution, was developed by the French company Airbus for military and civil applications within the scope of Peruvian territory. Perusat-1 has a spatial resolution of 0.7 m. in panchromatic and 2.8

m. in multispectral mode, performance under four spectral bands (blue, red, green, and NIR) (85).The use of radar growing, especially in Argentina through the SAOCOM 1A sensor, under the National Space Activities Commission CONAE, is an advantage to managing the EO principally, where cloud coverage makes acquiring valuable optical data challenging. Another satellite used in the region is Sentinel, which offers radar images (Sentinel-1) and interesting middle resolution optical data (Sentinel-2), which are used according to the applications and investigations needs.

2.2.2. Geographic Information System

Several definitions have been proposed for geographical information systems (GIS) based on their functions, components, and uses. Antenucci *et al*. (1991) defined GIS as a "computer system that stores and links nongraphic attributes or geographically referenced data with graphic map features to allow a wide range of information processing and display operations as well as map production, analysis, and modeling." Parr (1991) defined GIS according to its components, which include: (a) data input and editing, (b) data management, (c) data query and retrieval, (d) analysis, modeling, and synthesis, and (e) data display and output functions. Cowen (1990) viewed GIS as integrating spatial data for decision-support systems.

Geographical Information System (GIS) developed in the early 1980s, a computational approach to managing geographic data. This technological revolution generated a notable impact on geography by allowing the automation of numerous procedures. GIS has become a cornerstone of geoinformatics and has gained prestige as computational tool to manipulate geographic data from numerous sources (86,87).

Remote Sensing and GIS has benefited from significant development in recent years. Its applications have been directly linked to technological development (2). The potential of GIS in medical research is widely acknowledged. GIS inspires hypotheses because of its capability to (i) overlay and integrate spatial information and (ii) handle and process large amounts of data, thus substantiating quantitative analyses in disease ecology and health care delivery. While this technology requires strict measures to ensure data quality, it also creates new possibilities in ecological associative analysis (Verhasselt 1993). The ability to manage, manipulate, and analyze large quantities of spatial data more quickly and with less effort than conventional methods also opens up new avenues of research in public health, including health care management and environmental risk assessment (28).

2.2.2.1. Raster and vector models

Vector models focus on the precise location of the geographical elements in space, and they assume that the phenomena are discrete, i.e., they are precisely geographically delimited. In these models, each geographical feature can be represented by points, lines or polygons. Each point is spatially located by a pair of x and y coordinates in a given map projection system. Each feature is associated with a row in a database containing its attributes. A unique identifier links both databases, the spatial coordinates, and the descriptive table (2) (Figure 11-a)

Raster models rely on two-dimensional matrices of cells (commonly referred to as pixels). Each raster layer is associated with a parameter of interest, such as altitude or mean annual temperature, and each pixel from the layer is given a corresponding value. The parameter's value can be derived from aerial pictures or electromagnetic emissions of different wavelengths. To limit the file size, the result is often recorded as 8-bit images, which offer 256 levels for each output parameter (Figure 11-b)

Figure 11: Representation of (a) vectorial model (b) raster model, from the same area of visualization and analyzed. Adapted from the book Encyclopedia of GIS.

Advantages and disadvantages using vectorial or raster models.

Vector model

Advantages

- Uses few data volumes.
- Are more accurate than a raster when calculating surfaces and distances.
- Allows for more precise limits as they are polygons, lines, and points allowing neighborhood relationships between elements.
- Are more used and shared because of their economic cost.
- Vector files are easier to develop topological rules and conditions compared to raster files, but on the contrary, they generate topological problems more easily (overlaps between elements of the same layer).

Disadvantages:

- The location of each vertex needs to be stored explicitly.
- Continuous data e.g., elevation, and temperature, are not effectively represented in vector form. Interpolation is required for these data layers.

Raster model

Advantages

- Uses vast files.
- The inherent nature of raster maps, e.g., one attribute maps, is more suited to mathematical modeling and analyses.
  calculations, algorithms, and quantitative processing can be simple (map algebra).
- Simulate better reality three-dimensionally.
- Discrete data, e.g., forest coverage, is accommodated equally well as continuous data, facilitating the integration of various data types.

Disadvantages:

- It is especially difficult to adequately represent linear features depending on the cell resolution.
- Output maps from raster models do not conform to high-quality cartographic representation

Both models were used in this thesis, the raster to work with the images, and the vectorial to make maps.

2.3. The art of mapping

A map is a two-dimensional graphic representation of the earth. It is a schematic drawing or layout that represents the characteristics of a given region, such as its dimensions, coordinates, geographical elements, or other relevant aspects. Maps may represent areas of different sizes. They can be local, regional, or global (88).

“Maps are essential across a wide swathe of science, from ecology and anthropology to sociology and climatology, and today’s researchers have a rich variety of inexpensive or free tools to choose from” (89).

There are a lot of software currently available of computer aided mapping e.g., QGIS, GRASS-SIG, ArcGIS, SAGA-SIG, Open Jump, Maptitud, Geoda, SIGEPI Google Earth, etc. Some of them are commercial and others Open Source (OS). This software allows users to integrate information with spatial data. Since John Snow's cholera map (1855), many researchers have used geo-tools to study about the diseases, in which the event has a spatial position and it can be georeferenced enabling to know about the relationship between the event and its environment. Mapping could also be used to draw a point where the people work, to draw a line to describe their weekly commutes, or use a choropleth map to represents rates or indices.

One field in which mapping is currently growing, is maps cloud service, e.g., Carto, Fusion tables, Map box, ArcGIS online and, Google Earth Engine GEE. GEE combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. With this tool it is possible access to huge collections of remote sensing data and use them freely.

Mapping can show the spatial distribution of health infrastructure such us, location of hospitals, health centers, health posts, laboratories. Mapping also can show areas of contamination by pesticides, radiation, or other risk areas.

During the coronavirus pandemic, we witnessed the great utility of the maps, they were made available in different repositories worldwide, and we were able to observe from global maps to local maps, accompanied by statistical information. In many cases, this generates map-based alerts to indicate if there is an infected person in our area. The related technology for mapping such as, remote sensing, geographic information system, GPS and database makes the mapping and geospatial analysis more powerful. Smartphones also played an important role collecting and showing data, using the integrated tool (GPS) or through apps developed by the government or private efforts.

2.3.1. Types of maps

There are many types of maps, cartographic, topographic, thematic, etcetera. According its representation theses could be:

Punctual: These maps helps to represent schools, health establishments, localities, industries and an others at specific points or coordinates (x,y) on the map. Its representation could be simple points or proportional according to its magnitude.

Linear: These types of maps are helpful to represent flows, connections between points (networks), and maps of isolines or equal magnitude, e.g. (the same height, rainfall, temperature, pressure, or rainfall).

Surface: These types of maps show information of areas in two ways, such as, choropleth and chorochromatic

a. Choropleth shows the same values by geopolitical division, e.g., incidence maps by the department, province of districts.
b. Chorochromatic. The color is used to define homogeneous areas, e.g., physical maps.

2.3.1. Disease mapping

Map makes visible the development and nature of any phenomenon regarding its geographical distribution (90). Disease maps, like other types of map, “convey factual information” and stimulate the formation of causal hypotheses (Howe 1989). One of the most famous examples of map-driven hypothesis was based on the 1855 'spot map' produced by John Snow, which represented the distribution of cholera cases around the Broad Street (London) water pump during September 1854-55 (Figure 12).

At that time, it was thought that cholera was spread by ‘bad air’ emitted from rotting organic matter. Snow was skeptical about this hypothesis, and he wondered whether sewage dumped into rivers and cesspools in the vicinity of water pump could contaminate the water supplies and, cause cholera outbreaks. He used local hospital and public records and asked residents if they had drunk water from the suspicious pump. Using this information, he created a spot map to illustrate the cluster of cases around the pump. On September 7th, 1854, Snow presented the map to local officials and convinced them to remove the handle from the pump. The number of cases quickly dropped, and the cholera outbreak eventually ended. For this study, Snow is now considered the father of modern epidemiology (91).

Epidemiological maps may reveal spatial variations and distribution patterns that remain unsuspected from the examination of statistical tables (90). They can support decision making regarding the allocation of health services, and for the evaluation and monitoring of interventions. Maps are also a key research tool in analytical epidemiology and medical geography. Table 5 shows potential links between different types of epidemiological maps and diseases.

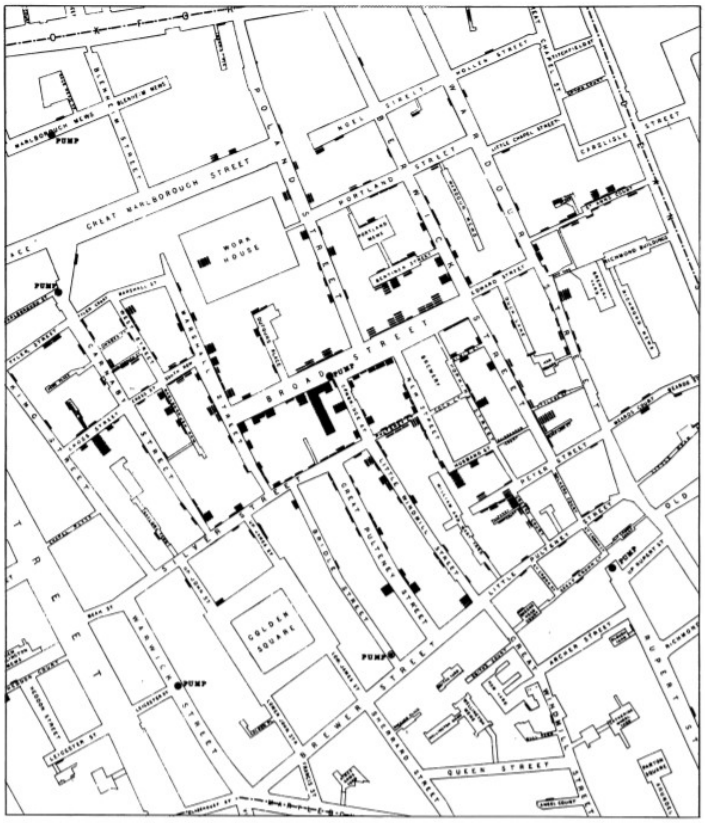


Figure 12: Snow's (1855) map of cholera cases and water pumps around Broad Street, London.

Types of maps in used in epidemiology and public health:

Maps of classification

Maps of buffers

Maps of location allocation

Maps of density

Maps of distribution

Maps of spatial-temporal data

Maps of spatial autocorrelation

Maps of risk

Table 5: Potential links between type of epidemiological maps and disease (92).

Factor	Disease	Mapping opportunity
Vegetation/crop type	Malaria	Breeding/resting/feeding habitats, crop pesticides vector resistance
	Chagas disease	Palm forest, dry and degraded woodland habitat for triatomines
	Leishmaniasis	Thick forests as vector/reservoir habitat in Americas
	Schistosomiasis	Agricultural association with snails, use of human fertilizer
	Trypanosomiasis	Glossina habitat (forests, around villages, depending on species)
	Yellow fiver	Reservoir (monkey) habitat
Deforestation	Yellow fever	Migration of infected human workers into forests where vectors exist
		Migration of disease reservoirs (monkeys) in search of new habitat
Forest patches	Yellow fever	Reservoir (monkey) habitat, migration routes
Flooding	Malaria	Mosquito habitat
	Schistosomiasis	Habitat creation for snails
Permanent water	Malaria	Breeding habitat for mosquitoes
	Onchocerciasis	Simulium larval habitat
	Schistosomiasis	Snail habitat
Flooded forest	Malaria	Mosquito habitat
Wetlands	Cholera	Vibrio cholerae associated with inland water
	Malaria	Mosquito habitat
	Schistosomiasis	Snail habitat
Soil moisture	Helminthiases	Worm habitat
	Malaria	Vactor breading habitat
	Schistosomiasis	Snail habitat
Canals	Malaria	Dry season mosquito-breeding habitat; ponding; leaking water
	Onchocerciasis	Simulium larval habitat
	Schistosomiasis	Snail habitat
Human settlements	Diseases	Source of infected humans; populations at risk for transmission in general
Ocean color	Cholera	Phytoplankton blooms, nutrients, sediments
	Red tides	
Sea surface temperature	Cholera	Plankton blooms (cold water upwelling in marine environment)
Sea surface height	Cholera	Inland movement of Vibrio-contaminated tidal water

2.4. Machine learning

Machine learning (ML) is area of artificial intelligence that uses computer programs to optimize a performance criterion using example data or past experience. ML uses statistics in building mathematical models, because the core task is making inference from a sample. Its models may be predictive, or descriptive, or both.

2.4.1. Boosted Regression Trees (BRT)

Boosted regression trees (or gradient boosting machine - GBM) is a semi-supervised machine learning ensemble model based on decision trees. Other similar ensembles are the random forest and bagging. These ensembles combine the outputs of several "weak building block models" (the decision trees), to obtaining a single powerful modeling engine.

Decision trees (DT) are conceptually simple procedures that can be used for regression and classification applications. In the context of regression, as this is the case for the current application, given a response variable Y, modeled by a function of many explanatory variables or predictors (X1, X2, ..., Xn), decision trees work by partitioning the regression predictor space (also known as the feature space) into distinct and non-overlapping regions via successive binary splits of the explanatory variables. It works on one predictor at each step and fits a constant value (a mean response value) for the (training) data observations in each of these regions. The splitting variables and split points are chosen to obtain the best fit to the data. This procedure continues repeated successively on the residuals left by the previous tree until some stopping rule is reached.

A Boosted Regression Tree model uses a boosting algorithm to sequentially generate and combine many successive trees, with each new tree being fitted to the residuals from the previous model, obtaining smaller residuals (and best fit to the data) at each stage. Thus, improving predictive performance as more trees are added to the ensemble (95,96). Given that each new tree explores predictor spaces left by previous trees, if enough trees are added, a BRT model has the flexibility to fit nonlinear functional forms of predictor variables and to automatically handle complex interactions between them (Elith, 2008). For example, compared to parametric regression or generalized linear models, BRT models have been increasingly used in an extensive range of studies, including in remote sensing of land cover/land use (LC/LU) classification and epidemiologic studies using ecological data (97,98). However, this

powerful modeling tool has some disadvantages that need to be properly addressed. The first is the risk of overfitting. Another difficulty with BRT is summarizing and communicating the results. Because BRT models, and other supervised machine learning models do not provide a formula like a regression equation and individual parameter estimations like the “betas” estimated in typical regression models, no significance or confidence intervals are associated with the individual predictors included in the model. The interpretability and confidence about the outputs or predictions obtained are compromised. To uncover the relationship between predictors and the response, some summaries of predictor importance and visualization tools need to be used such as partial dependence plots (PDP).

2.4.4. Partial Dependence Plots (PDPs)

Partial dependence plots help to visualize the functional form of the marginal effect on the response from a subset of predictor variables, usually one variable (less frequently two), keeping constant all other explanatory variables of the model. These marginal effects are generated by averaging the predicted values of the model at given fixed values in the range of a particular predictor, obtaining a straightforward function that links the prediction averages to that predictor, interactions with other explanatory variables included. The Partial dependence plots can show, for instance, whether the relationship between the response and a given predictor is linear, monotonic, or more complex. For instance, partial dependence plots display a linear (straight line) relationship if used in the context of a linear regression model.

It is useful to understand the PDPs in relation to the Individual conditional expectation plot (ICE plot). ICE plots display the estimated relationship between the predicted response and a predictor variable of interest for each unit of analysis in the data, resulting several lines, one per unit. Those lines represent all the heterogeneity of the predicted responses in the data for the range of values of the given predictor variable and the observed values of the other predictors. The PDP is just the averaging of the corresponding ICE curves across all those units see Figure 13.

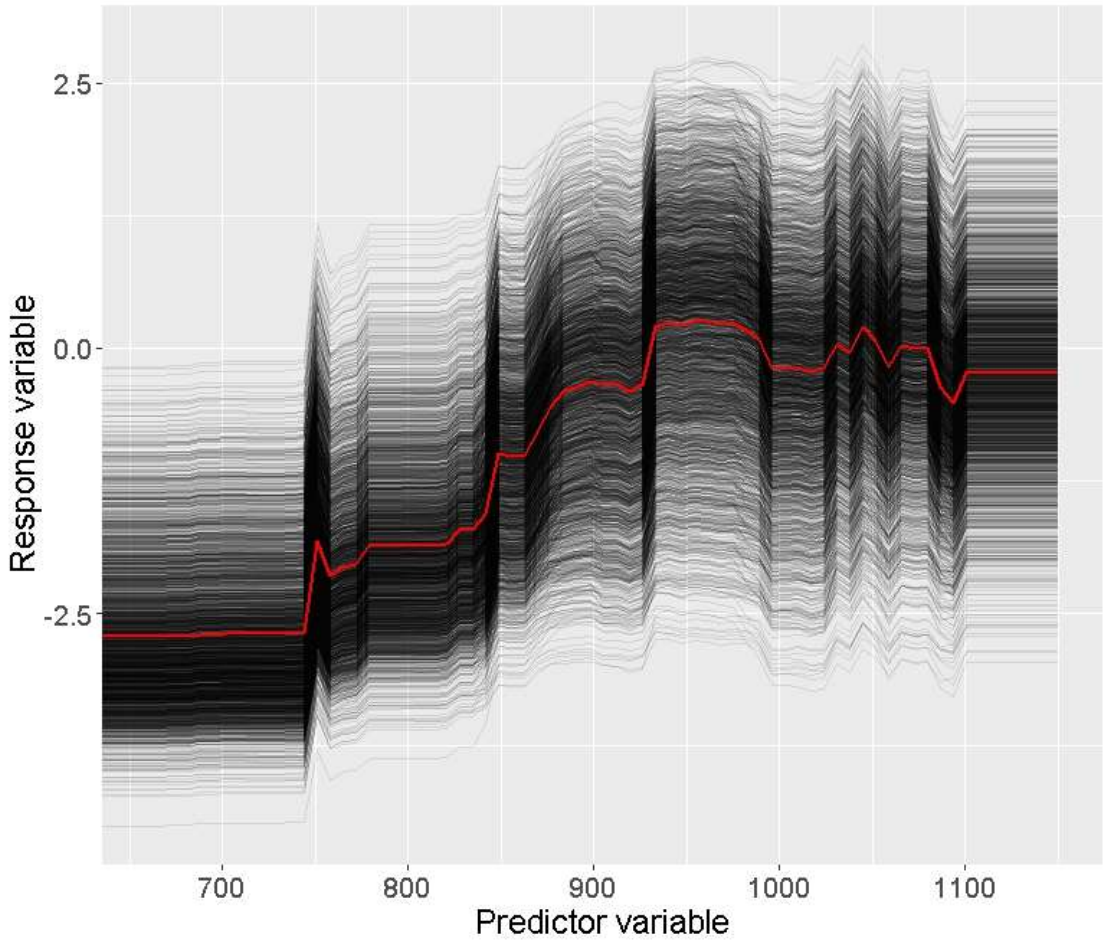


Figure 13: Partial dependence plot (red line) and ICE plot (black lines). PDP is the average of the predicted all the ICE lines for the range of values of the predictor variable on the x axis.

It is worth noting that PDP and ICE, although useful to visualize and possibly make causal interpretations about the black-box models, should not replace a randomized controlled experiment or a carefully designed observational study to establish causal relationships. When PDPs yield unexpected results, it is crucial to explore the data and look for the root of spurious associations such as unmeasured confounding factors (99).

2.4.5. Cross-validation

There are numerous procedures we can use to fine-tune model complexity. One of these methods is cross-validation (CV). We cannot calculate the bias and variance for a model, but we can calculate the total error.

CV is a procedure that employs rotating portions of the data to do repetitive training and testing of a model (Figure 14). It is frequently used in the context of supervised machine learning aimed at obtaining estimates of the predictive performance of the model. If there is a large validation dataset, CV is the best approach (1).

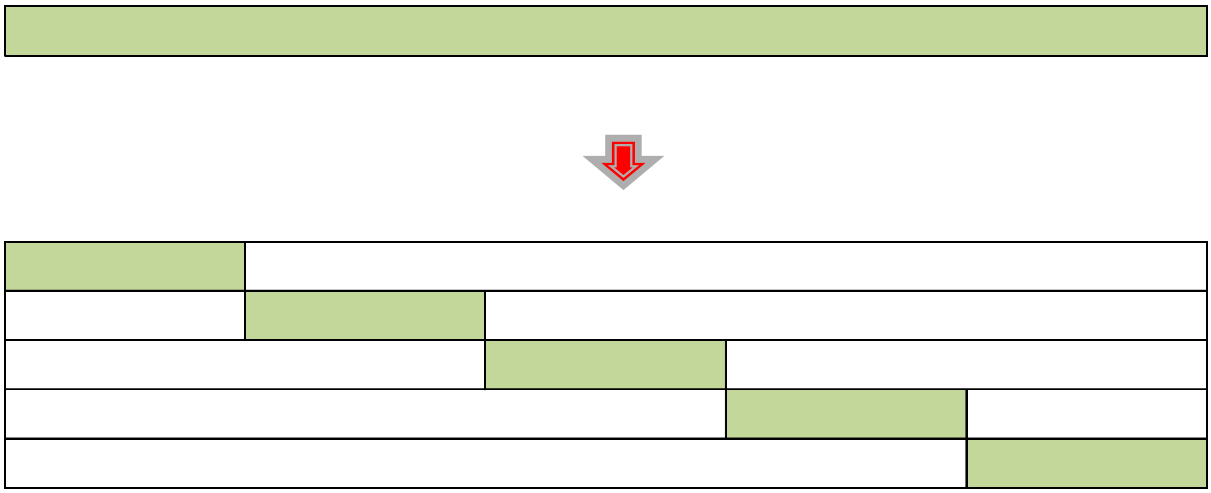


Figure 14: Example of the training and evaluation subsets generated in k-fold cross-validation. From a total of 25 elements. In the first iteration, we use the first 20 percent of data for evaluation and the remaining 80 percent for training.

Total instances: 25

Value of k: 5

No. Iteration	Training set observations	Testing set observations
1	[5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24]	[1 2 3 4 5]
2	[1 2 3 4 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25]	[6 7 8 9 10]
3	[1 2 3 4 5 6 7 8 9 15 16 17 18 19 20 21 22 23 24 25]	[11 12 13 14 15]
4	[1 2 3 4 5 6 7 8 9 10 11 12 13 14 20 21 22 23 24 25]	[16 17 18 19 20]
5	[1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]	[21 22 23 24 25]

Chapter 3: Methodology

This chapter describes the study area, the process to achieve the predictors, the use of boosted regression tree models and the risk map elaboration.

This ecological study developed in the Loreto department has the objective of assess, predict, and map the risk of co-endemic *P. vivax* and *P. falciparum* occurrence at the village level in the Peruvian Amazon, using boosted regression tree (BRT) models based on social and environmental predictors derived from satellite imagery.

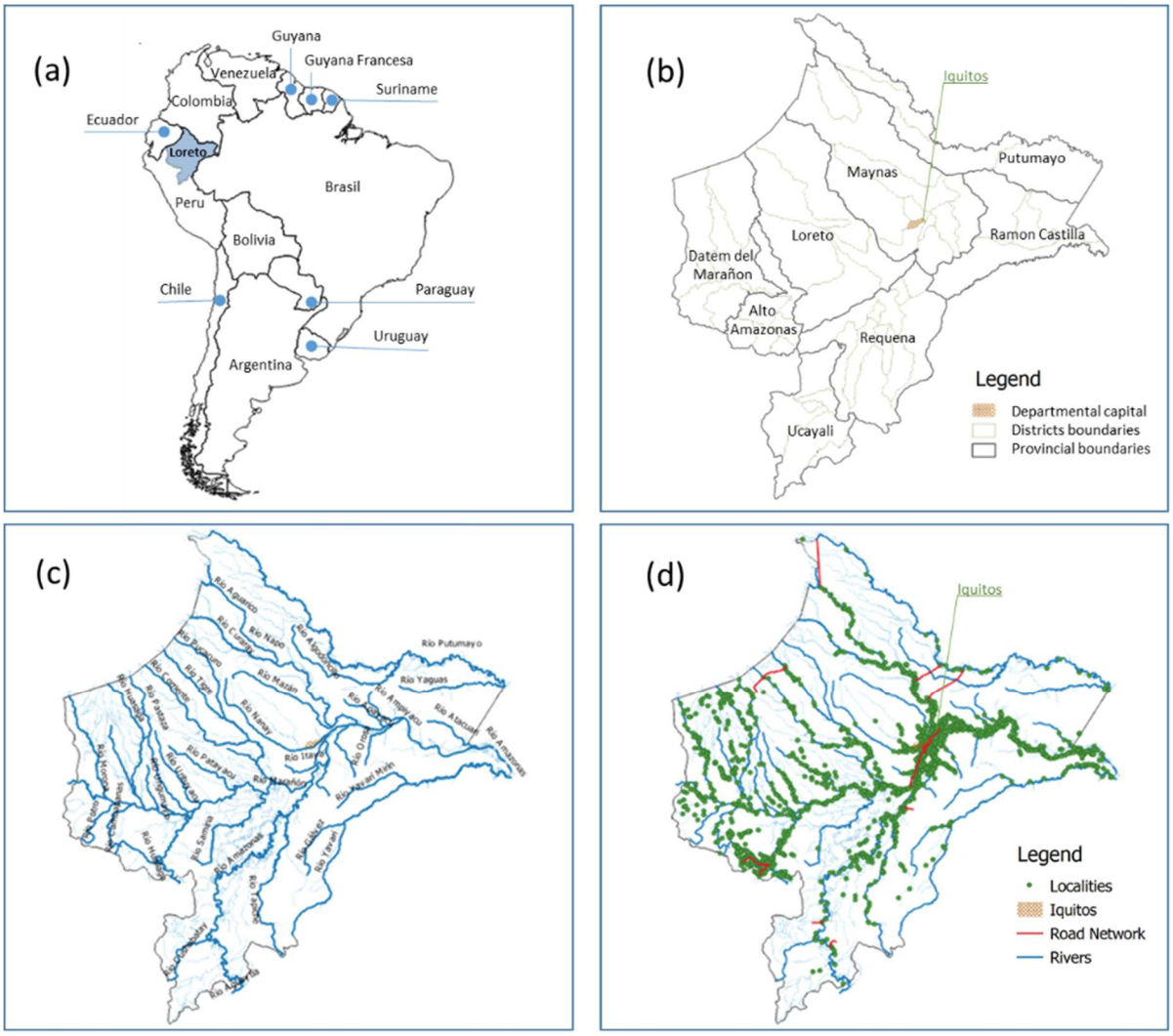
BRT model was used for both explanation and predict malaria risk. BRT is well known in the field of ecology. Some potentialities are that (1) it makes easier to identify which variables has a better performance into the model, (2) allows flexible modeling of the "functional form" between the explanatory variables and the result. Also, BRT allows modeling the interactions between explanatory variables.

3.1. Study area

Loreto, located in the northeast part of Peru between 61 to 220 meters above sea level, is the largest department in Peru (368,851.95 km^2, 28.7% of the national territory). Contradictorily, it has a low population, a total of 981,897 inhabitants lives in the department, according with the last Peruvian census (2017). The major part is concentrated in Iquitos, the capital. Approximately 40% of them aged below 15 years, and ~one third residing in rural areas (100). Geopolitically is divided in 8 provinces (Datem del Maranon, Alto Amazonas, Ucayali, Requena, Loreto, Maynas, Mariscal Ramon Castilla and Putumayo) (14). Loreto has an extensive fluvial network connecting to the Amazon River (Map 7) boats are the primary mode of transport. There is one paved road (Iquitos - Nauta 93Km.) and some other path roads (Saramiriza - Doce de Octubre 126 Km.) and (Yurimaguas - Jeberos 8.7Km.).

Loreto has a tropical climate, with a rainy season from November to May and a dry season from June to October. Precipitation is present throughout the year (cumulative average: 2,500 mm) with a peak in March (360 mm) and a minimum in July (50-100 mm). Malaria in Loreto

has a seasonal pattern, and higher transmission period is usually from February to July. Rural villages depend mainly on natural resource exploitation, while commercial activities are more important in cities.



Map 7: Study area: (a) geographical location of Loreto in South America; (b) administrative division of Loreto: department, provinces, and districts; (c) hydrographic map of Loreto; (d) road network, rivers, and georeferenced villages.

3.2. Malaria incidence and population

All cases confirmed by light microscopy (LM) that were reported by health facilities to the conventional surveillance system between 2010 and 2017 (101) were aggregated by village and year. Yearly overall and species-specific indices (API) were calculated for each village, using the formula: (number of confirmed cases/village population size) x 1,000. Using Peruvian Ministry of Health (MoH) thresholds, villages were classified as being or not being at high-risk (i.e., API>10 cases/1,000 people) or very-high-risk (API>50 cases/1,000 people). The population villages sizes, obtained from the exceptional National Census of 2012-2013 (102), was assumed to be uniform during the study period.

3.3. Georeferencing process

Malaria cases by species were aggregated from January 2010 to December 2017 for each village. Georeferenced data were verified by visual inspection of randomly selected village coordinates over populated areas identified in satellite images. The geographical location of villages that reported at least one malaria case during the study period was facilitated by the Regional Government of Loreto GOREL as standard village codes (ten-digit) in the malaria surveillance system and village georeferencing databases. In case village codes were not found in the georeferenced database or those had only village names, its locations were included such as matching villages and district names with the support of local health workers

Of the 2,843 villages included in the official Loreto cartography, we validated and georeferenced 2,766 village points in QGIS (103,104). Non-validation was primarily due to duplicated names and/or duplicated coordinates.

3.4. Download of predictors

Social and environmental variables previously associated with malaria transmission (105–107) were tested as factors for high-risk (IPA>10) or very-high-risk (IPA>50) (Table 6). These variables were derived from different satellite imagery and mixed products. JavaScript codes for Google Earth Engine (GEE)(108) were used to download and process this images at datum WGS84 zone 18 (EPSG 4326).

Table 6: Predictor and outcome variables used in BRT models.

Variable description				**Source information**				
Variable	**Descriptions**	**Units**	**Time-dependent variable**	**Data collection**	**Source**	**Spatial resolution**	**Temporal resolution**	**Units**
FC	Forest cover in a 2-km side square grid around village	%	yes (year)	UMD/hansen/global_forest_change_2015	year	30m	year	% per output grid cell
FL	Annual forest loss in a 2-km side square grid around village	%	yes (year)	UMD/hansen/global_forest_change_2015	Year	30m	Year	% per output grid cell
CAR	Cumulative annual rainfall. (average in a 2-km side square grid around village)	mm	yes (year)	TRMM/3B42	3hrs	~27km	3hrs	mm/3hrs x pp
LST	Land surface annual mean temperature	ºC	yes (year)	MODIS/006/MOD11A1/LST_Day_1km	1 day	1km	1 day	ºC
NDVI	Normalized difference vegetation index. (average in a 2-km side square grid around village)	index	yes (year)	MODIS/006/MOD13Q1	8 days	250m	8 days	index
NDWI	Normalized difference water index. (average in a 2-km side square gride around village)	index	yes (year)	LANDSAT/LC05C01/T1_T1 LANDSAT/LC08/C01/T1_T1 LANDSAT/LE07/C01/T1_T1	16 days	30m	16 days	index
SDR	Euclidean shortest distance to rivers	kilometers	no	JRC/GSW1_0/GlobalSurfaceWater (occurrence)	JRC/GSW Historical data	30m	once	kilometers
TPV	Travel time to major populated villages/towns	minutes	no	Oxford/MAP/accessibility_to_cities_2015_v1_0	Oxford (MAP), Google, (JRC) & Univ. Twente	1Km	once	minutes
POPD	Population in a 5-km side square grid around village)	log (number people)	no	WorldPop/POP	WorldPop 2015	100m	once	people in 100x100m grid cell
Malaria high-risk	Village with API** > 10 cases/1,000 people	binary (yes/no)	yes (year)	–	Surveillance system of Peruvian ministry of health	-	week	reported cases
Malaria very-high-risk	Village with API > 50 cases/1,000 people	binary (yes/no)						

** API = Confirmed cases in a year *1,000 / total population, * PP: Per-pixel

To download the images for the study area, we first evaluated the use of ENVI (remote sensing software) and Semi-Automatic classification plugin for QGIS. However, to cover Loreto extension with Landsat images 30-meter spatial resolution 25 scenes were needed (Figure 15), this is a big number for processing. Other difficulty was that this area is mostly rainy and cloudy and the acquiring clean images using optic sensor was a complex task. These difficulties pushed us to evaluate other alternatives. Thus, GEE was chosen because it consists of a multi-petabyte analysis-ready data catalog co-located with a high-performance through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) that enables processing in less time than others alternatives evaluated (108).

Using GEE, a feature collection encompassing Loreto coordinates was upload at code editor. The codes for each predictor were made using Java Script. The Image collections were managed assuming the data period, type of sensor. Calibration at Top of Atmosphere (TOA), and cloud masking were applied for Landsat images (81,109). The values were calculated according to the parameter required for the model according to the flow (Figure 16) after we clipped the area and rescaled and reprojected to obtain all images with the same parameters.

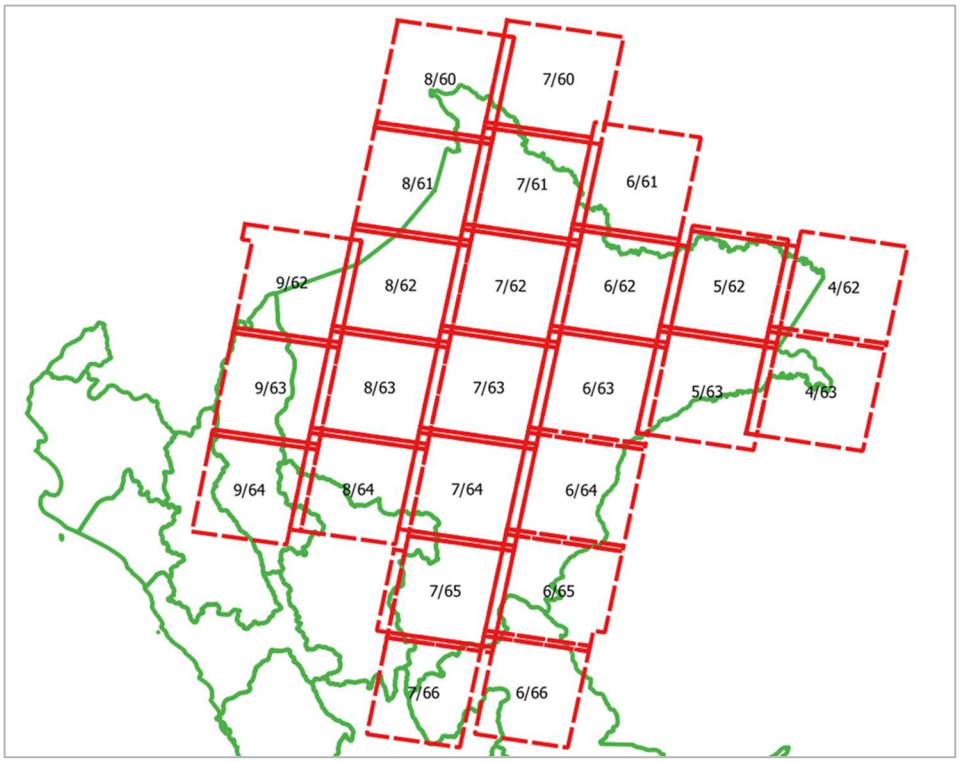


Figure 15: List of Landsat images (path and row) WRS-2 covering the study area.

