Up-To-Date Malaria Mapping: Prediction From Remote Imagery

Samuel Albanie
Dept. of Engineering
Oxford University

Stephen Roberts
Dept. of Engineering
Oxford University

***Abstract*—The creation of accurate, up-to-date malaria maps remains a key challenge in combatting the global threat posed by the presence of the disease. With limited up-to-date local site data available, the best use must be made of climate and human population information that has become available in near real-time through remote sensing techniques. In the past two decades, health surveys have gathered data on the prevalence of the deadly *plasmodium falciparum* (*pf.*) malaria parasite in a number of countries battling the disease. In this paper, prevalence datasets are used in combination with climate and human population data to investigate relationships between environmental factors and malaria prevalence. These relationships are then used to guide the development of Gaussian Process models for malaria prevalence prediction across three countries: Cambodia, Somalia and Mozambique. Finally, it is shown that vegetation volatility can be engineered as a simple feature to improve the accuracy of predictive models based on environmental data.**

I. Introduction

With an estimated 584,000 deaths caused by the disease globally in 2013 [33], the battle against malaria represents a significant health challenge. Malaria is both preventable and curable and it can be combatted effectively when preventative measures and treatment are provided where they are needed in a timely manner. For resources to be deployed effectively, accurate and up-to-date mapping of the disease is in great demand.

Malaria is most prevalent in many of the less economically developed regions of the world where accurate data about the distribution of *Anopheles* (the malaria carrying mosquito) populations is scarce. In contrast, measurements of other factors such as human population density and environmental variables (humidity, air temperature, cloud cover etc.) are readily available at high spatial resolutions across the globe through the use of satellite-based remote sensing techniques [5] but the relationships between these variables and malaria prevalence are complex and can vary from region to region. For areas in which the cost of making regular measurements of mosquito populations (a process which involves the use of fly traps and extensive human labour) is prohibitive, accurate malaria mapping requires localized models that take account of these changing relationships.

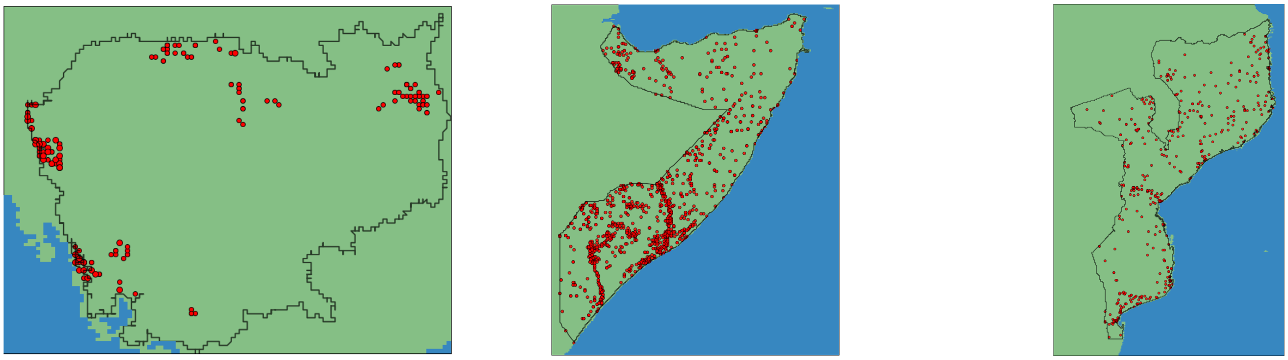
In this work, Gaussian Process Regression is used to develop environment- and population density-based models for the prevalence of the *plasmodium falciparum* malaria parasite in three countries with high clinical burdens: Cambodia, Somalia and Mozambique. Next, the strengths of the relationships between malaria prevalence and various environmental factors are investigated using straightforward correlation analysis and *Automatic Relevance Determination* kernels. Finally, vegetation volatility is incorporated as an additional feature into the GP models yielding improvements in predictive accuracy.

