# Covariate Update Overview

This document describes the current covariate set to be included for use in the updated risk model for *P. knowlesi*. Major changes from the 2015 risk model include:

- Addition of 338 new *Plasmodium knowlesi* occurrence/absence data points, bringing the total number of data points to 533.
- Updating the most recent year of the annual environmental covariates from 2013 to 2019.
- Replacement of **urban accessibility** (JRC-IES) with **access to healthcare** (Malaria Atlas Project)
- 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.

**Temperature Suitability Index**

We currently use the temperature suitability index for *P. falciparum* constructed by Gething et al. [1] as a proxy for the temperature suitability of *P. knowlesi*.

The temperature suitability index describes the relative effect of temperature on the basic reproduction number of a malaria parasite and vector combination. For differing *Plasmodium* species, this varies solely upon the extrinsic incubation period (or sporogony period), the period of time required for "parasites to develop in the mosquito from point of ingestion via an infected blood meal, through to the point at which sporozoites enter the salivary glands and the mosquito becomes infectious" [2]. We have not found sufficient data within existing literature to construct a model between temperature and extrinsic incubation period for *P. knowlesi* (see Chan and Johansson for similar work on constructing such a model for Dengue [3]).

**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 on the following page.

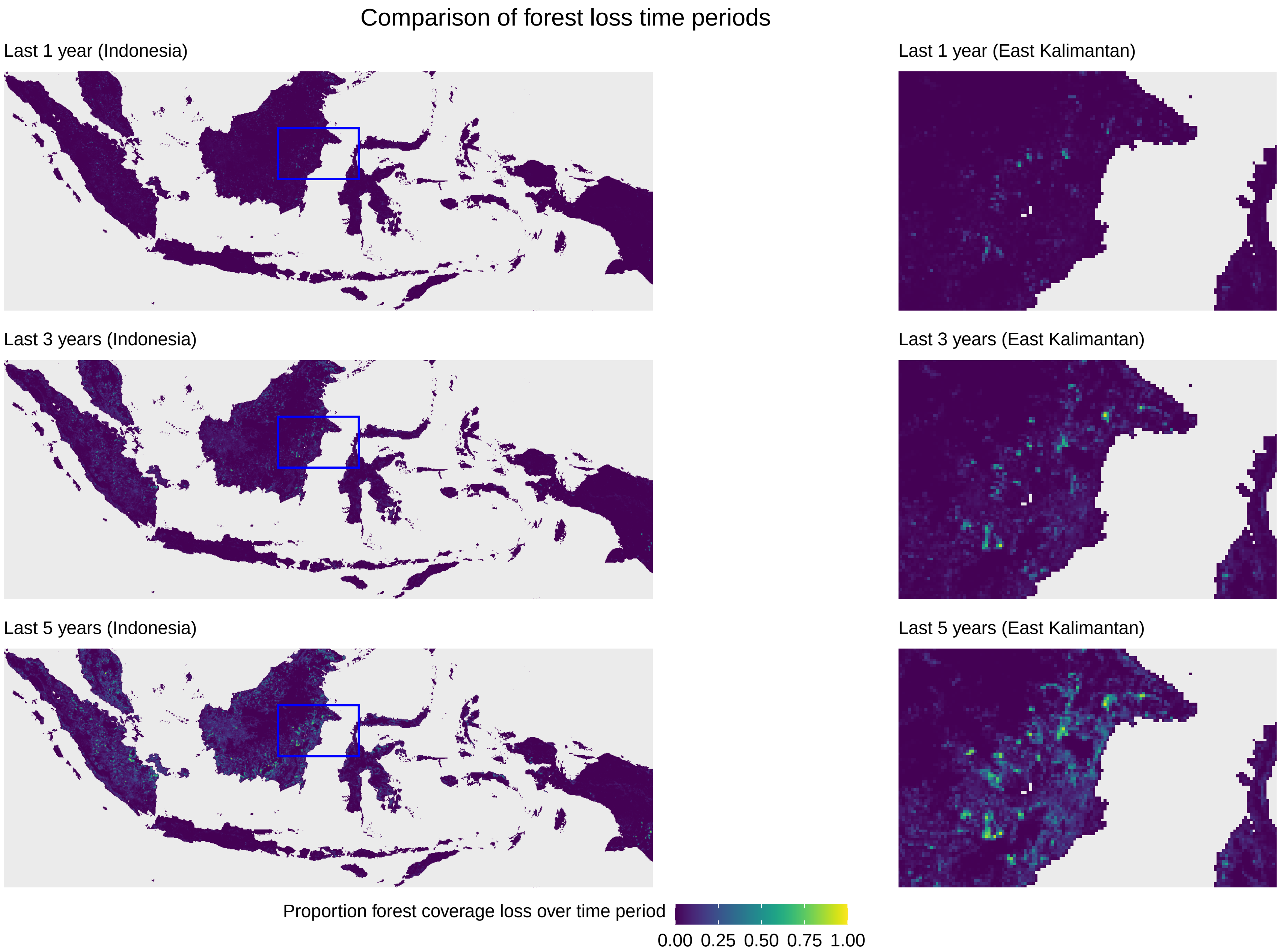
Comparison of forest loss time periods
Last 1 year (Indonesia)
Last 1 year (East Kalimantan)
Last 3 years (Indonesia)
Last 3 years (East Kalimantan)
Last 5 years (Indonesia)
Last 5 years (East Kalimantan)
Proportion forest coverage loss over time period
0.00 0.25 0.50 0.75 1.00