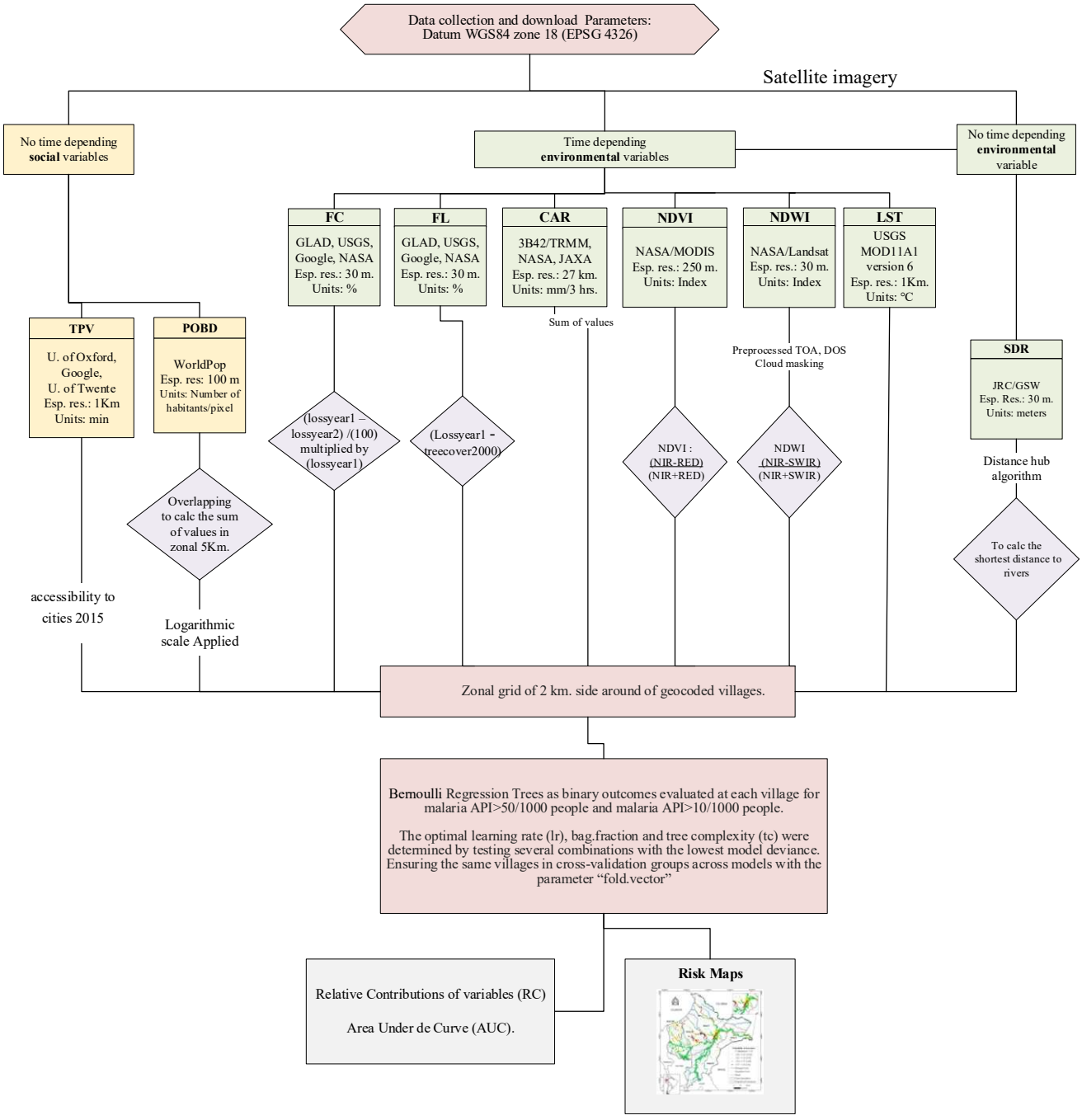


Figure 16: The flow adopted to capture values from satellite imagery (inputs) to analyze under Boosted Regression Trees (BRT) to obtain the malaria risk map.

3.5. Description of predictors

Time to major populated villages/towns (TPV): from the accessibility to cities 2015 developed by Malaria Atlas Project, University of Oxford, this map enumerates land-based travel time to the nearest densely-populated area (between 85 degrees north and 60 degrees south). Densely-populated areas defined as contiguous areas with 1,500 or more inhabitants per square kilometer or most built-up land cover types coincident with a population center of at least 50,000 inhabitants. The underlying datasets used to produce the friction map includes roads (comprising the first ever global-scale use of Open Street Map and Google roads datasets), railways, rivers, lakes, oceans, topographic conditions (slope and elevation), land cover types, and national borders. (20).

Population density (POPD): the estimated number of residents in each 100mx100m grid-cell in 2015, extracted from WorldPop (https://www.worldpop.org) (110).

Shortest distance to rivers (SDR), estimated using the Global Surface Water (occurrence) map (111) from JRC (Joint Research Centre) generated using scenes from Landsat 5, 7, and 8 acquired between 16 March 1984 and 10 October 2015. A 50% threshold mask was applied in GEE to select pixels with a presence of water at least half of the period 1984-2015, capturing main and secondary rivers. The shortest distance in (meters) between rivers and villages was calculated using the proximity algorithm QGIS v.3.4.2 (112,113).

Forest coverage (FC) and *forest loss (FL)*, time-dependent measures of the area covered by trees (%) in each year and the loss of tree coverage compared to the previous year (%), calculated from 2000–2018 Global Forest Change data. To register the 2017 forest coverage, we used the formula given by (lossyear2017 - lossyear2018) / (100) multiplied by (lossyear2017), this process was iterated to complete all the layers for the study period. The 2015 annual forest loss was calculated applying the formula (lossyear2017 - treecover2000) this process was iterated to complete all the layers for the study period (https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.6.html)(114).

Cumulative annual rainfall (CAR), generated from NASA GSFC TRMM. The Tropical Rainfall Measuring Mission (TRMM) is a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA) designed to monitor and study tropical rainfall. The estimated cumulative yearly rainfall (mm/year) calculated from daily 3-hour infrared precipitation estimates in product 3B42 of the (TRMM) (115). The 34B2 product contains a gridded, TRMM-adjusted, merged infrared precipitation (mm/hr) at 0.25-degree spatial resolution and 3 hours of temporal resolution. The band precipitation was selected to add all the pixels yearly.

Normalized Difference Water Index (NDWI), a satellite-derived index from the Near-Infrared (ρNIR) and Short Wave Infrared (ρSWIR) channels that estimates the amount of water in internal leaf structure (116,117). This index was processed from Landsat 5, 7 and 8 T1 (calibrated TOA, DOS and clouds masking), and calculated as NDWI=(ρNIR-ρSWIR)/(ρNIR+ρSWIR) (109,118). where ρRED and ρNIR correspond to the reflectance measured in the Red band and the Near Infrared band, respectively. Generally, NDVI ranges from −1 to +1, where water typically has an NDVI less than 0, bare soil between 0 and 0.1 and vegetation greater than 0.1.

Normalized Difference Vegetation Index (NDVI), the estimated fraction of radiation absorbed by the vegetation in the red (ρRED) and the near infrared (ρNIR) channels(119), from the MODIS/006/MOD13Q1 Moderate Resolution Imaging Spectroradiometer - National Aeronautics and Space Administration (NASA). The index was calculated at a spatial resolution of 250m. as: NDVI=(ρNIR-ρRED)/(ρNIR+ρRED) (118,120–122). where ρNIR and ρSWIR correspond to the reflectance measured in the Near Infrared band and the Shortwave Infrared band. In general, NDWI is positive for all water features and negative for all other land features.

Land Surface Temperature (LST), an estimate of the infrared emissivity of the earth in degrees Celsius (Emissivity Daily L3 Global, MOD11A1 version 6) (123). Mean values from 365 daily layers were used to produce raster images with yearly means.

3.6. Extraction of values

Raster images were resampled in 90-meter-pixels using R software version 3.5.1. Focal function that computes an output raster based on the neighborhood information (124) using a "moving window" around to the pinpointed village captured a 2 km-side square grid for forest coverage, forest loss, rainfall, vegetation index, water index and surface temperature, and in a 5 km-side square grid for density population. These 2 and 5 km-side thresholds were chosen to capture breeding and resting sites within the flight distance for *An. darlingi* mosquitoes (63,125). And to capture the population within and around the village, respectively. Raster values in square grids were aggregated as averages for environmental variables and as sums for POPD. For accessibility and distance to rivers, the value extracted was the village location grids.

3.7. Boosted regression tree analysis

BRT models(126,127) were created using R packages "gbm"(128,129) and "dismo"(130) to examine the relationship between potential predictors and the species-specific malaria risk status in villages for each study year. Fitting a Bernoulli response distribution to accommodate the binary data structure (i.e., villages with or without high-risk of malaria; or, villages with or without very-high-risk of malaria). The optimal learning rate (lr), bag.fraction and tree complexity (tc) were determined by testing several combinations with the lowest model deviance. Ten-fold cross-validation procedures using the "gbm.step" function enabled the selection of the optimal number of regression trees for each model, ensuring the same villages in cross-validation groups across models with the parameter "fold.vector". Model predictive performance was assessed using cross-validated estimates of the area under the curve (cvAUC) of the receiver operating characteristic (ROC), which is a measure of model discriminatory efficiency.

The relative contribution (RC) percentage was used to assess the relevance of each variable in BRT models. This metric measures how often the predictor is selected for partitioning, weighted by the squared model improvement resulting from successive partitions(130,131). Partial dependence plots (PDPs) were generated to describe the effect of one predictor on the malaria risk status of the villages,after accounting for the average effects of all other

predictors(132). In these plots, the vertical axis is the logit of cross-validated predicted probability (logit (p)) for high-risk or very-high-risk of malaria, and the horizontal axis is the variable predictor with corresponding units. BRT models built with data of a given year were assessed in terms of their ability to predict the malaria risk status of villages with data of the following year by estimating corresponding testing AUCs (tAUC).

The area under the curve (AUC) of the receiver operating characteristic (ROC) curve assessed the performance of BRT models for discriminating the malaria risk status in villages. Each cross-validation BRT model built with data of a given year yielded a cross-validated AUC (cvAUC), while its model predictions (i.e., predicted probabilities for villages at malaria risk) with testing data of the following year allowed for the estimation of a testing AUC (tAUC).

3.8. Risk maps elaboration

BRT models built with 2016 data were tested with 2017 data, obtaining predicted probabilities for villages at high-risk (API>10) and villages at very-high-risk (API>50). These predicted probabilities were linked to the pinpointed maps with the unique village code. The values were classified into four discrete categories (i.e., 0-0.25, 0.251-0.50, 0.511-0.75, 0.751-1.00), and to create species-specific risk maps for 2017 using QGIS v.3.4.2-Madeira.

Chapter 4: Results

This chapter shows the results of the analysis through the use of BRT. The relative contribution of variables, the partial dependence plots, the model's performance. This chapter also provides population-based malaria risk maps and risk zones.

4.1. Malaria occurrence

Of the 321,210 malaria cases reported to the epidemiological surveillance system in Loreto between 2010 and 2017, 311,128 cases (96.9%) occurred among our validated and georeferenced 2,766 villages. Non-pinpointed cases (3.1%) could correspond to infections in transient populations and/or inaccurate records of the place of the infection.

Malaria steadily increased in study villages from 10,994 cases in 2011 to 59,257 cases in 2014 and 58,679 in 2015, after which cases slightly declined to 51,663 in 2017 (Figure 17) *P. vivax* cases always predominated over *P. falciparum* cases with the maximum ratio in 2012 (Pv/Pf 5.5) and the minimum ratio in 2016 (Pv/Pf 2.6). The highest peak occurred earlier for *P. vivax* (2014) than for *P. falciparum* (2016). The number of villages at high-risk (API>10) and at very-high-risk (API>50) followed similar trends.

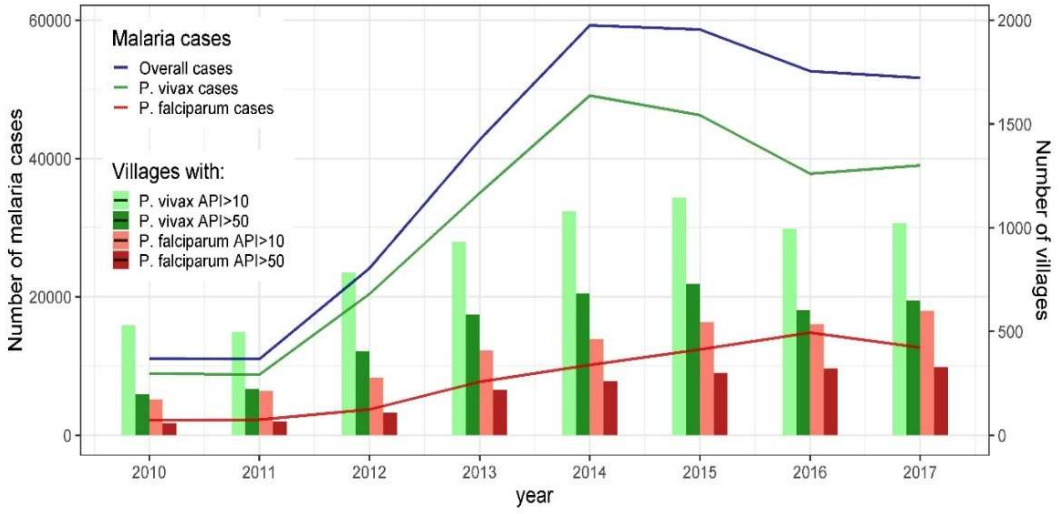


Figure 17: Reported malaria cases and number of villages at risk in Loreto from 2010 to 2017.

4.2. Relative contributions (RC) of variables

Table 7: Means (M), standard deviations (SD), medians (Mdn), and interquartile ranges (IQR) values of relative contributions of predictors for overall malaria, *P. vivax, and P. falciparum* risk, over the study period (2010-2017).

Predictors	Means (M)	Standard Deviations (SD)	Medians (Mdn)	Interquartile Ranges (IQR)
CAR	868.80	151.33	876.80	212.21
POPD	475.27	997.32	289.96	267.26
TPV	1236.28	902.57	1111.25	1508.26
NDWI	0.41	0.05	0.40	0.06
FC	77.05	16.26	79.10	24.57
NDVI	0.67	0.11	0.69	0.15
LST	27.76	0.93	27.74	1.14
SDR	2.12	4.74	0.30	1.28
FL	1.20	1.56	0.67	1.35

Table 8: Relative contributions (RC) of predictors obtained from yearly BRT models for the overall malaria risk over the study period (2010-2017).

Risk	**Variables**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
High risk (API>50)	CAR	16.3	20.8	20.4	39.1	41.3	49.3	31.8	32.6
	POPD	14.7	22.8	17.7	8.8	13.5	9.8	12.8	9.2
	TPV	10.4	10.6	11.1	12.1	8.0	9.7	9.2	10.6
	NDWI	18.9	4.8	10.4	7.9	4.8	3.4	10.1	12.7
	FC	12.3	7.4	6.8	5.8	5.5	3.3	6.0	8.7
	NDVI	4.4	13.6	10.9	6.5	5.0	7.2	5.1	6.6
	LST	8.8	8.4	6.8	5.6	7.7	6.5	11.5	7.6
	SDR	8.3	5.0	7.2	4.1	5.5	3.7	7.0	6.3
	FL	5.8	6.6	8.8	10.1	8.8	7.0	6.4	5.8
Very high risk (API>50)	CAR	14.4	17.1	20.4	32.0	29.0	36.5	26.4	27.2
	POPD	13.8	19.5	15.3	9.5	11.9	11.0	15.1	12.6
	TPV	10.7	12.7	11.5	13.3	11.2	12.8	11.5	13.4
	NDWI	17.3	7.4	13.1	10.7	9.7	6.2	10.5	10.4
	FC	12.5	6.8	6.8	5.1	7.9	5.4	7.8	8.4
	NDVI	5.9	10.2	9.7	9.4	5.8	9.4	5.6	8.3
	LST	10.9	11.0	9.0	8.4	10.5	8.2	10.5	7.0
	SDR	9.3	8.1	7.1	4.6	5.1	3.5	5.1	6.2
	FL	5.2	7.1	7.1	7.2	8.9	6.9	7.5	6.4

Table 9: Relative contributions (RC) of predictors obtained from yearly BRT models for the *P. vivax* malaria risk over the study period (2010-2017).

Risk	Variables	2010	2011	2012	2013	2014	2015	2016	2017
High risk (API>50)	CAR	17.0	18.5	20.6	38.1	41.8	48.4	33.2	31.7
	POPD	14.9	22.6	17.8	9.2	13.0	9.4	12.2	10.4
	TPV	10.3	11.3	11.0	11.6	7.4	9.7	10.3	11.1
	NDWI	18.9	5.3	11.5	8.0	5.3	3.8	9.5	12.0
	FC	11.9	7.1	6.3	5.8	5.4	3.1	5.5	8.2
	NDVI	5.1	12.1	10.5	6.9	4.8	7.4	4.4	7.2
	LST	8.2	9.3	6.5	5.7	7.3	6.8	11.7	7.6
	SDR	9.0	6.5	7.3	4.3	5.4	3.5	6.6	6.2
	FL	4.7	7.3	8.6	10.4	9.5	7.9	6.7	5.6
Very high risk (API>50)	CAR	14.3	17.3	19.9	30.8	27.7	35.5	25.6	28.1
	POPD	12.9	18.7	15.5	10.2	12.1	10.3	14.2	14.1
	TPV	12.0	12.0	11.0	12.7	11.5	12.5	11.0	13.1
	NDWI	16.8	6.9	13.1	11.2	10.5	7.1	10.6	10.1
	FC	12.2	7.1	7.8	5.6	7.7	5.4	8.4	8.2
	NDVI	6.7	11.1	10.3	9.5	5.1	9.7	6.5	7.8
	LST	11.6	11.1	8.3	7.3	10.5	8.8	10.2	6.8
	SDR	9.1	8.7	7.8	4.6	5.1	3.7	6.0	5.9
	FL	4.5	7.1	6.3	8.1	9.8	7.1	7.6	5.9

Table 10: Relative contributions (RC) of predictors obtained from yearly BRT models for the *P. falciparum* malaria risk over the study period (2010-2017).

Risk	Variables	2010	2011	2012	2013	2014	2015	2016	2017
High risk (API>50)	CAR	11.5	15.9	14.4	26.9	23.3	30.7	27.6	28.6
	POPD	12.6	18.0	16.4	11.7	20.6	15.1	16.8	12.1
	TPV	12.5	12.0	8.7	11.3	11.9	10.9	10.9	14.7
	NDWI	15.1	7.6	12.1	9.8	8.7	6.3	12.4	10.6
	FC	7.3	14.4	10.1	8.4	7.4	8.9	6.0	4.2
	NDVI	9.4	7.6	11.4	10.5	9.5	6.1	6.1	4.5
	LST	12.5	7.4	8.1	4.2	5.7	6.1	5.7	9.7
	SDR	5.6	7.4	6.9	7.0	5.0	5.1	5.2	4.4
	FL	8.2	5.7	6.6	5.5	3.6	4.0	3.7	5.1
Very high risk (API>50)	CAR	16.1	20.7	16.8	23.1	19.4	25.6	26.2	25.0
	POPD	13.0	10.0	15.3	16.7	21.4	16.2	16.2	12.5
	TPV	13.4	7.2	7.3	11.1	10.9	13.1	11.5	15.8
	NDWI	12.2	8.4	13.6	9.6	9.7	8.0	13.8	8.1
	FC	7.8	11.1	11.6	5.9	8.5	10.0	6.8	10.9
	NDVI	7.3	11.1	7.0	10.2	6.0	8.7	5.9	6.3
	LST	8.5	10.9	9.8	8.3	8.8	5.1	4.7	5.9
	SDR	6.4	13.8	10.5	5.9	4.6	3.7	4.2	6.0
	FL	10.2	4.2	4.7	5.4	5.8	3.4	6.1	3.9

4.3. Relative contributions (RC) of predictors from yearly BRT models

(a)

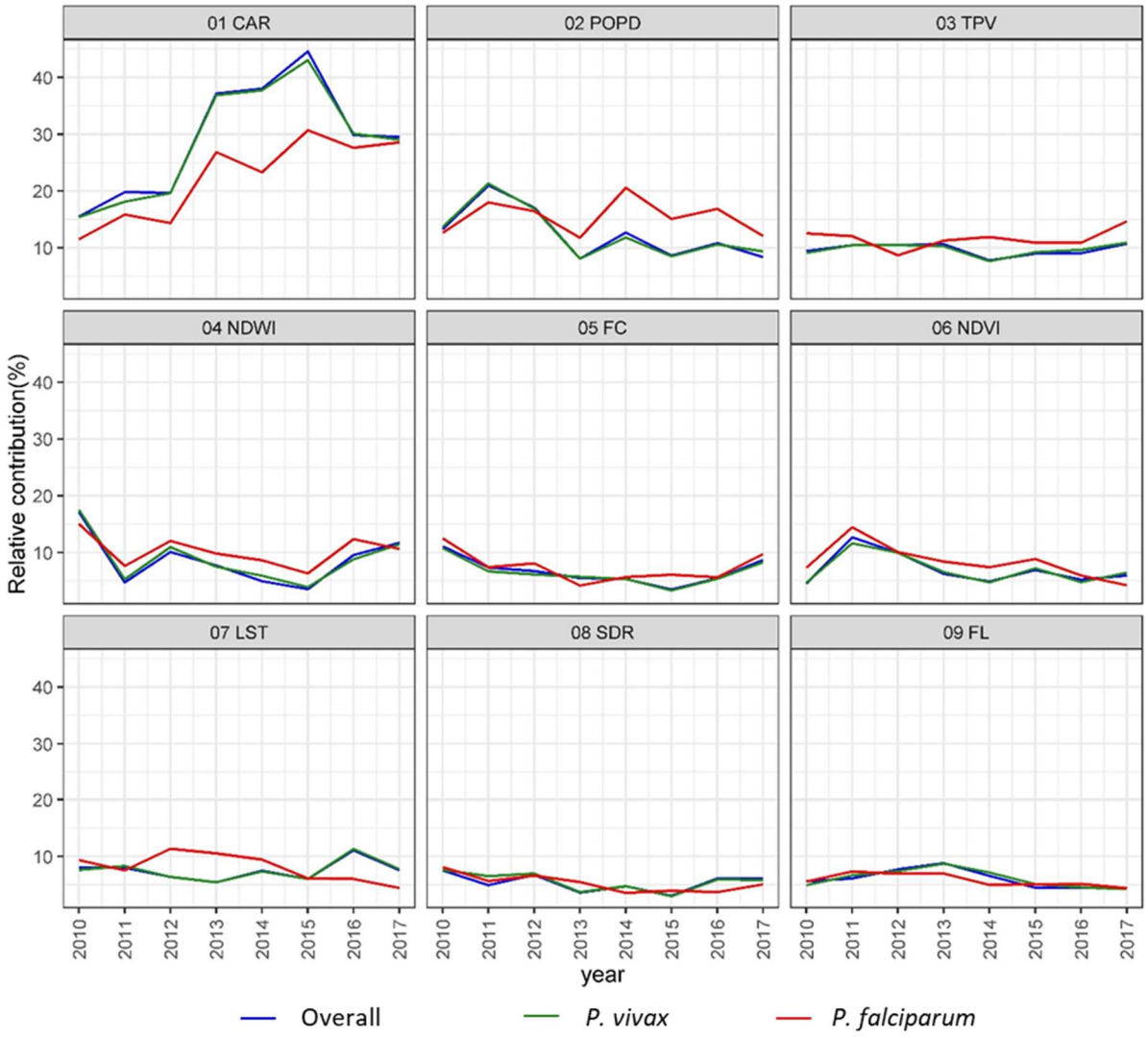


Figure 18: Relative contributions (RC) of predictors from yearly BRT models for malaria risk, overall and by species, years (2010-2017), high malaria risk (API>10 cases/1,000 people).

(b)

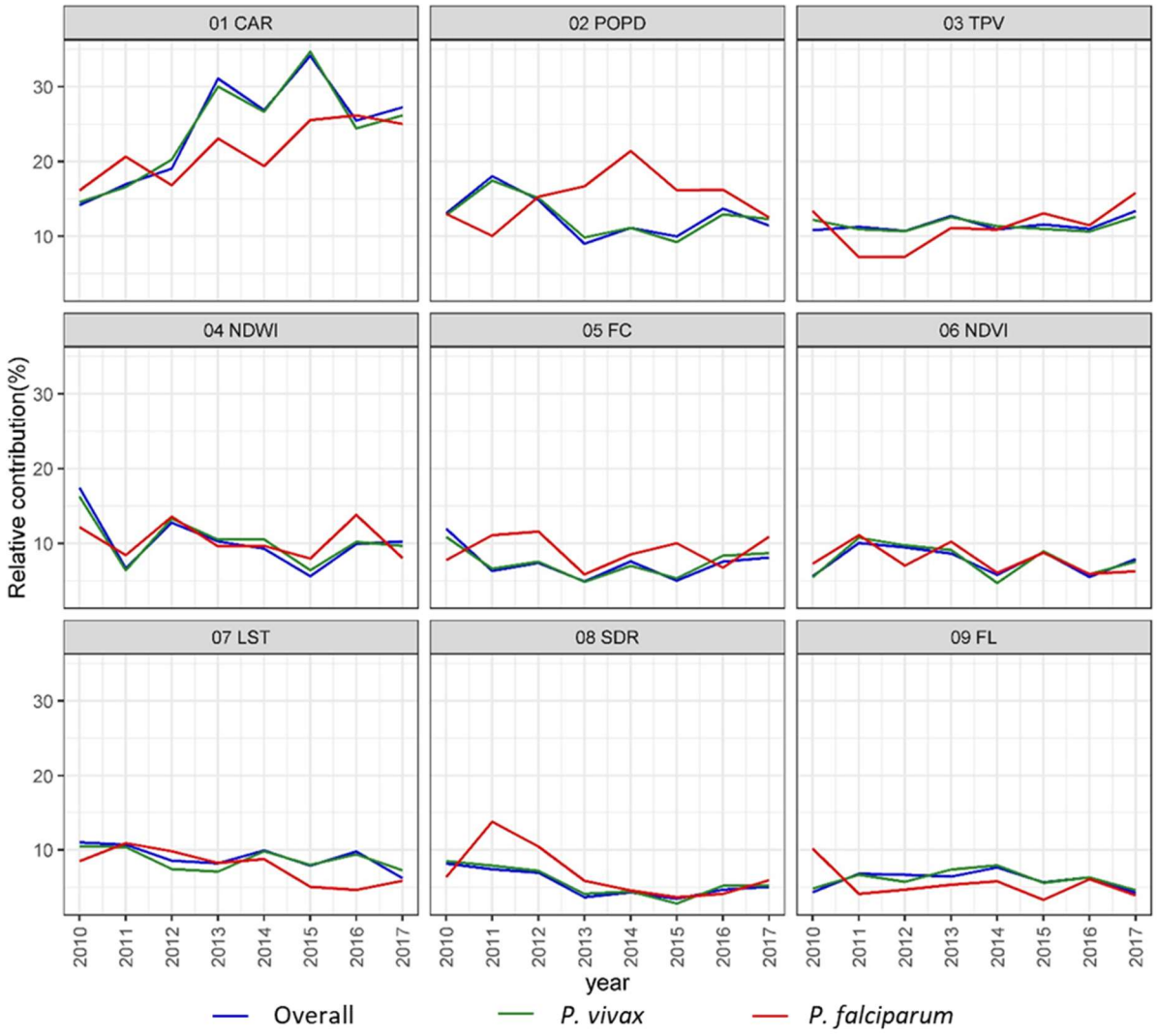


Figure 19: Relative contributions (RC) of predictors obtained from yearly BRT models for malaria risk, overall and by species, years (2010-2017), very high malaria risk (API>50 cases/1,000 people).

Most RC for overall malaria risk were similar to ones for *P. vivax* each year. CAR, POPD, and TPV, in that order, were always the top three predictors for increasing malaria risk in villages (overall and by species), with yearly RC medians exceeding 10% (Figure 18). NDWI was the fourth top predictor, but yearly RC medians only exceeded 10% when BRT models assessed villages at *P. vivax* very-high-risk of and villages at *P. falciparum* high-risk.

CAR was generally the most relevant predictor in yearly BRT models for both *P. vivax* and *P. falciparum*, and presented the widest variation in RC. The analysis for high-risk in villages showed a higher RC of CAR for *P. vivax* than for *P. falciparum* (Figure 19), with the lowest RC for both species in 2010 and the highest in 2015 (RC range for *P. vivax*: 17%-48.4%; RC range for *P. falciparum*: 11.5%-30.7%). The analysis for very-high-risk found the lowest RC in 2010 for both *P. vivax* (14.3%) and *P. falciparum* (16.1%), and the highest in 2015 for *P. vivax* (35.5%) and in 2016 for *P. falciparum* (26.2%). The estimated RC for CAR followed similar trends with increasing importance of CAR as malaria risk predictor until 2015 for *P. vivax* and until 2015-2016 for *P. falciparum*, followed by a decrease (*P. vivax*) or stabilization (*P. falciparum*).

Unlike CAR, the POPD RC in villages was slightly higher for *P. falciparum* than for *P. vivax* in most years (Figure 20). The relevance of POPD varied widely across years only when BRT models assessed for *P. vivax* high-risk (range: 9.2%-22.6%) and *P. falciparum* very-high-risk (range: 10.0%-21.4%). The highest RC for *P. vivax* malaria risk were found at the beginning of the study period (2010-2012) and then decreased and remained low with small variations; while RC for *P. falciparum* increased from 2010 to 2014 and then declined in the following years.

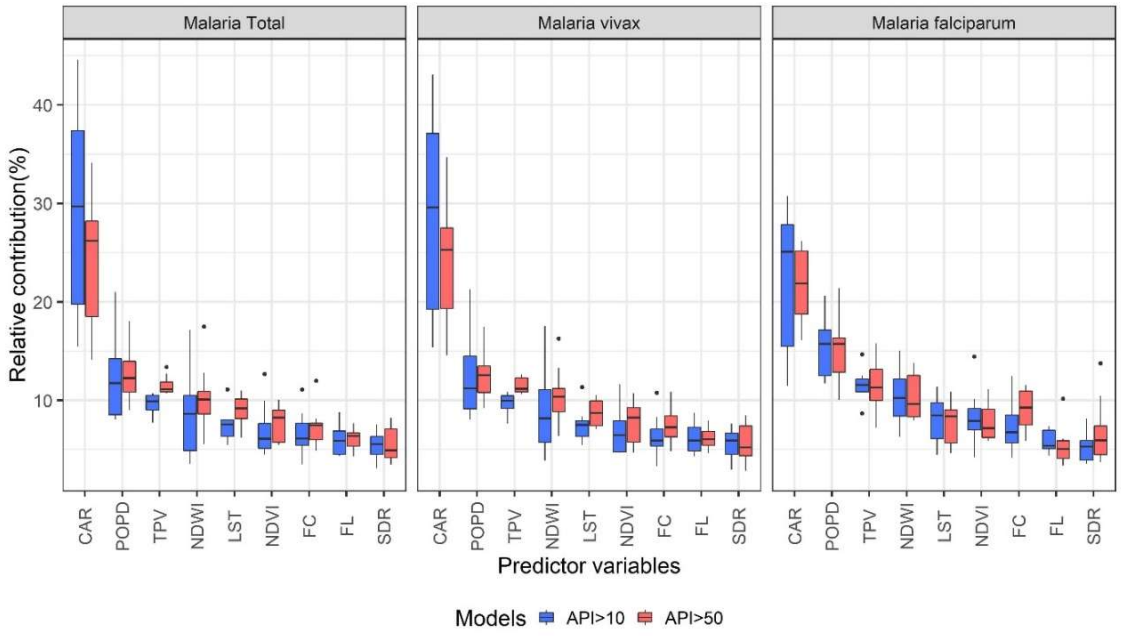


Figure 2019: Boxplots representing the relative contributions of predictors obtained from yearly BRT models for malaria risk, overall and by species, over the study period (2010-2017).

The importance of NDWI as species-specific malaria risk predictor varied during the study period. The difference between lowest and highest RC ranged between 3.8 and 18.9%, with an initial decrease in RC from 2010 to 2011 followed by peaks in 2012 and in 2016. The non-time dependent variable TPV was the predictor with the lowest RC variation across years for *P. vivax* risk in villages (7.4%-11.6% and 11%-13.1% for high-risk and very-high-risk respectively). It was also among the predictors with the lowest RC ranges when assessing the *P. falciparum* risk across time. For all other predictors in species-specific BRT models, the difference between lowest and highest RC did not exceed 10% across years.

4.4. Partial dependence plots

Figure 21 and Figure 213, presents partial dependence plots (PDP) (133) of the marginal effect of predictor variables on the probability of villages for being high-risk or very-high-risk, and shows the functional form of these effects. The probability for *P. vivax* high-risk in villages generally increases with: higher CAR from 800 mm/3hrs., POPD between 403 and 2,980 in the 5-km side square grid around the village, longer TPV from 700 minutes, NDWI around 0.4, higher FC in the 2-km side square grid around the village from levels around 50%, NDVI between 0.7 and 0.8, higher LST from 27ºC, shorter SDR, and higher FL in the in the 2-km

side square grid around the village in comparison with the previous year. The probability for *P. falciparum* high-risk increased with higher CAR from 800 mm, POPD ranging between 403 and 1096 inhabitants, longer TPV from 700 minutes, NDWI around 0.4, higher FC from about 60%, NDVI ranging between 0.5 and 0.8, higher LST from 26ºC, shorter SDR, and higher FL. PDPs for very-high-risk are shown in Figure 22, and Figure 24Figure .

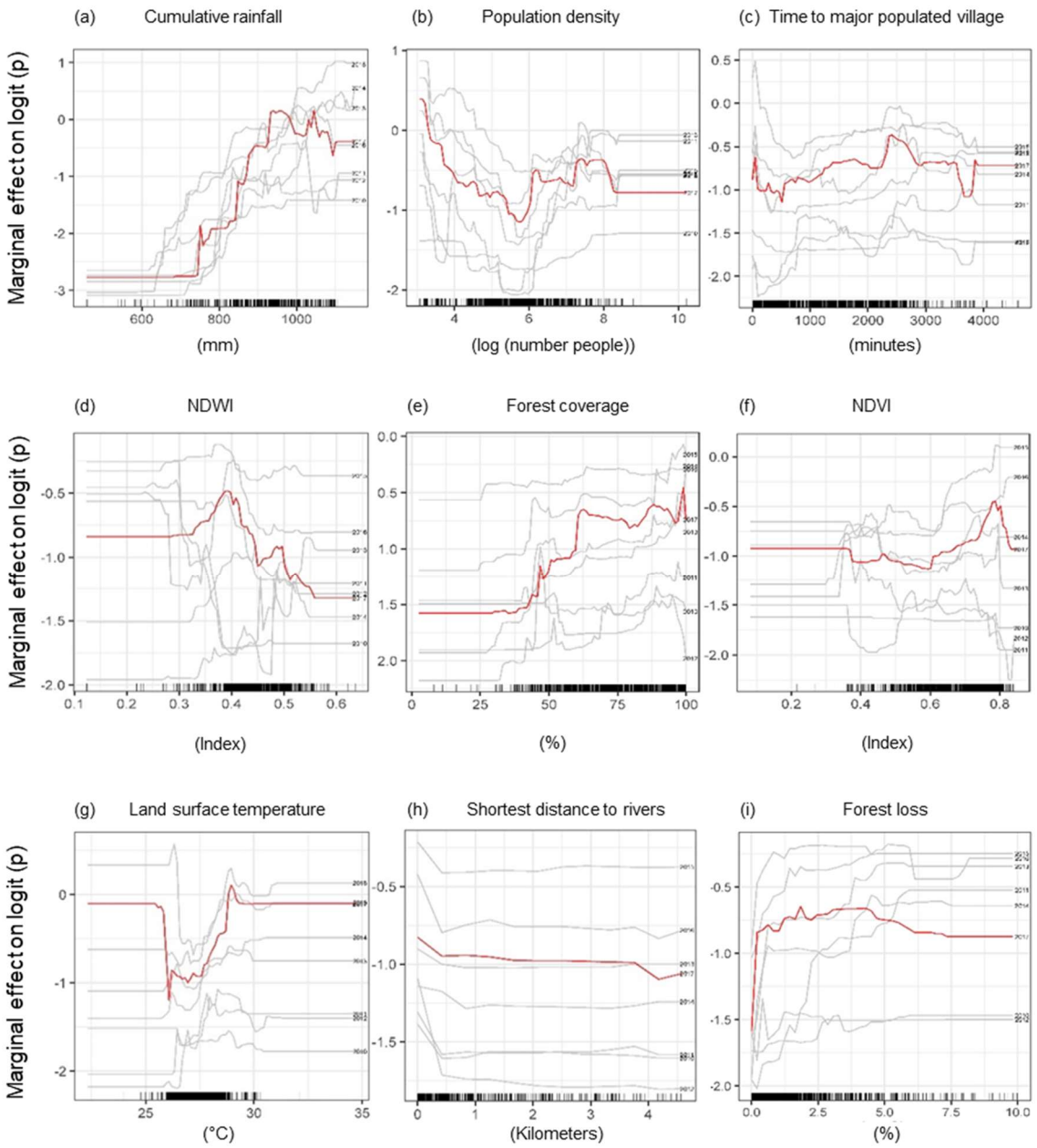


Figure 2120: Partial dependence plots indicating the marginal effect of predictor variables on the probability of villages for being a high malaria risk (API>10 malaria cases/1,000 people) for *P. vivax* Y-axes are on a logit scale. Red lines represent the predictions for 2017, while gray lines for other years. The distribution of variable values is indicated at the bottom of each plot.