II. Related Work

Climatic and environmental factors have received significant attention as potential indicator signals for the prevalence of malaria. Early work used rainfall, temperature records and the *Normalized Difference Vegetation Index* (NDVI) to forecast monthly malaria cases with a degree of success [11]. It was found that in certain regions malaria epidemics were often preceded by periods of abnormally high minimum temperatures [1] indicating the potential of temperature as a predictor variable. For this reason, the impact of global warming on malaria endemicity was investigated in [9] which concluded that while temperature is a relevant factor, it has less influence on malaria prevalence than interventions by health organizations. Strong correlations were found between peak rainfall and peak malaria incidence two to three months later in [15]. Other factors that have been considered as potential predictors of malaria prevalence include the presence of village health workers [19], immunological factors and migration [28]. An extensive review and categorization of malaria forecasting research is provided in [36].

Prediction and interpolation models based on Gaussian Processes regression have become popular for working with geographical information because they work naturally with spatial data and allow the confidence of model predictions to be estimated [27]. These models have been applied to the problem of malaria mapping and shown to offer improved prevalence interpolation over conventional logistic regression models at a local level [16]. The Gaussian Process framework can also be extended to include other types of information at different spatial and temporal resolutions as inputs to the model by using appropriate covariance structures [25].

Much of the literature has focused on the problem of predicting malaria prevalence in a single country or region. In this work, prediction models are developed in tandem for three countries to facilitate an investigation of geographical variation in the influences of different environmental covariates. Feature engineering remains underdeveloped in previous research and is investigated here as a possible approach to improve the predictive accuracy of the environment-based models.



(a) Cambodia (b) Somalia (c) Mozambique

Fig. 1: *Plasmodium falciparum* survey locations, 2000-2014

III. Data

A. Malaria prevalence surveys

The prevalence data consists of surveys of the *plasmodium falciparum* malaria parasite across Cambodia [14] [6] [12] [7], Somalia [26] [31] [34] and Mozambique [18] [8] taken between 2000 and 2014. The locations of these surveys are illustrated in Figure 1. It is worth noting that the surveys are far from uniformly distributed and are considerably sparser in Cambodia than in Somalia. In total, 220 surveys were conducted in Cambodia, 1591 in Somalia and 475 in Mozambique. Note that the number of surveys does in each country does not necessarily indicate the severity of malaria in that region, as demonstrated by Table I (Mozambique has a significantly higher incidence rate but far fewer surveys than Somalia).

B. Population Density Data

Population density data for each country is derived from the 2010 *worldpop* maps for Cambodia [35], Somalia and Mozambique [17] (and adjusted to meet UN estimates). These values are then scaled using the world bank annual estimates of the populations of each country between 2000 and 2014 [3].

C. Environmental Data

The environmental factors used in the models comprise four datasets: Enhanced Vegetation Index (EVI) [13] [21], air temperature [30] [20], cloud cover [2] [23], and the ratio of actual evapotranspiration to potential evapotranspiration (AET:PET ratio) [24] [22]. The Enhanced Vegetation Index is a measure of the vegetation in a region. Cloud cover is a measure of the portion of the sky occluded by clouds and is commonly used as a proxy for relative insolation. Air temperature refers to average monthly temperature at the surface (measured in degrees Celsius) and *AET:PET ratio* is used as a measure of aridity. *EVI*, cloud cover and *AET:PET ratio* take real values between 0 and 1. Each of the datasets consist of monthly time-series from 2000 to 2014 at 0.05 degree resolution.

Country	Population (million)[1]	Land Area (km^2)	Malaria cases [2]
Cambodia	15.14	181,000	21,309
Mozambique	25.83	802,000	2,998,874
Somalia	10.5	638,000	10,470

