

Supplementary appendix

Supplement to: Global malaria infection risk from global warming

Authors

Chao Li¹, Shunsuke Managi*¹

Affiliations

¹ Urban Institute, Kyushu University, Japan

* Correspondent to: Shunsuke Managi, managi@doc.kyushu-u.ac.jp, Kyushu University
744 Motoooka, Nishi-ku, Fukuoka 819-0395 Japan

Materials

Plasmodium falciparum Parasite Rate (PfPR)

The Malaria Atlas Project offers malaria infection prevalence annually, termed PfPR, in 2-10 olds globally from 2000 to 2019 ¹ (<https://malariaatlas.org/explorer/#/>). This data set is a series of spatial grid data with a 5-km resolution, covering the regions from 60° N to 60° S. Malaria is widely infected in Africa, though the infection prevalence has apparently decreased in the past 20 years ¹. To make the spatial resolution of PfPR concordant with the climatic data set, we upscale the spatial data set to 0.25 arc degree by the average method. **Figure S1** shows the spatial distribution of the temporally averaged value of PfPR in 2-10 olds (PfPR₂₋₁₀) from 2000 to 2019. In those past 20 years, although the coverage of the contemporary malaria control inventions was skyrocketing ¹, the public health regarding malaria infection in developing countries, especially in Sub-Saharan African countries, was still remarkably weak. For example, even in 2019, the PfPR₂₋₁₀ is still over 60% in many Sub-Sahara African regions ¹.

Temperature

Temperature data are extracted from the United States National Aeronautics and Space Administration (NASA) data set, NASA Global Land Data Assimilation System Version 2 (GLDAS-2.1). The GLDAS-2.1 system is based on a combination of model and observation data from 2000 to the present, whose products are a variety of monthly average climatic variables with a 0.25-arc-degree spatial resolution ². Firstly, we obtain several climatic variables from GLDAS-2.1, including temperature, air pressure, absolute

humidity, precipitation and wind speed. These variables are all monthly average values and have converted into annual average values. However, the correlations among these climatic variables are too high, which would lead to multicollinearity in the regressions. For example, the correlation coefficient between annual average air pressure and temperature is 0.745 (p -value $< 0.1\%$). Additionally, another critical independent variable, vegetation index, is significantly correlated with absolute humidity (correlation coefficient: 0.807, p -value $< 0.1\%$), precipitation (0.805, p -value $< 0.1\%$) and wind speed (-0.716, p -value $< 0.1\%$). Therefore, the only annual average temperature is kept among those five climatic variables. Secondly, evidence shows that malaria transmission through *Anopheles gambiae* peaks at 25°C ³. Hence, the relationship between malaria infection rate and temperature is non-linear and similar to the U-shape⁴⁻⁶. We take the temperature square into account. Furthermore, extreme low and high temperatures are unsuitable for transmission vector survival. In this way, the standard deviation of monthly average temperature in a specific year would also affect the PfPR. In terms of climate, three variables, involving annual average temperature, annual average temperature square, and standard deviation of monthly average temperature in a specific year, are incorporated into the analysis.

Vegetation Index

Local vegetation affects malaria vector population size, survival, biting, among others⁷⁻⁹. Therefore, the local vegetation status should be considered in the analysis. The normalized difference vegetation index (NDVI) is a widely used indicator to assess how much the observed grid contains live green vegetation^{10,11}. The NDVI ranges from -100% (no live green vegetation) to 100% (full of live green vegetation). The moderate-resolution

imaging spectroradiometer Terra and Aqua satellites provide monthly NDVI data from 2000 to the present with a 0.05-arc-degree spatial resolution. The MODIS vegetation index products used in the analysis are MOD13C2 and MYD13C2¹². To link the NDVI data set to the temperature and PfPR₂₋₁₀ data set, we upscale the NDVI data set to 0.25 arc degree by the average method. Then, these data are further converted into annual average values.

Population Density and GDP per Capita

The population density data are obtained from the WorldPop Project. The WorldPop Project provides spatial grid data set of population count from 2000 to 2020 with a 1-km resolution^{13,14}. Because the length of the latitudes decreases with the increasing latitude, the conversion of population count from 1-km to 0.25-arc-degree resolution is not perfect. At least, we cannot use the direct sum method. To solve this issue, we regard the population count in every grid as the population density, and the unit of this data set is regarded as capita per square kilometer (Cap/km²). Here, we upscale the data from 1-km to 0.25-arc-degree resolution by the average method. The data set eventually illustrates the annual population density with a 0.25-arc-degree resolution.

GDP per capita is associated with development level and public health investment. The higher development level and more public health investment, including a wide range of international aid, are linked with a relatively lower malaria infection rate^{1,15}. However, the spatial grid data set about GDP per capita is unavailable. We employ the country-level GDP per capita from the World Bank. The unit of GDP per capita is the current U.S. dollar. We assume that the GDP per capita in every grid is the same in a country.

81

82 *Climate Change Scenarios and Temperature Prediction*

83 Malaria infection rate is associated with temperature ^{3,4,6}. The different climate
84 change scenarios definitely lead to diverse malaria infection rates, assuming other
85 conditions are the same. The Intergovernmental Panel on Climate Change (IPCC) relies on
86 the Coupled Model Intercomparison Projects (CMIP) to predict the future temperature
87 based on several future emission scenarios ^{16,17}. The 2013 IPCC fifth assessment report
88 (AR5) features four Representative Concentration Pathways (RCPs) according to various
89 possible future greenhouse gas emissions ¹⁷. Based on the RCPs and the CMIP5 model, the
90 IPCC projects four scenarios, specifically RCP2.6, RCP4.5, RCP6.0, and RCP8.5, which
91 are well-known. In 2021, the IPCC sixth assessment report (AR6) is released. In the IPCC
92 AR6, updated emission scenarios are derived from different socioeconomic assumptions,
93 called the Shared Socioeconomic Pathways (SSPs) ¹⁶. These scenarios are the updated
94 version of the four scenarios in the IPCC AR5, so the new scenarios listed in the IPCC
95 AR6 are named SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 and projected by the CMIP6
96 model ¹⁶. Except SSP3-7.0, the other three scenarios are advanced from the previous
97 scenarios in the AR5. The SSP1-2.6 scenario predicts that the temperature increase
98 stabilizes at about 1.8°C by the end of the century. In the SSP2-4.5 scenario, the
99 temperatures rise 2.7°C by 2100. According to the SSP3-7.0 scenario, the temperatures go
100 up around 3.6°C by the end of the century. Finally, the SSP5-8.5 scenario predicts that the
101 temperatures grow 4.4°C or higher by 2100. The IPCC projects global temperatures in four
102 periods, including 2021-2040 (near term), 2041-2060 (medium term), and 2081-2100 (long
103 term), based on the different scenarios. In each period, the data set contains 12 grid data of

monthly average temperature. The IPCC provides open access data based on over 30-model stimulation with a 1-arc-degree resolution (<https://interactive-atlas.ipcc.ch/>). Furthermore, the IPCC also lists the predicted population density grid data produced by the models' stimulation with a 1-arc-degree resolution. We obtain the population density data under SSP2-4.5, SSP3-7.0, and SSP5-8.5, during three periods.

Grids

We integrate the abovementioned data set to build a long panel spatial grid data set. Because we only focus on the endemic areas, the grids are dropped, whose PfPR₂₋₁₀ is always 0. Additionally, every grid is required to have at least two-year complete records. Without enough records, the grid is also removed. Accordingly, 65,705 grids are kept. The coverage of our study is over 40 million km². According to statistical tests and data availability, our analysis includes temperature, NDVI, population density, and GDP per capita. Previous studies indicate that climate change affects vector-borne diseases¹⁸⁻²⁰, so temperature-related variables and NDVI are considered. Moreover, population density and prevention also play a critical role in malaria transmission^{1,21}. Although detailed prevention investments are unavailable, we replace them with GDP per capita.

Methods

Spatially Stationary Model

At first, we assume that relationships between PfPR₂₋₁₀ and other selected independent variables are the same in each grid. For example, the marginal effects of a 1%

increase in the NDVI in Africa and South America are the same, even though they are thousands of kilometers away. Without considering the spatial contexts, those relationships do not vary spatially^{22,23}. To estimate the links among variables, we apply the panel regression method. Because our data set is a panel data set, the hypotheses on the time-fixed effects of each grid should be carefully tested. Three typical panel regression models based on different time-fixed effect hypotheses, including fixed effects model (FEM), random effects model (REM), and pooled ordinary least square (POLS), are widely used. We perform several statistical tests to select the reasonable spatially stationary model. Firstly, we execute the F test for individual effects^{24,25}. The significant result indicates that the time-fixed effects exist since the null hypothesis is that no time-fixed effects are needed²⁵. Hence, the POLS is rejected because this model assumes that no time-fixed effects exist. Then, we implement the Hausman test to detect that either the REM or the FEM is reasonable²⁶. The null hypothesis of the Hausman test that the preferred model is the REM is rejected due to the significant test result. Therefore, the FEM is preferred in the spatially stationary analysis.

