Covariate Information

This document describes the current covariate set to be included for use in the updated risk model for *P. knowlesi*. Major changes from the original risk model include:

- Replacement of **urban accessibility** with **access to healthcare**
- Conversion of **human population** and **tasseled-cap transformed layers** to be temporally varying with year
- Addition of the **topographic diversity** covariate to describe the variety of habitats available within a region
- Replacement of the **intact forests** and **disturbed forests** covariates with **forest coverage** and **forest loss** derived from the *Global Forest Change* project.

It is not currently clear what the best method for describing forest loss in terms of the *Global Forest Change* data is. We present a number of potential options for different transformations of the data on page 5.

Table 1: Covariate candidates

Covariate name	In old model	Time varying	Description	Notes	Source
Species Distribution					
Macaca fascicularis suitability	Yes	No	Predicted suitability for inhabitation by macaques of species *M. fascicularis*		[1]
Macaca nemestrina suitability	Yes	No	Predicted suitability for inhabitation by macaques of species *M. nemestrina*		[1]
Leucosphyrus group suitability	Yes	No	Predicted suitability for inhabitation by mosquitos of the Leucosphyrus group		[1]
Various					
Human population	Yes	Yes	Gridded human population density from WorldPop data		[2]
Plasmodium falciparum temperature suitability	Yes	No	Temperature suitability for *P. falciparum* transmission	There does not currently appear to be enough data available to construct a temperature suitability index for *P. knowlesi*	[3]
Tasseled cap wetness standard deviation	Yes	Yes	Tasseled-cap transformed MODIS data		[4, 5]
Tasseled cap wetness mean	Yes	Yes	" " "		[4, 5]
Tasseled cap brightness standard deviation	Yes	Yes	" " "		[4, 5]

Table 1: Covariate candidates *(continued)*

Covariate name	In old model	Time varying	Description	Notes	Source
SRTM elevation	Yes	No	Mean elevation in a region		[6]
MODIS/IGBP Landcover					
Open shrublands	Yes	Yes	Proportion of land with given MODIS/IGBP land classifications		[7]
Woody savannas	Yes	Yes	" " "		[7]
Savannas	Yes	Yes	" " "		[7]
Grasslands	Yes	Yes	" " "		[7]
Permanent wetlands	Yes	Yes	" " "		[7]
Croplands	Yes	Yes	" " "		[7]
Cropland/natural vegetation mosaic	Yes	Yes	" " "		[7]
Urban and built up	No	Yes	" " "		[7]
Tree and Forest Coverage					
Forest coverage	No	Yes	Proportion of land with forest coverage in a given year		[8]
Forest loss	No	Yes	Proportion of land where forest coverage has been lost in the past year	See discussion below for questions on how best to represent this forest loss data	[8]

Table 1: Covariate candidates *(continued)*

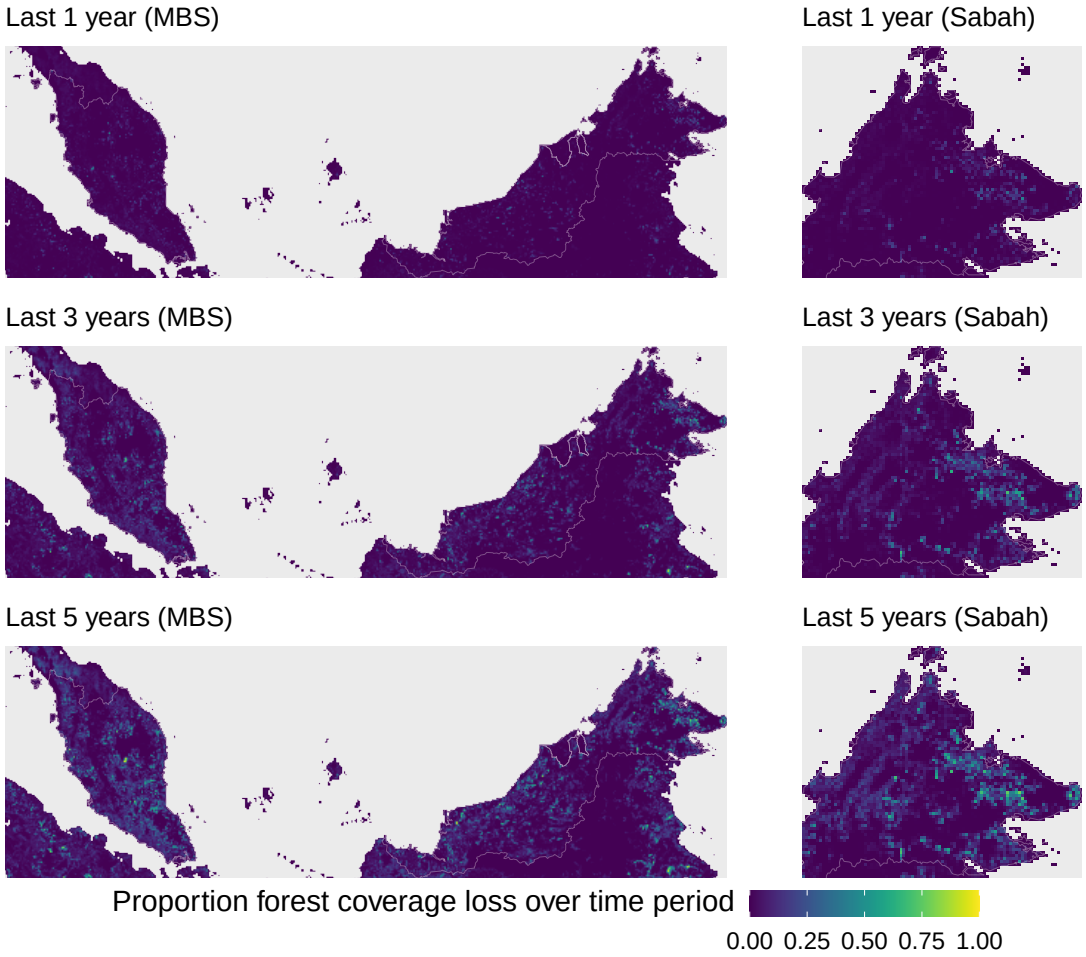
Covariate name	In old model	Time varying	Description	Notes	Source
New Covariates					
Topographic diversity	No	No	How diverse a region is regarding the variety of temperature and moisture environments present as possible habitats	This may offer useful biologically relevant information, with some evidence available that it may improve the performance of species distribution models	[9]
Access to healthcare	No	No	The accessibility to healthcare facilities, measured in duration travel time	This replaces the previous 'urban accessibility' map used in Shearer et al.	[10]