TABLE I: Country statistics

IV. Methods

A. Gaussian Processes Regression

Gaussian Process Regression (GPR) is a form of supervised learning that fits a Gaussian Process to a collection of inputs $\{\mathbf{x}_n\}$ and targets $\{t_n\}$. A Gaussian Process (GP) is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [29]. A GP, $\mathcal{GP}(\mu(\mathbf{x}), K(\mathbf{x}, \mathbf{x}'))$, is uniquely determined by its mean function $\mu(\mathbf{x})$ and its covariance function $K(\mathbf{x}, \mathbf{x}')$ which are defined for a real process $\mathbf{g}(\mathbf{x})$ by:

$$\mu(\mathbf{x}) = \mathcal{E}[\mathbf{g}(\mathbf{x})] \tag{1}$$

$$K(\mathbf{x}, \mathbf{x}') = \mathcal{E}[(\mathbf{g}(\mathbf{x}) - \mu(\mathbf{x}))(\mathbf{g}(\mathbf{x}') - \mu(\mathbf{x}'))] \tag{2}$$

The choice of covariance function has a significant influence on the character of the resulting process and its response to any discontinuities that may occur in the data. A popular choice of covariance function is the RBF kernel given by $K(\mathbf{x}, \mathbf{x}') = \exp(-c||\mathbf{x} - \mathbf{x}'||)$. By working with vector inputs the GP can combine multiple input variables (such as spatial information and environmental factors) to produce a predicted target variable value (in this case, malaria prevalence).

A measure of the predictive quality of each environmental factor can be determined by training and evaluating single-input GPs on each dataset. However, a more informative evaluation of the relative importance of each of variables can be obtained using Automatic Relevance Determination (ARD) [4]. This is achieved using an augmented kernel [10] (continuing with the RBF kernel as an example):

[1] World Bank 2013 midyear estimates[3]

[2] Reported confirmed cases, WHO World Malaria Report [32]

$$K(\mathbf{x}, \mathbf{x}') = \exp\left[-\sum_{a=1}^{d} \frac{(x^a - x'^a)^2}{2l_a^2}\right] \tag{3}$$

where d is the number of input dimensions and each dimension a separate lengthscale l_a. These lengthscales are then tuned as part of the process of hyperparameter optimisation. Inputs which exert greater predictive influence are given shorter lengthscales and the regression function is then more sensitive to data in the dimension of that input.

V. Correlation Analysis

Prior to developing prediction models, it is useful to explore the data for potential monotonic or linear relationships between each of the input variables and malaria prevalence. using *Spearman's rank* and *Pearson product-moment* correlation coefficients. The correlation coefficients and corresponding p-values for the most significant correlations (p-value less than 0.05) are given in Table II.

Cambodia				
Factor	Spearman	p-value	Pearson	p-value
AET:PET(t)	0.363	$4.27 \cdot 10^{-7}$	0.209	0.004
Pop density	-0.353	$9.65 \cdot 10^{-7}$	-0.187	0.011
Temp$(t-1)$	-0.319	$1.061 \cdot 10^{-5}$	-0.278	0.0001
EVI(t)	0.270	0.0002	0.240	0.001
EVI$(t-1)$	0.269	0.0002	0.260	0.0004
Somalia				
Factor	Spearman	p-value	Pearson	p-value
AET:PET$(t-3)$	0.260	$1.96 \cdot 10^{-14}$	0.289	$1.43 \cdot 10^{-17}$
EVI$(t-3)$	0.219	$1.46 \cdot 10^{-10}$	0.239	$2.41 \cdot 10^{-12}$
Mozambique				
Factor	Spearman	p-value	Pearson	p-value
Pop density	-0.238	$2.55 \cdot 10^{-6}$	-0.22	$1.45 \cdot 10^{-5}$
EVI(t)	0.235	$3.28 \cdot 10^{-6}$	0.204	$5.56 \cdot 10^{-5}$
EVI$(t-1)$	0.226	$7.81 \cdot 10^{-6}$	0.212	$2.80 \cdot 10^{-5}$
EVI$(t-2)$	0.231	$4.94 \cdot 10^{-6}$	0.215	$3.00 \cdot 10^{-5}$
AET:PET(t)	0.239	$2.25 \cdot 10^{-6}$	0.228	$6.83 \cdot 10^{-5}$

TABLE II: Correlation coefficients for input variables. An index of $(t-i)$ indicates the value of the environmental variable i months prior to the event of the parasite prevalence survey.