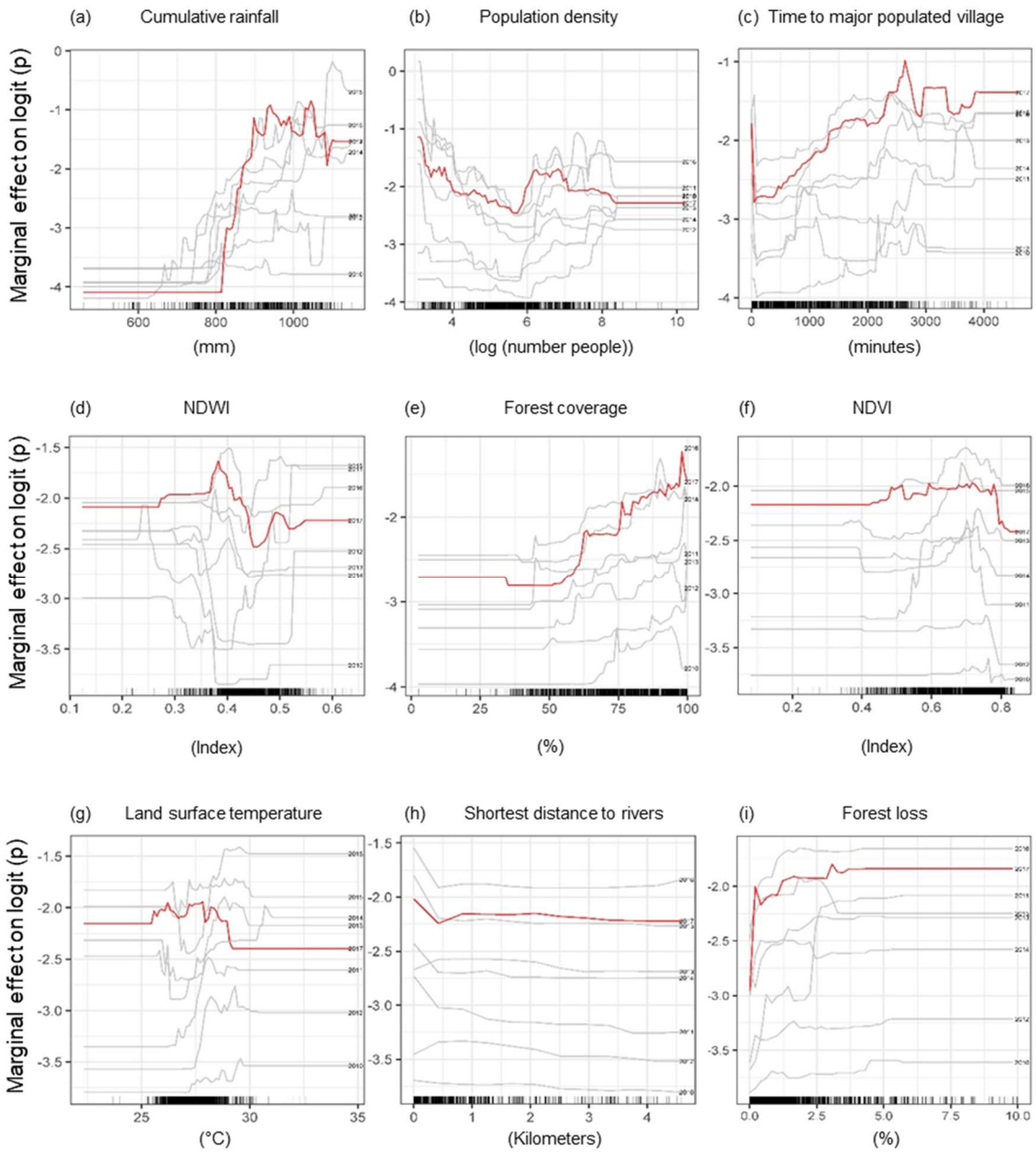


Figure 22: Partial dependence plots indicating the marginal effect of predictor variables on the probability of villages for being a high malaria risk (API>10 malaria cases/1,000 people) for *P. falciparum.* Y-axes are on a logit scale. Red lines represent the predictions for 2017, while gray lines for other years. The distribution of variable values is indicated at the bottom of each plot.

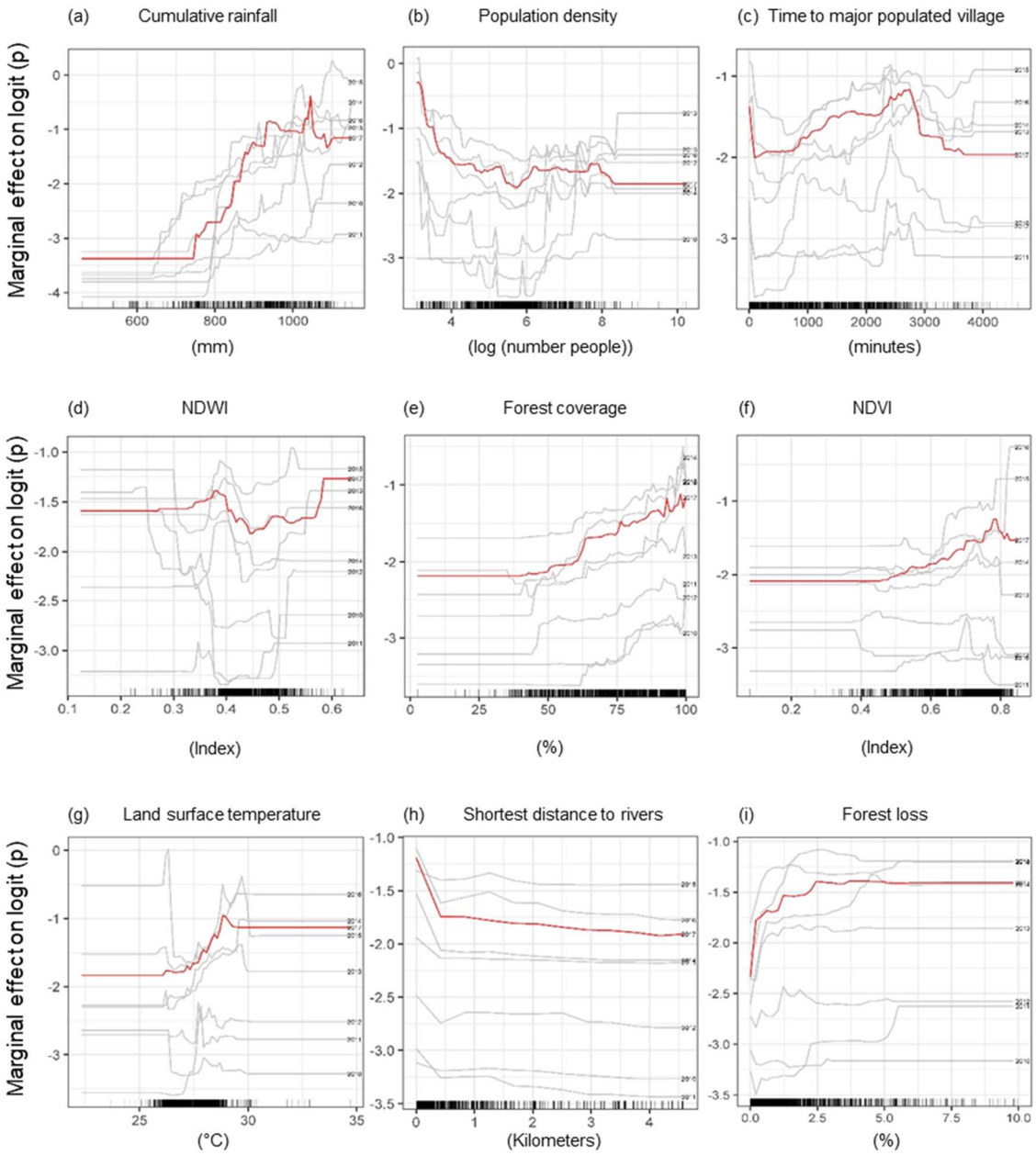


Figure 213: Partial dependence plots indicating the marginal effect of predictor variables on the probability of villages for being at very high malaria risk (API>50 malaria cases/1,000 people) for *P. vivax* (b). Y-axes are on a logit scale. Red lines represent the predictions for 2017, while gray lines for other years. The distribution of variable values is indicated at the bottom of each plot.

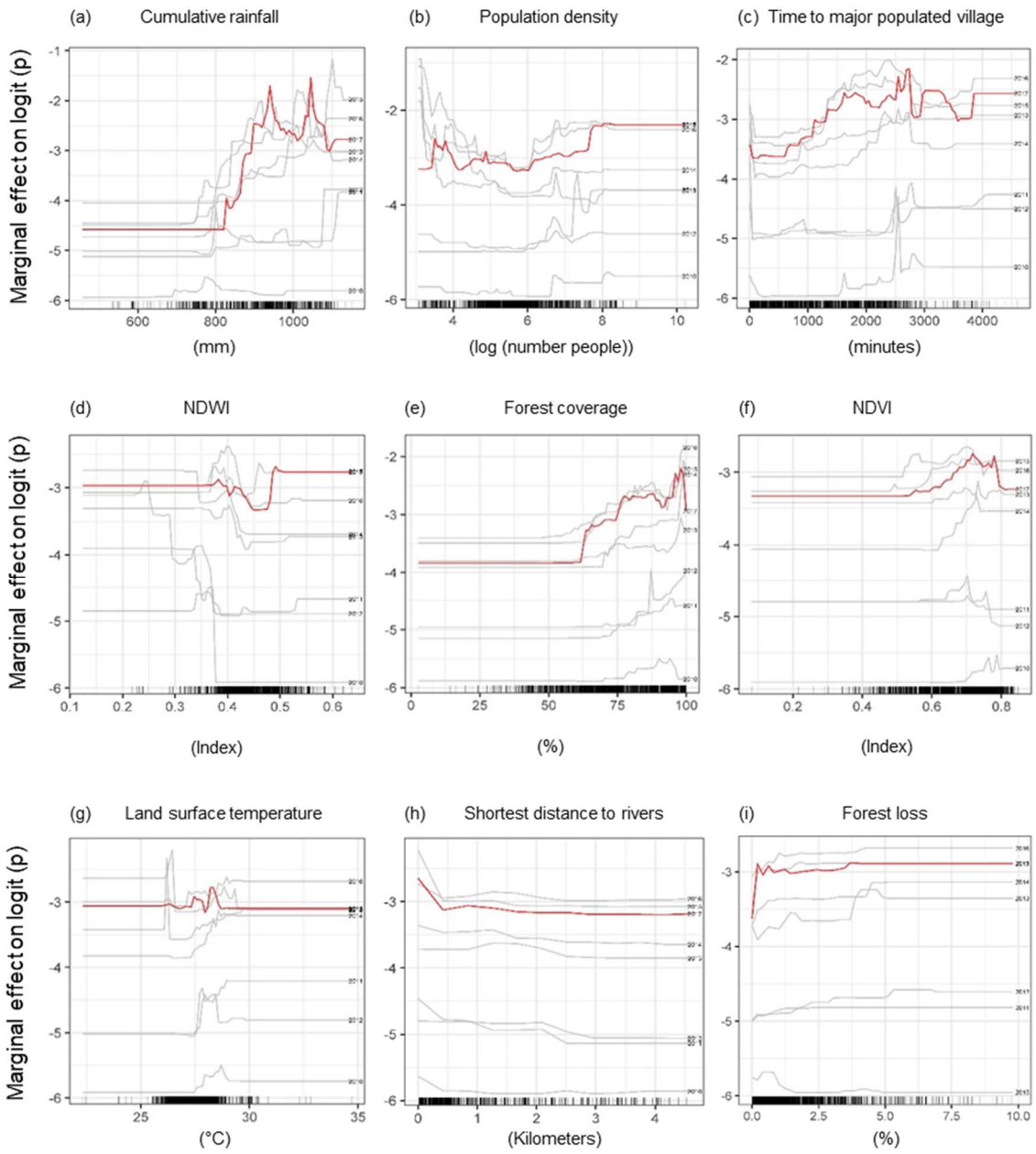


Figure 24: Partial dependence plots indicating the marginal effect of predictor variables on the probability of villages for being a very high malaria risk (API>50 malaria cases/1,000 people) for *P. falciparum.* Y-axes are on a logit scale. Red lines represent the predictions for 2017, while gray lines for other years. The distribution of variable values is indicated at the bottom of each plot.

4.5. The discriminatory capacity of the models

The cvAUCs shown in Table 11 suggested that most yearly BRT models efficiently discriminated the malaria risk in villages (cvAUC>0.80), with the exception of the 2010 models for villages at high-risk (cvAUC=0.72), and at *P. vivax* very-high-risk (cvAUC=0.77), as well as *P. falciparum* high-risk (cvAUC=0.78). The estimated tAUCs decreased slightly when the yearly models were tested with data for the following year, but most of the models efficiently predicted malaria risk in the following year. For example, 2016 species-specific models using 2017 data were able to efficiently discriminate high-risk (tAUC=0.81) and very-high-risk (tAUC=0.84) of *P. vivax*, and high-risk (tAUC=0.83) and very-high-risk of *P. falciparum* (tAUC=0.83) in 2017. Results from the discriminatory assessment of yearly BRT models with data from other years are presented in Figure 225.

Table 11: Assessment of the capacity of BRT models for discriminating malaria risk in villages.

	Model	**Overall**		***P. vivax***		***P.alciparum***	
		cvAUC	**tAUC**	**cvAUC**	**tAUC**	**cvAUC**	**tAUC**
High risk (API>10)	2010	0.72	0.70	0.72	0.70	0.78	0.76
	2011	0.80	0.76	0.8	0.75	0.86	0.74
	2012	0.82	0.80	0.82	0.80	0.84	0.81
	2013	0.84	0.80	0.83	0.80	0.87	0.80
	2014	0.83	0.82	0.82	0.82	0.85	0.84
	2015	0.82	0.79	0.82	0.79	0.87	0.85
	2016	0.82	0.80	0.82	0.81	0.87	0.84
	2017	0.82	-	0.83	-	0.87	-
Very high risk (API>50)	2010	0.76	0.76	0.77	0.76	0.82	0.78
	2011	0.85	0.81	0.84	0.80	0.88	0.81
	2012	0.85	0.82	0.85	0.81	0.87	0.76
	2013	0.86	0.80	0.86	0.80	0.89	0.84
	2014	0.84	0.84	0.84	0.83	0.89	0.83
	2015	0.85	0.83	0.85	0.82	0.89	0.86
	2016	0.84	0.83	0.84	0.83	0.88	0.83
	2017	0.85	-	0.84	-	0.89	-

While cvAUC are the cross-validated area under ROC curve using data of a given year, tAUC are the cross-validated area under ROC curve using data of the following year.

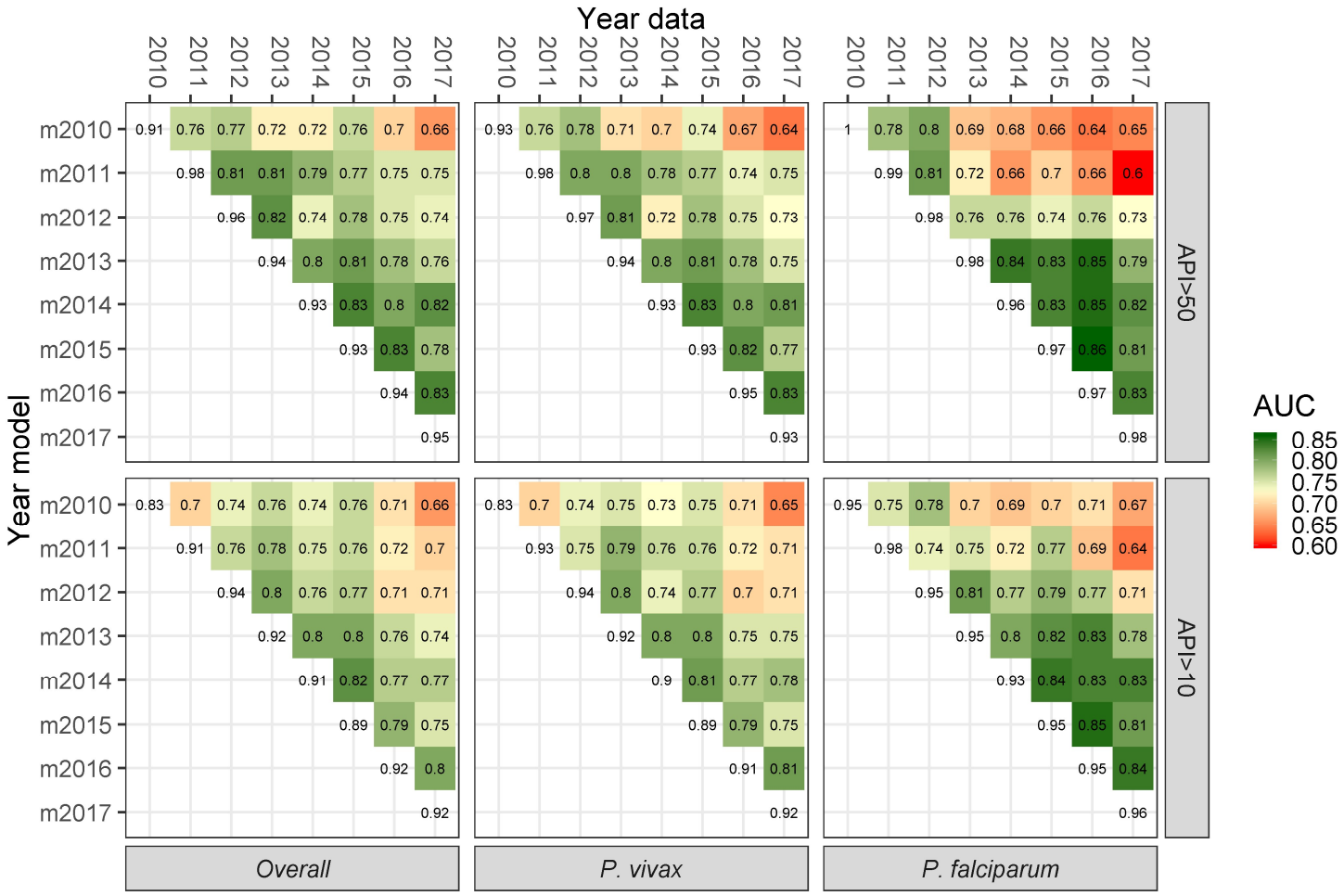
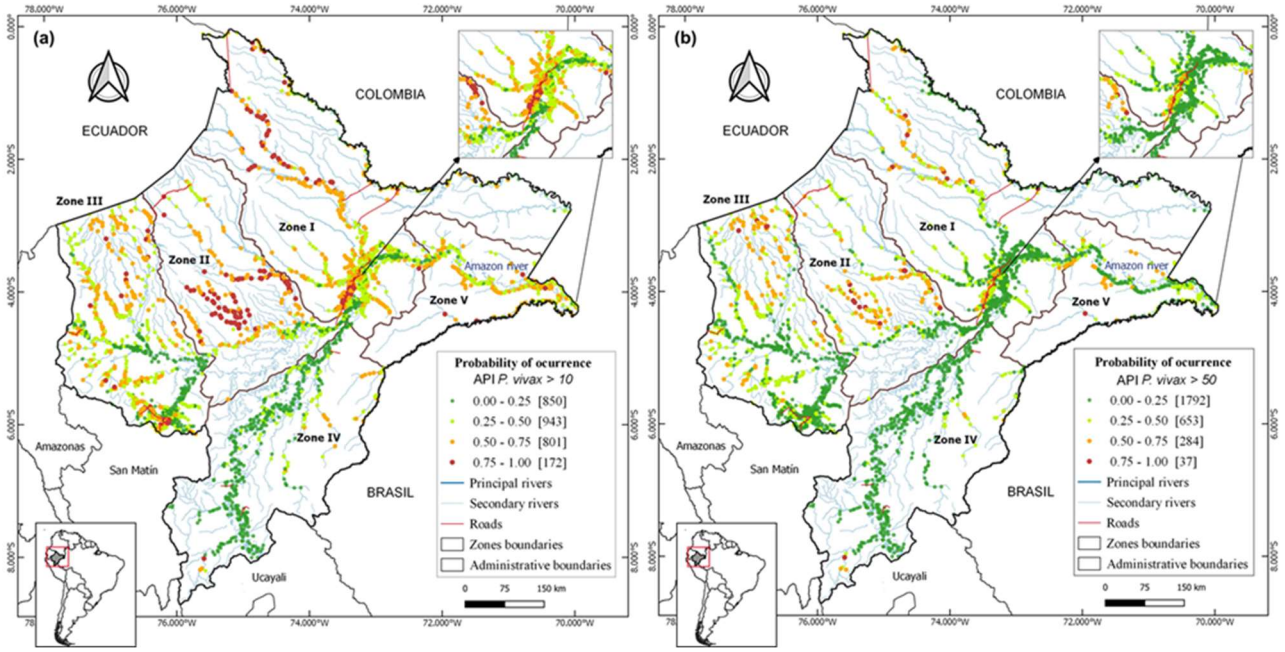


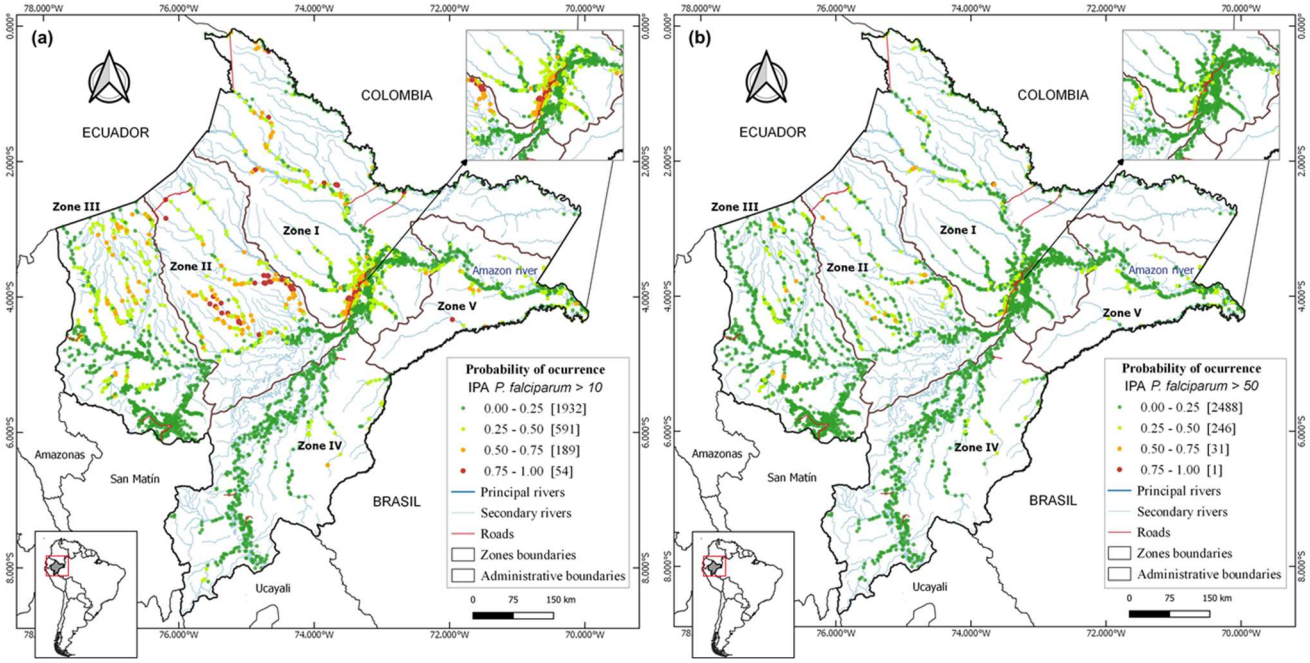
Figure 225: Discriminatory assessment of yearly BRT models with data from other years. The discriminatory efficiency is assessed with the area under the ROC curve (AUC).

4.6. Risk mapping

The study villages and their malaria risk in 2017 (estimated from 2016 BRT species-specific models) were mapped Map 8 and Map 9 in five distinct zones based on contiguity between districts, main networks of transport, shared river basins, and population density.



Map 8: Predicted *P. vivax* risk maps for the year 2017 using 2016´s BRT models showing: (a) villages at high *P. vivax* risk (API>10 cases/1,000 people), (b) villages at very high *P. vivax* risk (API>50 cases/1,000 people). Colors specify in villages (dots) their probability of being at risk.



Map 9: Predicted *P. falciparum* risk maps for the year 2017 using 2016´s BRT models showing: (a) villages at high *P. falciparum* risk (API>10 cases/1,000 people), (b) villages at very high *P. falciparum* (API>50 cases/1,000 people). Colors specify in villages (dots) their probability of being at risk.

4.7. Risk zones description

Table 12: Number of villages at high malaria risk and at very high malaria risk according to predictions of 2016 BRT models using 2017 data. Malaria risk with predicted probabilities > 0.5.

		High risk (IPA>10)			Very high risk (IPA>50)		
		N	N	%	n	N	%
P. vivax	Zone 1	369	860	42.9	90	860	10.5
	Zone 2	179	317	56.5	91	317	28.7
	Zone 3	294	853	34.5	105	853	12.3
	Zone 4	19	489	3.9	7	489	1.4
	Zone 5	112	247	45.3	28	247	11.3
P. falciparum	Zone 1	101	860	11.7	12	860	1.4
	Zone 2	88	317	27.8	11	317	3.5
	Zone 3	46	853	5.4	9	853	1.1
	Zone 4	1	489	0.2	0	489	0
	Zone 5	7	247	2.8	0	247	0

Zone I, includes Maynas province and is the largest, most densely populated and accessible area in the department. The risk maps showed that 42.9% of villages were at high-risk for *P. vivax*, among which 10.5% were at very-high-risk. Also, 11.7% of villages were at high-risk for *P. falciparum*, among which 1.4% were at very-high-risk.

Zone II, covers Loreto province; it is the least populated area. Most villages (56.5%), were at high-risk for *P. vivax*, and 28.7% at very-high-risk. For *P. falciparum*, 27.8% were at high-risk, and 3.5% at very-high-risk.

Zone III, is the second most densely populated area and covers Datem del Maranon and Alto Amazonas provinces, which include 30% of the study villages and 19% of the province's population. One third (34.5%) of the villages were at high-risk for *P. vivax*, and 12.3% at very-high-risk. For *P. falciparum*, 5.4% were at high-risk, and 1.1% at very-high-risk.

Zone IV, comprises Requena and Ucayali provinces, with 3.9% of villages at high-risk for *P vivax*, and 1.4% at very-high-risk. Only one village was at high-risk for *P. falciparum.*

Zone V, comprises Ramon Castilla province, and is the smallest zone, with 9% of the study villages and 6.7% of Loreto population. A significant proportion of villages (45.3%) were at *P. vivax* high-risk, and 11.3% at very-high-risk. Seven villages were at *P. falciparum* high-risk.

Chapter 5: Discussion, conclusions, limitations and perspectives.

This chapter presents the findings in perspective with the information presented in the introduction and literature review, describes the conclusions, contributions and avenues for future research.

5.1. Discussion

The Peruvian malaria fight against malaria has experienced a significant decline since 2012, after enormous achievements recorded in the previous years under the interventions of the PAMAFRO Project (2005–2011). In recent years, the Zero Malaria Plan (ZMP) a task force between MoH, Regional Health Direction (DIRESA - Loreto) started its interventions nationwide by the end of 2018, with the goal of Peruvian malaria eradication at the end of the year 2030. This new wave of financial support is strengthened by the global struggle, that pretends to reduce the global malaria burden by at least 90% of the number of cases compared to 2015 and malaria eradication from those countries where it is feasible (7). This is a huge challenge that requires malaria insight, uphill labor, and permanent actions that also embrace scientific innovations, new technologies, and tools to eliminate malaria more effectively and efficiently.

This thesis applied spatial tools to assess and predicted the co-endemic *Plasmodium vivax* and *Plasmodium falciparum* occurrence at villages in the Peruvian Amazon using Boosted Regression Trees.

Malaria infection/exposure was analyzed with environmental and social predictors deriving out of satellite imagery and products at village level between 2010 and 2017, and identified the most critical factors associated with this distribution. Yearly BRT models built with predictor data were able to efficiently discriminate the species-specific malaria risk in villages of such a given year, and most of these models also performed well when predictor data was used to discriminate malaria risk in the following year. The high discriminatory capacity of the 2016 BRT models supported the generation of maps pinpointing villages with a high probability of having high malaria incidence (API>10 cases x 1,000 people) and very high incidence (API>50 cases/1,000 people) in 2017.

Previous risk map studies (132,135–140) are consistent with our findings and recognize the mapping value to target resources for vector control or intensifying mosquito and disease surveillance, some of these researches suggest replicating the methodology in other geographical areas and using data of other diseases. The entomological inoculation rate (EIR) and the parasite prevalence are the best metrics of malaria transmission intensity but their estimations require several community surveys per year (to capture annually seasonal variations), very large sample sizes, and specialized laboratory techniques, making them cost-prohibitive for National Malaria Control Programs (NMCPs) (141). Instead, NMCPs have used the incidence of microscopically cases detected by routine surveillance to map the risk of malaria transmission, prioritize intervention areas, and monitor the impact of control interventions (142). However, these maps have coarser spatial resolution also are not necessarily correlated to condition factors limiting the usability to strategic plans elaboration or appropriate intervention within their national strategic agenda. The results we presented in this thesis can support more tailored intervention plans to improve the distribution of Insecticide-Treated Net (ITN) or using Focused Screening and Treatment (FSAT) or Mass Drug Administration (MDA) strategies to interrupt transmission. These massive treatment strategies are proposed in the WHO document to malaria control and eliminating 2016 - 2030 (31) also are contemplated in the implementation and elimination plan of the "Zero Malaria Plan", this alternative consists of treating all parasitemia in the population defined geographically. MDA can potentially reduce malaria mortality and morbidity interrupting various stages of the parasite life-cycle, and it can inhibit the sporogony cycle in the mosquito, reducing its vectorial capability (7). According to evidence, MDA is preferable in areas where populations are static and the risk of importation is low (143,144). Another consideration is that at least three "rounds" of administration are needed to affect transmission (145) and minimal risk for reintroduction of infection. Although, knowledge gaps remain, especially regarding the optimal size of the target population, methods to improve coverage, selection of drug-resistant parasites, and primaquine safety.

The mapping of malaria risk plays a key role in decision-making for designing and implementing malaria control measures (146), but it requires the use of metrics that ensure an accurate description of the malaria heterogeneity and the identification of changes in transmission intensity across time and geographical scales (141).

Our study confirmed that predicted malaria risk from BRT models built with RS-derived predictors was able to accurately characterize the spatial distribution of the malaria incidence in Loreto.

The ability of *P. vivax* parasites to relapse from persistent liver parasite stages (hypnozoites) weeks or months after a primary parasitaemia (147) may explain the lower discriminatory efficiency of *P. vivax* models compared with *P. falciparum* models in most study years. BRT models were built with available data for predictor variables reported to affect the malaria transmission. However, such variables are not yet proposed in the literature to be primary triggers of *P. vivax* relapses (147). When factors determining the hypnozoite reactivation are clearly identified, they could be incorporated to improve discriminatory capacity of the models for *P. vivax* risk. Improved models would be useful in the Peruvian Amazon, where epidemiological studies and parasite genotyping analyses suggest that relapses contribute substantially to the burden of *P. vivax* infections (148). Moreover, data on main control interventions could also be used during the model building process to account for the greater resilience of *P. vivax* to malaria control efforts in comparison with *P. falciparum* (149).

Environmental factors such as temperature and rainfall directly affect the lifecycles of both parasite and vector; higher temperatures accelerate parasitic *Plasmodium* growth within mosquitoes, while rainfall contributes to the accumulation of stagnant water that is ideal for mosquito breeding (63). Results from partial dependence plots (PDPs) showing higher malaria risk when the yearly average LST ranged between 26ºC and 29ºC are in good agreement with the reported optimal temperatures for the development of the exogenous *P. vivax* and *P. falciparum* cycles within the main malaria vector *An. darlingi* (63). However due to the low variability of LST across villages, this variable was not among the most important factors for discriminating the malaria risk between villages. Instead, the variable satellite-derived rainfall (CAR) was the best predictor for both species, with increased importance each year through 2015. Unusually heavy rains in the last trimester of each year between 2011 and 2014 generated abrupt increases of river levels, and flooded to villages along the Amazonas River and its tributaries between 2012 and 2015 (148,150–152). The river levels not only peaked earlier and higher compared to historical averages, but also, they remained high longer. Levels of Amazon river in 2015 exceeded the threshold for imminent flooding (117 metres above sea level) from March through June (153). River levels in 2011, 2012 and 2016 surpassed that threshold during 5, 14, and 4 weeks, respectively.

The proximity to breeding sites is an important determinant for the heterogeneity of both mosquito exposure and malaria occurrence in the Peruvian Amazon(154). After seasonal rains, permanent breeding sites around villages become more productive, and additional breeding sites arise (155). With severe flooding, breeding sites further enlarge and remain longer, leading to a wider dispersal of *An. darlingi* and increased vector-human contact rates. Villages with shorter distances to rivers (SDR) had increased malaria receptivity and consequently, more malaria incidence as shown in PDPs. Increased vector contact would explain both the higher contribution of CAR in *P. vivax* and *P. falciparum* models and the better discriminatory efficiency of *P. vivax* models in the years of severe floods (2012-2015). Therefore, our findings suggest that new infections contributed more than hypnozoite-triggered relapses to the rise in the *P. vivax* malaria incidence since 2012.

NDWI, recognized for appropriately identifying water bodies(156), was also shown an important factor in BRT models for both species (median in the study period slightly lower than 10%); but unlike with CAR, its RC did not present any identifiable temporal trend during the study period. Yearly average NDWI values around 0.4 associated with increased malaria risk suggest that most of the 2-km side square grid around villages with high malaria incidence was covered by open water, and/or wetlands, but were not able to characterize those that are suitable for the development of *An. darlingi*. A recent study conducted in rural villages in the Peruvian Amazon found that the *An. darlingi* larval habitats were significantly associated with water bodies in landscapes with more recent deforestation, lower light intensity, emergent vegetation and a lower vegetation index (154). The vegetation covering and surrounding these breeding sites could provide food for larvae, shelter from predators, and favourable oviposition conditions (157). The yearly average NDVI values in the 2-km side square grid around villages would not differentiate this vegetation as suggested by the low contribution of this variable to the malaria risk models, and the increased malaria incidence with NDVI values higher than 0.6.

Deforestation and environmental changes driven by human activity have been associated with *An. darlingi* breeding and malaria transmission (12); however, the different definitions of deforestation in these studies have prevented a clear conclusion (13). According to our BRT models, higher recent tree cover losses in a 2-km side square grid around villages (FL) were associated with higher malaria occurrence, however this forest-related variable was not among the top predictors for discriminating the malaria risk likely due to its limited variability across villages. The positive relationship between yearly deforestation and malaria risk is in line with

entomological studies showing that *An. darlingi* larvae were more likely to be found in water bodies with recent deforestation (12,154). Several studies in the Brazilian Amazon have found high densities of larval and adult *Anopheles* in forest fringes, as well as increased malaria morbidity in populations living or working near forest edges (62,158). A relationship between forest coverage (FC) and forest edges might explain why villages with lower FC have reduced malaria risk than those with higher FC. As deforested areas increase, the distance to forest edges also increases but malaria transmission remains high because of the quantity and extension of forest edges around villages. Reductions in FC would make forest edges more distant, thus reducing malaria risk unless residents need to engage in activities near forest edges. However, reduced FC can also indicate increased socio-economic development, which is associated with improvements in living conditions and malaria preventive practices (159).