Here, the FEM is illustrated as follows:

$$PfPR_{it} = \beta \mathbf{X}'_{it} + \alpha_i + \mu_{it} \quad (1)$$

where $PfPR_{it}$ denotes the annual PfPR₂₋₁₀ (ranging from 0 to 1) in grid i in t year, \mathbf{X}_{it} denotes a matrix of independent variables, including annual average temperature (°C), annual average temperature square (°C²), the standard deviation of monthly average temperature in t year (°C), NDVI, population density (Cap/km^2), and GDP per capita (USD/Cap) in measurement i in t year, α_i denotes the time-fixed effects, μ_{it} denotes an idiosyncratic error, and β is a vector of parameters to be estimated.

148

149 *Spatially Non-stationary Model*

150 Obviously, the marginal effects of the climate variable on malaria infection rate are
151 not always, even never, the same spatially in the real world. For instance, Giesen et al.
152 summarize the impacts of climate change on malaria in Africa in their review, and the
153 marginal effects of climate change are different country by country ²⁷. Two possible
154 reasons flash. Firstly, the relationship between climate change, especially global warming,
155 and the malaria infection rate is non-linear ³⁻⁶. To solve this issue, we employ the square
156 temperature term. Secondly, the association of malaria infection rate with temperature
157 spatially varies ²⁸. To grasp this spatial variability, the advanced spatially non-stationary
158 model should be employed.

159 Geographically weighted panel regression (GWPR) is an improvement of
160 geographically weighted regression, which allows the regression coefficients to vary
161 spatially ^{29,30} and is widely used in spatial analyses. Basically, the GWPR divides the total
162 data set into a large number of sub-datasets according to the optimal bandwidth. The
163 bandwidth of the GWPR is a threshold distance to judge whether two grids have a spatial
164 relationship ^{22,29,31}. The number of sub-datasets equals the number of grids. Every sub-
165 dataset takes one grid as the center and the bandwidth as the radius to select the other grids.
166 Of note, every grid could be utilized several times in the dividing process. The bandwidth
167 calibration is the essential step in the analysis. The mean square prediction error of the
168 regression is the critical index to calibrate the bandwidth. The lowest mean square
169 prediction error means the highest goodness of fit ³². In the GWPR bandwidth calibration,
170 the mean square prediction error is calculated as follows:

$$MSPE(b) = \frac{m \sum_j [y_j - \widehat{y_j(b)}]^2}{(m - p + 1)^2} \quad (2)$$

where $MSPE(b)$ is the mean square prediction error when the bandwidth is b , m is the data size, y_j is the dependent variable of the j th record, $\widehat{y_j(b)}$ is the predicted value of the j th record when the bandwidth is b , and p is the number of the parameters in the analysis. In the GWR, the basic version of the GWPR, in a way, previous studies generally assume that **Equation 2** is a U-shape function without statistical or mathematical evidence^{22,23,32}. This assumption helps to reduce calculating time but also brings enormous risks to the analysis.

We apply the step increment selection without assuming that **Equation 2** is a U-shape function. The bandwidth selection extent is from 0.25 to 20 arc degrees, and the examined bandwidth increases 0.25 arc degree every time because the spatial resolution of the grid data is 0.25 arc degree. The optimal bandwidth is 4.25 arc degrees (**Figure S2**). The mean square prediction error reaches the minimum when the bandwidth equals 0.75, but we reject it. According to the grid data set spatial resolution, every sub-dataset at most has 25 grids, but some of them might only have one grid. If the sub-datasets have only one grid, they do not fit FEM's statistical requirements. Moreover, some grids might not be spatially linked with other grids with a 0.75-arc-degree bandwidth, which leads to a not robust GWPR result. Therefore, the optimal bandwidth, 4.25 arc degrees, is calibrated.

The GWPR is performed after the optimal bandwidth has been calibrated. The GWPR model is written as follows:

$$PfPR_{it} = \beta_i X'_{it} + \alpha_i + \epsilon_{it} \quad (3)$$

where β_i denotes a vector of parameters in the regression with the sub-dataset taking the grid i as the center. There are 65,705 sub-datasets, so the number of regressions on sub-

191 datasets is also 65,705, and every estimation is based on the sub-dataset's spatial weight
 192 vector. For this reason, the parameters, β_i , vary spatially. The spatial weight vector is
 193 calculated as follows:

$$\mathbf{w}_i = \begin{cases} [1 - (\frac{\mathbf{d}}{b})^2]^2, & d_k \in \mathbf{d} \text{ if } d_k \leq b \\ 0, & d_k \notin \mathbf{d} \text{ if } d_k > b \end{cases} \quad (4)$$

194 where \mathbf{w}_i denotes a vector of spatial weights between the grid i and its neighbors, \mathbf{d}
 195 represents a vector of distances between the grid i and its neighbors, b is the calibrated
 196 optimal bandwidth, and d_k is the distance between the grid i and the grid k . Then, the
 197 GWPR estimates the parameters, based on the spatial weight vectors. Because the time-
 198 fixed effects term (α_i) in is unknown in **Equation 3**²⁴, we transform that equation as
 199 follows:

$$PfPR_{it} - \overline{PfPR}_i = \beta_i(\mathbf{X}_{it} - \overline{\mathbf{X}}_i)' + (\epsilon_{it} - \overline{\epsilon}_i) \quad (5)$$

200 where \overline{PfPR}_i denotes the mean of the PfPR₂₋₁₀ of the grid i , $\overline{\mathbf{X}}_i$ denotes a vector of means
 201 of independent variables of the grid i , and $\overline{\epsilon}_i$ denotes the mean of the error of the grid i . To
 202 simplify **Equation 5**, we define the matrix of transformed independent variables \mathbf{x}_{it} and
 203 the transformed dependent variable $pfpr_{it}$ as follows:

$$\mathbf{x}_{it} = \mathbf{X}_{it} - \overline{\mathbf{X}}_i \quad (6)$$

$$pfpr_{it} = PfPR_{it} - \overline{PfPR}_i \quad (7)$$

$$\sigma_{it} = \epsilon_{it} - \overline{\epsilon}_i \quad (8)$$

204 In the light of **Equation 6-8**, **Equation 3** is rewritten as follows:

$$pfpr_{it} = \beta_i \mathbf{x}'_{it} + \sigma_{it} \quad (9)$$

205 The parameter estimation could be displayed as follows:

$$\beta_i = [x_{it}^T W_i x_{it}]^{-1} x_{it}^T W_i pfp_{it} \quad (10)$$

206

207 *Statistical Indicators and 10-fold Cross Validation*

208 Because we need to make the prediction based on the results of IPCC stimulations,
 209 we must confirm the model's accuracy. Several statistical indicators, including R^2 , root
 210 mean square error (RMSE), mean absolute error (MAE), correlation coefficient (r) between
 211 observed and predicted values, and regression coefficients between observed and predicted
 212 values (intercept α and slope β), are applied to depict the accuracy of the model. R^2 is a
 213 critical statistical indicator describing the goodness of fit, which is expressed as follows:

$$R^2 = 1 - \frac{\sum_{k=1}^n (OPfPR_k - PPfPR_k)^2}{\sum_{k=1}^n (OPfPR_k - \overline{OPfPR})^2} \quad (11)$$

214 where n represents the observation number in the whole data set, $OPfPR_k$ represents the
 215 k th record of the observed PfPR₂₋₁₀, $PPfPR_k$ represents the k th record of the predicted
 216 PfPR₂₋₁₀, and \overline{OPfPR} represents the mean of the observed PfPR₂₋₁₀. The RMSE is sensitive
 217 to both systematic error and random error, expressed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (OPfPR_k - PPfPR_k)^2} \quad (12)$$

218 The MAE is computed as follows:

$$MAE = \text{mean}(|OPfPR_k - PPfPR_k|) \quad (13)$$

219 In the analysis, the MAE is expected to be lower. Additionally, the regression coefficients
 220 between observed and predicted values are estimated as follows:

$$OPfPR_k = \alpha + \beta PPfPR_k + \delta_k \quad (14)$$

where α is the intercept of the regression, the ideal value of α is 0, β is the slope, its ideal value is 1, and δ_k is a random error term.