Table 1: Candidate covariate set

| Covariate name | In 2015 model | Annually varying | Description | Notes | Source |
|---|---|---|---|---|---|
| **Species Distribution** | | | | | |
| *Macaca fascicularis* suitability | Yes | No | Predicted suitability for inhabitation by macaques of species *M. fascicularis* | | [4] |
| *Macaca nemestrina* suitability | Yes | No | Predicted suitability for inhabitation by macaques of species *M. nemestrina* | | [4] |
| Leucosphyrus group suitability | Yes | No | Predicted suitability for inhabitation by mosquitos of the Leucosphyrus group | | [4] |
| **Various** | | | | | |
| Human population | Yes | Yes | Gridded human population density from WorldPop data | | [5, 6] |
| 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* | [1] |
| Tasseled cap wetness standard deviation | Yes | Yes | Tasseled-cap transformed MODIS data | | [7, 8] |
| Tasseled cap wetness mean | Yes | Yes | “ “ “ | | [7, 8] |
| Tasseled cap brightness standard deviation | Yes | Yes | “ “ “ | | [7, 8] |

Table 1: Candidate covariate set *(continued)*

| Covariate name | In 2015 model | Annually varying | Description | Notes | Source |
|---|---|---|---|---|---|
| SRTM elevation | Yes | No | Mean elevation in a region | | [9] |
| **MODIS/IGBP Landcover** | | | | | |
| Open shrublands | Yes | Yes | Proportion of land with given MODIS/IGBP land classifications | | [10] |
| Woody savannas | Yes | Yes | “ “ “ | | [10] |
| Savannas | Yes | Yes | “ “ “ | | [10] |
| Grasslands | Yes | Yes | “ “ “ | | [10] |
| Permanent wetlands | Yes | Yes | “ “ “ | | [10] |
| Croplands | Yes | Yes | “ “ “ | | [10] |
| Cropland/natural vegetation mosaic | Yes | Yes | “ “ “ | | [10] |
| Urban and built up | No | Yes | “ “ “ | | [10] |
| **Tree and Forest Coverage** | | | | | |
| Forest coverage | No | Yes | Proportion of land with forest coverage in a given year | | [11] |

Table 1: Candidate covariate set *(continued)*

| Covariate name | In 2015 model | Annually varying | Description | Notes | Source |
|---|---|---|---|---|---|
| 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 | [11] |
| **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 | [12] |
| 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. | [13] |

[1] Annually varying covariates are those for which we have data available for each year between 2001 and 2019 (inclusive), which we assign

# References


[1] 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.

[2] Johanna R Ohm, Francesco Baldini, Priscille Barreaux, Thierry Lefevre, Penelope A Lynch, Eunho Suh, Shelley A Whitehead, and Matthew B Thomas. Rethinking the extrinsic incubation period of malaria parasites. *Parasites & vectors*, 11(1):1–9, 2018.

[3] Miranda Chan and Michael A Johansson. The incubation periods of dengue viruses. *PloS one*, 7(11):e50972, 2012.

[4] 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.

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

[6] Andrea E Gaughan, Forrest R Stevens, Catherine Linard, Peng Jia, and Andrew J Tatem. High resolution population distribution maps for southeast asia in 2010 and 2015. *PloS one*, 8(2):e55882, 2013.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.