Vector-human contact requires that mosquitoes fly from breeding sites and forest edges to access human blood meals (154). Therefore, contact rates and malaria transmission strongly depend on both the dispersal of *An. darlingi* and the population density (POPD) near breeding sites and forest edges. Our analysis highlighted the POPD as the second most important predictor for malaria risk. However, a positive relationship was only observed in the most densely populated villages. Further research is needed to confirm the negative relationship found in less densely populated villages. Nevertheless, the sparsely distributed population in the Peruvian Amazon may be at increased malaria risk because of precarious conditions, with limited access to health care, and exposure to mosquito bites due to subsistence farming and fishing, hunting and other activities near or within the forest (16,17). On the other hand, time to major populated villages (TPV) had a positive relationship with malaria risk since this variable can be a proxy for diminished access to health care facilities, reducing diagnosis and treatment, and sustaining malaria transmission, and hindering the delivery of malaria prevention interventions (19).

5.2. Conclusions

The Peruvian government is committed to eradicating malaria in the country by 2030 through the implementation of the Zero Malaria Plan (ZMP), which began nationwide interventions by the end of 2018. The success of this plan is supported by global efforts to reduce the global malaria burden and eradicate malaria from feasible countries, but it requires significant effort, innovation, and resources.

This thesis used spatial and machine learning tools to analyze and predict the occurrence of co-endemic Plasmodium vivax and Plasmodium falciparum malaria in villages in the Peruvian Amazon. The study analyzed environmental and social predictors using satellite imagery and products at the village level between 2010 and 2017. Yearly BRT models were built to discriminate the species-specific malaria risk in villages of a given year, and most of these models also performed well when predictor data was used to discriminate malaria risk in the following year. The study generated maps that identified villages with a high probability of having high malaria incidence and very high incidence in 2017.

Higher temperatures and rainfall positively affects the growth and development of the malaria parasite and mosquito vector. Partial dependence plots (PDPs) showed that satellite-derived rainfall (CAR) was the best predictor for both species *P. vivax* and *P. falciparum*, with increasing importance each year through 2015. Malaria risk is higher when the yearly average land surface temperature (LST) ranges between 26ºC and 29ºC, which is in agreement with the optimal temperatures for the development of *P. vivax* and *P. falciparum* cycles within the main malaria vector *An. darlingi.* However, due to the low variability of LST across villages, this variable was not among the most important factors for discriminating malaria risk between villages.

The proximity to breeding sites is recognized as an important determinant for mosquito exposure and malaria occurrence in the Peruvian Amazon. Severe flooding increases the number and enlarge existant breeding sites contributing to the dispersal of the vector. Normalized Difference Water Index (NDWI), a variable related to the presence of water bodies, was an important predictor in BRT models for both *P. vivax* and *P. falciparum*. Yearly average NDWI values around 0.4 suggest that a significant portion of the area surrounding villages with

high malaria incidence was covered by open water or wetlands. However, it was not able to characterize specific water bodies suitable for the development of *An. darlingi*, the main malaria vector.

Increased malaria incidence was observed with NDVI values higher than 0.6. An. darlingi larval habitats were found to be significantly associated with water bodies in landscapes with recent deforestation, lower light intensity, emergent vegetation, and a lower vegetation index (NDVI). These habitats provide food for larvae, shelter from predators, and favorable conditions for mosquito oviposition. Higher recent tree cover losses in a 2-km side square grid around villages were associated with higher malaria occurrence, but this forest-related variable was not among the top predictors for discriminating malaria risk likely due to its limited variability across villages. The positive relationship between yearly deforestation and malaria risk is consistent with entomological studies showing that An. darlingi larvae were more likely to be found in water bodies with recent deforestation.

Studies in the Brazilian Amazon have found high densities of larval and adult Anopheles in forest fringes, as well as increased malaria morbidity in populations living or working near forest edges. A relationship between forest coverage and forest edges might explain why villages with lower forest coverage have reduced malaria risk than those with higher forest coverage. Reductions in forest coverage would make forest edges more distant, thus reducing malaria risk unless residents need to engage in activities near forest edges. However, reduced forest coverage can also indicate increased socio-economic development, which is associated with improvements in living conditions and malaria preventive practices.

Population density (POPD) was found to be the second most important predictor for malaria risk, with a positive relationship observed only in the most densely populated villages. Time to major populated villages (TPV) had a positive relationship with malaria risk, as it can be a proxy for reduced access to healthcare facilities, hindering the delivery of malaria prevention interventions, and sustaining malaria transmission. The sparsely distributed population in the Peruvian Amazon may be at increased malaria risk due to limited access to healthcare, exposure to mosquito bites, and engagement in subsistence farming, fishing, hunting, and other activities near or within the forest.

About the contribution of this thesis

This study makes a significant contribution to the field of public health by presenting a novel method to improve our understanding of spatial and temporal malaria dynamics, identifying influential predictors and providing accurate risk assessment. This model accurately predicts the high-risk and very-high risk areas where intense malaria control activities should be concentrated. This work can be used as part of an early warning model.

The results presented in this thesis can support more tailored intervention plans to improve the distribution of Insecticide-Treated Net (ITN) or using Focused Screening and Treatment (FSAT) or Mass Drug Administration (MDA) strategies to interrupt transmission. MDA can potentially reduce malaria mortality and morbidity by inhibiting the sporogony cycle in the mosquito, reducing its vectorial capability. In countries like Peru that have limited public health resources, this model could help to direct those resources to the areas where the can be used most effectively.

As part of the contributions of this research, the results have been communicated to the scientific community through the article. https://www.nature.com/articles/s41598-019-51564-4

About the limitations of this thesis

First, the assumption of constant population size for villages across years could have reduced the discriminatory efficiency of the BRT models since they did not account for human migration, that is, people changing their residency from one place to another.

Secondly, NDWI and NDVI might not be the best proxies for environmental ground conditions that affect malaria transmission. NDWI and NDVI are yearly averages that might not capture the particular characteristics of seasonal variations that affect the breeding and resting sites of *An. darlingi*.

Thirdly, population density within mosquito dispersal ranges should include not only the population living near breeding sites or forest edges but also human mobility, that is, people

who temporarily enter these areas for economic activity. This information was not available at the village level.

Fourthly, data on the principal control interventions in each village during the study period was not available. These data would improve the discriminatory efficiency of the models and allow a better assessment of the differential contribution of predictor factors between species.

Fifthly, the resampling of predictor raster data to a higher resolution results in an output raster as precise as the coarsest inputs. This can add a systematic bias to models developed at high resolution. Higher resolution data for cumulative annual rainfall was not available for the study period. However, this variable has a low spatial variation in the Amazon region because of its flat topography, and the corresponding bias is expected to be limited.

About the perspectives of this thesis

This study highlights the need to consider the impact of human migration and mobility, which can change village population over time and affect malaria incidence and transmission. The integration of information about human migration and mobility would improve the accuracy of predictive models.

Future studies should account for seasonal variations in the characteristics of *An. darlingi* breeding and resting sites that are not captured by NDWI and NDVI annual averages.

Further investigations should include information about control intervention activities at the village level. This data would enhance the ability of models to distinguish between different factors and provide a more accurate assessment of their affects on different species.

The spatial resolution of satellite imagery has increased in recent years, and this will allow more accurate assessments of cumulative annual rainfall (CAR) in specific areas.

However, knowledge gaps remain, especially regarding the optimal size of the target population, methods to improve coverage, selection of drug-resistant parasites, and primaquine safety.

References

1. Alpaydm E, Aras N. Parametric distance functions vs . nonparametric neural networks for estimating road travel distances. 1996;2217(96).
2. Chuvieco E. Earth observation of global change. Vol. 53, Springer. 2013. 1–223 p.
3. Regional WHO. Terminology. 1998;
4. Xiaobai A. Yao. Georeferencing and Geocoding. Int Encycl Hum Geogr (Second Ed. 2020;1.
5. Rosas-Aguirre A, Gamboa D, Manrique P, Conn JE, Moreno M, Lescano AG, et al. Epidemiology of Plasmodium vivax malaria in Peru. Am J Trop Med Hyg. 2016;95(Suppl 6):133–44.
6. Soto-Calle V, Rosas-Aguirre A, Llanos-Cuentas A, Abatih E, Dedeken R, Rodriguez H, et al. Spatio-Temporal analysis of malaria incidence in the Peruvian Amazon Region between 2002 and 2013. Sci Rep. 2017;7.
7. Jamison DT, Nugent R, Gelband H, Horton S, Jha P, Laxminarayan R, et al. Major infectious diseases. Holmes KK, Bertozzi S, Bloom BR, Jha P, editors. World Bank Group; 2017. 300–340 p.
8. Hudson JE. Anopheles darlingi root (Diptera-Culicidae) in the Surinam rain forest.pdf. 1984;
9. Moreno M, Saavedra MP, Bickersmith SA, Lainhart W, Tong C, Alava F, et al. Implications for changes in Anopheles darlingi biting behaviour in three communities in the peri-Iquitos region of Amazonian Peru. Malar J. 2015;14(1):1–11.
10. Maheu-Giroux M, Casapía M, Soto-Calle VE, Ford LB, Buckeridge DL, Coomes OT, et al. Risk of malaria transmission from fish ponds in the Peruvian Amazon. Acta Trop. 2010 Jul;115(1–2):112–118.
11. Vittor AY, Gilman RH, Tielsch J, Glass G, Shields T, Lozano WS, et al. The effect of deforestation on the human-biting rate of Anopheles darlingi, the primary vector of Falciparum malaria in the Peruvian Amazon. Am J Trop Med Hyg. 2006 Jan;74(1):3–11.
12. Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, et al. Linking Deforestation to Malaria in the Amazon: Characterization of the Breeding Habitat of the Principal Malaria Vector, Anopheles darlingi. Am J Trop Med Hyg. 2009 Jul;81(1):5–12.
13. Tucker Lima JM, Vittor A, Rifai S, Valle D. Does deforestation promote or inhibit malaria transmission in the Amazon? A systematic literature review and critical appraisal of current evidence. Philos Trans R Soc B Biol Sci [Internet]. 2017;372(1722):20160125. Available from: http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2016.0125
14. CDC-MINSA/Peru. Análisis de la Situación de Salud del Perú. primera. Centro Nacional de Epidemiología, editor. Lima: CDC; 2017. 137 p.
15. S JC, Montenegro R, Castillo A, Che E, Muñoz A. Evolución de la malaria en la región Loreto. An la FMH-UNMSM. 2003;64(4):261–6.
16. Chuquiyauri R, Paredes M, Penataro P, Torres S, Marin S, Tenorio A, et al. Socio-demographics and the development of malaria elimination strategies in the low transmission setting. Acta Trop. 2012 Mar;121(3).
17. Rosas-Aguirre A, Guzman-Guzman M, Gamboa D, Chuquiyauri R, Ramirez R, Manrique P, et al. Micro-heterogeneity of malaria transmission in the Peruvian Amazon: A baseline assessment underlying a population-based cohort study. Malar J. 2017;16(1).

18. Casapía M, Vásquez LE, Rosas Á, Pinedo-Ríos N, Cabezas C, Chang J. Mejora en el diagnóstico y tratamiento oportuno de malaria con el uso de pruebas rápidas por promotores de salud en la Amazonía peruana. Rev Peru Med Exp Salud Publica. 2008 Oct;25(4):361–8.

19. Brierley CK, Suarez N, Arora G, Graham D. Healthcare Access and Health Beliefs of the Indigenous Peoples in Remote Amazonian Peru. Am J Trop Med Hyg. 2014 Jan;90(1):180–3.

20. Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nat Publ Gr [Internet]. 2018; Available from: http://dx.doi.org/10.1038/nature25181

21. Aramburú Guarda J, Ramal Asayag C, Witzig R. Malaria reemergence in the Peruvian Amazon Region. Emerg Infect Dis. 1999;5(2):209–15.

22. Grietens KP, Muela Ribera J, Soto V, Tenorio A, Hoibak S, Aguirre AR, et al. Traditional nets interfere with the uptake of long-lasting insecticidal nets in the Peruvian Amazon: the relevance of net preference for achieving high coverage and use. PLoS One. 2013;8(1):e50294.

23. Grietens KP, Soto V, Erhart A, Ribera ML, Toomer E, Tenorio A, et al. Adherence to 7-day Primaquine Treatment for the Radical Cure of *P*. vivax in the Amazon Region. AmJTropMed Hyg. 2015 Oct;82:1017–1023.

24. MINSA. Plan de implementacion del programa de eliminacion de la malaria cero en la region Loreto. 2017.

25. Torres K, Alava F, Soto-Calle V, Llanos-Cuentas A, Rodriguez H, Llacsahuanga L, et al. Malaria Situation in the Peruvian Amazon during the COVID-19 Pandemic. Am J Trop Med Hyg. 2020;103(5):1773–6.

26. Roshanravan B, Kari E, Gilman RH, Cabrera L, Lee E, Metcalfe J, et al. Endemic Malaria in the Peruvian Amazon Region of Iquitos. Am J Trop Med Hyg. 2003 Jul;69(1):45–52.

27. Branch O, Casapia WM, Gamboa D V., Hernandez JN, Alava FF, Roncal N, et al. Clustered local transmission and asymptomatic Plasmodium falciparum and Plasmodium vivax malaria infections in a recently emerged, hypoendemic Peruvian Amazon community. Malar J. 2005 Jun;4(1):27.

28. Solano-Villarreal E, Valdivia W, Pearcy M, Linard C, Pasapera-Gonzales J, Moreno-Gutierrez D, et al. Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. Sci Rep. 2019;9(1):1–12.

29. Carrasco-Escobar G, Gamboa D, Castro MC, Bangdiwala SI, Rodriguez H, Contreras-Mancilla J, et al. Micro-epidemiology and spatial heterogeneity of P. vivax parasitaemia in riverine communities of the Peruvian Amazon: A multilevel analysis. Sci Rep [Internet]. 2017;7(1):1–17. Available from: http://dx.doi.org/10.1038/s41598-017-07818-0

30. Miguel-Oteo M, Jiram AI, Ta-Tang TH, Lanza M, Hisam S, Rubio JM. Nested multiplex PCR for identification and detection of human Plasmodium species including Plasmodium knowlesi. Asian Pac J Trop Med. 2017;10(3):299–304.

31. World Health Organization. Global Technical Strategy for Malaria 2016-2030. Global Malaria Programme [Internet]. WHO, editor. WHO; 2015. 29 p. Available from: http://apps.who.int/iris/bitstream/handle/10665/176712/9789241564991_eng.pdf;jsessionid=66E6DA665C88369AF0BA3A99E8525283?sequence=1

32. Nilsson NJ. Introduction machine learning. Science (80-). 1996;1996–1996.

33. Bosque J. SIG y localización de instalaciones. 2013;

34. Bosque Sendra J, Palm Rojas FJ, Gómez Delgado M. LOCALIZA : una herramienta SIG para

resolver problemas de localización óptima . I Jornadas Sig Libr [Internet]. 2007;(1):16. Available from: http://www.sigte.udg.edu/jornadassiglibre2007/comun/2pdf/9.pdf

35. Daskin MS, Murray AT. Modeling Public Sector Facility Location Problems. Socioecon Plann Sci [Internet]. 2012;46(2):111-. Available from: http://www.sciencedirect.com/science/article/pii/S0038012112000171

36. Church RL. Geographical information systems and location science. 2002;29:541–62.

37. Weibel R, Heller M. Geographical Information Systems: Principles, Techniques, Management and Applications [Internet]. Geographical Information Systems: Principles and Applications. 1991. p. 269–97. Available from: http://www.wiley.com/legacy/wileychi/gis/volumes.html

38. Schintler LA, Kulkarni R, Gorman S, Stough R. Using Raster-Based GIS and Graph Theory to Analyze Complex Networks. 2007;301–13.

39. da Costa Gurgel H. A utilização das geotecnologias em estudos epidemiologicos: o exemplo da relação entre a malária e o NDVI em Roraima. An XI SBSR, Belo Horizonte, Bras 05 - 10 abril 2003, INPE. 2003;1303–10.

40. Flamm BJ. Suitable Housing Placement : A GIS-Based Approach. 2008;803–20.

41. Papinski D, Scott DM. A GIS-based toolkit for route choice analysis. J Transp Geogr [Internet]. 2011;19(3):434–42. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2010.09.009

42. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. A Review of Spatial Methods in Epidemiology , 2000 – 2010. 2012;

43. Nisanci R. GIS based fire analysis and production of fire-risk maps : The Trabzon experience. 2010;5(9):970–7.

44. Yang D-H, Goerge R, Mullner R. Comparing GIS-based methods of measuring spatial accessibility to health services. J Med Syst. 2006 Feb;30(1):23–32.

45. Bosque-Sendra JB, Jimenez AM. SIG y Localización óptima de Instalaciones y Equipamientos [Internet]. Ra-Ma; 2004. Available from: http://books.google.com.pe/books?id=aKSLAAAACAAJ

46. Bosque Sendra J, Gómez Delgado M, Moreno Jiménez A, dal Pozzo F. Hacia un sistema de ayuda a la decisión espacial para la localización de equipamientos. Estud Geogr. 2000;(241):567–98.

47. Bosque Sendra J. PLANIFICACIÓN Y GESTIÓN DEL TERRITORIO. De los SIG a los Sistemas de ayuda a la decisión espacial (SADE). El campo las ciencias y las artes [Internet]. 2001;138:137–74. Available from: https://geogra.uah.es/joaquin/pdf/SIG-y-SADE.pdf

48. Salonen M, Toivonen T, Cohalan J, Coomes OT. Critical distances : Comparing measures of spatial accessibility in the riverine landscapes of Peruvian Amazonia. Appl Geogr [Internet]. 2012;32(2):501–13. Available from: http://dx.doi.org/10.1016/j.apgeog.2011.06.017

49. Toivonen T, Mäki S. Accessibility of lowland rain forests of Western Amazonia via rivers - An analysis using Landsat TM mosaic and GIS. Proc SPIE - Int Soc Opt Eng [Internet]. 2002;4545:145–54. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-0036395573&partnerID=40&md5=7938bf47ec02bb2c34d250da8a8c9400

50. Vercellis C. Business Intelligence: Data Mining and Optimization for Decision Making. Business Intelligence: Data Mining and Optimization for Decision Making. 2009. 1–417 p.

51. Morgan-Forster AH. Climate, environment and malaria during the prehistory of mainland Greece. 2010;(September).

52. Rollinson D, Hay S. Advances in parasitology: Reflections on a Century of Malaria Biochemistry-Academic Press (2008).

53. Krishna R. Dronamraju, Paolo Arese. Malaria: Genetic and Evolutionary Aspects. 2005.

54. Pimenta PFP, Orfano AS, Bahia AC, Duarte APM, Ríos-Velásquez CM, Melo FF, et al. An overview of malaria transmission from the perspective of amazon anopheles vectors. Mem Inst Oswaldo Cruz. 2015;110(1):23–47.

55. Olson SH, Gangnon R, Elguero E, Durieux L, Guégan JF, Foley JA, et al. Links between climate, malaria, and wetlands in the amazon basin. Emerg Infect Dis. 2009;15(4):659–62.

56. Reinbold-Wasson DD, Sardelis MR, Jones JW, Watts DM, Fernandez R, Carbajal F, et al. Determinants of Anopheles seasonal distribution patterns across a forest to periurban gradient near Iquitos, Peru. Am J Trop Med Hyg. 2012 Mar;86(3):459–463.

57. Vargas J. Prevención y control de la malaria y otras enfermedades transmitidas por vectores en el Perú. Rev Peru Epidemiol. 2003;11(1):1–14.

58. Hiwat H, Bretas G. Ecology of Anopheles darlingi Root with respect to vector importance: A review. Parasites and Vectors [Internet]. 2011;4(1):177. Available from: http://www.parasitesandvectors.com/content/4/1/177

59. Moreno M, Saavedra MP, Bickersmith SA, Lainhart W, Tong C, Alava F, et al. Implications for changes in Anopheles darlingi biting behaviour in three communities in the peri-Iquitos region of Amazonian Peru. Malar J. 2015;14(1):1–11.

60. Manguin S, Wilkerson RC, Conn JANE, Rubio-palis Y, Danoff-burg JA, Roberts DR. Population structure of the primary malaria vector in South America, anopheles darlingi , using isozyme , random amplified polymorphic DNA, internal transcribed spacer 2 , and morphologic markers. 1999;60(3):364–76.

61. Charlwood J. Capture-recapture studies with the South American Malaria vector Anopheles darlingi , Root. 2016;(April).

62. Barros FSM, Honório NA. Deforestation and Malaria on the Amazon Frontier : Larval Clustering of Anopheles darlingi (Diptera : Culicidae) Determines Focal Distribution of Malaria. 2015;93(5):939–53.

63. Hiwat H, Bretas G. Ecology of Anopheles darlingi Root with respect to vector importance: A review. Parasites and Vectors. 2011;4(1):1–13.

64. Rufalco-moutinho P, Schweigmann N, Pimentel D, Anice M, Sallum M. Larval habitats of Anopheles species in a rural settlement on the malaria frontier of southwest Amazon , Brazil. Acta Trop [Internet]. 2016;164:243–58. Available from: http://dx.doi.org/10.1016/j.actatropica.2016.08.032

65. World Health Organization. Word Malaria Report 2021. Word Malaria report Geneva: World Health Organization. (2021). Licence: CC. 2021. 2013–2015 p.

66. WHO and Roll Back Malaria Partnership, Geneva. Malaria Control Achievements, Problems, and Strategies.". 1999.

67. Nabarro D. N. and E. M. Taylor. The Roll Back Malaria Campaign.

68. PAHO. Actualización Epidemiológica. Aumento de malaria en las Américas, noviembre 2019 [Internet]. 2018. Available from: https://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=43437&lang=es%0Ahttp://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=43437&lang=es

69. Frédéric Ariey, Frédérick Gay RM. Malaria Control and Elimination. 2013. 197 p.

70. Griffing SM, Gamboa D, Udhayakumar V. The history of 20th century malaria control in Peru. Malar J. 2013;12(1):1–7.

71. Durand S, Lachira-Alban A, Cabezas C. Impacto de diferentes esquemas terapéuticos sobre la malaria en la costa y amazonía peruana, en el marco de una política de medicamentos antimaláricos, 1994-2017. 2018;35(3):497–504.

72. Fernández R, Vera H, Calderón G. Revisión histórica de la distribución de Anopheles (Nyssorhynchus) darlingi (Diptera: Culicidae) en la Amazonía Peruana. Rev Peru Med Exp Salud Publica. 2014;31(2):310–8.

73. PAHO. Impacto del Niño sobre la salud de la población 1998 - Crónicas de Desastres - Fenómeno El Niño [Internet]. 2000. 288 p. Available from: http://helid.digicollection.org/en/d/Jwho70s/3.3.4.2.html

74. Soto Tarazona A, Solari Zerpa L, Mendoza Requena D, Llanos-Cuentas A, Magill A. Evaluation of the rapid diagnostic test OptiMAL for diagnosis of malaria due to Plasmodium vivax. Braz J Infect Dis. 2004;8(2):151–5.

75. Cossio MLT, Giesen LF, Araya G, Pérez-Cotapos MLS, Vergara RL, Manca M, et al. Norma técnica para la atencion curativa de la malaria esquemas terapéuticos en el tratamiento de la malaria no complicada y malaria grave en el Peru [Internet]. Vol. XXXIII, Uma ética para quantos? 2012. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15003161%5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991%5Cnhttp://www.scielo.cl/pdf/udecada/v15n26/art06.pdf%5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-84861150233&partnerID=tZOtx3y1

76. Hussein MIH, Albashir AAD, Elawad OAMA, Homeida A. Malaria and COVID-19: unmasking their ties. Malar J [Internet]. 2020;19(1):1–10. Available from: https://doi.org/10.1186/s12936-020-03541-w

77. MINSA. Resolución Ministerial N° 244-2017-MINSA. 2017 p. 12–3.

78. Ministry of Health Peru. Analisis de Situacion de Salud del departamento Loreto. Cent Nac Epidemiol Prevención y Control Enfermedades [Internet]. 2015;8. Available from: http://www.dge.gob.pe/portal/Asis/indreg/asis_loreto.pdf

79. Lillesand TM, Kiefer RW. Remote sensing and image interpretation. Remote sensing and image interpretation. 1979. 1–190 p.

80. Emmanuel NN, Loha N, Okolo MO, Ikenna OK. Landscape epidemiology: An emerging perspective in the mapping and modelling of disease and disease risk factors. Asian Pacific J Trop Dis. 2011;1(3):247–50.

81. Jensen JR. Remote sensing of the environment: an earth resource perspective. Vol. 1, Prentice Hall, Upper Saddle River, NJ. 2000. 333–378 p.

82. Thenkabail PS. Remote sensing handbook: Remote sensing of water resources, disasters, and urban studies. Vol. 3, Remote Sensing of Water Resources, Disasters, and Urban Studies. 2015. 289–310 p.

83. Sensing) C (Centre for R. Fundamentals of Remote Sensing [Internet]. Forestry. 1999. Available from: http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/index_e.php

84. Maître H. Processing of Synthetic Aperture Radar Images. 2008. 23–36 p.

85. (MTC) M de T y comunicaciones. PERUSAT-1 Y CLASIFICACIÓN DE LA CARRETERA CENTRAL. Vol. 000. 2017.

86. Eastman J. Multi-criteria evaluation and GIS. Geogr Inf Syst [Internet]. 1999;493–502. Available from: http://www.geos.ed.ac.uk/%7B~%7Dgisteac/gis_book_abridged/files/ch35.pdf

87. Daash A, Srivastava A, Nagpal BN, Saxena R, Gupta SK. Geographical information system (GIS) in decision support to control malaria--a case study of Koraput district in Orissa, India. J Vector Borne Dis. 2009 Mar;46(1):72–4.

88. Buzai GD, Baxendale CA. Modelos de localización-asignación aplicados a servicios públicos urbanos: Análisis espacial de escuelas EGB en la ciudad de Luján. Rev Univ Geogr. 2008;17:233–54.

89. Perkel JM. Map-making on a budget toolbox. Nature. 2018;558(7708):147–8.

90. Brown T, Mclafferty S, Moon G. A Companion to Health and Medical Geography. 2010. 96–108 p.

91. Brody H, Rip MR, Vinten-johansen P, Paneth N, Rachman S. Map-making and myth-making in Broad Street : the London cholera epidemic , 1854. Lancet. 2000;356:64–8.

92. Beck LR, Lobitz BM, Wood BL. Remote sensing and human health: New sensors and new opportunities. Emerg Infect Dis. 2000;6(3):217–27.

93. Loh W. Classification and regression trees. WIREs Data Min Knowl Discov. 2011;1(February):14–23.

94. Leo Breiman, Jerome H. Friedman, Richard A. Olshen CJS. Clasification and regression trees. 1998.

95. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802–13.

96. Elith J, Leathwick JR, Hastie T, R. Leathwick J. Elith, Leathwick & Hastie A working guide to boosted regression trees - Online Appendices Page 1. J Anim Ecol [Internet]. 2008;77(4):802–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18397250

97. Ahmed A. GIS and remote sensing for malaria risk mapping, Ethiopia. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2014;40(8):155–61.

98. Albert DP, Gesler WM, Levergood B. Spatial Analysis, GIS and Remote Sensing: Applications in the Health Sciences. Sleeping bear press; 2000.

99. Zhao Q, Hastie T. Causal Interpretations of Black-Box Models. J Bus Econ Stat. 2019;

100. INEI. Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas. 2017;43. Available from: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1437/libro.pdf

101. CDC-MINSA/Peru. Boletín epidemiológico del perú. 2017.

102. INEI. Ficha técnica empadronamiento distrital de población y vivienda 2013 Empadronamiento Distrital de Población y Vivienda 2013. Vol. 2013. 2013. p. 3–4.

103. Tanser F. innova t ions (GIS) Innovations For Primary Health Care in Developing Countries. 2006;106–22.

104. Eastman J. Multi-criteria evaluation and GIS. Geogr Inf Syst [Internet]. 1999;493–502. Available from: http://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch35.pdf

105. Omumbo JA, Hay SI, Snow RW, Tatem AJ, Rogers DJ. Modelling malaria risk in East Africa at high-spatial resolution. Trop Med Int Heal. 2005;10(6):557–66.

106. Hay SI, Snow RW, Rogers DJ. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Trans R Soc Trop Med Hyg. 1998;92(1):12–20.

107. Patz JA, Strzepek K, Lele S, Hedden M, Greene S, Noden B, et al. Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. Trop Med Int Heal. 1998;3(10):818–27.

108. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Remote Sensing of Environment Google Earth Engine : Planetary-scale geospatial analysis for everyone. Remote Sens Environ [Internet]. 2017;202:18–27. Available from: https://doi.org/10.1016/j.rse.2017.06.031

109. Hantson S, Chuvieco E, Pons X, Domingo C, Cea C, Moré G, et al. Cadena de pre-procesamiento estándar para las imágenes Landsat del Plan Nacional de Teledetección. Rev Teledetección. 2011;36:51–61.

110. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. 2015;1–22.

111. Pekel JF, Cottam A, Gorelick N, Belward AS. High-resolution mapping of global surface water and its long-term changes. Nature. 2016;540(7633):418–22.

112. Bruy A, Svidzinska D. QGIS by example. PACKT; 2015.

113. Menke K, Davis P, Ebooks Corporation. Mastering QGIS : go beyond the basics and unleash the full power of QGIS with practical, step-by-step examples [Internet]. Community experience distilled. 1 online resource. Available from: http://libaccess.mcmaster.ca/login?url=http://www.MCMU.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_2005217_0

114. Hansen MC, Potapov P V, Moore R, Hancher M, Turubanova SA, Tyukavina A. High-Resolution Global Maps of of 21st-Century Forest Cover Change. 2013;134(November):850–4.

115. Simpson J, Adler RF, North GR. A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite. Bull Am Meteorol Soc [Internet]. 1988;69(3):278–95. Available from: http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%281988%29069%3C0278%3AAPTRMM%3E2.0.CO%3B2

116. Cuéllar AC. Uso De Sensores Remotos Para La Predicción De Casos De Malaria En El Departamento Orán, Salta, Argentina. 2014;95.

117. Szabó S, Gácsi Z, Balázs B. Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. Landsc Environ [Internet]. 2016;10(3–4):194–202. Available from: http://landscape.geo.klte.hu/pdf/agd/2016/2016v10is3_4_13.pdf

118. Jensen JR. Remote sensing of the environment: an earth resource perspective. Vol. 1, Prentice Hall, Upper Saddle River, NJ. 2000. 333–378 p.

119. Myneni RB, Ganapol BD, Asrar G. Remote sensing of vegetation canopy photosynthetic and stomatal conductance efficiencies. Remote Sens Environ. 1992;42(3):217–38.

120. da Costa Gurgel H. A utilização das geotecnologias em estudos epidemiologicos: o exemplo da relação entre a malária e o NDVI em Roraima. An XI SBSR, Belo Horizonte, Bras 05 - 10 abril 2003, INPE. 2003;1303–10.

121. Tourre YM, Jarlan L, Lacaux J-P, Rotela CH, Lafaye M. Spatio-temporal variability of NDVI–precipitation over southernmost South America: possible linkages between climate signals and epidemics. Environ Res Lett. 2008;3:044008.

122. Goward SN, Markham B, Dye DG, Dulaney W, Yang J. Normalized difference vegetation

index measurements from the advanced very high resolution radiometer. Remote Sens Environ. 1991;35(2–3):257–77.

123. Gillespie A. Land Surface Temperature. Encycl Remote Sens [Internet]. 2014;314–9. Available from: http://link.springer.com/10.1007/978-0-387-36699-9_79

124. Saha SK. Retrieval of Agrometeorological parameters using satellite data. In: Satellite Remote Sensing and GIS Applications in Meteorology [Internet]. 2004. p. 151–74. Available from: http://www.bishensinghbooks.com

125. Zeilhofer P, Soares E, Ribeiro ALM, Miyazaki RD, Atanaka M. Habitat suitability mapping of Anopheles darlingi in the surroundings of the Manso hydropower plant reservoir , Mato Grosso , Central. 2007;14:1–14.

126. Gareth James, Daniela Witten, Trevor Hastie RT. An Introduction to Statistical Learning with Applications in R.

127. Elith A. Boosted Regression Trees in R. Biometrics [Internet]. 2008;2008(January):1–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17688511

128. Greenwell B, Boehmke B, Cunningham J, GBM Developers. Generalized Boosted Regression Models. 2018; Available from: https://github.com/gbm-developers/gbm%0D

129. Hijmans RJ, Phillips S, Leathwick JR, Elith J. Dismo package for R, version 1.1-4. 2017; Available from: http://cran.r-project.org/package=dismo (

130. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802–13.

131. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, et al. Predictive analysis across spatial scales links zoonotic malaria to deforestation. 2019;

132. Artois J, Jiang H, Wang X, Qin Y, Pearcy M, Lai S, et al. Changing geographic patterns and risk factors for avian influenza A(H7N9) infections in humans, China. Emerg Infect Dis. 2018;24(1):87–94.

133. Greenwell BM. pdp : An R Package for Constructing Partial Dependence Plots. 2017;9(June):421–36.

134. WHO. World Malaria Report 2018 [Internet]. 2018. 1–146 p. Available from: https://www.who.int/malaria/publications/world-malaria-report-2018/en/

135. Moss WJ, Hamapumbu H, Kobayashi T, Shields T, Kamanga A, Clennon J, et al. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community survey. Malar J [Internet]. 2011;10(1):163. Available from: http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-10-163

136. van der Hoek W, Konradsen F, Amerasinghe PH, Perera D, Piyaratne MK, Amerasinghe FP. Towards a risk map of malaria for Sri Lanka: The importance of house location relative to vector breeding sites. Int J Epidemiol. 2003;32(2):280–5.

137. Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, et al. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. PLoS Pathog. 2011;7(3).

138. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. Elife. 2016;5:1–19.

139. Ashby J, Moreno-Madriñán M, Yiannoutsos C, Stanforth A. Niche Modeling of Dengue Fever Using Remotely Sensed Environmental Factors and Boosted Regression Trees. Remote Sens [Internet]. 2017;9(4):328. Available from: http://www.mdpi.com/2072-4292/9/4/328

140. Pigott DM, Millear AI, Earl L, Morozoff C, Han BA, Shearer FM, et al. Updates to the zoonotic niche map of Ebola virus disease in Africa. Elife. 2016;5(2016JULY):1–13.

141. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in Plasmodium falciparum transmission: Precision, accuracy and costs of metrics. Adv Parasitol. 2016;1–48.

142. Roll Back Malaria &, Partnership. Framework for Monitoring Progress and Evaluating Outcomes and Impact. Geneva: RBM; 2000.

143. Cohen JM, Dlamini S, Novotny JM, Kandula D, Kunene S, Tatem AJ. Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. 2013;1–12.

144. Gosling RD, Okell L, Mosha J, Chandramohan D. The role of antimalarial treatment in the elimination of malaria. Clin Microbiol Infect [Internet]. 2011;17(11):1617–23. Available from: http://dx.doi.org/10.1111/j.1469-0691.2011.03660.x

145. Maude RJ, Socheat D, Nguon C, Saroth P, Dara P, Li G, et al. Optimising strategies for Plasmodium falciparum Malaria elimination in Cambodia: Primaquine, mass drug administration and Artemisinin resistance. PLoS One. 2012;7(5).

146. Sturrock HJW, Bennett AF, Midekisa A, Gosling RD, Gething PW, Greenhouse B. Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. Trends Parasitol [Internet]. 2016;32(8):635–45. Available from: http://dx.doi.org/10.1016/j.pt.2016.05.001

147. White NJ. Determinants of relapse periodicity in Plasmodium vivax malaria Determinants of relapse periodicity in Plasmodium vivax malaria. 2011;297(October).

148. Rosas-Aguirre A, Gamboa D, Manrique P, Conn JE, Moreno M, Lescano AG, et al. Epidemiology of Plasmodium vivax malaria in Peru. Am J Trop Med Hyg. 2016;95(November):133–44.

149. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, et al. Global epidemiology of Plasmodium vivax. Am J Trop Med Hyg. 2016;95(Suppl 6):15–34.

150. COEN-INDECI. Inundación afecta a la provincia de Putumayo - Loreto. 2016. p. 1–10.

151. PAHO. Inundaciones en Loreto: Respuesta del Sector de Agua, Saneamiento e Higiene - Experiencias y aprendizajes. PAHO. 2013.