To check the reliability of the GWPR model, we perform the 10-fold cross validation. In the 10-fold cross validation, the whole data set is randomly divided into ten subsets. In every individual fold cross validation, nine subsets are used to train the GWPR model, while the left one subset is employed to test the accuracy and ability of prediction. All statistical indicators, involving R^2 , RMSE, MAE, r , α , and β , are reported based on the training and testing results.

Effects of Temperature Change on PfPR₂₋₁₀

The effects of temperature change on PfPR₂₋₁₀ cannot directly be estimated because other necessary variables, such as NDVI, population, and GDP per capita, cannot be precisely predicted. The errors from the other variable predictions would drastically affect the estimation of PfPR₂₋₁₀. However, if we assume that other variables are the same in the various scenarios, we could extract the difference among the impacts of temperature change in different periods. The difference between the two temperature scenarios is calculated as follows:

$$\Delta PfPR_{it} = S2PfPR_{it} - S1PfPR_{it} \quad (15)$$

where $\Delta PfPR_{it}$ denotes the difference between the estimation of PfPR₂₋₁₀ with the different temperature scenarios in the grid i during the period t , $S1PfPR_{it}$ denotes the estimation of PfPR₂₋₁₀ based on the temperature scenario $S1$ in the grid i during the period t , and $S2PfPR_{it}$ denotes the estimation of PfPR₂₋₁₀ based on the temperature scenario $S2$

242 in the grid i during the period t . We hope to obtain the impacts of the higher greenhouse
 243 gas emission on PfPR₂₋₁₀, so the SSP1-2.6 scenario is always set as the temperature
 244 scenario $S1$ and the temperature scenario $S2$ is one of the other three scenarios, including
 245 SSP2-4.5, SSP3-7.0, and SSP5-8.5.

246 To estimate $S1PfPR_{it}$ and $S2PfPR_{it}$, we expand **Equation 3** as follows:

$$S1PfPR_{it} = \beta_{1i}TeS1'_{it} + \beta_{2i}TeS1^{2'}_{it} + \beta_{3i}TSDS1'_{it} + \beta_{4i}NDVI'_{it} \quad (16)$$

$$+ \beta_{5i}POP'_{it} + \beta_{6i}GDP'_{it} + \alpha_i$$

$$S2PfPR_{it} = \beta_{1i}TeS2'_{it} + \beta_{2i}TeS2^{2'}_{it} + \beta_{3i}TSDS2'_{it} + \beta_{4i}NDVI'_{it} \quad (17)$$

$$+ \beta_{5i}POP'_{it} + \beta_{6i}GDP'_{it} + \alpha_i$$

247 where β_{1i} , β_{2i} , β_{3i} , β_{4i} , β_{5i} , and β_{6i} are parts of the estimated GWPR parameters β_i
 248 from **Equation 10**, $TeS1_{it}$ denotes the stimulated annual average temperature under the
 249 SSP1-2.6 scenario in the grid i during the period t , $TSDS1_{it}$ denotes the standard
 250 deviation of stimulated monthly average temperature under the SSP1-2.6 scenario in the
 251 grid i during the period t , $TeS2_{it}$ denotes the stimulated annual average temperature under
 252 another scenario in the grid i during the period t , $TSDS2_{it}$ denotes the standard deviation
 253 of stimulated monthly average temperature under another scenario in the grid i during the
 254 period t , and $NDVI_{it}$, POP_{it} , and GDP_{it} denote the NDVI, population density, and GDP
 255 per capita in the grid i during the period t , respectively.

256 In view of **Equations 16** and **17**, **Equation 15** is rewritten and simplified as follows:

$$\Delta PfPR_{it} = \beta_{1i}(TeS2_{it} - TeS1_{it})' + \beta_{2i}(TeS2^2_{it} - TeS1^2_{it})' \quad (18)$$

$$+ \beta_{3i}(TSDS2_{it} - TSDS1_{it})'$$

257

Difference in Infection Cases between Different Scenarios

We estimate the infection case changes due to the shifts in development scenarios based on the predicted differences of PfPR₂₋₁₀ between different scenarios. In this study, the predicted dependent variable is the PfPR₂₋₁₀, but the IPCC does not provide the population in 2-10 olds. Hence, we assume the PfPR in all ages is the same as PfPR₂₋₁₀ to predicate the infection case changes. The infection cases changes can be estimated as follows:

$$\Delta ICC_{itk} = \Delta PfPR_{itk} \times Population_{itk} \quad (19)$$

where ΔICC_{itk} denotes the infection case change due to the k shift of development scenarios in the grid i during the period t , $\Delta PfPR_{itk}$ denotes the difference due to the k shift of development scenarios in the grid i during the period t estimated by **Equation 18**, and $Population_{itk}$ denotes the IPCC predicted population in the grid i during the period t under the higher-temperature scenario in the k shift. The shifts of development scenarios include three types: from SSP1-2.6 to SSP2-4.5, from SSP1-2.6 to SSP3-7.0, and from SSP1-2.6 to SSP5-8.5.

Results

The R^2 of the GWPR model is 94.75%, indicating the high accuracy of the estimation. The outcomes of 10-fold cross validation are listed in **Table S1**. In the 10-fold cross validation, the R^2 s of the training and testing data sets are over 94%, which hints at the model's stability. Besides, the RMSEs, MEAs, α s, β s, and r s are close to their ideal

values, further proposing the high reliability of the estimated impacts of temperature change on PfPR₂₋₁₀.

Estimated Effects of Temperature Change on PfPR₂₋₁₀

Figures S3-S5 illustrate the difference between the estimation of PfPR₂₋₁₀ with the different temperature scenarios during 2021 – 2040, including the difference between SSP2-4.5 and SSP1-2.6 ($\Delta PfPR_{2021-2040}^{SSP245-SSP126}$), the difference between SSP3.7-0 and SSP1-2.6 ($\Delta PfPR_{2021-2040}^{SSP370-SSP126}$), and the difference between SSP5.8-5 and SSP1-2.6 ($\Delta PfPR_{2021-2040}^{SSP585-SSP126}$). **Figure S6** demonstrates the statistical distribution of $\Delta PfPR_{2021-2040}^{SSP245-SSP126}$, assuming that all conditions are the same except temperatures. The higher temperature scenario SSP2-4.5 positively impacts the malaria infection rate in many grids, resulting in a potential fall in PfPR₂₋₁₀ during 2041 - 2060. Nevertheless, most grids show an increasing trend. Manifest in Africa, the rising trends in Congo (CD), Central African Republic (CF), Cameroun (CM), Burkina Faso (BF), Nigeria (NG), etc., mark the adverse effects of climate change. Since the difference in temperature change between SSP2-4.5 and SSP1-2.6 is tiny, the impacts of climate change are not potent, but they appear in the worst malaria-affected areas (**Figure S1**). According to **Figures S7** and **S8** that illuminate the statistical distributions of $\Delta PfPR_{2021-2040}^{SSP370-SSP126}$ and $\Delta PfPR_{2021-2040}^{SSP585-SSP126}$, respectively, the influences of climate change shift partially. In **Figure S7**, the impacts of temperature change from SSP1-2.6 to SSP3-7.0 on PfPR₂₋₁₀ range from -1.68% to 2.66%, and the grid number of positive impacts outstrips the number of grids with negative effects. **Figure S8**'s distribution is similar to **Figure S7**'s. Moreover, we compare the geographical distributions of average PfPR₂₋₁₀ and temperature change's

impacts. Notably, compared with other countries, CD, CF, CM, BF, and NG are the worst hit by malaria in the past 20 years. Worse, they might face the most harmful impacts of climate change on the malaria infection rates during 2021 - 2040.

With further increase in temperature during 2041-2060, the distribution and intention of negative impacts shift. **Figures S9-S11** show the difference between SSP2-4.5 and SSP1-2.6 ($\Delta PfPR_{2041-2060}^{SSP245-SSP126}$), the difference between SSP3.7-0 and SSP1-2.6 ($\Delta PfPR_{2041-2060}^{SSP370-SSP126}$), and the difference between SSP5.8-5 and SSP1-2.6 ($\Delta PfPR_{2041-2060}^{SSP585-SSP126}$) during 2041 – 2060. $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$ becomes minus in most areas of CD, CF, CM, BF, and NG. The positive impacts of the increase in temperature on $PfPR_{2-10}$ also rise, according to $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$ and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$ because in SSP3.7-0 and SSP5.8-5, the predicted temperatures are higher than the scenario SSP2-4.5. The negative impacts mainly appear in Gabon (GA), Angola (AO), and Zambia (ZM), which are at the boundaries of high malaria-affected regions. The possible result of climate change, especially in Sub-Saharan Africa, is that the increased temperature would spur the high-risk area to spread to the higher latitude regions and coastlines. Moreover, the original high-risk areas are no longer suitable for human living, owing to extreme temperatures. **Figures S12-S14** demonstrate the statistical distribution of $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$, $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$, and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$, respectively. Compared with differences among the scenarios during 2021- 2040, the values of $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$, $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$, and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$ are absolutely larger. Moreover, the impacts of the increase in temperature are positive in most grids of Africa, Asia, and South America.