Interestingly, the composition of variables that exhibit correlation with prevalence is very different in each country. It was noted in previous studies that rainfall two to three months prior to the parasite survey correlated with malaria prevalence and this appears to be supported by the results for Somalia in which the most significant relationships were with EVI and AET:PET ratio (proxies for vegetation and humidity) time-lagged by three months. However, very different factors emerge as significant for Cambodia and Mozambique, in particular population density which has not received as much attention in the literature. While no single factor exhibits a particularly strong correlation with malaria prevalence, it is promising that in each country multiple variables are signficantly weakly correlated with the target variable.

The correlation coefficients can now be compared to the results of fitting an ARD kernel to the data, shown in Table III. The factors are ranked in order of shortest lengthscale (most relevant) for each country. For comparison, random noise is also included as an input (any factor which is relevant should have a lengthscale significantly shorter than the lengthscale selected for the noise).

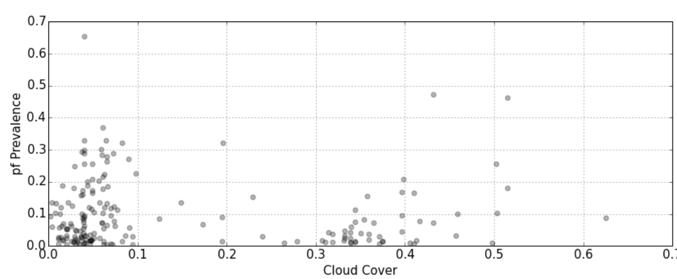
For Cambodia, cloud cover from the previous month proves to be the most relevant factor in a manner that appears at odds with the correlation values. However, we see from Figure 2 that while there is a pattern present in the data, the relationship is neither linear or monotonic (as a consequence of the wet and dry seasons in Cambodia) and so this relationship does not appear in the correlation analysis. Furthermore, in contrast to the correlation values population density is almost completely disregarded as irrelevant by the GP in both Cambodia and Mozambique with lengthscales much larger than those generated by random noise inputs. Despite these outcomes, there is agreement on the importance of other factors such as *EVI*$(t-1)$. Somalia produces more consistent results with both *AET:PET*$(t-3)$ and *EVI*$(t-3)$ viewed as relevant by the ARD kernel.

Cambodia			
Factor	Lengthscale	Factor	Lengthscale
Clouds$(t-1)$	0.08	*EVI*(t)	11.648
EVI$(t-2)$	0.511	*AET:PET*(t)	13.271
EVI$(t-1)$	0.570	*Random Noise*	26.010
EVI$(t-3)$	4.320	*Pop density*	358.134
Somalia			
Factor	Lengthscale	Factor	Lengthscale
AET:PET$(t-3)$	0.313	*Clouds*$(t-1)$	0.802
Clouds$(t-2)$	0.513	*Clouds*(t)	1.233
EVI$(t-3)$	0.794	*Random Noise*	24.093
Mozambique			
Factor	Lengthscale	Factor	Lengthscale
Clouds$(t-3)$	0.362	*AET:PET*$(t-2)$	0.948
AET:PET(t)	0.499	*AET:PET*$(t-3)$	3.804
AET:PET$(t-1)$	0.611	*Random Noise*	53.942
EVI$(t-1)$	0.670	*Pop density*	1407.969

TABLE III: ARD lengthscales for input variables

In summary, it is clear from both the correlation analysis and ARD kernels that each country will need a model with carefully tailored inputs to be effective in forming accurate predictions.

Fig. 2: Cloud cover vs. *pf.* prevalence, Cambodia, 2000-2014



VI. Prevalence Prediction

To evaluate models, 40% of the data in each of the three countries was held out for testing and the remaining 60% was used for training. The first predictive models were developed using individual environmental factors. It was noted in previous work [15] that there is often a lag of two to three months between an environmental factor and a corresponding change in malaria levels. Consequently, measurements of the three months of the variable up to (and including) the month in which the survey was taken are provided as inputs to a Gaussian Process. Of the several forms of covariance function that were considered, the *Matérn* family of kernels often performed best, perhaps due to the sharply varying nature of the malaria prevalence data. The performance (measured in terms of *Normalized Root-Mean-Square-Error*) of GP predictors trained on individual environmental factors is given in Table IV.