152. Reach-Initiative. 2012 Flood Events, PERU. 2012.

153. SEHINAV. Boletín de avisos a los navegantes fluviales - SERVICIO DE HIDROGRAFÍA Y NAVEGACIÓN DE LA AMAZONÍA. 2015.

154. Prussing C, Saavedra MP, Bickersmith SA, Alava F, Guzmán M, Manrique E, et al. Malaria vector species in Amazonian Peru co-occur in larval habitats but have distinct larval microbial communities. PLoS Negl Trop Dis. 2019;13(5):e0007412.

155. Rosas-Aguirre A, Speybroeck N, Llanos-Cuentas A, Rosanas-Urgell A, Carrasco-Escobar G, Rodriguez H, et al. Hotspots of malaria transmission in the Peruvian amazon: Rapid assessment through a parasitological and serological survey. PLoS One. 2015;10(9):1–21.

156. Guo M, Li J, Sheng C, Xu J, Wu L. A Review of Wetland Remote Sensing. 2017;1–36.

157. Rejmánková E, Grieco J, Achee N RD. Ecology of Larval Habitats. In: Manguin S, editor. Anopheles mosquitoes—New insights into malaria vectors. 397–446 p.

158. Arruda ME, Gurgel H. Spatial clustering and longitudinal variation of Anopheles darlingi (Diptera : Culicidae) larvae in a river of the Amazon : the importance of the forest fringe and of obstructions to flow in frontier malaria. 2011;(April 2014).

159. Id AD, Fennie K, Degarege D, Id SC, Madhivanan P. Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa : A systematic review and meta-analysis. 2019;1–26.

160. Bloland PB, Williams HA. Malaria control during mass population movements and natural disasters. 2003.

Appendix A: Article *Sci Rep* 9, 15173 (2019)

www.nature.com/scientificreports

SCIENTIFIC REPORTS
nature research

OPEN

Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon

Elisa Solano-Villarreal[1,2,3*], Walter Valdivia[4], Morgan Pearcy[2], Catherine Linard[5,6], José Pasapera-Gonzales[7], Diamantina Moreno-Gutierrez[2,8,9], Philippe Lejeune[1], Alejandro Llanos-Cuentas[3], Niko Speybroeck[2], Marie-Pierre Hayette[1] & Angel Rosas-Aguirre[2,3,10]

This is the first study to assess the risk of co-endemic *Plasmodium vivax* and *Plasmodium falciparum* transmission in the Peruvian Amazon using boosted regression tree (BRT) models based on social and environmental predictors derived from satellite imagery and data. Yearly cross-validated BRT models were created to discriminate high-risk (annual parasite index API > 10 cases/1000 people) and very-high-risk for malaria (API > 50 cases/1000 people) in 2766 georeferenced villages of Loreto department, between 2010–2017 as other parts in the article (graphs, tables, and texts). Predictors were cumulative annual rainfall, forest coverage, annual forest loss, annual mean land surface temperature, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), shortest distance to rivers, time to populated villages, and population density. BRT models built with predictor data of a given year efficiently discriminated the malaria risk for that year in villages (area under the ROC curve (AUC) > 0.80), and most models also effectively predicted malaria risk in the following year. Cumulative rainfall, population density and time to populated villages were consistently the top three predictors for both *P. vivax* and *P. falciparum* incidence. Maps created using the BRT models characterize the spatial distribution of the malaria incidence in Loreto and should contribute to malaria-related decision making in the area.

In spite of the investment in control and prevention allocated by the Peruvian government over the last decades, malaria due to both *Plasmodium vivax* and *P. falciparum* remains a significant public health issue in the country. In 2015, 61,856 malaria cases were reported[1], representing ~15% of the total reported cases in the Americas and showing a continuous increase since 2012[2,3]. The Amazonian department of Loreto is the most affected area, accounting for 95% of national cases despite being home of only 3.5% of the Peruvian population[2].

Loreto is a hypoendemic malaria area[4,5], but malaria transmission in the department is highly heterogeneous, with some areas in remote Amazonia having entomological inoculation rates (EIRs) near those reported in Africa. This heterogeneity creates opportunities for targeted interventions, provided high-risk hotspots are first identified[3,6]. Since the 1960s, malaria risk maps have been essential for planning interventions[7–9]. Recently, electronic geographic information systems and remote sensing have improved these maps by providing an array of environmental and social data collected through passive and active satellite sensors[10–13].

[1]Université de Liège, 4000, Liège, Belgium. [2]Research Institute of Health and Society (IRSS), Université catholique de Louvain, 1200, Brussels, Belgium. [3]Institute of Tropical Medicine Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, 15102, Peru. [4]Ministry of Development and Social Inclusion, Lima, 15047, Peru. [5]Namur Research Institute for Life Sciences (Narilis), Université de Namur, 5000, Namur, Belgium. [6]Institute of Life-Earth-Environment (ILEE), 5000, Namur, Belgium. [7]National Aerospace Development Commission, Lima, 15046, Peru. [8]University of Antwerp, 2000, Antwerp, Belgium. [9]Faculty of Human Medicine, Universidad Nacional de la Amazonía Peruana, Loreto, 160, Peru. [10]Fonds de la Recherche Scientifique (FNRS), 1000, Brussels, Belgium. *email: elitayoan@gmail.com

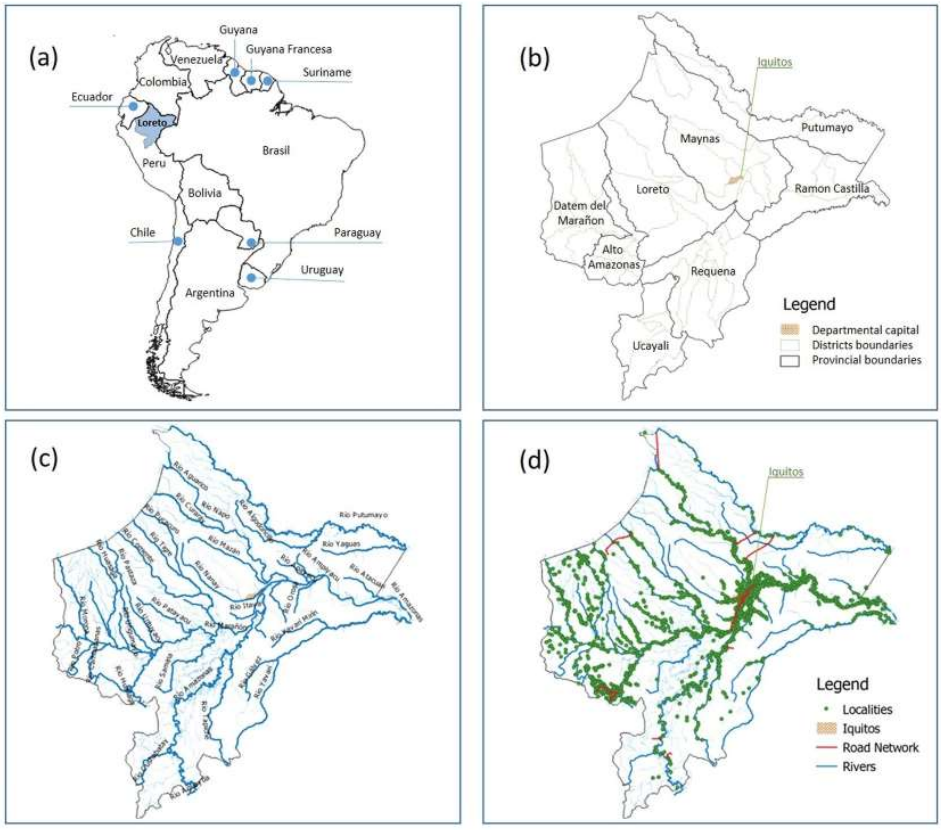


Figure 1. Study area: (**a**) geographical location of Loreto in South America; (**b**) administrative division of Loreto: department, provinces, and districts; (**c**) hydrographic map of Loreto; (**d**) road network, rivers, and georeferenced villages.

The main factors responsible for the variation in malaria transmission in the Amazon are environmental characteristics facilitating larvae breeding sites (temperature, precipitation, natural and human-made water bodies) and mosquito resting places (surrounding vegetation, deforestation) for the primary malaria vector of *P. vivax* and *P. falciparum*, *Anopheles darlingi*[14–17]. Social factors also influence malaria transmission by increasing exposure (forest-based economic activities)[18,19], by contributing to delayed diagnosis and treatment (poor geographical accessibility to health facilities)[20–22], or by limiting effectiveness of malaria interventions (lower availability and utilization of control resources and preventive measures, and inappropriate treatment-seeking behaviors)[23–25].

Satellite data[26,27] were used to derive environmental and social potential predictors[28–30] of malaria risk in villages of Loreto during the period 2010–2017. The association of these predictors with malaria occurrence in the villages was analyzed using boosted regression trees (BRTs), a method that accounts for nonlinearities and interactions of variables with high predictive accuracy[31–33], and allows for disease mapping[32,34–36]. Yearly analyses were performed to determine the relevance of predictors at local scale across the years. Maps identifying villages at highest risk were created to support the malaria decision making in Loreto.

Methods

Study area. Loreto, located in the northeast part of Peru between 61 to 220 meters above sea level, is the largest department in Peru (368,851.95 km2, 28.7% of the national territory). A total of 883,510 inhabitants live in the department, with ~40% of them aged below 15 years, and ~one third residing in rural areas[37]. Loreto has a tropical climate, with a rainy season from November to May and a dry season from June to October. Precipitation is present throughout the year (cumulative average: 2,500 mm) with a peak in March (360 mm) and a minimum in July (50–100 mm). Rural villages mainly depend on natural resource exploitation, while commercial activities are more important in cities and large villages. Loreto has an extensive fluvial network connected to the Amazon River (Fig. 1), and boats are the primary mode of transportation.

Population and malaria incidence. Of the 2,843 villages included in the official cartography of Loreto, we validated and georeferenced 2,766 villages in Loreto as points in QGIS[38,39]. Non-validation was primarily due to duplicated names and/or duplicated coordinates. The population of villages, obtained from the exceptional National Census of 2012–2013[40], was assumed to be uniform during the study period.

Variable description					Source information				
Type	Variable	Descriptions	Units	Time-dependent variable	Data collection	Source	Spatial resolution	Temporal resolution	Units
Environment predictor variable	FC	Forest cover in a 2-km side square grid around village	%	yes (year)	UMD/hansen/global_forest_change_2015	Univ. Maryland	30 m	year	% per output grid cell
	FL	Annual forest loss in a 2-km side square grid around village	%	yes (year)	UMD/hansen/global_forest_change_2015	Univ. Maryland	30 m	year	% per output grid cell
	CAR	Cumulative annual rainfall. (average in a 2-km side square grid around village)	mm	yes (year)	TRMM/3B42	TRMM	~27 km	3hrs	mm/3hrs x pp
	LST	Land surface annual mean temperature	°C	yes (year)	MODIS/006/MOD11A1/LST_Day_1 km	NASA, MODIS, LST	1km	1 day	°C
	NDVI	Normalized difference vegetation index. (average in a 2-km side square grid around village)	index	yes (year)	MODIS/006/MOD13Q1	NASA, MODIS, Vegetation Index	250 m	8 days	index
	NDWI	Normalized difference water index. (average in a 2-km side square gride around village)	index	yes (year)	LANDSAT/LC05C01/T1_T1 LANDSAT/LC08/C01/T1_T1 LANDSAT/LE07/C01/T1_T1	NASA, LANDSAT	30 m	16 days	index
	SDR	Euclidean shortest distance to rivers	kilometers	no	JRC/GSW1_0/GlobalSurfaceWater (occurrence)	JRC/GSW Historical data	30 m	once	kilometers
Social predictor variable	TPV	Travel time to major populated villages/towns	minutes	no	Oxford/MAP/accessibility_to_cities_2015_v1_0	Oxford (MAP), Google, (JRC) & Univ. Twente	1Km	once	minutes
	POPD	Population in a 5-km side square grid around village)	log (number people)	no	WorldPop/POP	WorldPop 2015	~100 m	once	people in ~100 × 100 m grid cell
Outcome variable	Malaria high-risk	Village with API** > 10 cases/1000 people	binary (yes/no)	yes (year)	—	Surveillance system of Peruvian ministry of health	—	week	reported cases
	Malaria very-high-risk	Village with API > 50 cases/1000 people	binary (yes/no)						

Table 1. Predictor and outcome variables used in BRT models. **API = Confirmed cases in a year *1000 / total population. *PP: Per-pixel.

Malaria is a mandatory notifiable disease in Peru. Malaria cases are confirmed by light microscopy in health facilities and attributed to the apparent place of infection using a village's alphanumeric code, then reported to the surveillance system of Ministry of Health (MoH)[41]. Since malaria has been associated with travel- and forest-related activities in Loreto, the place of infection provides a more accurate information than the place of residence when modelling malaria transmission. For the purpose of our study, all reported cases in Loreto between 2010 and 2017 were aggregated by village and by year. Yearly overall and species-specific indices (API) were calculated for each village using the formula: (number of confirmed cases/village population size) x 1,000. Villages were classified as being or not at high-risk (i.e. API > 10 cases/1,000 people) or very-high-risk (API > 50 cases/1,000 people) based on MoH thresholds.

Social and environmental predictors. Social and environmental variables previously associated with malaria transmission[7,42,43] were tested as factors for high-risk (IPA > 10) or very-high-risk (IPA > 50) (Table 1). These variables were derived from different satellite imagery and mixed products. JavaScript codes for Google Earth Engine (GEE)[44] were used to download and process this images at Datum WGS84 zone 18 (EPSG 4326). Below is a brief description of the variables:

Time to major populated villages/towns (TPV), social variable indicating the land-based travel time in minutes to the nearest densely-populated area for the year 2015. Underlying datasets included roads, railways, rivers, lakes, oceans, slope and elevation, land cover types, and national borders[22].

Population density (POPD), social variable representing the estimated number of residents in each 100mx100m grid-cell in 2015, extracted from WorldPop (https://www.worldpop.org)[29].

Shortest distance to rivers (SDR), environmental variable estimated using the Global Surface Water (occurrence) map[45] from JRC (Joint Research Centre). A 50% threshold mask was applied in GEE to select pixels with a presence of water at least half of the period 1984–2015, capturing main and secondary rivers. The shortest distance in (meters) between rivers and villages was calculated using the proximity algorithm QGIS v.3.4.2[46,47].

Forest coverage (FC) and *forest loss (FL)*, time-dependent measures of the area covered by trees (%) in each year and the loss of tree coverage compared to the previous year (%), calculated from 2000–2018 Global Forest Change data (https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.6.html)[48].

Cumulative annual rainfall (CAR), the estimated cumulative yearly rainfall (mm/year) calculated from daily 3-hour infrared precipitation estimates in product 3B42 of the Tropical Rainfall Measuring Mission (TRMM)[49].

Normalized Difference Water Index (NDWI), a satellite-derived index from the Near-Infrared (ρNIR) and Short Wave Infrared (ρSWIR) channels that estimates the amount of water in internal leaf structure[50,51]. This index was processed from Landsat 5, 7 and 8 T1 (calibrated TOA, DOS and clouds masking), and calculated as NDWI = (ρNIR-ρSWIR)/(ρNIR + ρSWIR)[52,53].

Normalized Difference Vegetation Index (NDVI)), the estimated fraction of radiation absorbed by the vegetation in the red (ρRED) and the near infrared (ρNIR) channels[54], from the MODIS/006/MOD13Q1 Moderate Resolution Imaging Spectroradiometer - National Aeronautics and Space Administration (NASA). The index was calculated at a spatial resolution of 250 m. as: NDVI = (ρNIR-ρRED)/(ρNIR + ρRED)[52,55–57].

Land Surface Temperature (LST), an estimate of the infrared emissivity of the earth in degrees Celsius (Emissivity Daily L3 Global, MOD11A1 version 6)[58]. Mean values from 365 daily layers were used to produce raster images with yearly means.

Extraction of values. Raster images were re-sampled to 90-meter-pixels using R software version 3.5.1. The focal function generates an output raster based on the neighborhood information[59] using "moving windows" around pinpointed villages. It was used here to capture a 2 km-side square grid for FC, FL, CAR, NDVI, NDWI and LST, and a 5 km-side square grid for POPD. The 2 km-side threshold was chosen to cover breeding and resting sites within the flight distance for *An. darlingi* mosquitoes[60,61], while a 5 km-side threshold capturing the population within and around the village. Raster values in square grids were aggregated as averages for environmental variables and as sums for POPD. For TPV and SDR, the extracted values were at village location. Corresponding R codes are presented in Supplementary Text S1.

Boosted regression tree analysis. BRT models[62,63] were created using R packages "gbm"[64,65] and "dismo"[31] to examine the relationship between potential predictors and the species-specific malaria risk status in villages for each study year. The binary data (i.e. villages with or without high-risk of malaria; or, villages with or without very-high-risk of malaria) was fitted using a Bernouilli response distribution. The optimal learning rate, bag. fraction and tree complexity were determined by testing several combinations with the lowest model deviance. Ten-fold cross-validation procedures using the "gbm.step" function enabled the selection of the optimal number of regression trees for each model, ensuring the same villages in cross-validation groups across models with the parameter "fold.vector".

The relative contribution (RC) percentage was used to assess the relevance of each variable in BRT models. This metric measures how often the predictor is selected for partitioning, weighted by the squared model improvement resulting from successive partitions[31,66]. Partial dependence plots were generated to describe the effect of one predictor on the malaria risk status of the villages, after accounting for the average effects of all other predictors[67]. In these plots, the vertical axis is the logit of cross-validated predicted probability (logit (p)) for high-risk or very-high-risk of malaria, and the horizontal axis is the variable predictor with corresponding units.

The area under the curve (AUC) of the receiver operating characteristic (ROC) curve assessed the performance of BRT models for discriminating the malaria risk status in villages. Each cross-validation BRT model built with data of a given year yielded a cross-validated AUC (cvAUC), while its model predictions (i.e. predicted probabilities for villages at malaria risk) with testing data of the following year allowed for the estimation of a testing AUC (tAUC).

Risk maps elaboration. BRT models built with 2016 data were tested with 2017 data, obtaining predicted probabilities for villages at high-risk (API > 10) and villages at very-high-risk (API > 50). These predicted probabilities were then used to classify villages into four categories (i.e. 0–0.25, 0.251–0.50, 0.511–0.75, 0.751–1.00), and to create species-specific risk maps for 2017 using QGIS v.3.4.2-Madeira.

Ethical considerations. This study is part of the research project "Spatial tools and optimization techniques to support case detection strategies for the malaria control in the Peruvian Amazon" approved by Universidad Peruana Cayetano Heredia, SIDISI number 101052.

Results

Malaria occurrence. Of the 321,210 malaria cases reported to the epidemiological surveillance system in Loreto between 2010 and 2017, 311,128 cases (96.9%) occurred among our validated and georeferenced 2,766 villages. Non-pinpointed cases (3.1%) could correspond to infections in transient populations and/or inaccurate records of the place of the infection.

Malaria steadily increased in study villages from 10,994 cases in 2011 to 59,257 cases in 2014 and 58,679 in 2015, after which cases slightly declined to 51,663 in 2017 (Fig. 2). *P. vivax* cases always predominated over *P. falciparum* cases with the maximum ratio in 2012 (Pv/Pf 5.5) and the minimum ratio in 2016 (Pv/Pf 2.6). The highest peak occurred earlier for *P. vivax* (2014) than for *P. falciparum* (2016). The number of villages at high-risk (API > 10) and at very-high-risk (API > 50) followed similar trends.

Relative contribution (RC) of variables. Mean (M), standard deviation (SD), median (Mdn), and interquartile range (IQR) values of predictor variables are presented in Supplementary Table S1. The RC of predictors obtained from yearly cross-validated BRT models are shown in Supplementary Tables S2, S3 and S4 and Fig. S1. Most RCs for overall malaria risk were similar to ones for *P. vivax* each year. CAR, POPD, and TPV, in that order, were always the top three predictors for increasing malaria risk in villages (overall and by species), with yearly RC medians exceeding 10% (Fig. 3). NDWI was the fourth top predictor, but yearly RC medians only exceeded 10% when BRT models assessed villages at *P. vivax* very-high-risk of and villages at *P. falciparum* high-risk.

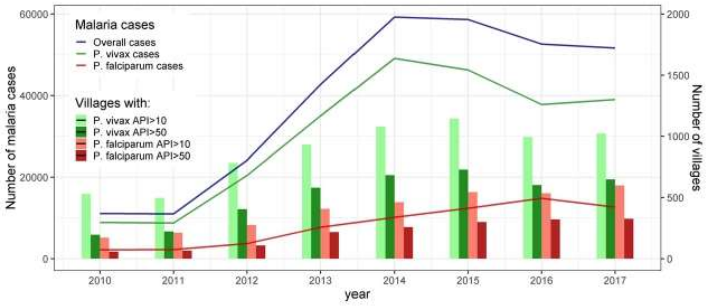


Figure 2. Reported malaria cases and number of villages at risk in Loreto from 2010 to 2017.

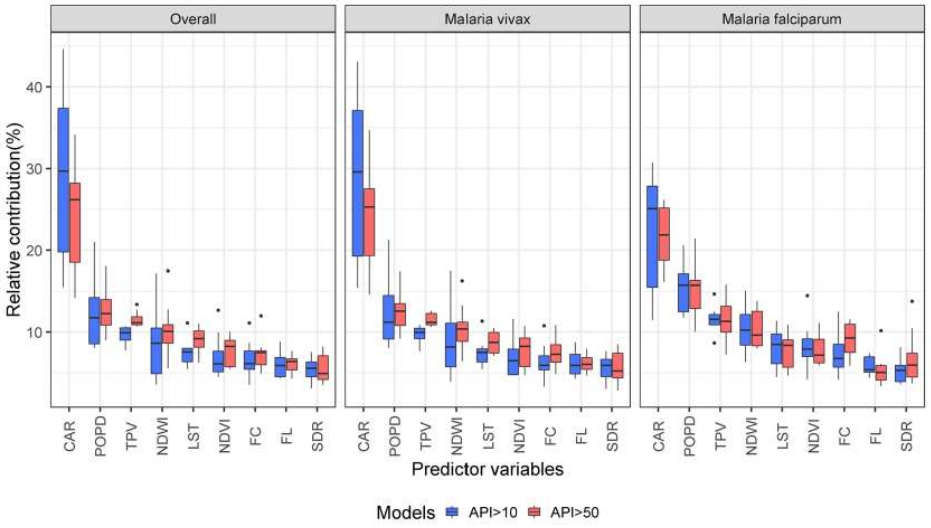


Figure 3. Relative contributions of predictors obtained from yearly BRT models for malaria risk, overall and by species, over the study period (2010–2017).

CAR was generally the most relevant predictor in yearly BRT models for both *P. vivax* and *P. falciparum*, and presented the widest variation in RC. The analysis for high-risk in villages showed a higher RC of CAR for *P. vivax* than for *P. falciparum* (Supplementary Fig. S1), with the lowest RC for both species in 2010 and the highest in 2015 (RC range for *P. vivax*: 17%-48.4%; RC range for *P. falciparum*: 11.5%-30.7%). The analysis for very-high-risk found the lowest RC in 2010 for both *P. vivax* (14.3%) and *P. falciparum* (16.1%), and the highest in 2015 for *P. vivax* (35.5%) and in 2016 for *P. falciparum* (26.2%). The estimated RC for CAR followed similar trends with increasing importance of CAR as malaria risk predictor until 2015 for *P. vivax* and until 2015–2016 for *P. falciparum*, followed by a decrease (*P. vivax*) or stabilization (*P. falciparum*).

Unlike CAR, the POPD RC in villages was slightly higher for *P. falciparum* than for *P. vivax* in most years (Supplementary Fig. S1). The relevance of POPD varied widely across years only when BRT models assessed for *P. vivax* high-risk (range: 9.2–22.6%) and *P. falciparum* very-high-risk (range: 10.0–21.4%). The highest RCs for *P. vivax* malaria risk were found at the beginning of the study period (2010–2012) and then decreased and remained low with small variations; while RCs for *P. falciparum* increased from 2010 to 2014 and then declined in the following years.

The importance of NDWI as species-specific malaria risk predictor varied during the study period. The difference between lowest and highest RCs ranged between 3.8 and 18.9%, with an initial decrease in RC from 2010 to 2011 followed by peaks in 2012 and in 2016. The non-time dependent variable TPV was the predictor with the lowest RC variation across years for *P. vivax* risk in villages (7.4%-11.6% and 11%-13.1% for high-risk and very-high-risk respectively). It was also among the predictors with the lowest RC ranges when assessing the *P. falciparum* risk across time. For all other predictors in species-specific BRT models, the difference between lowest and highest RCs did not exceed 10% across years.

		Overall		*P. vivax*		*P. falciparum*	
	Model	cvAUC	tAUC	cvAUC	tAUC	cvAUC	tAUC
High risk (API > 10)	2010	0.72	0.70	0.72	0.70	0.78	0.76
	2011	0.80	0.76	0.80	0.75	0.86	0.74
	2012	0.82	0.80	0.82	0.80	0.84	0.81
	2013	0.84	0.80	0.83	0.80	0.87	0.80
	2014	0.83	0.82	0.82	0.82	0.85	0.84
	2015	0.82	0.79	0.82	0.79	0.87	0.85
	2016	0.82	0.80	0.82	0.81	0.87	0.84
	2017	0.82	—	0.83	—	0.87	—
Very high risk (API > 50)	2010	0.76	0.76	0.77	0.76	0.82	0.78
	2011	0.85	0.81	0.84	0.80	0.88	0.81
	2012	0.85	0.82	0.85	0.81	0.87	0.76
	2013	0.86	0.80	0.86	0.80	0.89	0.84
	2014	0.84	0.84	0.84	0.83	0.89	0.83
	2015	0.85	0.83	0.85	0.82	0.89	0.86
	2016	0.84	0.83	0.84	0.83	0.88	0.83
	2017	0.85	—	0.84	—	0.89	—

Table 2. Assessment of the discriminating power of BRT models for malaria risk in villages. Each cross-validation BRT model built with data of a given year yielded a cross-validated AUC (cvAUC), while its model predictions with testing data of the following year allowed for the estimation of a testing AUC (tAUC).

Partial dependence plots. Supplementary Fig. S2 presents partial dependence plots (PDPs)[68] of the marginal effect of predictor variables on the probability of villages to be at high-risk or very-high-risk, and shows whether the relationship is linear, monotonous or complex. The probability for a village to be at high-risk of *P. vivax* malaria generally increases with: CAR over 800 mm/3 hrs, POPD between 403 and 2,980 in the 5-km side square grid around the village, TPV superior to 700 minutes, NDWI around 0.4, FC higher than 50% in the 2-km side square grid around the village, NDVI between 0.7 and 0.8, LST higher than 27 °C, shorter SDR, and higher FL in the in the 2-km side square grid around the village. The probability for a village to be at high-risk for *P. falciparum* malaria increased with CAR over 800 mm/3 hrs, POPD between 403 and 1096 inhabitants, longer TPV, NDWI around 0.4, FC superior to 60%, NDVI ranging between 0.5 and 0.8, LST higher than 26 °C, shorter SDR, and higher FL. PDPs for very-high-risk are shown in Supplementary Fig S3.

The discriminatory capacity of the models. The cvAUCs shown in Table 2 suggest that most yearly BRT models efficiently discriminate malaria risk in villages (cvAUC > 0.80), with the exception of the 2010 models for villages at high-risk (cvAUC = 0.72), and very-high-risk of *P. vivax* (cvAUC = 0.77), as well as *P. falciparum* high-risk (cvAUC = 0.78). The estimated tAUCs decreased slightly when the yearly models were tested with data corresponding to the following year, but most of the models still efficiently predicted malaria risk in the following year. For example, 2016 species-specific models using 2017 data were able to efficiently discriminate villages at high-risk (tAUC = 0.81) and very-high-risk (tAUC = 0.84) of *P. vivax*, and high-risk (tAUC = 0.83) and very-high-risk of *P. falciparum* (tAUC = 0.83) in 2017. Results from the discriminatory assessment of yearly BRT models with data from other years are presented in Supplementary Fig. S4.

Risk mapping. The study villages and their malaria risk in 2017 (estimated from 2016 BRT species-specific models) were mapped (Figs 4 and 5) in five distinct zones based on contiguity between districts, main networks of transport, shared river basins, and population density.

Zone I, includes Maynas province and is the largest, most densely populated and accessible area in the department. The risk maps showed that 42.9% of villages were at high-risk of *P. vivax*, among which 10.5% were at very-high-risk. Also, 11.7% of villages were at high-risk of *P. falciparum*, among which 1.4% were at very-high-risk.

Zone II, covers Loreto province; it is the least populated area. Most villages (56.5%), were at high-risk of *P. vivax*, and 28.7% at very-high-risk. For *P. falciparum*, 27.8% were at high-risk, and 3.5% at very-high-risk.

Zone III, is the second most densely populated area and covers Datem del Maranon and Alto Amazonas provinces, which include 30% of the study villages and 19% of the province's population. One third (34.5%) of the villages were at high-risk of *P. vivax*, and 12.3% at very-high-risk. For *P. falciparum*, 5.4% were at high-risk, and 1.1% at very-high-risk.

Zone IV, comprises Requena and Ucayali provinces, with 3.9% of villages at high-risk of *P vivax*, and 1.4% at very-high-risk. Only one village was at high-risk for *P. falciparum*.

Zone V, comprises Ramon Castilla province, and is the smallest zone, with 9% of the study villages and 6.7% of Loreto population. A significant proportion of villages (45.3%) were at high-risk of *P. vivax*, and 11.3% at very-high-risk. Seven villages were at high-risk of *P. falciparum*.

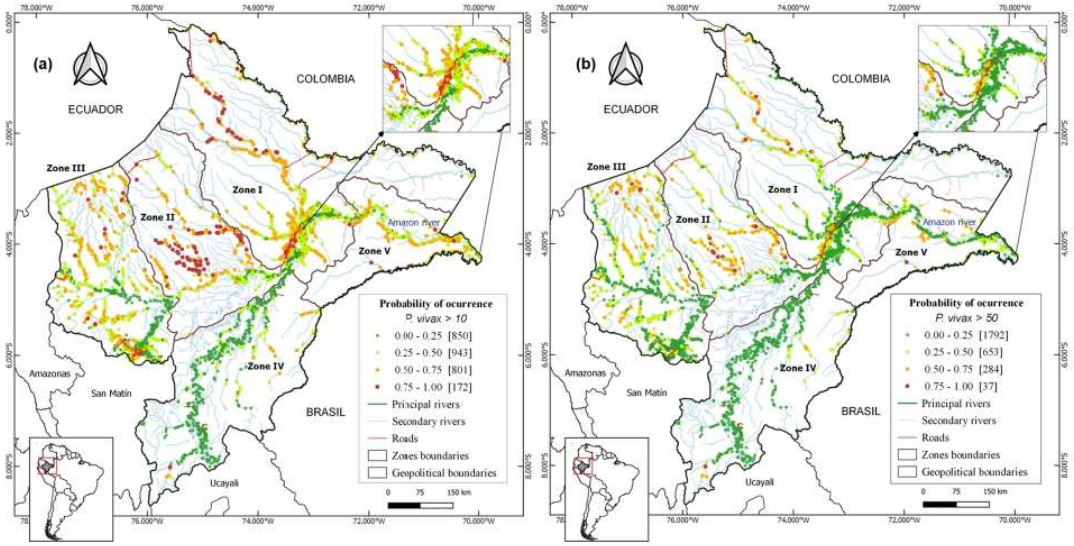


Figure 4. Predicted *P. vivax* risk maps for the year 2017 using 2016´s BRT models, showing: (**a**) villages at high *P. vivax* risk (API > 10 cases/1,000 people), (**b**) villages at very high *P. vivax* risk (API > 50 cases/1,000 people). Colors indicate the probability of a village (dots) of being at risk.

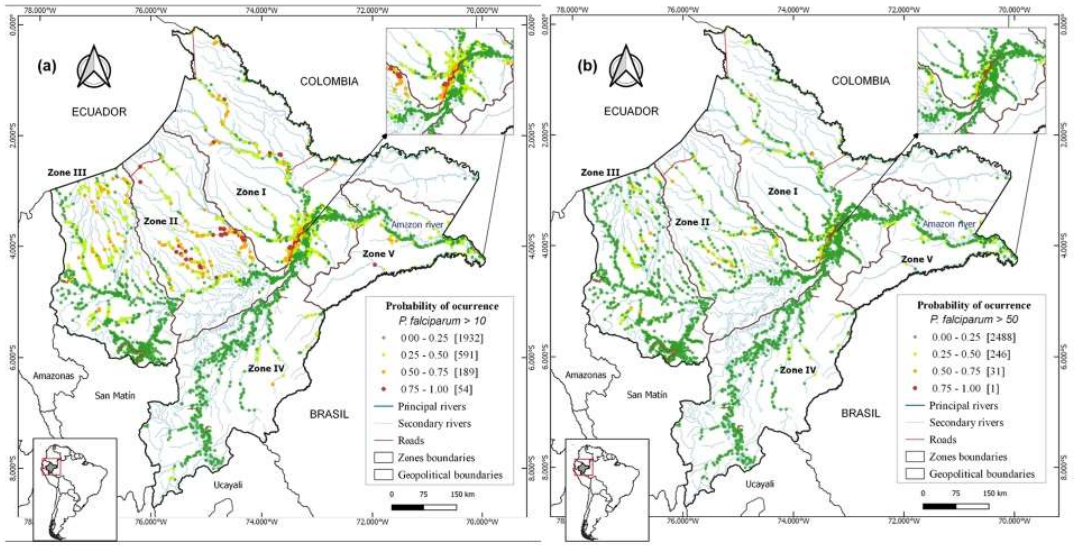


Figure 5. Predicted *P. falciparum* risk maps for the year 2017 using 2016´s BRT models, showing: (**a**) villages at high *P. falciparum* risk (API > 10 cases/1,000 people), (**b**) villages at very high *P. falciparum* (API > 50 cases/1,000 people). Colors indicate the probability of a village (dots) of being at risk.

Discussion

Using cross-validated BRT and remote-sensing (RS) satellite data, we modelled the distribution of malaria incidence by *Plasmodium* species in the Peruvian Amazon at village level between 2010 and 2017, and identified the most critical factors associated with this distribution. Yearly BRT models built with predictor data were able to efficiently discriminate the species-specific malaria risk in villages of the same year, and most of these models also performed well when predictor data was used to discriminate malaria risk in the following year. The high discriminatory capacity of the 2016 BRT models supported the generation of maps pinpointing villages with a high probability of having high malaria incidence (API > 10 cases x 1000 people) and very high incidence (API > 50 cases/1000 people) in 2017.

The mapping of malaria risk plays a key role in decision-making for designing and implementing malaria control measures[69], but it requires the use of metrics that ensure an accurate description of malaria heterogeneity and the identification of changes in transmission intensity across time and geographical scales[70]. The entomological inoculation rate (EIR) and the parasite prevalence are the best metrics of malaria transmission intensity but their estimations require several community surveys per year (to capture annually seasonal variations), very large sample sizes, and specialized laboratory techniques, making them cost-prohibitive for National Malaria Control Programs (NMCPs)[70]. Instead, NMCPs have used the incidence of microscopically confirmed malaria cases detected by routine surveillance to map the risk of malaria transmission, prioritize intervention areas, and monitor the impact of control interventions[71]. Our study confirmed that predicted malaria risk from BRT models built with remote sensing derived predictors can accurately characterize the spatial distribution of the malaria incidence in Loreto.

The ability of *P. vivax* parasites to relapse from persistent liver parasite stages (hypnozoites) weeks or months after a primary parasitaemia[72] may explain the lower discriminatory efficiency of *P. vivax* models compared with *P. falciparum* models in most study years. BRT models were built with available data from predictor variables reported to affect the malaria transmission. However, such variables are not yet proposed in the literature to be primary triggers of *P. vivax* relapses[72]. Once factors determining the hypnozoite reactivation are clearly identified, they could be incorporated to improve discriminatory capacity of the models for *P. vivax* risk. Improved models would be useful in the Peruvian Amazon, where epidemiological studies and parasite genotyping analyses suggest that relapses contribute substantially to the burden of *P. vivax* infections[2]. Moreover, data on main control interventions could also be used during the model building process, to account for the greater resilience of *P. vivax* to malaria control efforts in comparison with *P. falciparum*[73].