By the end of the 21st century, the four scenarios end in different temperature predictions. **Figures S15-S17** demonstrate the difference between SSP2-4.5 and SSP1-2.6 ($\Delta PfPR_{2081-2100}^{SSP245-SSP126}$), the difference between SSP3.7-0 and SSP1-2.6 ($\Delta PfPR_{2081-2100}^{SSP370-SSP126}$), and the difference between SSP5.8-5 and SSP1-2.6 ($\Delta PfPR_{2081-2100}^{SSP585-SSP126}$) during 2081 – 2100. Because the difference in temperatures of the four scenarios becomes more significant, the impacts of temperature change on PfPR₂₋₁₀ are intensified. The spatial distributions of $\Delta PfPR_{2081-2100}^{SSP245-SSP126}$, $\Delta PfPR_{2081-2100}^{SSP370-SSP126}$, and $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$ are similar to the situation of $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$, but the values of the difference become larger during 2080 – 2100. **Figures S18-S20** propose the statistical distribution of $\Delta PfPR_{2081-2100}^{SSP245-SSP126}$, $\Delta PfPR_{2081-2100}^{SSP370-SSP126}$, and $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$. In fact, $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$ ranges -77.20% to 101.54%.

Figure S21 illustrates the shapes of the relationship between annual average temperature and PfPR₂₋₁₀ in each grid. The basic shapes of the relationship are two types, specifically, U-shape and inverted U-shape. The coefficient of the square temperature term decides the basic shape of the relationship. Additionally, together with the coefficient of temperature term, the symmetry lines of the relationship of each grid can be pinned down. In **Figure S21**, we set the interested temperature range from 0°C to 50°C. If the symmetry lines of the relationships of some grids do not belong to the interested range, we consider those relationships are either monotonically decreasing or monotonically increasing in the range, mathematically. Therefore, the other four shapes are derived from the basic types, including increase based on inverted U-shape (Increase Inverted U), decrease based on U-shape (Decrease U), decrease based on inverted U-shape (Decrease Inverted U), and increase based on U-shape (Increase U). Previous studies generally put forward that the

relationship at national or regional levels is an inverted U-shape, and the temperature impacts would peak at a specific temperature. However, the relationship at the grid level is obviously more complicated and various. Hence, we must treat the impacts of climate change carefully. Although we accept the assumption that the relationship between temperature and malaria infection rates is non-linear, we did not force to set the malaria-transmission peaking temperature to 25°C or other values, as the previous studies mentioned^{4,6}. In our results, most grids' transmission peaking temperatures are not 25°C. We employ the annual average temperature to probe the relationship, whereas previous studies use laboratory environments and real-time or shorter-term temperatures. As mentioned before, even though the annual average temperature is 25°C, how many days and hours are at that temperature remains unknown. For this reason, in many grids, the transmission peaking temperatures are not 25°C. The shapes in each grid are even different, though either U-shape or inverted U-shape. Furthermore, real-world environments are significantly more complex than laboratory environments. The ignored factors by the laboratory environments are also related to the malaria-transmission peaking temperature. These ignored variables lead to the difference in the transmission peaking temperatures between our result and previous studies.

Estimated Difference in Infection Cases between Different Scenarios

Figures S22-S24 demonstrate the spatial distributions of the estimated differences in infection cases between different scenarios ($\Delta DIC_{2021-2040}$) during 2021 – 2040. Under three scenario shifts, including from SSP1-2.6 to SSP2-4.5, from SSP1-2.6 to SSP3-7.0, and from SSP1-2.6 to SSP5-8.5, the $\Delta DIC_{2021-2040}$ are expressed as $\Delta DIC_{2021-2040}^{SSP126-SSP245}$,

$\Delta DIC_{2021-2040}^{SSP126-SSP370}$, and $\Delta DIC_{2021-2040}^{SSP126-SSP585}$, respectively. The spatial distribution of $\Delta DIC_{2021-2040}$ is affected by the population density pattern and the difference between estimated PfPR₂₋₁₀ with different temperature scenarios. Hence, the PfPR₂₋₁₀ most-increased regions might not have the largest infection case increment. However, infection cases are predicted to increase most in NG, CD, and BF, if the scenario changes from SSP1-2.6 to SSP2-4.5 in the near term. In some areas of these three countries, the infection cases increase by over 10 cases/km² (**Figure S22**). Slight rising temperatures aggravate the situation in the countries severely hit by malaria. Under the shifts from SSP1-2.6 to SSP2-4.5 and SSP5-8.5, NG, CD, and BF are still the most affected, but the impact degree decreases (**Figures S23 and S24**). This reduction in the negative impacts of temperatures is because the temperatures in these countries are close to the most suitable value for malaria transmission. **Figures S25-S27** illustrate the statistical distribution of $\Delta DIC_{2021-2040}^{SSP126-SSP245}$, $\Delta DIC_{2021-2040}^{SSP126-SSP370}$, and $\Delta DIC_{2021-2040}^{SSP126-SSP585}$, respectively. In Africa, most grids suffer from the negative impact of global warming on malaria transmission in the near term.

Figures S28-S30 demonstrate the spatial distributions of $\Delta DIC_{2041-2060}^{SSP126-SSP245}$, $\Delta DIC_{2041-2060}^{SSP126-SSP370}$, and $\Delta DIC_{2041-2060}^{SSP126-SSP585}$, respectively. During 2041 - 2060, the impacts of temperature increase become positive in most regions because the temperatures in those areas exceed the most suitable value for malaria transmission. For example, in the Southern part of NG, $\Delta DIC_{2041-2060}$ is less than -20 case/km² under all three scenario shifts, which indicates a decrease in malaria infection cases due to temperature increase. Compared with **Figures S22-S24**, the temperature impacts show an opposite direction in most regions in **Figures S28-S30**. **Figures S31-S33** illustrate the statistical distribution of

$\Delta DIC_{2041-2060}^{SSP126-SSP245}$, $\Delta DIC_{2041-2060}^{SSP126-SSP370}$, and $\Delta DIC_{2041-2060}^{SSP126-SSP585}$, respectively. In most regions, the temperature increase positively impacts the public health issue.

With the temperature increase during 2081 - 2100, the negative impacts of the temperature increase move to high-latitude and coastline areas. **Figures S34-S36** demonstrate the spatial distributions of $\Delta DIC_{2081-2100}^{SSP126-SSP245}$, $\Delta DIC_{2081-2100}^{SSP126-SSP370}$, and $\Delta DIC_{2081-2100}^{SSP126-SSP585}$, respectively. The most negative impacts locate in severely infected borders, as shown in **Figure S1**. **Figures S37-S39** show the statistical distributions of $\Delta DIC_{2081-2100}^{SSP126-SSP245}$, $\Delta DIC_{2081-2100}^{SSP126-SSP370}$, and $\Delta DIC_{2081-2100}^{SSP126-SSP585}$. Although the ΔDIC becomes negative in most areas, high-risk regions shift to areas that are less severely infected during 2000 - 2019. Therefore, the effect of rising temperatures must be taken into account.

Data Availability

All data sources used in the analyses, along with fully reproducible code, are publicly available at https://github.com/MichaelChaoLi-cpu/Malaria_And_Climate_Change.