Country	Air Temp	Cloud Cover	AET:PET	EVI	Pop Density
Cambodia	0.220	**0.204**	0.224	0.218	0.231
Somalia	0.154	**0.150**	**0.150**	0.155	0.165
Mozambique	0.243	**0.241**	0.249	0.253	0.263

TABLE IV: *NMRSE* scores for individual predictive factors

Cloud cover proves to be the most useful individual factor for making predictions across each of the three countries (jointly with AET:PET for Somalia). Unsurprisingly, these results align much more closely with the ARD kernel rankings of relevance than the correlation analysis (recalling that the cloud cover data was neither linear nor monotonic).

A. Fused Factors

Instead of building models around environmental factors, an alternative approach is to fuse these together as grouped inputs to the GP. Three techniques were used to combine factors:

- *All* - combines the locations of the surveys, population density and the three month history of each environmental variable.
- *TopCor* - combines the locations of the surveys with the factors that exhibited the strongest correlation with prevalence.
- *TopARD* - combines the locations of the surveys with the factors that were ranked as most relevant by the ARD kernels.

For comparison, a benchmark prediction score is calculated using multivariate polynomial regression over the survey locations. The order of the polynomial was optimised using 5-fold cross validation on the training data. The predictive accuracies of the fused models and the benchmark prediction models are given in Table V.

The most accurate models were those formed by combining the factors ranked as most relevant by the ARD kernels. The improvement in accuracy is significant for Cambodia, but is no better than using the best individual factor in Somalia. A small improvement of the benchmark predictor is achieved in Mozambique. Combining all possible factors and using them as inputs to a GP does not perform well, perhaps because there is not enough data to be effective in the higher number of input dimensions.

Country	Single	Poly	ALL	TopCorr	TopARD
Cambodia	0.204	0.179	0.206	0.207	**0.157**
Somalia	**0.150**	0.157	0.164	0.151	**0.150**
Mozambique	0.241	0.238	0.253	0.249	**0.234**

TABLE V: Comparison of *NMRSE* scores for factor combinations. *Poly* represents the multivariate polynomial regression predictions. *Simple* represents the best scores achieved by a single factor model.

B. Volatility

The previous results suggest that the GP is not making sense of all the available raw data. It is possible that it will make better sense of more abstract representations of the data. One approach is to use the ”volatility” of an environmental factor as an input. Define the *volatility* of an environmental factor, $ENV(t)$, as:

$$\mathrm{Vol(ENV)} = \begin{cases} 1 & \text{if } \frac{1}{3}\sum_{i=0}^{2} \mathrm{Diff(ENV}, i) > \mathrm{AvgDiff(ENV)} \\ 0 & \text{if otherwise} \end{cases} \quad (4)$$

where we define AvgDiff(ENV) to be the average of the absolute differences between time lagged values of the ENV variable over all training values and the $\mathrm{Diff}(\cdot)$ function is defined by:

$$\mathrm{Diff(ENV}, i) = |\mathrm{ENV}(t-i) - \mathrm{ENV}(t-(i+1))| \quad (5)$$

The intuition behind this feature is that rapidly changing environmental factors might make it more difficult for the mosquito populations that carry the *.pf* parasite to survive.

Country	TopARD	TopARD + Vol(AET:PET)
Cambodia	0.157	**0.156**
Somalia	**0.150**	**0.150**
Mozambique	0.234	**0.227**

TABLE VI: *NMRSE* scores for volatility

Of the four possible environmental factors, *AET:PET* volatility as a factor yielded improvements in predictive accuracy for Cambodia and Mozambique when combined with the *TopARD* features (see VI) and did not affect the performance of *TopARD* in Somalia.