Environmental factors such as temperature and rainfall directly affect the lifecycles of both parasite and vector; higher temperatures accelerate parasitic *plasmodium* growth within mosquitoes, while rainfall contributes to the accumulation of stagnant water that is ideal for mosquito breeding[61]. Results from partial dependence plots showing higher malaria risk when the yearly average LST ranged between 26 °C and 29 °C are in good agreement with the reported optimal temperatures for the development of the exogenous *P. vivax* and *P. falciparum* cycles within the main malaria vector *An. darlingi*[61]. However, due to the low variability of LST across villages, this variable was not among the most important factors for discriminating the malaria risk between villages. Instead, the variable satellite-derived rainfall was the best predictor for both species, with increased importance each year until2015. Unusually heavy rains in the last trimester of each year between 2011 and 2014 generated abrupt increases of river levels, and flooded to villages along the Amazonas River and its tributaries between 2012 and 2015[2,74–76]. The river levels not only peaked earlier and higher compared to historical averages, but also they remained high for a longer period. Levels of Amazon river in 2015 exceeded the threshold for imminent flooding (117 metres above sea level) from March through June[77]. River levels in 2011, 2012 and 2016 surpassed that threshold during 5, 14, and 4 weeks, respectively.

The proximity to breeding sites is an important determinant for the heterogeneity of both mosquito exposure and malaria occurrence in the Peruvian Amazon[78]. After seasonal rains, permanent breeding sites around villages become more productive, and additional breeding sites arise[79]. With severe flooding, breeding sites further enlarge and remain longer, leading to a wider dispersal of *An. darlingi* and increased vector-human contact rates. Villages with shorter distances to rivers had increased malaria receptivity and consequently, more malaria incidence as shown in PDPs. Increased vector contact would explain both the higher contribution of CAR in *P. vivax* and *P. falciparum* models and the better discriminatory efficiency of *P. vivax* models in the years of severe floods (2012–2015). Therefore, our findings suggest that new infections contributed more than hypnozoite-triggered relapses to the rise in the *P. vivax* malaria incidence since 2012.

In the literature, NDWI has been successfully used to identify bodies of water[80]. This predictor was an important factor in BRT models for both species (median in the study period slightly lower than 10%); but unlike with rainfall, its relative contribution did not present any identifiable temporal trend during the study period. Yearly average NDWI values around 0.4 were associated with increased malaria risk, suggesting that most of the 2-km side square grid around villages with high malaria incidence was covered by open water and/or wetlands, although NDWI did not allow to characterize those water bodies that are suitable for the development of *An. darlingi*. A recent study conducted in rural villages in the Peruvian Amazon found that the *An. darlingi* larval habitats were significantly associated with water bodies in landscapes with more recent deforestation, lower light intensity, emergent vegetation and a lower vegetation index[78]. The vegetation covering and surrounding these breeding sites could provide food for larvae, shelter from predators, and favourable oviposition conditions[81]. The yearly average NDVI values in the 2-km side square grid around villages would not differentiate this vegetation as suggested by the low contribution of this variable to the malaria risk models, and the increased malaria incidence associated with NDVI values above 0.6.

Deforestation and environmental changes driven by human activity have been associated with *An. darlingi* breeding and malaria transmission[16]; however, the heterogeneous definitions of deforestation in these studies precludes us from drawing a firm conclusion[17]. According to our BRT models, substantial recent tree cover loss in a 2-km side square grid around villages, forest loss was associated with higher malaria occurrence. However, this forest-related variable was not among the top predictors for discriminating the malaria risk, likely due to its limited variability across villages. The positive relationship between yearly deforestation and malaria risk is in line with entomological studies showing that *An. darlingi* larvae were more likely to be found in water bodies with recent deforestation[16,78]. Several studies in the Brazilian Amazon have found high densities of larval and adult *Anopheles* in forest fringes, as well as increased malaria morbidity in populations living or working near forest edges[82,83]. A relationship between forest coverage and forest edges might explain why villages with lower forest coverage have reduced malaria risk than those with higher coverage. As deforested areas increase, the distance to forest edges also increases but malaria transmission remains high because of the quantity and extension of

forest edges around villages. Reduction in forest coverage would make forest edges more distant, hereby reducing malaria risk, unless residents engage in activities near forest edges. However, reduced forest coverage can also indicate increased socio-economic development, which is often associated with improvements in living conditions and malaria preventive practices[84].

Vector-human contact requires that mosquitoes fly from breeding sites and forest edges to access to human blood meals[78]. Therefore, contact rates and malaria transmission strongly depend on both the dispersal of *An. darlingi* and the population density near breeding sites and forest edges. Our analysis highlighted the population density as the second most important predictor for malaria risk; yet, a positive relationship was only observed in the most densely populated villages. Further research will be required to confirm the negative relationship found in less densely populated villages. The sparsely distributed population in the Peruvian Amazon may be at increased malaria risk because of precarious conditions, with limited access to health care, and exposure to mosquito bites due to subsistence farming, fishing, hunting and other activities near or within the forest[18,19]. Besides, the time to major populated villages had a positive relationship with malaria risk, consistent with the idea that this variable can represent a valid proxy for diminished access to health care facilities, hereby reducing diagnosis and treatment, and hindering the delivery of malaria prevention interventions[21].

The limitations of this study must also be acknowledged. First, the assumption of constant population size for villages across years could have reduced the discriminatory efficiency of the BRT models since they did not account for migration. Second, NDWI and NDVI might not be among the best proxies for environmental ground conditions that affect malaria transmission. Yearly means of these indexes might not capture the particular characteristics of breeding and resting sites of *An. darlingi*. Future research accounting for annual and seasonal variations in malaria risk and predictor variables will indicate to what extent predictive models can be improved. Third, population density within mosquito dispersal ranges does not only refer to the population living near breeding sites or forest edges, but also to individuals approaching these places for economic activities (human mobility). Unfortunately, the information was not available at the village level. Fourth, data of main control interventions per village during the study period was not available. These data would improve the discriminatory efficiency of models and allow a better assessment of the differential contribution of predictor factors between species. Fifth, the resampling of predictor raster data to a higher resolution always results in output raster as precise as the coarsest inputs, and it can add a systematic bias to developed models at high resolution. A higher resolution data for cumulative annual rainfall was not available for the study period. However, this variable has a low spatial variation in the Amazon region principally for its extensive plain topography[85], and the corresponding bias is therefore expected to be limited.

In this study, we demonstrated that the probability of malaria occurrence in villages of Loreto in the Peruvian Amazon could be estimated using machine learning BRT models and RS-derived variables associated with the complex malaria host-vector-parasite relationships, resulting in predictive malaria risk maps that accurately characterized the spatial distribution of malaria incidence. Although further validation with better metrics of malaria transmission (i.e. EIR, parasite prevalence) is required, we hypothesize that the performance in discriminating the malaria transmission risk of model-predicted maps could surpass that of incidence maps in areas of reduced transmission and predominant asymptomatic infections not detected by traditional surveillance systems. This scenario will likely be reached shortly with the implementation of the 2017 governmental initiative "Zero Malaria Plan"[86] aimed at reducing the burden of clinical malaria cases in the short-term, and eliminating the disease in the long-term.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and/or its supplementary materials. Additional data is available from the corresponding author E.S-V. on request.

Received: 23 July 2019; Accepted: 2 October 2019;
Published online: 23 October 2019

References

1. CDC-MINSA/Peru. *Análisis de la Situación de Salud del Perú*. (CDC, 2017).
2. Rosas-Aguirre, A. *et al*. Epidemiology of Plasmodium vivax malaria in Peru. *Am. J. Trop. Med. Hyg.* **95**, 133–144 (2016).
3. Soto-Calle, V. *et al*. Spatio-Temporal analysis of malaria incidence in the Peruvian Amazon Region between 2002 and 2013. *Sci. Rep.* **7** (2017).
4. Roshanravan, B. *et al*. Endemic Malaria in the Peruvian Amazon Region of Iquitos. *Am. J. Trop. Med. Hyg.* **69**, 45–52 (2003).
5. Branch, O. *et al*. Clustered local transmission and asymptomatic Plasmodium falciparum and Plasmodium vivax malaria infections in a recently emerged, hypoendemic Peruvian Amazon community. *Malar. J.* **4**, 27 (2005).
6. Bousema, T. *et al*. Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control. *J. Infect. Dis.* **201**, 1764–1774 (2010).
7. Hay, S. I., Snow, R. W. & Rogers, D. J. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Trans. R. Soc. Trop. Med. Hyg.* **92**, 12–20 (1998).
8. Adeola, A. M., Olwoch, J. M., Botai, J. O., Kalumba, A. M. & Adisa, O. M. Landsat satellite derived environmental metric for mapping mosquitoes breeding habitats in the Nkomazi municipality, Mpumalanga Province, South Africa. **6245** (2015).
9. Sewe, M. O., Ahlm, C. & Rocklv, J. Remotely sensed environmental conditions and malaria mortality in three malaria endemic regions in western kenya. *PLoS One* **11**, 1–16 (2016).
10. Moss, W. J. *et al*. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community survey. *Malar. J.* **10**, 163 (2011).
11. Alimi, T. O. *et al*. A multi-criteria decision analysis approach to assessing malaria risk in northern South America. *BMC Public Health* **16**, 1–10 (2016).
12. Li, Q., Lu, L., Weng, Q., Xie, Y. & Guo, H. Monitoring Urban Dynamics in the Southeast USA 13–15, https://doi.org/10.3390/rs8070578 (2016).

13. Mellander, C., Lobo, J., Stolarick, K. & Matheson, Z. Night-Time Light Data: A Good Proxy Measure for Economic Activity? *PLoS One* 1–18, https://doi.org/10.1371/journal.pone.0139779 (2015).
14. Maheu-Giroux, M. *et al*. Risk of malaria transmission from fish ponds in the Peruvian Amazon. *Acta Trop.* **115**, 112–118 (2010).
15. Vittor, A. Y. *et al*. The effect of deforestation on the human-biting rate of Anopheles darlingi, the primary vector of Falciparum malaria in the Peruvian Amazon. *Am. J. Trop. Med. Hyg.* **74**, 3–11 (2006).
16. Vittor, A. Y. *et al*. Linking Deforestation to Malaria in the Amazon: Characterization of the Breeding Habitat of the Principal Malaria Vector, Anopheles darlingi. *Am. J. Trop. Med. Hyg.* **81**, 5–12 (2009).
17. Tucker Lima, J. M., Vittor, A., Rifai, S. & Valle, D. Does deforestation promote or inhibit malaria transmission in the Amazon? A systematic literature review and critical appraisal of current evidence. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160125 (2017).
18. Chuquiyauri, R. *et al*. Socio-demographics and the development of malaria elimination strategies in the low transmission setting. *Acta Trop.* **121** (2012).
19. Rosas-Aguirre, A. *et al*. Micro-heterogeneity of malaria transmission in the Peruvian Amazon: A baseline assessment underlying a population-based cohort study. *Malar. J.* **16** (2017).
20. Casapía, M. *et al*. Mejora en el diagnóstico y tratamiento oportuno de malaria con el uso de pruebas rápidas por promotores de salud en la Amazonía peruana. *Rev. Peru. Med. Exp. Salud Publica* **25**, 361–368 (2008).
21. Brierley, C. K., Suarez, N., Arora, G. & Graham, D. Healthcare Access and Health Beliefs of the Indigenous Peoples in Remote Amazonian Peru. *Am. J. Trop. Med. Hyg.* **90**, 180–183 (2014).
22. Weiss, D. J. *et al*. inequalities in accessibility in 2015. *Nat. Publ. Gr.*, https://doi.org/10.1038/nature25181 (2018).
23. Aramburú Guarda, J., Ramal Asayag, C. & Witzig, R. Malaria reemergence in the Peruvian Amazon Region. *Emerg. Infect. Dis.* **5**, 209–215 (1999).
24. Grietens, K. P. *et al*. Traditional nets interfere with the uptake of long-lasting insecticidal nets in the Peruvian Amazon: the relevance of net preference for achieving high coverage and use. *PLoS One* **8**, e50294 (2013).
25. Grietens, K. P. *et al*. Adherence to 7-day Primaquine Treatment for the Radical Cure of *P*. vivax in the Amazon Region. *Am.J.Trop. Med Hyg.* **82**, 1017–1023 (2015).
26. Markham, B. Forty-Year Calibrated Record of Earth-Surface Reflected Radiance from Landsat: A Review. 1–52 (2019).
27. Woodcock, C. E. *et al*. The global Landsat archive: Status, consolidation, and direction Remote Sensing of Environment The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **185**, 271–283 (2016).
28. Hall, O. Remote sensing in social science research. *Open Remote Sens. J.* **3**, 1–16 (2010).
29. Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. 1–22, https://doi.org/10.1371/journal.pone.0107042 (2015).
30. Thenkabail, P. S. *Remote sensing handbook: Remote sensing of water resources, disasters, and urban studies. Remote Sensing of Water Resources, Disasters, and Urban Studies* **3** (2015).
31. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
32. Ashby, J., Moreno-Madriñán, M., Yiannoutsos, C. & Stanforth, A. Niche Modeling of Dengue Fever Using Remotely Sensed Environmental Factors and Boosted Regression Trees. *Remote Sens.* **9**, 328 (2017).
33. Kabaria, C. W. *et al*. Mapping intra - urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. *Int. J. Health Geogr.* 1–12, https://doi.org/10.1186/s12942-016-0051-y (2016).
34. Messina, J. P. *et al*. Mapping global environmental suitability for Zika virus. *Elife* **5**, 1–19 (2016).
35. Pigott, D. M. *et al*. Updates to the zoonotic niche map of Ebola virus disease in Africa. *Elife* **5**, 1–13 (2016).
36. Cheong, Y. L., Leitão, P. J. & Lakes, T. Assessment of land use factors associated with dengue cases in Malaysia using boosted regression trees. *Spat. Spatiotemporal. Epidemiol.* **10**, 75–84 (2014).
37. INEI. Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas. 43 (2017).
38. Tanser, F. Innova t ions (GIS) Innovations For Primary Health Care in Developing Countries. 106–122 (2006).
39. Eastman, J. Multi-criteria evaluation and GIS. *Geogr. Inf. Syst.* 493–502 (1999).
40. INEI. Ficha técnica empadronamiento distrital de población y vivienda 2013 Empadronamiento Distrital de Población y Vivienda 2013. **2013**, 3–4 (2013).
41. CDC-MINSA/Peru. *Boletín epidemiológico del perú*. (2017).
42. Omumbo, J. A., Hay, S. I., Snow, R. W., Tatem, A. J. & Rogers, D. J. Modelling malaria risk in East Africa at high-spatial resolution. *Trop. Med. Int. Heal.* **10**, 557–566 (2005).
43. Patz, J. A. *et al*. Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. *Trop. Med. Int. Heal.* **3**, 818–827 (1998).
44. Gorelick, N. *et al*. Remote Sensing of Environment Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
45. Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
46. Bruy, A. & Svidzinska, D. *QGIS by example*. (PACKT, 2015).
47. Menke, K., Davis, P. & Ebooks Corporation. *Mastering QGIS: go beyond the basics and unleash the full power of QGIS with practical, step-by-step examples. Community experience distilled*.
48. Hansen, M. C. *et al*. High-Resolution Global Maps of of 21st-Century Forest Cover Change. **134**, 850–854 (2013).
49. Simpson, J., Adler, R. F. & North, G. R. A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite. *Bull. Am. Meteorol. Soc.* **69**, 278–295 (1988).
50. Cuéllar, A. C. Uso De Sensores Remotos Para La Predicción De Casos De Malaria En El Departamento Orán, Salta, Argentina. 95 (2014).
51. Szabó, S., Gácsi, Z. & Balázs, B. Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Landsc. Environ.* **10**, 194–202 (2016).
52. Jensen, J. R. *Remote sensing of the environment: an earth resource perspective*. *Prentice Hall, Upper Saddle River, NJ* 1 (2000).
53. Hantson, S. *et al*. Cadena de pre-procesamiento estándar para las imágenes Landsat del Plan Nacional de Teledetección. *Rev. Teledetección* **36**, 51–61 (2011).
54. Myneni, R. B., Ganapol, B. D. & Asrar, G. Remote sensing of vegetation canopy photosynthetic and stomatal conductance efficiencies. *Remote Sens. Environ.* **42**, 217–238 (1992).
55. da Costa Gurgel, H. A utilização das geotecnologias em estudos epidemiologicos: o exemplo da relação entre a malária e o NDVI em Roraima. *An. XI SBSR, Belo Horizonte, Bras. 05 - 10 abril 2003, INPE* 1303–1310 (2003).
56. Tourre, Y. M., Jarlan, L., Lacaux, J.-P., Rotela, C. H. & Lafaye, M. Spatio-temporal variability of NDVI–precipitation over southernmost South America: possible linkages between climate signals and epidemics. *Environ. Res. Lett.* **3**, 044008 (2008).
57. Goward, S. N., Markham, B., Dye, D. G., Dulaney, W. & Yang, J. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sens. Environ.* **35**, 257–277 (1991).
58. Gillespie, A. Land Surface Temperature. *Encycl. Remote Sens.* 314–319, https://doi.org/10.1007/978-0-387-36699-9_79 (2014)
59. Saha, S. K. Retrieval of Agrometeorological parameters using satellite data. In *Satellite Remote Sensing and GIS Applications in Meteorology* 151–174 (2004).
60. Zeilhofer, P., Soares, E., Ribeiro, A. L. M., Miyazaki, R. D. & Atanaka, M. Habitat suitability mapping of Anopheles darlingi in the surroundings of the Manso hydropower plant reservoir. *Mato Grosso, Central.* **14**, 1–14 (2007).

61. Hiwat, H. & Bretas, G. Ecology of Anopheles darlingi Root with respect to vector importance: A review. *Parasites and Vectors* **4**, 1–13 (2011).
62. Gareth James, Daniela Witten, Trevor Hastie, R. T. *An Introduction to Statistical Learning with Applications in R*.
63. Elith, A. Boosted Regression Trees in R. *Biometrics* **2008**, 1–15 (2008).
64. Greenwell, B., Boehmke, B., Cunningham, J. & GBM Developers. Generalized Boosted Regression Models. (2018).
65. Hijmans, R. J., Phillips, S., Leathwick, J. R. & Elith, J. Dismo package for R, version 1.1-4, https://doi.org/10.1016/j.jhydrol.2011.07.022 (2017).
66. Brock, P. M. *et al*. Predictive analysis across spatial scales links zoonotic malaria to deforestation (2019).
67. Artois, J. *et al*. Changing geographic patterns and risk factors for avian influenza A(H7N9) infections in humans, China. *Emerg. Infect. Dis.* **24**, 87–94 (2018).
68. Greenwell, B. M. pdp: An R Package for Constructing Partial Dependence Plots. **9**, 421–436 (2017).
69. Sturrock, H. J. W. *et al*. Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. *Trends Parasitol.* **32**, 635–645 (2016).
70. Tusting, L. S., Bousema, T., Smith, D. L. & Drakeley, C. Measuring changes in Plasmodium falciparum transmission: Precision, accuracy and costs of metrics. *Adv Parasitol* 1–48, https://doi.org/10.1016/B978-0-12-800099-1.00003-X.Measuring (2016).
71. Roll Back Malaria & & Partnership. *Framework for Monitoring Progress and Evaluating Outcomes and Impact. Geneva: RBM; 2000*.
72. White, N. J. Determinants of relapse periodicity in Plasmodium vivax malaria Determinants of relapse periodicity in Plasmodium vivax malaria. **297** (2011).
73. Howes, R. E. *et al*. Global epidemiology of Plasmodium vivax. *Am. J. Trop. Med. Hyg.* **95**, 15–34 (2016).
74. COEN-INDECI. Inundación afecta a la provincia de Putumayo - Loreto. 1–10 (2016).
75. PAHO. *Inundaciones en Loreto: Respuesta del Sector de Agua, Saneamiento e Higiene - Experiencias y aprendizajes*. *PAHO* (2013).
76. Reach-Initiative. 2012 Flood Events, PERU. (2012).
77. SEHINAV. Boletín de avisos a los navegantes fluviales - SERVICIO DE HIDROGRAFÍA Y NAVEGACIÓN DE LA AMAZONÍA. (2015).
78. Prussing, C. *et al*. Malaria vector species in Amazonian Peru co-occur in larval habitats but have distinct larval microbial communities. *PLoS Negl. Trop. Dis.* **13**, e0007412 (2019).
79. Rosas-Aguirre, A. *et al*. Hotspots of malaria transmission in the Peruvian amazon: Rapid assessment through a parasitological and serological survey. *PLoS One* **10**, 1–21 (2015).
80. Guo, M., Li, J., Sheng, C., Xu, J. & Wu, L. A Review of Wetland Remote Sensing. 1–36, https://doi.org/10.3390/s17040777 (2017).
81. Rejmánková E, Grieco J, Achee N R. D. *Ecology of Larval Habitats. In: Manguin S, editor. Anopheles mosquitoes—New insights into malaria vectors*.
82. Barros, F. S. M. & Honório, N. A. Deforestation and Malaria on the Amazon Frontier: Larval Clustering of Anopheles darlingi (Diptera: Culicidae) Determines Focal Distribution of Malaria. **93**, 939–953 (2015).
83. Arruda, M. E. & Gurgel, H. Spatial clustering and longitudinal variation of Anopheles darlingi (Diptera: Culicidae) larvae in a river of the Amazon: the importance of the forest fringe and of obstructions to flow in frontier malaria, https://doi.org/10.1017/S0007485311000265 (2011).
84. Id, A. D., Fennie, K., Degarege, D., Id, S. C. & Madhivanan, P. Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa: A systematic review and meta-analysis. 1–26 (2019).
85. Zhang, L. V. A. X. *et al*. Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *Int. J. Climatol.* **29**, 317–319 (2009).
86. MINSA. *Resolución Ministerial N° 244-2017-MINSA*. 12–13 (2017).

Acknowledgements

This study was funded by the Peruvian National Council of Science - CONCYTEC (008-2014-FONDECYT) and the Académie de Recherche et d'Enseignement Supérieur - Commission de la Coopération au Développement of Belgium (ARES-CCD, PRD-Peru 2014–2019). Thanks to Nathalie Malève and Yulissa Vasquez for general support, and to Bruce Millies for English corrections.

Author contributions

Conceptualization, E.S.-V. and A.R.-A.; Remote sensing data collection, E.S.-V. and J.J.P.-G.; Remote Sensing Processing, E.S.-V. and A.R.-A.; Geo-codification process, E.S.-V. W.V., D.M-G. and A.R.-A.; Methodology, E.S.-V., W.V. and A.R.-A.; Codes elaboration, E.S.-V., W.V. and A.R.-A.; Formal analysis, E.S.-V., W.V. and A.R.-A.; Visualization, E.S.-V. and V.W.; Validation, C.L., P.L, M.-P.H., N.S. and A.R.-A.; Project administration, A.L.-C. and A.R.-A.; Funding acquisition, M.-P.H., N.S. and A.R.-A. Supervision, M.-P.H.; N.S. and A.R.-A.; Writing-original draft, E.S.-V. and A.R.-A.; Writing-review & editing, E.S.-V. W.V., P.M., C.L. and A.R.-A. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-51564-4.

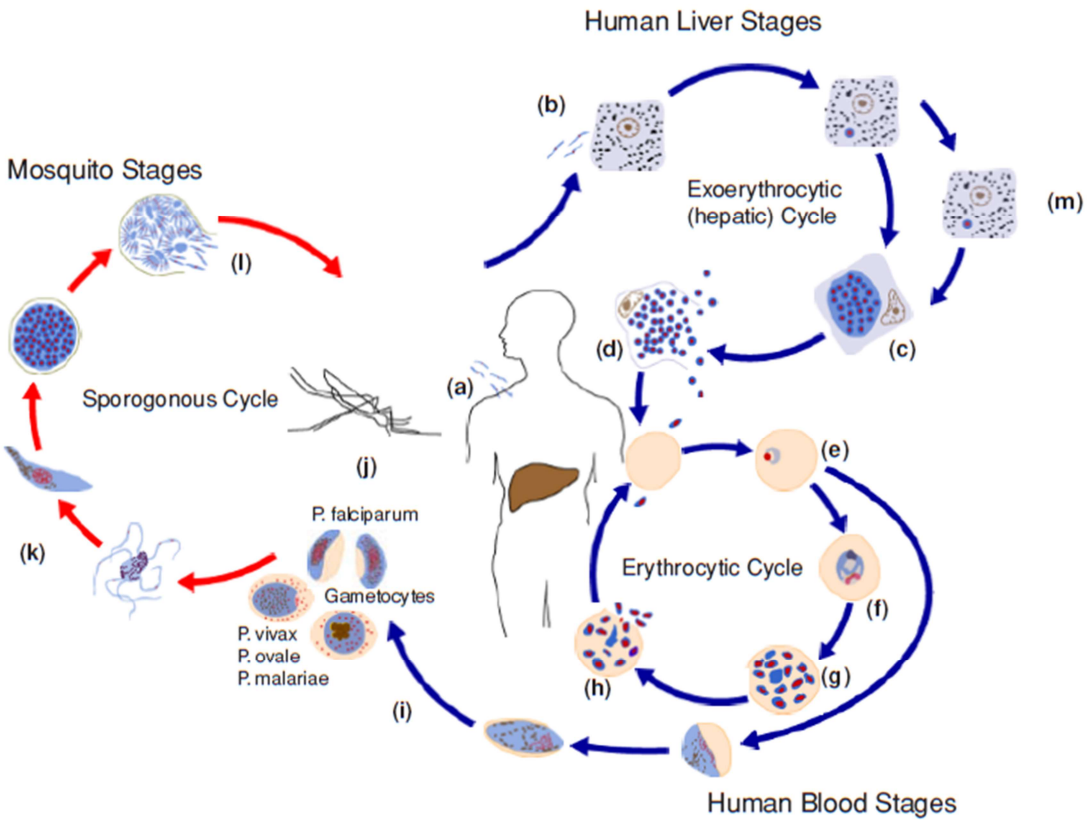
Correspondence and requests for materials should be addressed to E.S.-V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix B: Complementary information

1. *Plasmodium* life cycle



(Appendix) Figure 1: *Plasmodium* life cycle, from (160)

Infective mosquitoes inject sporozoites into the bloodstream during feeding. These sporozoites infect liver cells (b) where they undergo asexual reproduction (exoerythrocytic schizogony), producing schizonts (c). In 6 to 14 days (sometimes longer), the schizonts rupture, releasing merozoites into the bloodstream (d). Merozoites invade red blood cells and undergo a second phase of asexual reproduction (erythrocytic schizogony), developing into rings (e), trophozoites (f), and finally blood stage schizonts (g). The schizonts rupture, destroying the red blood cell and releasing more merozoites into the bloodstream, starting another cycle of asexual development and multiplication (h). This erythocytic cycle will continue until the infected individual is successfully treated, mounts an immune response that clears the infection, or dies. During this cycle, sexual forms called gametocytes are produced (i) and can be ingested by a mosquito during a blood meal (j). Sexual reproduction occurs in the mosquito (k). Sporozoites are formed (l), which migrate to the salivary glands, making the mosquito infective to humans. The timing of events in the life cycle of malaria parasites and the number of merozoites produced during schizogony vary by species. Additionally, two species of malaria, *P. vivax* and *P. ovale*, have a form, "hypnozoites" (m), that can persist in the liver for months to years, causing periodic relapses of peripheral parasitemia and illness" adapted from Malaria Cycle life (Malaria control during mass population movements and natural disasters).

2. List of *Anopheles*

(Appendix) Table 1: Primary and secondary vectors currently recognized in the Region of the Americas.

Subregion	Primary vectors	Secundary vectors
Mesoamérica	*Anopheles albimanus* *An. pseudopunctipennis*	*An. vestitipennis* *An. darlingi* *An. puntimacula* *An. apicimacula* *An. pseudopunctipennis*
Areas no Amazon	*An. albimanus* *An. darlingi* *An. nuneztovari* *An. aquasalis*	*An. pseudopunctipennis* *An. punctimacula*
Amazon basin	*An. darlingi*	*An. benarrochi* *An. oswaldoi* *An. rangeli* *An. triannulatus* *An. marajoara* *An. aquasalis* *An. deaneorum* *An. janconnae* *An. nuñeztovari* *An. braziliensis* *An. triannulatus* *An. peryassui*

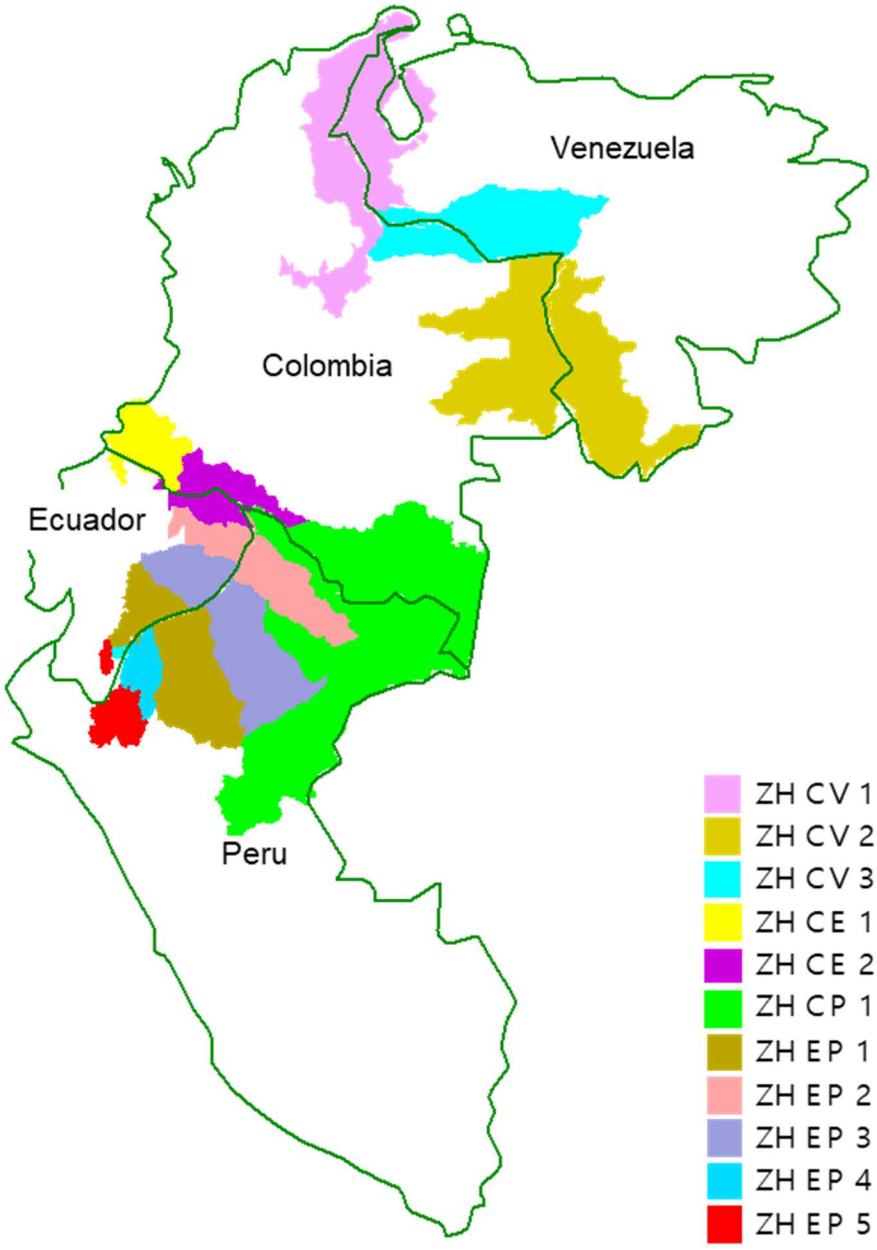
From: Strategic guidance document for the surveillance and control of malaria vectors in Latin America and the Caribbean.

3. Sensors and applications

(Appendix) Table 2: 'Earth Science Satellite Applications: Current and Future. Modified from (Flores et al. 2012).

EO-derived measurement and/or product	**Sensor, satellite**
Elevation data	SRTM ASTER, Terra
Rainfall	Imager, GOES series MVIRI, Meteosat-7/5 VISSR, GMS TMI, TRMM SSM/I, DMSP series AMSR-E, Aqua AMSU-B, NOAA series
Multispectral VNIR reflectance/radiance	AVHRR, POES series MODIS, Aqua/Terra ASTER, Terra ALI, EO-1 Landsat series SPOT GeoEye-1 Ikonos QuickBird Worldview-1 & 2 Formosat-2 Sentinel-2
Hyperspectral VNIR reflectance/radiance	Hyperion, EO-1
Fire and thermal anomalies	MODIS, Aqua/Terra
Burned area	MODIS, Aqua/Terra
Surface temperature	Landsat Thermal Infrared (TIR)
Land cover, tree cover maps	MODIS, Aqua/Terra
	MERIS, Envisat Landsat series ASTER, Terra SPOT L-Band SAR (PALSAR), ALOS
Relative Humidity	AIRS and AMSU, Aqua
Synthetic Aperture Radar (SAR) images	Sentinel-1 C-band SAR, Radarsat-1 & Radarsat-2 C-band SAR, ERS-2 L-Band SAR (PALSAR), ALOS
SRTM-water body	SRTM
Gridded population distribution	DMS series
Near real-time cloud movement	Imager, GOES series

4. PAMAFRO map scope



(Appendix) Map 1:Homogeneous zones for PAMAFRO project, considering contiguity in borders or similar socioeconomic processes. Similarities of climate and landscape. Endemic or potentially endemic status for malaria.

5. gbm package

Greg Ridgeway, Brian Kriegler, Stefan Schroedl.

Description

Workhorse function providing the link between R and the C++ gbm engine. gbm is a front-end to gbm.fit that uses the familiar R modeling formulas. However, model.frame is very slow if there are many predictor variables. For power-users with many variables use gbm.fit. For general practice gbm is preferable.

Usage

gbm.fit(x, y, offset = NULL, misc = NULL, distribution = "bernoulli", w = NULL, var.monotone = NULL, n.trees = 100, interaction.depth = 1, n.minobsinnode = 10, shrinkage = 0.001, bag.fraction = 0.5, nTrain = NULL, train.fraction = NULL, keep.data = TRUE, verbose = TRUE, var.names = NULL, response.name = "y", group = NULL)

Arguments

x A data frame or matrix containing the predictor variables. The number of rows in x must be the same as the length of y.

y A vector of outcomes. The number of rows in x must be the same as the length of y.

offset A vector of offset values.

misc An R object that is simply passed on to the gbm engine. It can be used for additional data for the specific distribution. Currently it is only used for passing the censoring indicator for the Cox proportional hazards model.

distribution Either a character string specifying the name of the distribution to use or a list with a component name specifying the distribution and any additional parameters needed. If not specified, gbm will try to guess: if the response has only 2 unique values, bernoulli is assumed; otherwise, if the response is a factor, multinomial is assumed; otherwise, if the response has class "Surv", coxph is assumed; otherwise, gaussian is assumed. gbm.fit 13 Currently

available options are "gaussian" (squared error), "laplace" (absolute loss), "tdist" (t-distribution loss), "bernoulli" (logistic regression for 0-1 outcomes), "huberized" (huberized hinge loss for 0-1 outcomes), classes), "adaboost" (the AdaBoost exponential loss for 0-1 outcomes), "poisson" (count outcomes), "coxph" (right censored observations), "quantile", or "pairwise" (ranking measure using the LambdaMart algorithm).

If quantile regression is specified, distribution must be a list of the form list(name = "quantile",alpha = 0.25) where alpha is the quantile to estimate. The current version's quantile regression method does not handle nonconstant weights and will stop. If "tdist" is specified, the default degrees of freedom is 4 and this can be controlled by specifying distribution = list(name = "tdist",df = DF) where DF is your chosen degrees of freedom. If "pairwise" regression is specified, distribution must be a list of the form list(name="pairwise",group=...,metric=...,max.rank=...) (metric and max.rank are optional, see below). group is a character vector with the column names of data that jointly indicate the group an instance belongs to (typically a query in Information Retrieval applications). For training, only pairs of instances from the same group and with different target labels can be considered. metric is the IR measure to use, one of list("conc") Fraction of concordant pairs; for binary labels, this is equivalent to the Area under the ROC Curve : Fraction of concordant pairs; for binary labels, this is equivalent to the Area under the ROC Curve list("mrr") Mean reciprocal rank of the highest-ranked positive instance : Mean reciprocal rank of the highest-ranked positive instance list("map") Mean average precision, a generalization of mrr to multiple positive instances : Mean average precision, a generalization of mrr to multiple positive instances list("ndcg:") Normalized discounted cumulative gain. The score is the weighted sum (DCG) of the user-supplied target values, weighted by log(rank+1), and normalized to the maximum achievable value. This is the default if the user did not specify a metric. ndcg and conc allow arbitrary target values, while binary targets 0,1 are expected for map and mrr. For ndcg and mrr, a cut-off can be chosen using a positive integer parameter max.rank. If left unspecified, all ranks are taken into account. Note that splitting of instances into training and validation sets follows group boundaries and therefore only

approximates the specified train.fraction ratio (the same applies to cross-validation folds). Internally, queries are randomly shuffled before training, to avoid bias. Weights can be used in conjunction with pairwise metrics, however it is assumed that they are constant for instances from the same group. For details and background on the algorithm, see e.g. Burges (2010).