Figure:

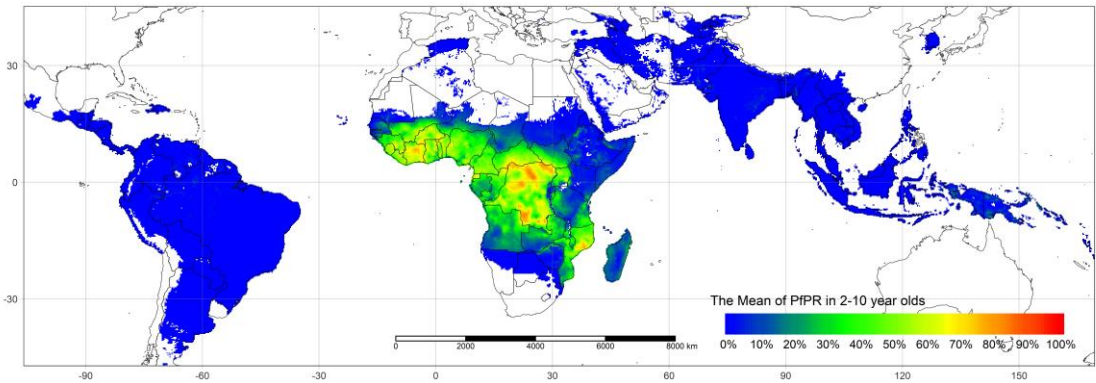


Figure S1: The mean of PfPR₂₋₁₀ from 2000 to 2019

(Note: if the PfPR₂₋₁₀ in a grid is always 0 from 2000 to 2019, the grids are blank)

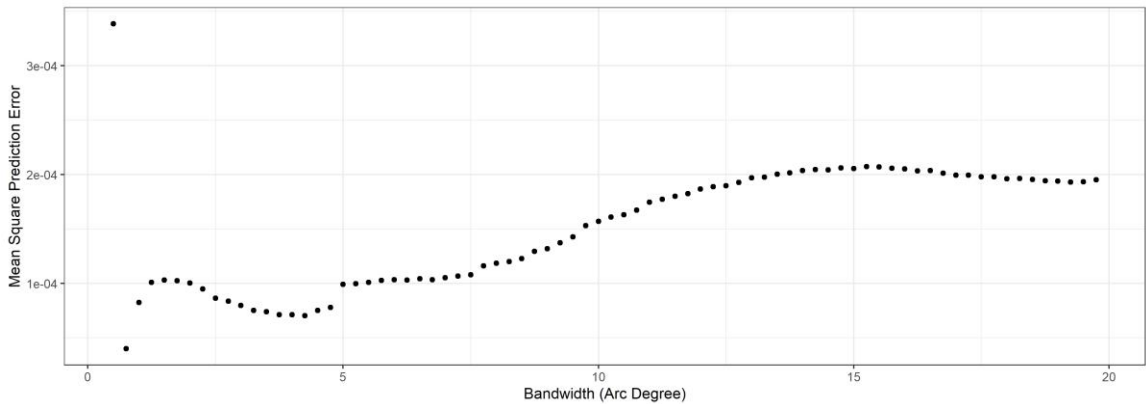
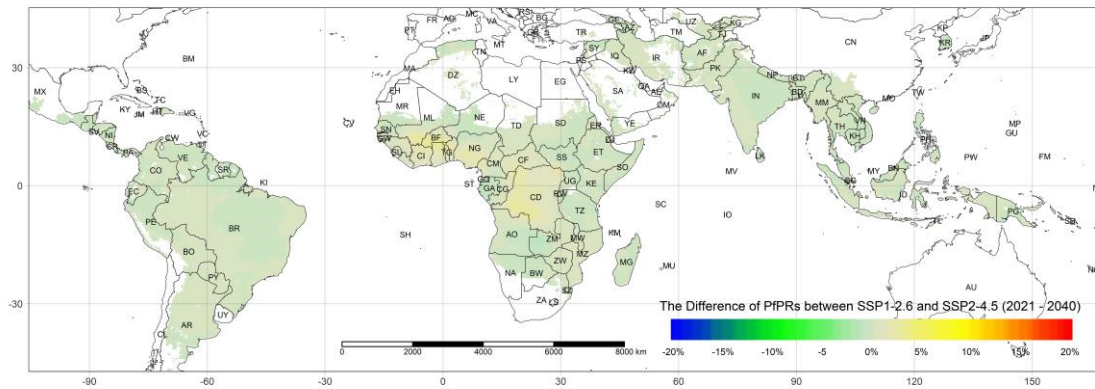
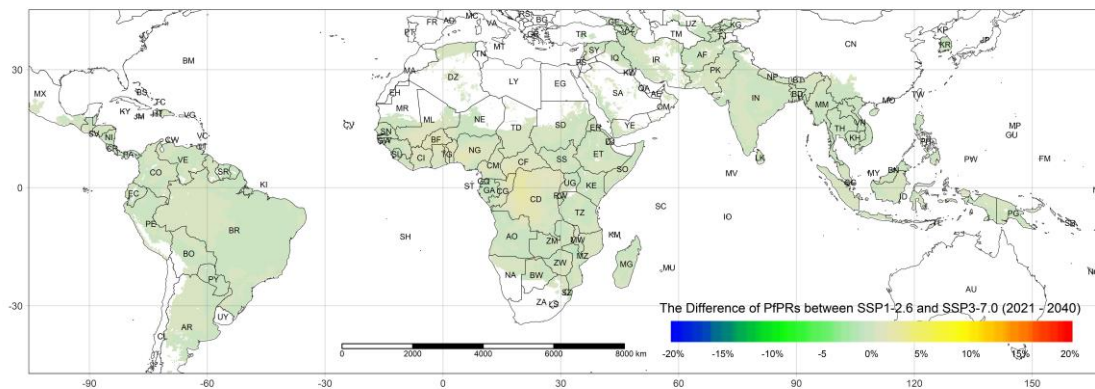


Figure S2: Step Increment Bandwidth Selection



416

417 **Figure S3: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2041 - 2060)**



418

419 **Figure S4: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2041 - 2060)**

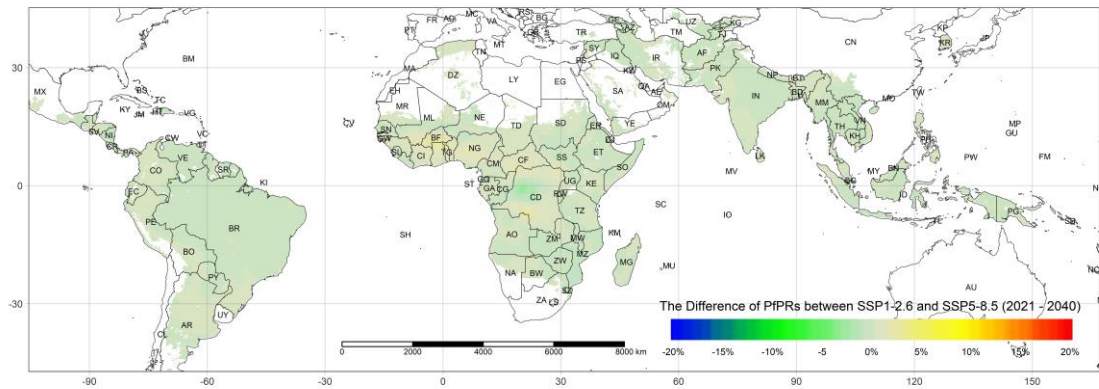


Figure S5: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2041 - 2060)

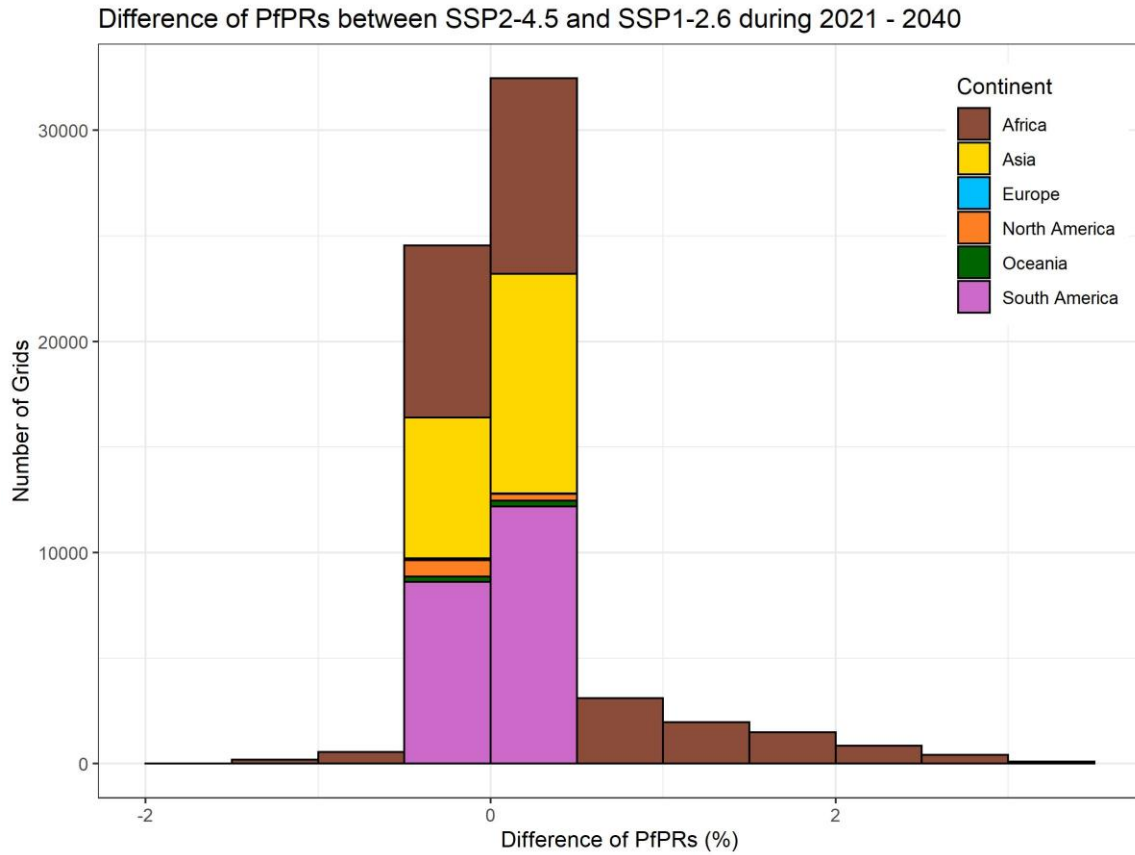


Figure S6: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2041 - 2060)