VII. Conclusion

Predictive models based on environmental factors and population density have been presented for three different countries. The relationships between these factors and prevalence were often not monotonic or linear as evidenced by the disparity between the results of correlation analysis and ARD kernel feature selection. In each country, cloud cover offered the best performance as an individual factor, but improvements were found by combining the top ranked ARD-selected features as inputs. This led to improvements in predictive accuracy over a benchmark model that did not incorporate environmental or population density data. Further improvements were achieved by including AET:PET volatility as a feature, possibly indicating that this has an impact on the mosquito population. In future work, more sophisticated representations could be combined with new streams of remote imagery to improve performance further.

REFERENCES

[1] Tarekegn Abose Abeku, Gerrit J van Oortmarssen, Gerard Borsboom, Sake J de Vlas, and JDF Habbema. Spatial and temporal variations of malaria epidemic risk in ethiopia: factors involved and implications. *Acta tropica*, 87(3):331–340, 2003.

[2] Steve Ackerman, Kathleen Strabala, Paul Menzel, Richard Frey, Chris Moeller, and Liam Gumley. Discriminating clear-sky from cloud with modis algorithm theoretical basis document (mod35. In *MODIS Cloud Mask Team, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin*. Citeseer, 2010.

[3] The World Bank. World development indicators, 2014.

[4] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[5] Stephen J Connor, Madeleine C Thomson, Stephane P Flasse, and Anita H Perryman. Environmental information systems in malaria risk mapping and epidemic forecasting. *Disasters*, 22(1):39–56, 1998.

[6] J. Cox. Remote sensing as a tool for malaria stratification in cambodia: A feasibility study in ratanakiri. *London School of Hygiene and Tropical Medicine*, 2005.

[7] Lek Dysoley, Akira Kaneko, Hideaki Eto, Toshihiro Mita, Doung Socheat, Anders Börkman, and Takatoshi Kobayakawa. Changing patterns of forest malaria among the mobile adult male population in chumkiri district, cambodia. *Acta tropica*, 106(3):207–212, 2008.

[8] Sonia Enosse, Pascal Magnussen, Fatima Abacassamo, Xavier Gómez-Olivé, Anita M Ronn, Ricardo Thompson, and Michael Alifrangis. Rapid increase of plasmodium falciparum dhfr/dhps resistant haplotypes, after the adoption of sulphadoxine-pyrimethamine as first line treatment in 2002, in southern mozambique. *Malar J*, 7:115, 2008.

[9] Peter W Gething, David L Smith, Anand P Patil, Andrew J Tatem, Robert W Snow, and Simon I Hay. Climate change and the global malaria recession. *Nature*, 465(7296):342–345, 2010.

[10] Carl Gold, Alex Holub, and Peter Sollich. Bayesian approach to feature selection and parameter tuning for support vector machine classifiers. *Neural Networks*, 18(5):693–701, 2005.

[11] Alberto Gomez-Elipe, Angel Otero, Michel Van Herp, and Armando Aguirre-Jaime. Forecasting malaria incidence based on monthly case reports and environmental factors in karuzi, burundi, 1997–2003. *Malaria Journal*, 6(1):129, 2007.

[12] K. Hewitt. A survey of plasmodium falciparum prevalence among phnong villagers in mondulkiri, cambodia. *London School of Hygiene and Tropical Medicine*, 2003.

[13] A Huete, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao, and Laerte G Ferreira. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, 83(1):195–213, 2002.

[14] Sandra Incardona, Sirenda Vong, Lim Chiv, Pharath Lim, Sina Nhem, Rithy Sem, Nimol Khim, Socheat Doung, Odile Mercereau-Puijalon, and Thierry Fandeur. Large-scale malaria survey in cambodia: novel insights on species distribution and risk factors. *Malaria journal*, 6(1):37, 2007.

[15] AHD Kilian, P Langi, A Talisuna, and G Kabagambe. Rainfall pattern. el niño and malaria in uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(1):22–23, 1999.

[16] I Kleinschmidt, M Bagayoko, GPY Clarke, M Craig, and D Le Sueur. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, 29(2):355–361, 2000.