Forest Coverage Loss Data

The original *P. knowlesi* risk model used data from the *Intact Forest Landscapes* project (intactforests.org) consisting of one **intact forests** layer and one **disturbed forests** layer. This dataset was manually constructed with a strict definition for what defines an 'intact' forest, where no signs of human activity are visible whilst still remaining contiguously large enough that a diversity of species could inhabit it. However, the strict definition of intact forest and lack of temporal variation when describing disturbed forests likely means we do not capture as much predictive power from forest change as we otherwise could.

Instead, we propose using the *Global Forest Change* (doi.org/10.1126/science.1244693) dataset that describes yearly observed forest loss over the last 20 years at a 25m resolution, where forest coverage is defined as any vegetation observed over 5m in height. This has the advantage of being well defined across both our landscape and our time period of interest. From this we can construct, for each of the years 2000-2019, the proportion of forest coverage observed on that year at our $5km^2$ resolution (our **forest coverage** covariate).

Similarly we can construct a **forest loss** covariate consisting of the proportion of land where a loss in forest coverage has occurred recently. However, we are not sure as to the most biologically appropriate time-frame to define as 'recent', where the ideal time-frame best captures the effect of deforestation on *P. knowlesi* transmission (for example, a time-frame of 3 years would mean we assume that deforestation that has occurred at any point over the last three years could have contributed to a current infection event). We have constructed three potential time-frames and present them below:



References

[1] Catherine L Moyes, Freya M Shearer, Zhi Huang, Antoinette Wiebe, Harry S Gibson, Vincent Nijman, Jayasilan Mohd-Azlan, Jedediah F Brodie, Suchinda Malaivijitnond, Matthew Linkie, et al. Predicting the geographical distributions of the macaque hosts and mosquito vectors of plasmodium knowlesi malaria in forested and non-forested areas. *Parasites & vectors*, 9(1):1–12, 2016.

[2] WorldPop. Global high-resolution population denominators project. 2018.

[3] Peter W Gething, Thomas P Van Boeckel, David L Smith, Carlos A Guerra, Anand P Patil, Robert W Snow, and Simon I Hay. Modelling the global constraints of temperature on transmission of plasmodium falciparum and p. vivax. *Parasites & vectors*, 4(1):1–11, 2011.

[4] Sarah E Lobser. Vegetation characteristics expressed through transformed modis data: a modis tasseled cap. 2004.

[5] Daniel J Weiss, Peter M Atkinson, Samir Bhatt, Bonnie Mappin, Simon I Hay, and Peter W Gething. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98:106–118, 2014.

[6] Tom G Farr, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. The shuttle radar topography mission. *Reviews of geophysics*, 45(2), 2007.

[7] Mark A Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, 114(1):168–182, 2010.

[8] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.

[9] David M Theobald, Dylan Harrison-Atlas, William B Monahan, and Christine M Albano. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PloS one*, 10(12):e0143619, 2015.

[10] DJ Weiss, A Nelson, CA Vargas-Ruiz, K Gligorić, S Bavadekar, E Gabrilovich, A Bertozzi-Villa, J Rozier, HS Gibson, T Shekel, et al. Global maps of travel time to healthcare facilities. *Nature Medicine*, 26(12):1835–1838, 2020.