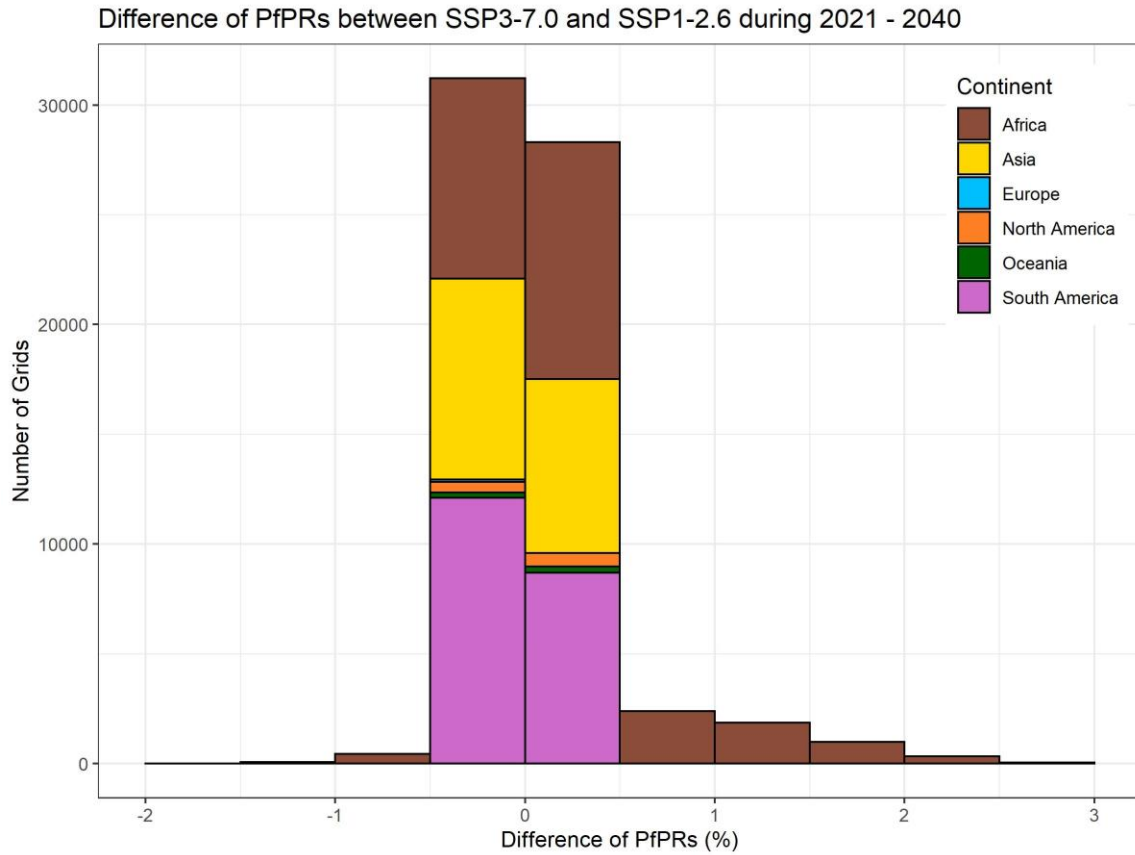


Figure S7: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2041 - 2060)

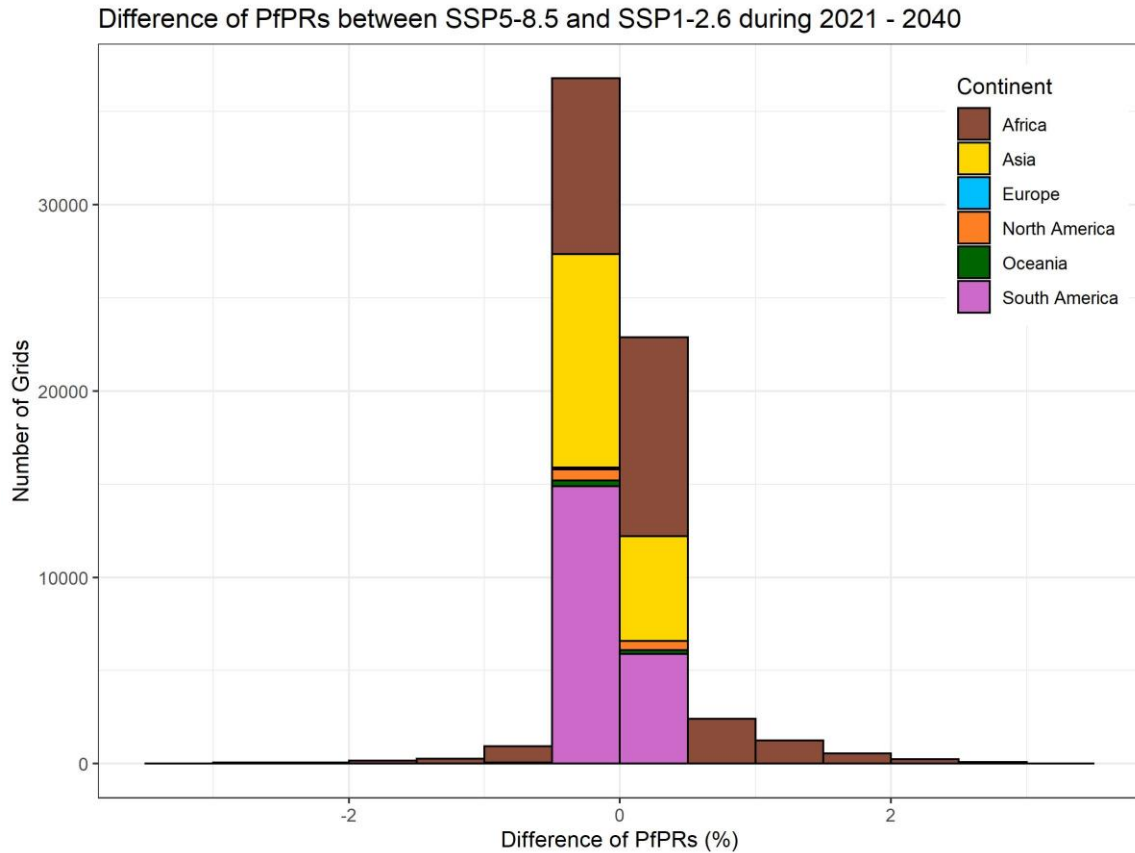
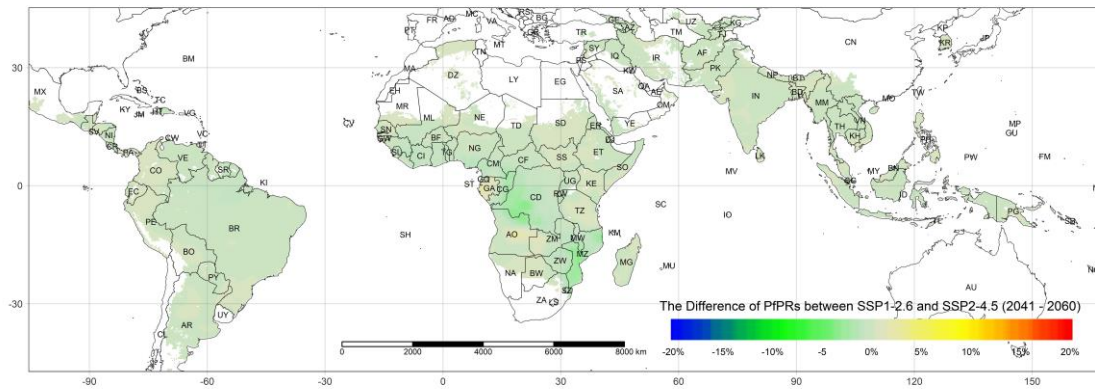
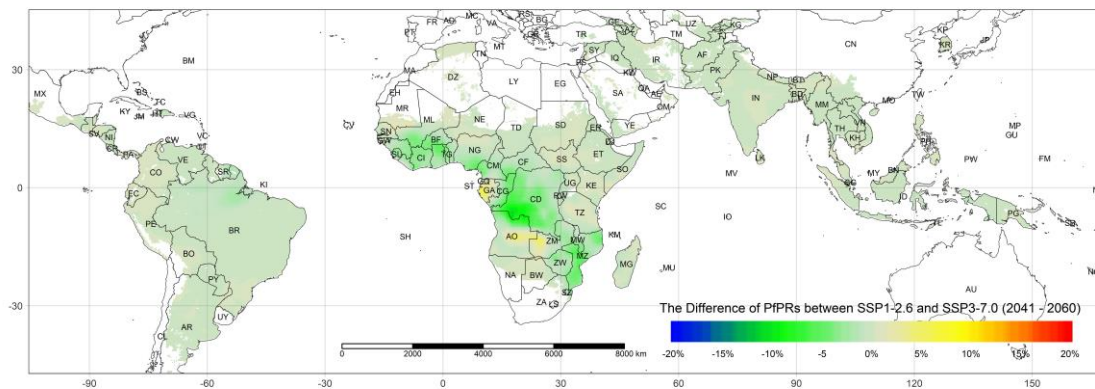