[17] Catherine Linard, Marius Gilbert, Robert W Snow, Abdisalan M Noor, and Andrew J Tatem. Population distribution, settlement patterns and accessibility across africa in 2010. *PloS one*, 7(2):e31743, 2012.

[18] Samuel Mabunda, Sónia Casimiro, Llorenç Quinto, and Pedro Alonso. A country-wide malaria survey in mozambique. i. plasmodium falciparum infection in children in different epidemiological settings. *Malaria journal*, 7(1):216, 2008.

[19] Richard J Maude, Chea Nguon, Po Ly, Tol Bunkea, Pengby Ngor, Sara E Canavati de la Torre, Nicholas J White, Arjen M Dondorp, Nicholas PJ Day, Lisa J White, et al. Spatial and temporal epidemiology of clinical malaria in cambodia 2004-2013. *Malaria journal*, 13(1):385, 2014.

[20] USGS Earth Resources Observation MOD07_L2 and South Dakota Science (EROS) Center, Sioux Falls. Atmosphere product, dataset: Nasa eosdis land processes daac. https://lpdaac.usgs.gov.

[21] USGS Earth Resources Observation MOD13C2 and South Dakota Science (EROS) Center, Sioux Falls. Vegetation indices monthly l3 global 0.05deg cmg, dataset: Nasa eosdis land processes daac. https://lpdaac.usgs.gov.

[22] USGS Earth Resources Observation MOD16A2 and South Dakota Science (EROS) Center, Sioux Falls. Global evapotranspiration project, dataset: Nasa eosdis land processes daac. https://lpdaac.usgs.gov.

[23] USGS Earth Resources Observation MOD35_L2 and South Dakota Science (EROS) Center, Sioux Falls. Cloud mask, dataset: Nasa eosdis land processes daac. https://lpdaac.usgs.gov.

[24] Qiaozhen Mu, Faith Ann Heinsch, Maosheng Zhao, and Steven W Running. Development of a global evapotranspiration algorithm based on modis and global meteorology data. *Remote Sensing of Environment*, 111(4):519–536, 2007.

[25] Martin Mubangizi, Ricardo Andrade-Pacheco, Michael Smith, John A Quinn, and Neil Lawrence. Malaria surveillance with multiple data sources using gaussian process models.

[26] Abdisalan M Noor, Archie CA Clements, Peter W Gething, Grainne Moloney, Mohammed Borle, Tanya Shewchuk, Simon I Hay, and Robert W Snow. Spatial prediction of plasmodium falciparum prevalence in somalia. *Malaria journal*, 7(1):159, 2008.

[27] Margaret A Oliver and R Webster. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, 4(3):313–332, 1990.

[28] E Onori and B Grab. Indicators for the forecasting of malaria epidemics. *Bulletin of the World Health Organization*, 58(1):91, 1980.

[29] Carl Edward Rasmussen. Gaussian processes for machine learning. 2006.

[30] Suzanne Seemann, Eva E Borbas, J Li, W PAUL Menzel, and Liam E Gumley. Modis atmospheric profile retrieval algorithm theoretical basis document. 2006.

[31] M Warsame, H Perlmann, S Ali, H Hagi, S Farah, M Lebbad, and A Björkman. The seroreactivity against pf155 (resa) antigen in villagers from a mesoendemic area in somalia. *Tropical medicine and parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ)*, 40(4):412–414, 1989.

[32] Who. *World malaria report 2013*. World Health Organization, 2014.

[33] WHO. World malaria report 2013. world health organization, 2014.

[34] WHO-MERLIN. National malaria prevalence survey, somalia, january - february 2005, final report, somalia, 2005.

[35] WorldPop. Khm_ppp_v2b_2010_unadj. www.worldpop.org.uk.

[36] Kate Zinszer, Aman D Verma, Katia Charland, Timothy F Brewer, John S Brownstein, Zhuoyu Sun, and David L Buckeridge. A scoping review of malaria forecasting: past work and future directions. *BMJ open*, 2(6):e001992, 2012.