Data an optional data frame containing the variables in the model. By default the variables are taken from environment(formula), typically the environment from which gbm is called. If keep.data=TRUE in the initial call to gbm then gbm stores a copy with the object. If keep.data=FALSE then subsequent calls to gbm.more must resupply the same dataset. It becomes the user's responsibility to resupply the same data at this point.

weights an optional vector of weights to be used in the fitting process. Must be positive but do not need to be normalized. If keep.data=FALSE in the initial call to gbm then it is the user's responsibility to resupply the weights to gbm.more

Value A gbm.object object.

var.monotone an optional vector, the same length as the number of predictors, indicating which variables have a monotone increasing (+1), decreasing (-1), or arbitrary (0) relationship with the outcome.

n.trees Integer specifying the total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. Default is 100. interaction.depth Integer specifying the maximum depth of each tree (i.e., the highest level of variable interactions allowed). A value of 1 implies an additive model, a value of 2 implies a model with up to 2-way interactions, etc. Default is 1.

n.minobsinnode Integer specifying the minimum number of observations in the terminal nodes of the trees. Note that this is the actual number of observations, not the total weight.

shrinkage a shrinkage parameter applied to each tree in the expansion. Also known as the learning rate or step-size reduction; 0.001 to 0.1 usually work, but a smaller learning rate typically requires more trees. Default is 0.1.

bag.fraction the fraction of the training set observations randomly selected to propose the next tree in the expansion. This introduces randomnesses into the model fit. If bag.fraction < 1 then running the same model twice will result in similar but different fits. gbm uses the R random number generator so set.seed can ensure that the model can be reconstructed. Preferably, the user can save the returned gbm.object using save. Default is 0.5.

train.fraction The first train.fraction * nrows(data) observations are used to fit the gbm and the remainder are used for computing out-of-sample estimates of the loss function.

cv.folds Number of cross-validation folds to perform. If cv.folds>1 then gbm, in addition to the usual fit, will perform a cross-validation, calculate an estimate of generalization error returned in cv.error.

keep.data a logical variable indicating whether to keep the data and an index of the data stored with the object. Keeping the data and index makes subsequent calls to gbm.more faster at the cost of storing an extra copy of the dataset.

verbose Logical indicating whether or not to print out progress and performance indicators (TRUE). If this option is left unspecified for gbm.more, then it uses verbose from object. Default is FALSE.

class.stratify.cv Logical indicating whether or not the cross-validation should be stratified by class. Defaults to TRUE for distribution = "multinomial" and is only implemented for "multinomial" and "bernoulli". The purpose of stratifying the cross-validation is to help avoiding situations in which training sets do not contain all classes.

n.cores The number of CPU cores to use. The cross-validation loop will attempt to send different CV folds off to different cores. If n.cores is not specified by the user, it is guessed using the detectCores function in the parallel package. Note that the documentation for detectCores makes clear that it is not failsafe and could return a spurious number of available cores.

6. Codes implemented

Codes implemented for R

```
################################## GBM models ###################################
##    tuning
##---------------------------------------------------------------------
##-- (0) Do not tuning (default)
##--     Lists must contain 1 value
##------------------------------
##-- (1) for doing TUNING (disable). //Enable for tuning//
##--     then the lists must contain >=1 value (The more values in the lists, the more models
to evaluate)
if (tuning==0){
    tree_cmplx <- NO.TUNING_tree_complexity        # 5
    learn_rate <- NO.TUNING_learning_rate    # 0.0005
    bag_fracc  <- NO.TUNING_bag_fraction
    pdp1  <- 1    ## Specifies that PDPs are to be obtained
     print("%%%%%%%%%%---------------- MODELS WITH FIXED PARAMETERS ARE RUN ---------------%
%%%%%%%%%%")
}else if (tuning==1){
    tree_cmplx <- TUNING_tree_complexity
    learn_rate <- TUNING_learning_rate
    bag_fracc  <- TUNING_bag_fraction
    pdp1 <- 0    ## Specifies to don't obtain PDPs
   print("%%%%%%----------  We run de models with variable parameters (Tuning) ----------
%%%%")
}else{tuning <- 0}

##----------------------  End of PARAMETERS ----------------------

#--- Tibble (dataframe) "tb_vdep" associated with Dependent Variables (Total of 6 dependent
variables)
num_vdep <- c(34:39)
nombres_vdep <- c("ctot_IPA50", "cviv_IPA50", "cfal_IPA50","ctot_IPA10", "cviv_IPA10",
"cfal_IPA10" )
nombres_modelos <- c("Malaria Total IPA>50", "Malaria vivax IPA>50", "Malaria falciparum
IPA>50","Malaria Total IPA>10", "Malaria Vivax IPA>10", "Malaria Falciparum IPA>10")

tb_vdep <- tibble(num_vdep, nombres_vdep, nombres_modelos)

#--- Generates list of predictor var names from input
k <- tibble(1:ncol(data),  colnames(data))
a <- c()
text_p  <- ""
for (i in lista_predictoras) {   a[length(a)+1]<- k[i,2]  }
for (i in lista_predictoras) {   text_p<- paste(text_p, k[i,2], sep=",")  }
ltexto_predictoras <- unlist(a)
rm(a)

##---------- Defining models

## Specifies family variable (type of model)familia="bernoulli"
## simple modelo
callm <- paste("gbm.step(data=data0, gbm.x=lista_predictoras, gbm.y=vd, distribution =
as.character(familia), "  ,
                "learning.rate=lr, tree.complexity=tc, bag.fraction = bf, max.trees =
50000, plot.main=FALSE, verbose=FALSE)", sep="")

## model keep.vector=TRUE (save fold.vector)
callm_keep <- paste("gbm.step(data=data0, gbm.x=lista_predictoras, gbm.y=vd, distribution =
as.character(familia), "  ,
                    "learning.rate=lr, tree.complexity=tc, bag.fraction = bf, max.trees =
50000, plot.main=FALSE, verbose=FALSE, " ,
                   "keep.fold.vector=TRUE)", sep="")

## model using fold.vector of a previously saved model (model "mkeep")
callm_foldv <- paste("gbm.step(data=data0, gbm.x=lista_predictoras, gbm.y=vd, distribution =
as.character(familia), "  ,
                     "learning.rate=lr, tree.complexity=tc, bag.fraction = bf, max.trees =
50000, plot.main=FALSE, verbose=FALSE, " ,
                     "fold.vector=keep_fold_vector)", sep="")

##-- Lists to capture the relevant data from the models
lista_modelo  <- list()
lista_vdep  <- list()
lista_year  <- list()
lista_bf  <- list()
```

```
lista_lr  <- list()
lista_tc  <- list()
##
l_y  <- list()
##
l_cv.dev        <- list()
l_cv.dev.se     <- list()
l_cv.auc1       <- list()
l_cv.cor        <- list()
l_cv.cor.se     <- list()
l_cv.auc        <- list()
l_cv.auc.se     <- list()
l_ntrees        <- list()
l_tiempo        <- list()
l_sf.auc     <- list()
l_sf.cor     <- list()
l_sf.null    <- list()
l_sf.m.null  <- list()
l_sf.resid   <- list()
l_sf.res.dv  <- list()
l_cv.stat    <- list()
l_self.stat  <- list()
l_call       <- list()
l_contrib    <- list()

#-- Containers for crossprediction
e_auc <- c()
e_cor <- c()
e_tp <- c()
e_fp <- c()
e_fn <- c()
e_tn <- c()
y_    <- c()
d_    <- c()
l_vd  <- c()
l_nvd <- c()
e_trh<- c()
e_tss <- c()
e_sensibility <- c()
e_specificity <- c()
e_fpr <- c()

# Container for the predictions (from crosspred) to map (7 columns)
data_pr <- data_2017[,1:7]

##-- Container object resulting table of RCs (relative contributions of #the explanatory V.
of the models)ocontributions <- tibble()

#sink("Modelos_BRT_20190313_v0.txt")

year <- sort(year)  ## Order the years so that the models are executed in #a predictable
sequence (important for keep.fvector//fvector)
  print("---- Data are modeled separately for the years ")
  print( year )
  cat("\n")

for (vd in vdep){
    nomb_modelo  <- tb_vdep %>%   filter(num_vdep == vd)  %>%  dplyr::select(nombres_modelos)
%>% pull()
    nomb_vdep  <- tb_vdep %>%   filter(num_vdep == vd)  %>%  dplyr::select(nombres_vdep) %>%
pull()
    print(paste("=======----------------%%%  MODELOS  ",nomb_modelo,  "  (var_depend =", vd
,")  %%%----------------======="))
    print("---- Predictor variables are used ")
    print( ltexto_predictoras )
    cat("\n")

    for (y in year) {
        cat("\n")
        print(paste("======= model year  =====>>>    ",y, "    <<<======"))
        ## Set seed for each model corresponding to the INITIAL YEAR
        if (y==min(year) & abs(nseed)>=0  & tuning==0)  {
            set.seed(nseed)
            print(paste("Apply SET.SEED =", nseed, " al a?o INICIAL (", y, ")"))
            } else{ print(paste("No se aplica set.seed al a?o ", y)) }
        if (y==min(year) & y==max(year))  {
            print("Se obtiene el modelo de un UNICO a?o")
            } else{ print("") }
        ##Construye nombre del objeto donde se almacenar?_cada_modelo
        data0 <- eval(parse(text= paste("data_",y,sep="")))
        for (bf in bag_fracc) {
            for (lr in learn_rate){
                for (tc in tree_cmplx){
                    print(paste("-- Bag fraction: ",bf,", Learning rate: ",lr,", Tree
complexity: ",tc))
                    ##--Se construye el nombre del modelo
```

```
                    nmodelo <- paste("m_", nomb_vdep,"_",y,"_",sprintf("%0.05f",
bf),"_",sprintf("%0.04f", lr),"_",tc, sep="")
                    ##-- Se generan las listas que conformaran la tabla/dataframe donde se
almacenaran los resultados
                    lista_modelo[(length(lista_modelo) + 1)] <- nmodelo
                    lista_vdep[(length(lista_vdep) + 1)]  <- nomb_vdep
                    lista_year[(length(lista_year) + 1)]  <- y
                    lista_bf[(length(lista_bf)     + 1)]  <- bf
                    lista_lr[(length(lista_lr)     + 1)]  <- lr
                    lista_tc[(length(lista_tc)     + 1)]  <- tc
                    print(paste("Codigo del modelo ---->",nmodelo))
                    ###==========  Se ejecuta la llamada al modelo GBM
                    if (k_f_vector==1 & tuning==0 ) {
                        ## Se captura el fold vector del a?o inicial de la lista de a?os
(ejemplo 2010, si 2010)
                        if (y==min(year) ) {
                            print(paste("Mediante keep.fold.vector se captura el fold.vector
del modelo del a?o  --> ", y))
                            cat("\n-----------------------------")
                            m <- eval(parse(text = callm_keep))
                            keep_fold_vector <-  m$fold.vector
                        } else if (y>min(year) ) {
                            print(paste("Se aplica fold.vector del a?o ", min(year), " al
modelo del a?o  --> ", y))
                            cat("\n-----------------------------")
                            m <- eval(parse(text = callm_foldv))
                        } else {print("dont apply keep.fold.vector ...")}
                    } else {
                        print("The Model call is executed randomly without using
'keep.fold.vector' ni 'fold.vector'")
                        cat("\n---------------------------")
                        m <- eval(parse(text = callm))
                    }

#==================================CROSSPREDICTABILITY=====================================

                    #======================== CrossPredictability if models run ok ==
                        cat("\n-------------------------\n")
                    if (crosspredict==1) {
                        anos07 <- 2017-y   #-- numbers of years until 2017
                        sumy=0
                        for (py in 0:anos07) {
                                sumy <- sumy+1
                                my <- y + py        # years since (y+0) until (y+anos07), x
ej. if y=2011, the range (2011+0)=2011 hasta (2011+6)=2017
                                datamy <- eval(parse(text= paste("data_", my, sep="")))
                                #crpr_m <- paste0("crpr_m_", my)
                                #==========>>   The call to predict is made, creating the
container object "crpr_m"
                                crpr_call <- paste0(' predict.gbm(m, datamy,
n.trees=m$gbm.call$best.trees, type="response")')

############################==========================================================
                                if(!is.null(m$cv.statistics$deviance.mean)==TRUE) {
                                #-------------------------------------------------
                                    print(paste("Se obtiene Cross-predicts Modelo a?o ",y, "
en Datos a?o",my))
                                    crpr_m <- eval(parse(text = crpr_call))
                                    ##--------------------------
                                    ####----------- CONTAINER for the results of the
PREDICTIONS -> Columns of the df "data_pr" (for MAPS)-------------####
if (tuning==0) {
                                        var_pr <- paste0("data_pr$pr_", vd, "_m",y, "_d",my)
                                        ## The predictions are transferred to var_pr which
represents a columna del df "data_pr"
                                        eval(parse(text= paste(var_pr," <-  crpr_m")))
                                    } else {
                                        print("No predictions for maps (tuning activated)")
                                    }

                                    #---   Asses (evaluate)
                                    #-- Predictions are identified in data with a positive
and negative indicator (eg IPA>50 or IPA<=50 respectively)
                                    crpr_ipositivas <- crpr_m[datamy[,vd]==1]   # predictions
on positive data in datamy
                                    crpr_inegativas <- crpr_m[datamy[,vd]==0]   # predictions
on negative data in datamy                                  ##-------
                                    e <- evaluate(crpr_ipositivas, crpr_inegativas)
                                    ##-------
                                    l_vd[length(l_vd)+1] <- vd              #--( const inside
model) Indicates the vd (dependent variable) of the model
                                    l_nvd[length(l_nvd)+1] <- nomb_vdep     #--( const inside
model) Indicates name of the vd of the model
                                    y_ [length(y_)+1] <- paste0("m",y)     #--( const inside
model) Model year list (and 2010-and 2017)
                                    d_ [length(d_)+1] <- paste0("d",my)    #--( Var inside
model) List year of the data in which the model is crs-predicted (d2010-d2017)
```

```
                                    #-- to obtain AUC and COrrelation
                                    e_auc[length(e_auc)+1] <- e@auc     #--(Var within model)
AUC of the crs-predict of the model with data from different years >=Year of the model
                                    e_cor[length(e_cor)+1] <- e@cor
                                    #-- Get optimal threshold probability
                                    trh <- e@t[which(e@TPR+e@TNR==max(e@TPR+e@TNR))]
                                    trh <- max(trh)
                                    e_trh[length(e_trh)+1] <- trh
                                    #-- Get confusion matrix for optim threshold
                                    #    (  mc[1]=tp,  mc[2]=fp,  mc[3]= fn,  mc[4]= tn )
                                    mc <- e@confusion[which(e@TPR+e@TNR==max(e@TPR+e@TNR)), ]
                                    e_tp[length(e_tp)+1] <- mc[1]
                                    e_fp[length(e_fp)+1] <- mc[2]
                                    e_fn[length(e_fn)+1] <- mc[3]
                                    e_tn[length(e_tn)+1] <- mc[4]
                                    #--   Get Sensitivity or TRUE Positive rate (TPR) in
(treshold) optimo (tp/(tp+fn))
                                    tpr <-  mc[1] / (mc[1] + mc[3])    # <- Denominator (true
positives)
                                    e_sensibility[length(e_sensibility)+1]  <- tpr
                                    #--   Obtains Specificity or TRUE Negative rate (TNR) at
the optimal threshold (treshold)                                        tnr <- mc[4] / (mc[4] +
mc[2])    # <- Denominator (true negatives)
                                    e_specificity[length(e_specificity)+1]  <- tnr
                                    #--   Get False Positive Rate (fp/(tn+fp))
                                    fpr <- 1-tnr
                                    e_fpr[length(e_fpr)+1] <- fpr
                                    #--  Obtiene TSS (True Skill Statistic)   ## do not use
                                    num_tss <- (mc[1]*mc[4])-(mc[3]*mc[2])
                                    den_tss <- (mc[1]+mc[2])*(mc[3]+mc[4])
                                    tss <- num_tss/den_tss
                                    e_tss[length(e_tss)+1] <- tss
                                } else {
                                    l_vd[length(l_vd)+1] <- NA
                                    l_nvd[length(l_nvd)+1] <- NA
                                    y_[length(y_)+1]  <- NA
                                    d_[length(d_)+1]  <- NA
                                    e_auc[length(e_auc)+1]  <- NA
                                    e_cor[length(e_cor)+1]  <- NA
                                    e_trh[length(e_trh)+1]  <- NA
                                    e_tp[length(e_tp)+1]  <- NA
                                    e_fp[length(e_fp)+1]  <- NA
                                    e_fn[length(e_fn)+1]  <- NA
                                    e_tn[length(e_tn)+1]  <- NA
                                    e_sensibility[length(e_sensibility)+1]  <- NA
                                    e_specificity[length(e_specificity)+1]  <- NA
                                    e_fpr[length(e_fpr)+1]  <- NA
                                    e_tss[length(e_tss)+1]  <- NA
                                    print(paste("does NOT generate Cross-preds Model year
",y, " in Data year",my, "--> lr large. NA is stored"))                                  }
                                }
                    }   else{ print("Cross Predictions are not obtained between models of
different years")                        }

#=========================================================================
                    if(!is.null(m$cv.statistics$deviance.mean)==TRUE) {
                        #============================================
                        l_y[   (length(l_y)              + 1)]    <- m$gbm.call$gbm.y
#Var dependiente (extraida del modelo)
                        l_cv.dev[   (length(l_cv.dev)    + 1)]    <-
m$cv.statistics$deviance.mean          # CV deviance mean
                        l_cv.dev.se[(length(l_cv.dev.se) + 1)]    <-
m$cv.statistics$deviance.se            # CV deviance standard error
                        l_cv.cor[   (length(l_cv.cor)    + 1)]    <-
m$cv.statistics$correlation.mean       # CV correlation mean
                        l_cv.cor.se[(length(l_cv.cor.se) + 1)]    <-
m$cv.statistics$correlation.se         # CV correlation standard error
                        l_cv.auc1[  (length(l_cv.auc1)   + 1)]    <- mean(m$cv.roc.matrix)
# CV AUC mean (average folds)
                        l_cv.auc[   (length(l_cv.auc)    + 1)]    <-
m$cv.statistics$discrimination.mean    # CV AUC mean (must match above)
                        l_cv.auc.se[(length(l_cv.auc.se) + 1)]    <-
m$cv.statistics$discrimination.se      # CV AUC standard error
                        l_sf.auc[   (length(l_sf.auc)    + 1)]    <-
m$self.statistics$discrimination       # AUC mean (sin cv)
                        l_sf.cor[   (length(l_sf.cor)    + 1)]    <-
m$self.statistics$correlation          # Correlation (sin cv)
                        l_sf.null[  (length(l_sf.null)   + 1)]    <- m$self.statistics$null
# Total deviance
                        l_sf.m.null[(length(l_sf.m.null) + 1)]    <-
m$self.statistics$mean.null            # Mean total deviance
                        l_sf.resid[ (length(l_sf.resid)  + 1)]    <- m$self.statistics$resid
# Resid deviance
                        l_sf.res.dv[(length(l_sf.res.dv) + 1)]    <-
m$self.statistics$mean.resid           # Mean Resid deviance
```

```
                        l_ntrees[   (length(l_ntrees) + 1)]    <- m$gbm.call$n.trees
# Num of trees
                        l_tiempo[   (length(l_tiempo) + 1)]    <-
m$gbm.call$elapsed.time.minutes     # Tiempo

                        #l_cv.stat[  (length(l_cv.stat) + 1)]   <-
ifelse(!is.null(m$cv.statistics)==TRUE ,m$cv.statistics, NA)             #cv
                        #l_self.stat[(length(l_self.stat) + 1)] <-
ifelse(!is.null(m$self.statistics)==TRUE ,m$self.statistics, NA)             #Self

                        con <- as_tibble(m$contributions)
                        con <- add_column(con, cmodelo = nmodelo                        ,
.before="var")  # Add as first column (antes de "var")
                        con <- add_column(con, nvdep  = vd                              ,
.before="var")  # Add as first column (antes de "var")
                        con <- add_column(con, vdepend= nomb_vdep                       ,
.before="var")  # Add as first column (antes de "var")
                        con <- add_column(con, malaria_api= nomb_modelo                 ,
.before="var")  # Add as first column (antes de "var")
                        con <- add_column(con, malaria = str_sub(con$malaria_api, start= 1,
end=-8))       # Add columna derivaded "malaria"
                        con <- add_column(con, api     = str_sub(con$malaria_api, start= -6))
# Add columna derivated  "api"
                        con <- add_column(con, year   = y                               ,
.before="var")  # Add a column (antes de "var")
                        con <- add_column(con, bf     = bf                              ,
.before="var")  # Add as column (antes de "var")
                        con <- add_column(con, lr     = lr                              ,
.before="var")  # Add as column (antes de "var")
                        con <- add_column(con, tc     = tc                              ,
.before="var")  # Add as column (antes de "var")
                        con <- add columna (con, cv.auc = m$cv.statistics$discrimination.mean
, .before="var")  # Add last column (antes de "var")
                        con <- con %>%
                         #select(var) %>%
                            mutate(var1 = case_when(
                                        var == "rain"      ~"CAR"  ,
                                        var == "logpop5k"  ~"POPD" ,
                                        var == "access"    ~"TPV"  ,
                                        var == "ndwi2k"    ~"NDWI" ,
                                        var == "covfor2k"  ~"FC"   ,
                                        var == "ndvi2k"    ~"NDVI" ,
                                        var == "temp"      ~"LST"  ,
                                        var == "distriver" ~"SDR"  ,
                                        var == "loss2k"    ~"FL"   ,
                                        var == "logpop"    ~"POP"  ,
                                        TRUE               ~"otra" ) ,
                                          order = case_when(
                                         var == "rain"      ~1  ,
                                         var == "logpop5k"  ~2 ,
                                         var == "access"    ~3 ,
                                         var == "ndwi2k"    ~4 ,
                                         var == "covfor2k"  ~5 ,
                                         var == "ndvi2k"    ~6 ,
                                         var == "temp"      ~7 ,
                                         var == "distriver" ~8 ,
                                         var == "loss2k"    ~9 ,
                                         var == "logpop"    ~10  )   )
                         #-------------------------------------------------------------------
                         #------ The partial predictions of the models are obtained (logit
form: type="link" by default)
                          # those that are stored in the "pmarginal" df
                        cat("\n-------------------\n")
                        if (gpdp==1 & exists("pmarginal") & tuning==0) {
                            print("tuning=0 & existe el insumo 'pmarginal' ----> OK  ")
                            print("Se obtienen las predicciones para los PDPs ... ")
                            print("-----")
                            ##--- predictor variable name containers (For fixed model)
                            lvars_yy  <- c()
                            ##  The predictions will be stored in the table "PMARGINAL"
                             ## The use of the same values of predictors and covariates for
the estimation of the PDPs
                             ## ensures that the magnitudes of the predictions of PDPs from
different years can be compared.
                            for (pv in ltexto_predictoras){
                                lvars_yy[length(lvars_yy) + 1]  <- paste0("yy_", pv)   ##
lista de variables y_17
                               #COn objeto de obtener  marginal predicts for "vpred" Change
the names of the variables "yy_vpred" to "vpred"
                                # the original "vpred" is in turn renamed to "vpred_1", to
temporarily override it                                 names(pmarginal)[names(pmarginal)==
pv]                <- paste0(pv, "_1")
                                names(pmarginal)[names(pmarginal)== paste0("yy_", pv)] <- pv
                               # Is the text string (syntax) to call generated? to the
prediction
                                # The predictions of the 2010-2017 models are obtained
(logit form: type="link", which is? tb by default)
```

```
                                pr_call <- paste0("pmarginal$pr_",vd,"_",pv,"_",y,' <-
predict.gbm(m, pmarginal, n.trees=m$gbm.call$best.trees, type="link")')
                                print(paste0("Se obtuvo la prediccion->
pr_",vd,"_",pv,"_",y))
                               # Se ejecutan los predicts
                                eval(parse(text = pr_call))
                               # Las variables renombradas previamente retoman su nominaci?n
original
                                names(pmarginal)[names(pmarginal)== pv] <- paste0("yy_", pv)
                                names(pmarginal)[names(pmarginal)== paste0(pv, "_1")] <- pv
                            }
                          }else{print("NO predictions are obtained for PDPs || There is NO
pmarginal to store the marginal predictions of the models !!")}                    }
                     else{
                        l_y[   (length(l_y)              + 1)]    <- NA   #Var dependiente
(extraida del modelo)
                        l_cv.dev[   (length(l_cv.dev)    + 1)]    <- NA   # CV deviance mean
                        l_cv.dev.se[(length(l_cv.dev.se) + 1)]    <- NA   # CV deviance
standard error
                        l_cv.cor[   (length(l_cv.cor)    + 1)]    <- NA   # CV correlation
mean
                        l_cv.cor.se[(length(l_cv.cor.se) + 1)]    <- NA   # CV correlation
standard error
                        l_cv.auc1[  (length(l_cv.auc1)   + 1)]    <- NA   # CV AUC mean
(average folds)
                        l_cv.auc[   (length(l_cv.auc)    + 1)]    <- NA   # CV AUC mean
                        l_cv.auc.se[(length(l_cv.auc.se) + 1)]    <- NA   # CV AUC standard
error
                        l_sf.auc[   (length(l_sf.auc)    + 1)]    <- NA   # AUC mean (sin cv)
                        l_sf.cor[   (length(l_sf.cor)    + 1)]    <- NA   # Correlation (sin
cv)
                        l_sf.null[  (length(l_sf.null)   + 1)]    <- NA   # Total deviance
                        l_sf.m.null[(length(l_sf.m.null) + 1)]    <- NA   # Mean total
deviance
                        l_sf.resid[ (length(l_sf.resid)  + 1)]    <- NA   # Resid deviance
                        l_sf.res.dv[(length(l_sf.res.dv) + 1)]    <- NA   # Mean Resid
deviance
                        l_ntrees[   (length(l_ntrees) + 1)]    <- NA     # Num of trees
                        l_tiempo[   (length(l_tiempo) + 1)]    <- NA     # Tiempo
                        con <- NA
                     }
                     ocontributions <- rbind(ocontributions, con)
                     #print(contr)
}}}}}

ocrosspred <- tibble("vardep"=l_vd, "nvardep"=l_nvd, "model_year"=y_, "data"=d_, e_auc,
e_cor, e_trh, e_tp, e_fp, e_fn, e_tn, e_sensibility, e_specificity, e_fpr, e_tss)


ostatistics <- data.frame(unlist(lista_modelo),unlist(l_y),
unlist(lista_vdep),unlist(lista_year),unlist(lista_bf),unlist(lista_lr),unlist(lista_tc),
                          unlist(l_cv.dev), unlist(l_cv.auc1), unlist(l_cv.auc),
unlist(l_cv.auc.se), unlist(l_cv.cor), unlist(l_cv.cor.se),
                          unlist(l_sf.auc), unlist(l_sf.cor), unlist(l_sf.null),
unlist(l_sf.m.null), unlist(l_sf.resid ), unlist(l_sf.res.dv),
                          unlist(l_ntrees), unlist(l_tiempo))

                                       #print(lista_modelo)
names(ostatistics) <- c("modelo","vdep.n", "vdepend","year","bf","lr","tc",
                        "cv.dev", "cv.auc", "cv.auc1", "cv.auc.se", "cv.cor","cv.cor.se",
                        "sf.auc", "sf.cor", "sf.null", "sf.m.null", "sf.resid" , "sf.res.dv",
                        "n.trees","time.min")

## MARGINAL PREDICTIONS on values selected from the range of each predictor variable.
## The marginal prediction for a predictor variable "v1" is the average of the
## individual predictions (from each ccpp) for various values of v1 ( values
## equally spaced from the range of v1) keeping the others as is
## predictor variables of each CCPP. The average of these predictions for
## each ccpp for each selected value of "v1" results in the marginal effect ?
## average effect of "v1" on the response variable, which averages over the
## effects of the other predictor variables.

#-------------------------------------------------------------------------------------------
oice <-  pmarginal %>%
               dplyr::rename_at(vars(starts_with('yy_')), list(~sub('yy_', 'x_', .))) %>%
                dplyr::select(COD_OBJ, year, starts_with("x_"), starts_with("pr_"))

#-------------------------------------------------------------------------------------------

lvars_yy <- unique(lvars_yy, incomparables = FALSE)

if (pdp1==1 & exists("pmarginal") ){

     grouping_vars <- syms(lvars_yy)
     pmarginal_mean <-
         pmarginal %>%                  ### <<<<<<---------  input "PMARGINAL"
         group_by( !!!grouping_vars )  %>%
```

```
         summarise_at( vars( starts_with("pr_")), mean)

     lpdp <- pmarginal_mean %>%
       gather(key, value, -lvars_yy )  %>%
         tidyr::extract(col = key, into = c("variable", "year"), "(pr_3[4-9]_[a-
z]+[25]?k?)_(201[0-7])$") %>%
         spread(variable, value)

     l2pdp <- pmarginal_mean %>%
       gather(key, value, -lvars_yy )  %>%
         tidyr::extract(col = key, into = c("variable", "year"), "(pr_3[4-9]_[a-
z]+[25]?k?)_(201[0-7])$")

     #-------------------------------------------------------------

    olpdp <- lpdp %>%
                dplyr::rename_at(vars(starts_with('yy_')), list(~sub('yy_', 'x_', .))) %>%
                 dplyr::select(year, starts_with("x_"), starts_with("pr_"))

     #-------------------------------------------------------------

     } else{print("Predictions are not processed for PDPs ")}

##------------------------------ End  ------------------------------##
```

Codes implemented for Google Earth Engine (GEE)

Rainfall_Anual

```
//Area of study
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-
Rog');//Nuevoextend
Map.addLayer(region);
//var region = Loreto;

Map.setCenter(-75,-4, 6); //lat, long, zoom

//Definición de Trimestres
var Calendar= ee.Filter.calendarRange(1,365);
//var trimestreB = ee.Filter.calendarRange(91,180);
//var trimestreC = ee.Filter.calendarRange(181,270);
//var trimestreD = ee.Filter.calendarRange(271,366);

// Collect bands and scale
var collection = ee.ImageCollection('TRMM/3B42').select('precipitation');

var func = function(image){
  return image.clip(region);//maybe create a property to group the reducer
};

//var sum = collection.map(func).mean();

//============================================================================
//2010
//============================================================================
var FromDate='2010-01-01';
var ToDate='2010-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2010';
var RR2010t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2010 = RR2010t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2010, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2010');

//============================================================================
//2011
//============================================================================
var FromDate='2011-01-01';
var ToDate='2011-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2011';
var RR2011t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2011 = RR2011t.map(func).reduce(ee.Reducer.sum());
```

```
Map.addLayer (RR2011, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2011');

//============================================================================
//2012
//============================================================================
var FromDate='2012-01-01';
var ToDate='2012-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2012';
var RR2012t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2012 = RR2012t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2012, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2012');

//============================================================================
//2013
//============================================================================
var FromDate='2013-01-01';
var ToDate='2013-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2013';
var RR2013t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2013 = RR2013t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2013, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2013');
//============================================================================
//2014
//============================================================================
var FromDate='2014-01-01';
var ToDate='2014-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2014';
var RR2014t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2014 = RR2014t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2014, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2014');

//============================================================================
//2015
//============================================================================
var FromDate='2015-01-01';
var ToDate='2015-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2015';
var RR2015t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
print(RR2015t);
var RR2015 = RR2015t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2015, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2015');

//============================================================================
//2016
//============================================================================
var FromDate='2016-01-01';
var ToDate='2016-12-31';
//============================================================================
//TrimestreA
//============================================================================
var etiqueta='RR2016'
var RR2016t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2016 = RR2016t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2016, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2016');

//============================================================================
//2017
//============================================================================
var FromDate='2017-01-01';
var ToDate='2017-12-31';
//============================================================================
//TrimestreA
```

```
//=========================================================================
var etiqueta='RR2017'
var RR2017t = ee.ImageCollection(collection.filterDate(FromDate,ToDate).filter(Calendar));
var RR2017 = RR2017t.map(func).reduce(ee.Reducer.sum());
Map.addLayer (RR2017, {'min': 0, 'max': 3000,
'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000"}, 'RR2016');

//=========================================================================

var
RRtrimestral2010=RR2010;//.addBands(RR2010tB).addBands(RR2010tC).addBands(RR2010tD).addBands(
RR2011tA).addBands(RR2011tB).addBands(RR2011tC).addBands(RR2011tD).addBands(RR2012tA).addBand
s(RR2012tB).addBands(RR2012tC).addBands(RR2012tD).addBands(RR2013tA).addBands(RR2013tB).addBa
nds(RR2013tC).addBands(RR2013tD).addBands(RR2014tA).addBands(RR2014tB).addBands(RR2014tC).add
Bands(RR2014tD).addBands(RR2015tA).addBands(RR2015tB).addBands(RR2015tC).addBands(RR2015tD).a
ddBands(RR2016tA).addBands(RR2016tB).addBands(RR2016tC).addBands(RR2016tD);

Export.image.toDrive({
  image: RRtrimestral2010,
  description: 'TRMM_AcumuladoANual2010',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});

//=========================================================================

var RRtrimestral2011=RR2011;

Export.image.toDrive({
  image: RRtrimestral2011,
  description: 'TRMM_AcumuladoANual2011',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});

//=========================================================================

var RRtrimestral2012=RR2012;

Export.image.toDrive({
  image: RRtrimestral2012,
  description: 'TRMM_AcumuladoANual2012',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});
//=========================================================================

var RRtrimestral2013=RR2013;

Export.image.toDrive({
  image: RRtrimestral2013,
  description: 'TRMM_AcumuladoANual2013',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});

//=========================================================================

var RRtrimestral2014=RR2014;

Export.image.toDrive({
  image: RRtrimestral2014,
  description: 'TRMM_AcumuladoANual2014',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});

//=========================================================================

var RRtrimestral2015=RR2015;

Export.image.toDrive({
  image: RRtrimestral2015,
  description: 'TRMM_AcumuladoANual2015',
  scale: 27000,
```

```
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var RRtrimestral2016=RR2016;

Export.image.toDrive({
  image: RRtrimestral2016,
  description: 'TRMM_AcumuladoANual2016',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var RRtrimestral2017=RR2017;

Export.image.toDrive({
  image: RRtrimestral2017,
  description: 'TRMM_AcumuladoANual2017',
  scale: 27000,
  region: region,
  maxPixels: 1e12,
  //driveFolder: 'MODELOS_BRT',
});
```

Accessibility

```
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevo extend
Map.addLayer(region);

var dataset = ee.Image('Oxford/MAP/accessibility_to_cities_2015_v1_0');
var accessibility = dataset.select('accessibility');

var accessibilityVis = {
  min: 0.0,
  max: 41556.0,
  gamma: 4.0,
};

Map.setCenter(18.98, 6.66, 2);
Map.addLayer(accessibility.clip(region), accessibilityVis, 'Accessibility');

//Map.addLayer(accessibility, accessibilityVis, 'Accessibility');

Export.image.toDrive({
  image: accessibility,
  description: 'Accesibilidad_Geo',
  //crs: 'EPSG:32718',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//Map.addLayer(singleImage.clip(geometry.buffer(1000)), {min: 5000.0, max: 25000.0}, 'False Color -
Circle');
```