Figure S8: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2041 - 2060)



436

437 **Figure S9: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2061 - 2080)**



438

439 **Figure S10: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2061 - 2080)**

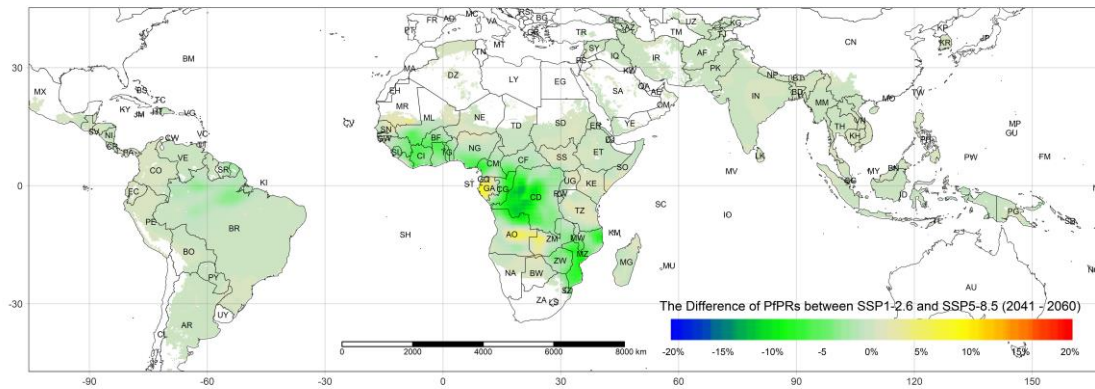


Figure S11: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2061 - 2080)

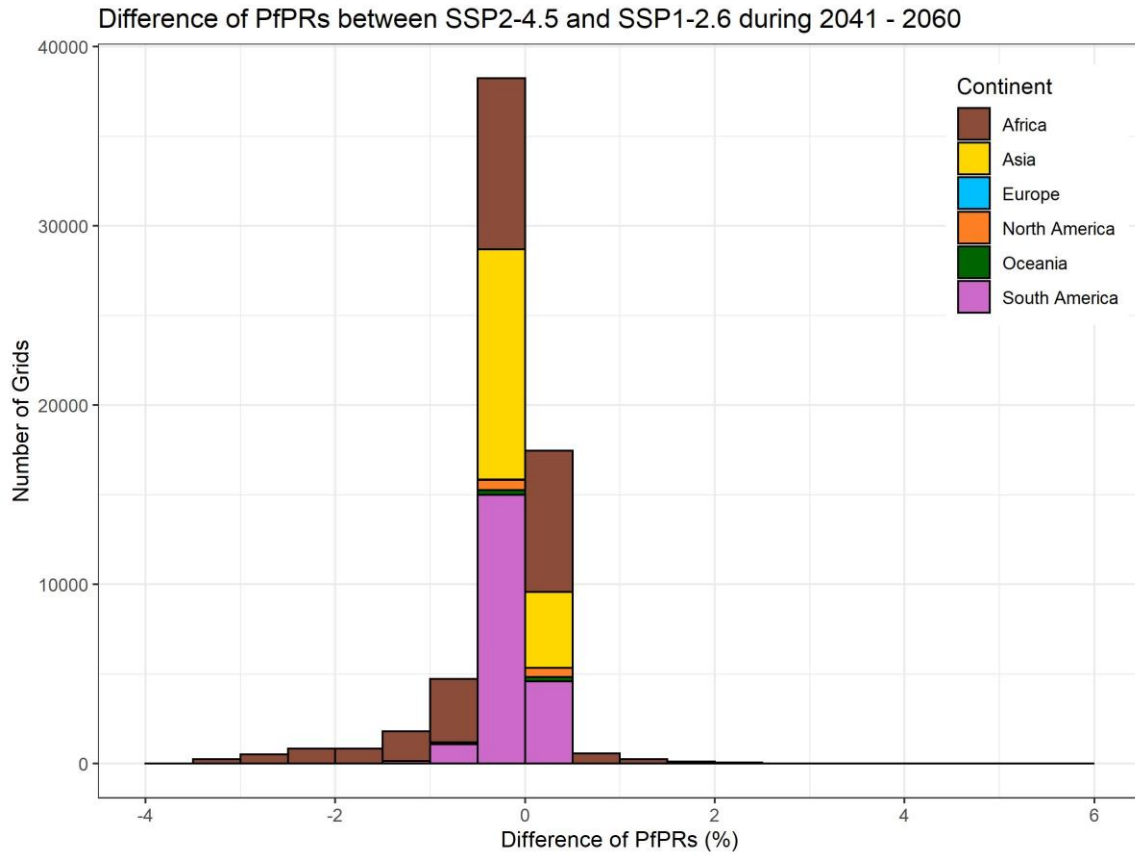


Figure S12: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2061 - 2080)

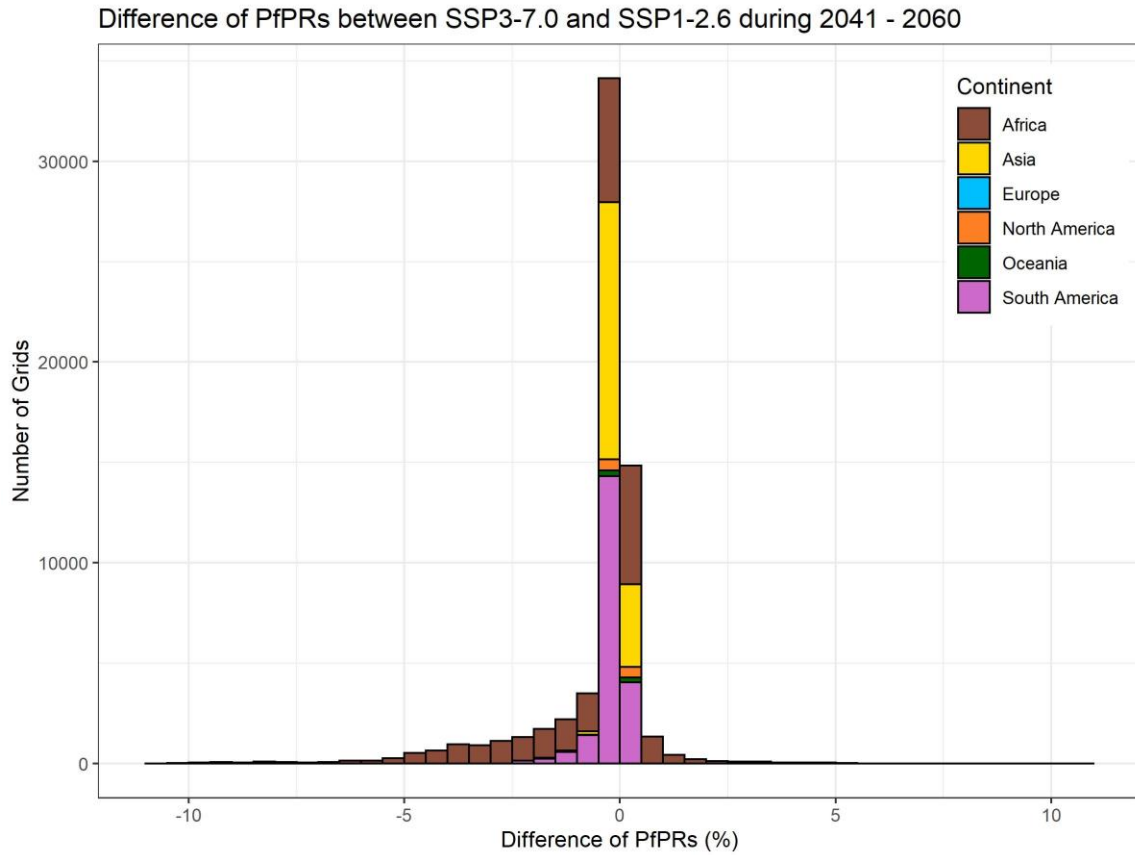


Figure S13: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2061 - 2080)