Forest 2015 (loss year)

```
//Area de Estudio
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevoextend
Map.addLayer(region);
//var region = Loreto;

var dataset = ee.Image('UMD/hansen/global_forest_change_2015');
var treeCanopyCover = dataset.select('treecover2000');
var treeCover = treeCanopyCover.clip(region);
var treeCoverVis = {
  min: 0.0,
  max: 100.0,
  palette:
      ['3d3d3d', '080a02', '080a02', '080a02', '106e12', '37a930', '03ff17'],
};
Map.setCenter(-72, -3.0, 6);
Map.addLayer(treeCover.clip(region), treeCoverVis, 'treecover2000');

Export.image.toDrive({
  image: treeCover,
  description: 'treecover2000_data2015',
  folder: "_Finales_ResOriginal",
  scale: 30,
  region: region,
```

```
  maxPixels: 1e12,
});

//*********************************************************************

var dataset = ee.Image('UMD/hansen/global_forest_change_2017_v1_5');
var lossyear16 = dataset.select('lossyear');
var lossyear16r = lossyear16.clip(region);
var lossyear16rc = lossyear16r
.remap([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17],
          [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1]);
Map.addLayer(lossyear16rc, {min: 0, max: 1}, 'lossyear2016');

Export.image.toDrive({
  image: lossyear16r,
  description: 'lossyear_data2015',
  scale: 30,
  folder: "_Finales_ResOriginal",
  region: region,
  maxPixels: 1e12,
});
```

Forest 2015

```
//Area de Estudio
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevoextend
Map.addLayer(region);
//var region = Loreto;

var Calendar= ee.Filter.calendarRange(1,365);
var FromDate='2015-01-01';
var ToDate='2015-12-31';

var Calendar= ee.Filter.calendarRange(1,365);

var dataset = ee.Image('UMD/hansen/global_forest_change_2017_v1_5');
var treeCanopyCover = dataset.select('treecover2000');
var treeCover = treeCanopyCover.clip(region);
var treeCoverVis = {
  min: 0.0,
  max: 100.0,
  palette:
      ['3d3d3d', '080a02', '080a02', '080a02', '106e12', '37a930', '03ff17'],
};
Map.setCenter(-72, -3.0, 6);
Map.addLayer(treeCover.clip(region), treeCoverVis, 'treecover2000');

Export.image.toDrive({
  image: treeCover,
  description: 'treecover2000_Data2017',
  folder: "_Finales_ResOriginal",
  scale: 30,
  region: region,
  maxPixels: 1e12,
});

//*********************************************************************
```

Forest 2017

```
var dataset = ee.Image('UMD/hansen/global_forest_change_2017_v1_5');
var lossyear16 = dataset.select('lossyear');
var lossyear16r = lossyear16.clip(region);
var lossyear16rc = lossyear16r
.remap([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17],
           [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1]);
Map.addLayer(lossyear16rc, {min: 0, max: 1}, 'lossyear2016');

Export.image.toDrive({
  image: lossyear16r,
  description: 'lossyear_Data2017',
  scale: 30,
  folder: "_Finales_ResOriginal",
  region: region,
  maxPixels: 1e12,
});

/*
Export.image.toDrive({
  image: lossyear16rc,
  description: 'lossyear2016base',
  scale: 30,
  region: Loreto,
  maxPixels: 1e12,
});
```

LST annual

```
//Area de Estudio Loreto más 5km. de borde.
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevo
extend
Map.addLayer(region);

//Definición de Trimestres
var trimestreA= ee.Filter.calendarRange(1,365);
//var trimestreB = ee.Filter.calendarRange(91,180);
//var trimestreC = ee.Filter.calendarRange(181,270);
//var trimestreD = ee.Filter.calendarRange(271,366);

// Collect bands and scale
var modisLSTday = ee.ImageCollection('MODIS/006/MOD11A1').select('LST_Day_1km');

var modLSTday = modisLSTday.map(function(img){
return
img.multiply(0.02).subtract(273.15).copyProperties(img,['system:time_start','system:time_end'
]);
});

Map.setCenter(-80,-4, 4); //lat, long, zoom

//========================================================================
//2010
//========================================================================
var FromDate='2010-01-01';
var ToDate='2010-12-31';
//========================================================================
//========================================================================
var etiqueta='LST2010'
var LST2010t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2010 = LST2010t.mean().clip(region)
Map.addLayer (LST2010, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2011
//========================================================================
var FromDate='2011-01-01';
var ToDate='2011-12-31';
//========================================================================
//========================================================================
var etiqueta='LST2011'
var LST2011t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2011 = LST2011t.mean().clip(region)
Map.addLayer (LST2011, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2012
//========================================================================
var FromDate='2012-01-01';
var ToDate='2012-12-31';
//========================================================================
//========================================================================
var etiqueta='LST2012'
var LST2012t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2012 = LST2012t.mean().clip(region)
Map.addLayer (LST2012, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2013
//========================================================================
var FromDate='2013-01-01';
var ToDate='2013-12-31';
//========================================================================
//========================================================================
var etiqueta='LST2013'
var LST2013t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2013 = LST2013t.mean().clip(region)
Map.addLayer (LST2013, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2014
//========================================================================
var FromDate='2014-01-01';
```

```
var ToDate='2014-12-31';
//=============================================================================
//=============================================================================
var etiqueta='LST2014'
var LST2014t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2014 = LST2014t.mean().clip(region)
Map.addLayer (LST2014, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//=============================================================================
//2015
//=============================================================================
var FromDate='2015-01-01';
var ToDate='2015-12-31';
//=============================================================================
//=============================================================================
var etiqueta='LST2015'
var LST2015t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
var LST2015 = LST2015t.mean().clip(region)
Map.addLayer (LST2015, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//=============================================================================
//2016
//=============================================================================
var FromDate='2016-01-01';
var ToDate='2016-12-31';
//=============================================================================
//=============================================================================
var etiqueta='LST2016'
var LST2016t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
print(LST2016t);
var LST2016 = LST2016t.mean().clip(region)

Map.addLayer (LST2016, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//=============================================================================
//2017
//=============================================================================
var FromDate='2017-01-01';
var ToDate='2017-12-31';
//=============================================================================
//=============================================================================
var etiqueta='LST2017'
var LST2017t = ee.ImageCollection(modLSTday.filterDate(FromDate,ToDate).filter(trimestreA));
print(LST2017t);
var LST2017 = LST2017t.mean().clip(region)

Map.addLayer (LST2017, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//=============================================================================
//=============================================================================

//var
LSTtrimestral20102016=LST2010tA.addBands(LST2010tB).addBands(LST2010tC).addBands(LST2010tD).a
ddBands(LST2011tA).addBands(LST2011tB).addBands(LST2011tC).addBands(LST2011tD).addBands(LST20
12tA).addBands(LST2012tB).addBands(LST2012tC).addBands(LST2012tD).addBands(LST2013tA).addBand
s(LST2013tB).addBands(LST2013tC).addBands(LST2013tD).addBands(LST2014tA).addBands(LST2014tB).
addBands(LST2014tC).addBands(LST2014tD).addBands(LST2015tA).addBands(LST2015tB).addBands(LST2
015tC).addBands(LST2015tD).addBands(LST2016tA).addBands(LST2016tB).addBands(LST2016tC).addBan
ds(LST2016tD);

var
LSTAnual2010=LST2010;//.addBands(RR2010tB).addBands(RR2010tC).addBands(RR2010tD).addBands(RR2
011tA).addBands(RR2011tB).addBands(RR2011tC).addBands(RR2011tD).addBands(RR2012tA).addBands(R
R2012tB).addBands(RR2012tC).addBands(RR2012tD).addBands(RR2013tA).addBands(RR2013tB).addBands
(RR2013tC).addBands(RR2013tD).addBands(RR2014tA).addBands(RR2014tB).addBands(RR2014tC).addBan
ds(RR2014tD).addBands(RR2015tA).addBands(RR2015tB).addBands(RR2015tC).addBands(RR2015tD).addB
ands(RR2016tA).addBands(RR2016tB).addBands(RR2016tC).addBands(RR2016tD);

Export.image.toDrive({
  image: LSTAnual2010,
  description: 'LSTAnual2010',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});
```

```
//=============================================================================

var LSTAnual2011=LST2011;

Export.image.toDrive({
  image: LSTAnual2011,
  description: 'LSTAnual2011',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2012=LST2012;

Export.image.toDrive({
  image: LSTAnual2012,
  description: 'LSTAnual2012',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2013=LST2013;

Export.image.toDrive({
  image: LSTAnual2013,
  description: 'LSTAnual2013',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2014=LST2014;

Export.image.toDrive({
  image: LSTAnual2014,
  description: 'LSTAnual2014',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2015=LST2015;

Export.image.toDrive({
  image: LSTAnual2015,
  description: 'LSTAnual2015',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2016=LST2016;

Export.image.toDrive({
  image: LSTAnual2016,
  description: 'LSTAnual2016',
  scale: 1000,
  region: region,
  maxPixels: 1e12,
});

//=============================================================================

var LSTAnual2017=LST2017;

Export.image.toDrive({
  image: LSTAnual2017,
  description: 'LSTAnual2017',
  scale: 1000,
  region: region,
```

```
  maxPixels: 1e12,
});
```

NDVI

```
//Area of study Loreto more 5km. of buffer.
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevo
extend
Map.addLayer(region);

//Defining trimestre
var Calendar= ee.Filter.calendarRange(1,365);

// Collect bands and scale
var modisNDVI = ee.ImageCollection('MODIS/006/MOD13Q1').select('NDVI');
print(modisNDVI)


Map.setCenter(-80,-4, 4); //lat, long, zoom


//============================================================================
//2010
//============================================================================
var FromDate='2010-01-01';
var ToDate='2010-12-31';
//============================================================================
//============================================================================
var etiqueta='NDVI2010';
var NDVI2010t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
print(NDVI2010t);
var NDVI2010 = NDVI2010t.mean().clip(region);
var NDVI2010d = NDVI2010.multiply(0.0001);
Map.addLayer (NDVI2010d, {'min': 0, 'max':1000, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//============================================================================
//2011
//============================================================================
var FromDate='2011-01-01';
var ToDate='2011-12-31';
//============================================================================
//============================================================================
var etiqueta='NDVI2011'
var NDVI2011t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
var NDVI2011 = NDVI2011t.mean().clip(region)
var NDVI2011d = NDVI2011.multiply(0.0001);
Map.addLayer (NDVI2011d, {'min': -1, 'max':1, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//============================================================================
//2012
//============================================================================
var FromDate='2012-01-01';
var ToDate='2012-12-31';
//============================================================================
//============================================================================
var etiqueta='NDVI2012'
var NDVI2012t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
var NDVI2012 = NDVI2012t.mean().clip(region)
var NDVI2012d = NDVI2012.multiply(0.0001);
Map.addLayer (NDVI2012d, {'min': -1, 'max':1, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//============================================================================
//2013
//============================================================================
var FromDate='2013-01-01';
var ToDate='2013-12-31';
//============================================================================
//============================================================================
var etiqueta='NDVI2013'
var NDVI2013t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
var NDVI2013 = NDVI2013t.mean().clip(region)
var NDVI2013d = NDVI2013.multiply(0.0001);
Map.addLayer (NDVI2013d, {'min': -1, 'max':1, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//============================================================================
//2014
```

```
//========================================================================
var FromDate='2014-01-01';
var ToDate='2014-12-31';
//========================================================================
//========================================================================
var etiqueta='NDVI2014'
var NDVI2014t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
var NDVI2014 = NDVI2014t.mean().clip(region)
var NDVI2014d = NDVI2014.multiply(0.0001);
Map.addLayer (NDVI2014d, {'min': -1, 'max':1, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2015
//========================================================================
var FromDate='2015-01-01';
var ToDate='2015-12-31';
//========================================================================
//========================================================================
var etiqueta='NDVI2015'
var NDVI2015t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
var NDVI2015 = NDVI2015t.mean().clip(region);
var NDVI2015d = NDVI2015.multiply(0.0001);
Map.addLayer (NDVI2015d, {'min': -1, 'max':1, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2016
//========================================================================
var FromDate='2016-01-01';
var ToDate='2016-12-31';
//========================================================================
//========================================================================
var etiqueta='NDVI2016'
var NDVI2016t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
print(NDVI2016t);
var NDVI2016 = NDVI2016t.mean().clip(region)
var NDVI2016d = NDVI2016.multiply(0.0001);
Map.addLayer (NDVI2016d, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});

//========================================================================
//2017
//========================================================================
var FromDate='2017-01-01';
var ToDate='2017-12-31';
//========================================================================
//========================================================================
var etiqueta='NDVI2017'
var NDVI2017t = ee.ImageCollection(modisNDVI.filterDate(FromDate,ToDate).filter(Calendar));
print(NDVI2017t);
var NDVI2017 = NDVI2017t.mean().clip(region)
var NDVI2017d = NDVI2017.multiply(0.0001);
Map.addLayer (NDVI2017d, {'min': 0, 'max': 40, 'palette':"0000ff,32cd32,ffff00,ff8c00,ff0000
"});
//========================================================================
//========================================================================
//var
LSTtrimestral20102016=LST2010tA.addBands(LST2010tB).addBands(LST2010tC).addBands(LST2010tD).a
ddBands(LST2011tA).addBands(LST2011tB).addBands(LST2011tC).addBands(LST2011tD).addBands(LST20
12tA).addBands(LST2012tB).addBands(LST2012tC).addBands(LST2012tD).addBands(LST2013tA).addBand
s(LST2013tB).addBands(LST2013tC).addBands(LST2013tD).addBands(LST2014tA).addBands(LST2014tB).
addBands(LST2014tC).addBands(LST2014tD).addBands(LST2015tA).addBands(LST2015tB).addBands(LST2
015tC).addBands(LST2015tD).addBands(LST2016tA).addBands(LST2016tB).addBands(LST2016tC).addBan
ds(LST2016tD);

var
NDVIAnual2010=NDVI2010d;//.addBands(RR2010tB).addBands(RR2010tC).addBands(RR2010tD).addBands(
RR2011tA).addBands(RR2011tB).addBands(RR2011tC).addBands(RR2011tD).addBands(RR2012tA).addBand
s(RR2012tB).addBands(RR2012tC).addBands(RR2012tD).addBands(RR2013tA).addBands(RR2013tB).addBa
nds(RR2013tC).addBands(RR2013tD).addBands(RR2014tA).addBands(RR2014tB).addBands(RR2014tC).add
Bands(RR2014tD).addBands(RR2015tA).addBands(RR2015tB).addBands(RR2015tC).addBands(RR2015tD).a
ddBands(RR2016tA).addBands(RR2016tB).addBands(RR2016tC).addBands(RR2016tD);

Export.image.toDrive({
  image: NDVIAnual2010,
  description: 'NDVIAnual2010',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
```

```
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2011=NDVI2011d;

Export.image.toDrive({
  image: NDVIAnual2011,
  description: 'NDVIAnual2011',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2012=NDVI2012d;

Export.image.toDrive({
  image: NDVIAnual2012,
  description: 'NDVIAnual2012',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2013=NDVI2013d;

Export.image.toDrive({
  image: NDVIAnual2013,
  description: 'NDVIAnual2013',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2014=NDVI2014d;

Export.image.toDrive({
  image: NDVIAnual2014,
  description: 'NDVIAnual2014',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2015=NDVI2015d;

Export.image.toDrive({
  image: NDVIAnual2015,
  description: 'NDVIAnual2015',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});

//============================================================================

var NDVIAnual2016=NDVI2016d;

Export.image.toDrive({
  image: NDVIAnual2016,
  description: 'NDVIAnual2016',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});
```

```
//=========================================================================

var NDVIAnual2017=NDVI2017d;

Export.image.toDrive({
  image: NDVIAnual2017,
  description: 'NDVIAnual2017',
  folder: "_Finales_ResOriginal",
  scale: 250,
  region: region,
  maxPixels: 1e12,
});
```

Population

```
//Area de Estudio
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-
Rog');//Nuevoestend
Map.addLayer(region);
//var region = Loreto;

var Calendar= ee.Filter.calendarRange(1,365);
var FromDate='2015-01-01';
var ToDate='2015-12-31';

//Map.addLayer(dataset);
var dataset = ee.ImageCollection("WorldPop/POP")
print(dataset)
Map.addLayer(dataset)
var pop = ee.ImageCollection(dataset.filterDate(FromDate,ToDate).filter(Calendar));
var population = pop.select('population');
var populationr = population.mean().clip(region);
var populationrVis = {
  min: 1.0,
  max: 10.0,
  palette: ['24126c', '1fff4f', 'd4ff50'],
};
Map.setCenter(-75, -3);
Map.addLayer(populationr, populationrVis, 'Population');

Export.image.toDrive({
  image: populationr,
  description: 'Population2015',
  folder: "_Finales_ResOriginal",
  scale: 100,
  region: region,
  maxPixels: 1e12,
});
```

Water Mask

```
var region = ee.FeatureCollection('ft:1WEJQbaw4txw5f66-lBStlILHCIvjX2OZnVYV-Rog');//Nuevo
extend
Map.addLayer(region);



////////////////////////////////////////////////////////
// Asset List
////////////////////////////////////////////////////////

var gsw = ee.Image('JRC/GSW1_0/GlobalSurfaceWater');
//print(gsw);
var occurrence = gsw.select('occurrence');
var occurrencer = occurrence.clip(region);

var occurrencerVis = {
  min: 0.0,
  max: 41556.0,
  gamma: 4.0,
};
```

```
//print(occurrencer);
Map.addLayer(occurrencer.clip(region), occurrencerVis, 'ocurrencia');

////////////////////////////////////////////////////////////
// Constants
////////////////////////////////////////////////////////////

var occurrenceVis = {
  min:0,
  max:100,
  palette: ['red', 'blue']
};

var water_mask50 = occurrencer.gt(50).unmask(0);
//Map.addLayer(water_mask50);

Export.image.toDrive({
  image: water_mask50,
  description: 'WaterMask30m',
  //crs: 'EPSG:32718',
  scale: 30,
  region: region,
  maxPixels: 1e12,
});
```

Appendix C: Curriculum Vitae

El Cardenal Street 186 apartment 102.
Santiago de Surco, Lima - Peru
Phone (+511) 240.8160
Mobile (+511) 980.502.139
elitayoan@gmail.com

Elisa Solano

Ph.D. Student, M.Sc., Secretary of a former associate of Peruvian Geographers.

CURRICULUM VITAE

Education

2016-2020 ULg-Belgium	**Student of PhD in Biomedical Sciences and Pharmaceutics** Université de Liége in cooperation with Université Catholic Louvain and Universidad Peruana Cayetano Heredia.
2011-2013 UNC - Argentine	**Master in Science (Space Applications on Early Warning and Response to Emergencies),** National Commission for Space Activities of Argentine (CONAE), Mario Gulich Institut (IG) – Universidad Nacional de Cordoba (UNC) - Argentine.
2012 FBK - Italy	**Training in informatics tools,** Predictive Models for Biomedicine & Environment (MBPA) Fondazione Bruno Kessler (FBK). Trento (Italy).
2008 UW - USA	**Short-term training in health Informatics Research,** The University of Washington. Seattle (USA).
2005-2007 UNMSM - Peru	**Master in Epidemiology,** National University of San Marcos (UNMSM) Lima - Peru.
2007 IG - Argentine	**Advanced Training School on Landscape Epidemiology,** Mario Gulich Institut (IG) - National Commission for Space Activities of Argentine (CONAE), Universidad Nacional de Cordoba (UNC) - Argentine.
2003 PUCP - Peru	**Diploma in Applied Statistics,** Pontificia Universidad Catolica del Peru (PUCP). Lima Peru.
1993-1998 UNFV - Peru	**Geographical Engineering.** Universidad Nacional Federico Villarreal (UNFV). Lima Peru.
Additional information	**Thesis to qualify for the degree of Master of Science**, in the field of Space Applications of Early Warning and Response to Emergencies. **Title:** Geospatial tools and mathematical algorithms for territorial planning and improvement of geographic access to health services, Cusco (Peru). Tutors: Dr. Cesar Furlanello (FBK / Italy) Dr. Marcelo Scavuzzo (CONAE / Argentine) http://ig.conae.unc.edu.ar/wp-content/uploads/sites/68/2017/08/2011_Solano-Elisa.pdf

Teacher experience (selection)

2021 November 11 -17 CONIDA	Spatial analysis Using QGIS in public health. National Aerospace Research and Development Commission CONIDA Lima-Peru.
From 2016 – 2021 (UPCH)	Visiting professor for the Geographical Information Systems (GIS) course at the Master's degree in Global Health - Universidad Peruana Cayetano Heredia.
2015 – 2016 UPCH	Development of Massive Open Online Course (MOOC) for the application of Geographic Information Systems in Public Health. (Developing).
November 2014 UPCH	Guest professor for the Geographic Information Systems (GIS) course. Víctor Alzamora Castro - Universidad Peruana Cayetano Heredia.
June 2001 UPCH	Guest professor for the GIS Workshop: Information for the new Millenium: An introduction to health informatics, Universidad Peruana Cayetano Heredia and Washington University, Lima - Peru.
July 2003 DIGESA	Guest lecturer for GIS application in the epidemiological study of the baseline prevalence of respiratory diseases in children aged 3 to 14 years associated with air quality - DIGESA - Peru.
September 2000 USIL	Guest Professor of Geographic Information Systems in Health. Specialization Program in Field Epidemiology (PREC). Stage IV Prevention and Control. San Ignacio de Loyola University. Lima - Peru.
April to December 2000 UNFV	Professor assistant to the course Human and Economic Geography and Environmental Engineering at the Universidad Nacional Federico Villarreal.

Professional experience

2016 NIH	Specialized technical assistance in the preparation of ethiological maps for public health events, aimed at professionals from the Public Health Intelligence System (SISP) of the Peruvian National Institute of Health (NIH).
2015 INS	Application of georeferencing systems using GIS, for interactive consultations and analysis of geographic information for the reporting of information on public health events, at the laboratories of the regional and national levels of the Peruvian National Institute of Health (NIH).
2015 MoH	Geospatial mapping of indicators and capacity building of the personnal of the National Health Strategy for the Prevention and Control of Damages from Non-communicable diseases. National Center for Epidemiology, Disease Prevention, and Control (CDC - MINSA).
2015 MoH	Exploration of tabular and spatial data and preparation of comparative tuberculosis maps for the years 2013 and 2014 by health facilities and by TB types, through the kernel density technique. Ministry of Health - Peru.

2014 CESTAS	Application of the Geographic Information System (GIS) on health in an emergency situation and public health events in border areas between Ecuador, Colombia, and Peru, using Opensource GIS. CESTAS/ Andean Health Organization ORAS-CONHU.
2014 CESTAS	Application of the Geographic Information System (GIS) on health in emergency situation and public health events in border areas between Bolivia and Peru, using Opensource GIS. CESTAS/ORAS-CONHU
2013 MIDIS	Development of a proposal for indicators of geographic isolation at the population in villages level - Ministry of Social Inclusion (MIDIS - Peru).
2009-2011 ORAS/CONHU	Coordination of the PASAFRO.ORAS / CAF project to fulfill specific objective: Diagnosis and analysis of the health situation of border villages (Venezuela, Colombia, Ecuador, Peru, Bolivia, Chile). Andean Health Organization – Convenio Hipolito Unanue.
2009 PARSALUD	Implementation of Geographic Information System - GIS for the Health Reform Support Program (PARSALUD II).
2006-2008 PAMAFRO	Responsible for the implementation and development of geographic information systems (GIS) with a community vision for Malaria control Project in the frontiers of (Venezuela, Colombia, Ecuador, Peru). -PAMAFRO / ORAS-CONHU
2006 WVI.	Development of an information system for health and nutritional food safety indicators for World Vision International projects.
2004-2006 NIH	Collaboration and coordination for research with a spatial approach and application of Geographic Information Systems, the flow of activities, staff training, and monitoring of indicators for the National Institute of Health (NIH). Lima - Peru.
2000-2004 (CDC-MINSA)	Responsible for the implementation, supervision, and strengthening and development of the Geographic Information Systems (GIS) of the Executive General Directorate of Health Situation Analysis (ASIS) of the National Center for Disease Control (CDC) of the Ministry of Health (MINSA) Lima Peru.
1999 MINSA	Professional practices digitizing national cartography in the Health Services Strengthening Program (PFSS) and preparation of epidemiological maps in the North Lima Health Directorate - MINSA - Peru.
1998 (enero a marzo) CTAR - Ancash	Specialist in Cartography Temporary Regional Administration Council (CTAR - ANCASH). Scheme of Territorial Organization of the provinces of Marañon, Mariscal Luzuriaga, and Bolognesi - Department of Ancash. Transitory Council of Regional Administration (CTAR - ANCASH).

Language skills

Spanish	Native language
English	Speaking, reading and writing
Italian	Speaking and reading
Others	Quechua (Some areas of Ancash)

Computer skills

Operative Systems	Windows, Linux
Microsoft Office	Word, Excel, Power Point, Project.
Software GIS comercial y Open Source	ArcGIS, Maptitude, AutoCAD Map, MapInfo, SIGEPI, otros. GVSIG, GRASS GIS, Quantum GIS, Geoda, SpatStat, CrimeStat, and others.
Remote sensing software	IDRISI, ENVI, Nest. Google Earth Engine.
L. programación	IDL
Software estadístico	R, SPSS
Typesetting system	LaTeX
References	Mendeley, JabRef

Selection of oral presentations

October 8 al 10 - 2019 Davis, California USA	Malaria risk assessment and mapping using satellite imagery and boosted regression tres (BRT) in the Peruvian Amazon. Geovet 2019 conference.
Noviembre 2018 Liège, Belgium	Malaria risk assessment at the local level using satellite imagery and Boosted Regression Trees (BRT) in the Peruvian Amazon. Doctoral School Day on Public Health at Université de Liège. Belgium.
November 2018 Liège, Belgium	Geospatial tools, to improve geographic access to health services. Cusco - Peru. Spatial epidemiology symposium at Argentine Congress of Zoonosis, La Plata - Argentine.
June 2014 La Plata, Argentine	Geographic Access to Health Services, from the internship period in Italy 2012. for the third International School for Advanced Training in Panoramic Epidemiology "Dra. Sonia Blanco", Mario Gulich Institute National Space Commission Argentine Cordoba - Argentine.
May 2013 CONAE Cordoba, Argentine	Analysis of Malaria in the department of Loreto - Peru, using LANDSAT and MODIS images. Mario Gulich Institute, Argentine National Space Commission (CONAE). Cordoba - Argentine
June 2007 CONAE Cordoba, Argentine	Space-time analysis of the dengue outbreak in the district of Comas, Lima - Peru. Workshop on Space Applications to human health for the benefit of the countries of America and the Caribbean, Cordoba - Argentine.
November 2006 Cartagena, Colombia	Regional analysis of the spread of Dengue in South America. XII SELPER Colombia. Cartagena de Indias - Colombia.
Septiembre 2005 Cordoba, Argentine	Análisis Espacio – Tiempo del brote de dengue en el distrito de Comas, Lima - Perú. Workshop on Space Applications to human health for the benefit of the countries of America and the Caribbean, Cordoba - Argentine.

Selección de participación a eventos científicos

May to June 2007 Córdoba, Argentine	Advanced Training School on Landscape Epidemiology, at Instituto Mario Gulich National Space Commission Argentine Cordoba – Argentine.
November 2006 Cartagena, Colombia	SELPER Colombia Geomatics Meeting, Cartagena de Indias - Colombia
Septiembre 2005 Córdoba - Argentine	a). Course on applications of space technology to human health for the benefit of Latin American and Caribbean countries, Mario Gulich/CONAE Institute, Cordoba – Argentine. b). Workshop on applications of space technology to human health for the benefit of Latin American and Caribbean countries, Mario Gulich/CONAE Institute, Cordoba – Argentine.
November 2004 Santiago de Chile	SELPER Latin American Society in Remote Sensing a Spatial Information Systems Meeting. SELPER chapter Chile.
November 2003 Rio de Janeiro - Brazil	Interamerican Workshop on the use of Remote to control, infectious diseases. Fundacao Oswaldo Cruz (Fiocruz), National Institute of Spaciais Research (INPE) -Rio de Janeiro, Brazil.

Selección de Publicaciones

Octubre 2019 Nature Scientific Reports	Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. https://www.nature.com/articles/s41598-019-51564-4
Octubre 2008 UW – USA	Audiovisual material in Spanish, ten videos for the use of Maptitude. http://www.caliper.com/training/maptitudespanishtutorials.htm. Supervisor: Dr. Dick Hoskins, University of Washington (Seattle-USA).
Revista peruana de medicina experimental y salud publica 2009	Climate and healthy climate in the Andean region, Oscar Feo, Elisa Solano, Luis Beingolea, Marilyn Aparicio, Mario Villagra, María José Prieto, Jairo García, Patricia Jiménez, Óscar Betancourt, Marcelo Aguilar, Johannes Beckmann, María del Carmen Gastañaga; Alejandro Llanos - Cuentas, Ana Elisa Osorio, Raul Silveti.
SELPER 2006 Journal SELPER	Regional analysis of the dissemination of Dengue in South America. Rotela1 C., Quevedo2 S., Solano3 E., Lamfri1 M., Scavuzzo1 C.M. Latin American Society in Remote Perception and Space Information Systems.
Mayo 2004 CDC - MINSA	Development of thematic maps for Health Indicators, National and Regional and Health Situation Analysis (ASIS) Center for Epidemiology, Disease Prevention and Control (CDC - MINSA).
Enero 2001 CDC - MINSA	Material Pedagógico Virtual de los Sistemas de Información Georeferencial (SIG) con aplicación al Programa de Especialización en Epidemiología de Campo (PREC). Dirección General de Epidemiología.

Other researches

2012 Remote Sensing to Assessment Damages Post Earthquakes, focusing in Urban Structures - CONAE – Argentine. Url: http://aulavirtual.ig.conae.gov.ar/moodle/mod/page/view.php?id=194

2012 Detection of areas with oil spillage, from the use of radar imagery (SAR) in a Section of the Galician Sea using Open Source (Nest) for remote sensing CONAE - Argentine.

2011 Geographic access to health services, use of mathematical algorithms and geospatial tools with free software.
Predictive Models for Biomedicine & Environment (MBPA) Fondazione Bruno Kessler (FBK) - Italy.

2008 Analysis of the spatial distribution of chronic malnutrition at the community level for the prioritization of vulnerable groups. Url: http://eventos.cicese.mx/10/selper/pub/revisor.php?&idioma=Eng&login=&op=lP&tipo=66&modulo=paper&idEvento=2

2006 Field trial for comparison of three strategies for the control of Aedes aegypti and the Dengue virus in Lima. National Institute of Health (INS) Lima – Peru. Georeferencing of productive hatcheries of Aedes aegypti within an intervention of larval chemical control DISA III Lima Norte. National Institute of Health (INS) Lima - Peru.

ACTIVITIES DURING THE Ph.D

Publications

1. **Article:** Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon
Authors: Elisa Solano-Villarreal, Walter Valdivia, Pearcy Morgan, Catherine Linard, J. J. Pasapera-Gonzales, Philippe Lejeune, Alejandro Llanos-Cuentas, Niko Speybroeck, Marie-Pierre Hayette, Angel Rosas-Aguirre.
Journal: https://www.nature.com/articles/s41598-019-51564-4
Average IF for the last five years: 4.96
Citations 2020: 20 in total 15 web services
CiteScore: This article is in the 79th percentile

2. **Abstract: Malaria risk assessment at the local level using satellite imagery and Boosted Regression Trees (BRT) in the Peruvian Amazon.**
Elisa Solano-Villarreal, Walter Valdivia, Catherine Linard, J. J. Pasapera-Gonzales, Philippe Lejeune, Alejandro Llanos-Cuentas, Niko Speybroeck, Marie-Pierre Hayette, Angel Rosas-Aguirre. Archives of Public Health 2019, 77(Suppl 1):7 https://doi.org/10.1186/s13690-019-0333-5.

Oral presentations

1. **Elisa Solano-Villarreal,** Malaria risk assessment and mapping using satellite imagery and boosted regression tres (BRT) in the Peruvian Amazon. Geovet 2019 conference. October from 8 to 10 – 2019 Davis, California, United States of America

2. **Elisa Solano-Villarreal**, Assessment of the risk of malaria in the Peruvian Amazon.
WORKSHOP: PERUSAT-1: Lessons learned - 2019 celebrating the third year of launch of the Peruvian satellite Perusat. November from26 to 28 (2019), CNOIS-PUCUSANA. Lima – Peru.

3. **Elisa Solano-Villarreal,** Walter Valdivia, Pearcy Morgan, Catherine Linard, J. J. Pasapera-Gonzales, Philippe Lejeune, Alejandro Llanos-Cuentas, Niko Speybroeck, Marie-Pierre Hayette, Angel Rosas-Aguirre. Malaria risk assessment and mapping using satellite imagery and boosted regression tres (BRT) in the Peruvian Amazon. November 13th., 2018. At the University of Liege, Belgium.

4. **Elisa Solano-Villareal**, Facundo Casasola. Use of Spatial Information in Optimization Problems. Workshop: Exploring New Frontiers in the Processing and Analysis of Images in Remote Sensing. Dates: June 27, 28, and 30, 2018. Presented by: Eduardo Romero – CONAE (Argentina).

Courses attended

1. **Longitudinal surveys, that follow individuals over several periods, are becoming increasingly popular in social sciences.** Université Catholique de Louvain. Panel-data analysis using Stata - 3 semi journeys May 18th and 19th. 2017 (S. Fontenay and A. Guillet).
Longitudinal surveys, that follow individuals over several periods, are becoming increasingly popular in social sciences. The main reason is that they offer a particular setting that researchers can use to control for some types of unobserved variables bias.

1. **Course online. PostgreSQL.** Two days of 8 hours August 2018**.** *PostgreSQL is a general purpose and object-relational database management system.*

2. **Using software R for modeling and distribution of species and introduction to de use of super-computer of IIAP.** Research Institute of Peruvian Amazon (IIAP) 3 days 21 academic hours, December 5th. to 7th. 2017. *The objective of the course is to disseminate the new techniques of information processing from remote sensors, with emphasis on modeling the distribution of species and the use of computer clusters.*

3. **Course online of Google Earth Engine** – A planetary-scale platform for Earth science data and analysis. January 2018. https://www.edmodo.com/post/663462590/ *A planetary-scale platform for Earth science data & analysis. Enables users to compute petabytes of data on the fly without having them navigate the complexities of cloud-based parallelization. Enhancing inclusive access has spurred the growth of environmental projects conducted at scales previously unimaginable.*

4. **Analysis of satellite images with the environment Google Earth Engine (GEE).** September 17 to 21, 2018 from 09 to 18 hours. Place: Faculty of Natural Sciences of the National University of Tucumán – Argentina. *Required intermediate knowledge of remote sensing and spatial information. (i.e. knowledge of image corrections, geographic projections, and basic digital classification algorithms) minimum knowledge to read and elaborate command lines in programming with objects (the GEE environment works in Javascript). The program is oriented to a community of students that combines advanced and postgraduate students from different careers, in addition to professionals with a technical profile oriented to work in government agencies, NGOs, and private businesses.*

Teaching courses

1. **Spatial analysis Using QGIS in public health.**
National Aerospace Research and Development Commission CONIDA Lima- Peru. Understanding geographical aspects in public health. Manage GIS tools for the generation of new maps.
Apply statistics and geostatistics for data analysis and representation. Use of raster data for the solution to problems related to Health Sciences. From November 11 to 17, 2021.

2. **Geographic Information System workshop** using QGIS 2.18 Las Palmas. *Course to work with vector and raster geographic information focused on operational research and Epidemiological data analysis for professionals from the Regional Directorate of Health Loreto, CIRCULOS project, and Peruvian Amazon University.* December 12th. to 19th. 2016.

3. **Course in Geographic Information System for Bioinformatics Master on Global Health at the University of Cayetano Heredia.** Two weeks from February 6th. to February 17th. 2017.

4. **Massive online open courses MOOC Geographic Information System on Public Health using QGIS 2.18 Las Palmas.** For TED Miriada-plataform and Universidad Cayetano Heredia. Four modules in Spanish. July 2017

5. **Course in Geographic Information System for Bioinformatics Master on Global Health at the University of Cayetano Heredia.** Two weeks, from January 29th. to February 09th. 2018.

6. **Course of Geographic Information System for Bioinformatics Master on Global Health at the University of Cayetano Heredia.** Two weeks, from January 29th. to February 09th. 2019.

Short fellow trainee

National Aerospace Research and Development Commission CONIDA Lima – Peru.
The National Commission for Research and Aerospace Development is the institution that addresses the research and development of aerospace activities in Peru and hosted the Peruvian Space Agency.
Training in semi-automatic classification into algorithm QGIS V.2.18 to search, download, and pre-process satellite imagery. June 26th – July 7th, 2017.