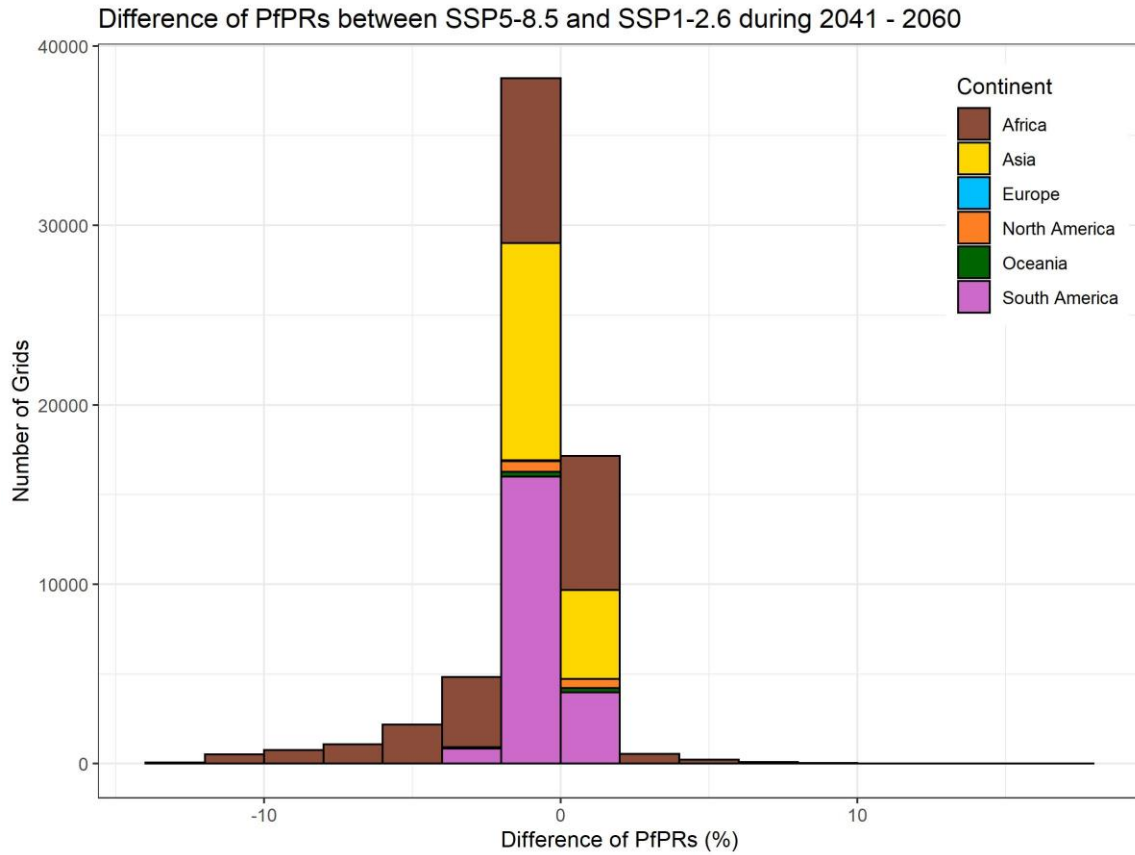
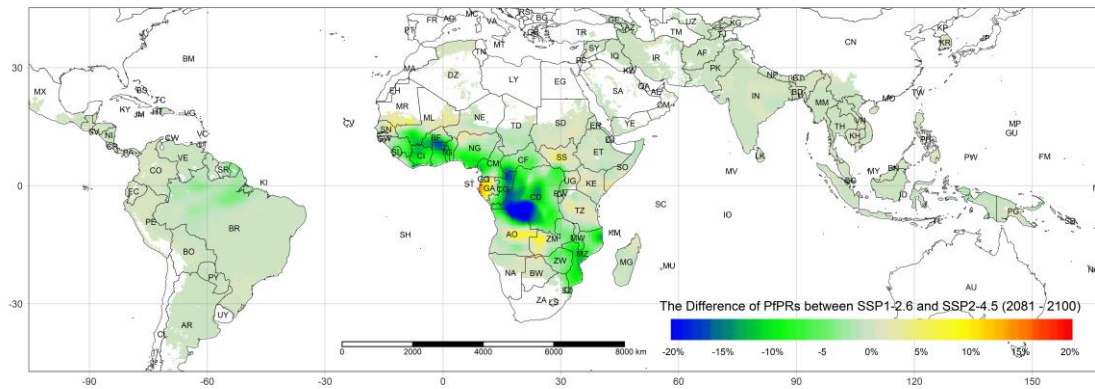
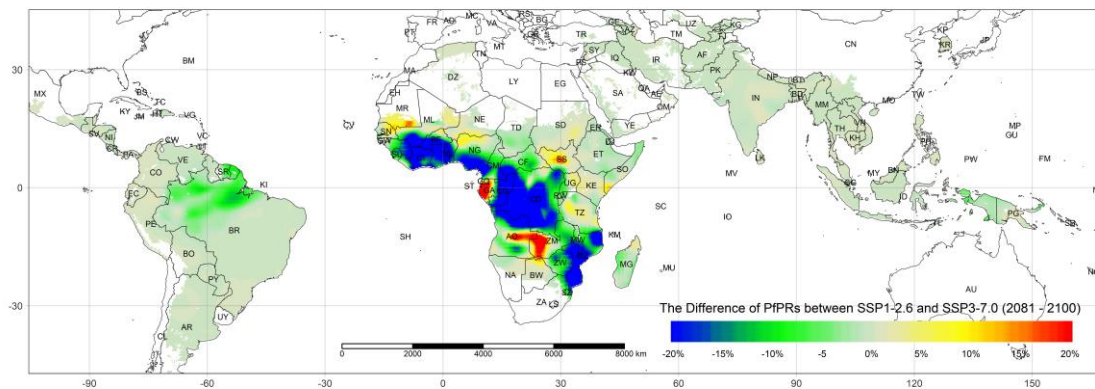


Figure S14: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2061 - 2080)



456

457 **Figure S15: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2081 - 2100)**



458

459 **Figure S16: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2081 - 2100)**

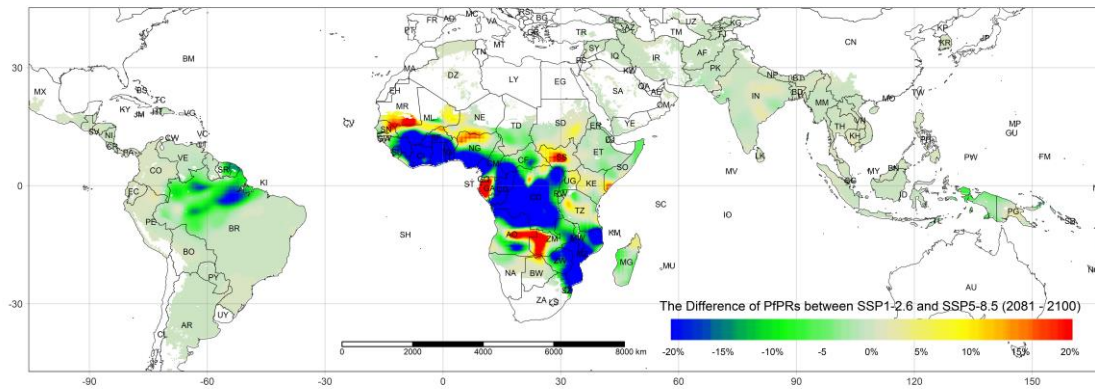


Figure S17: The Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2081 - 2100)

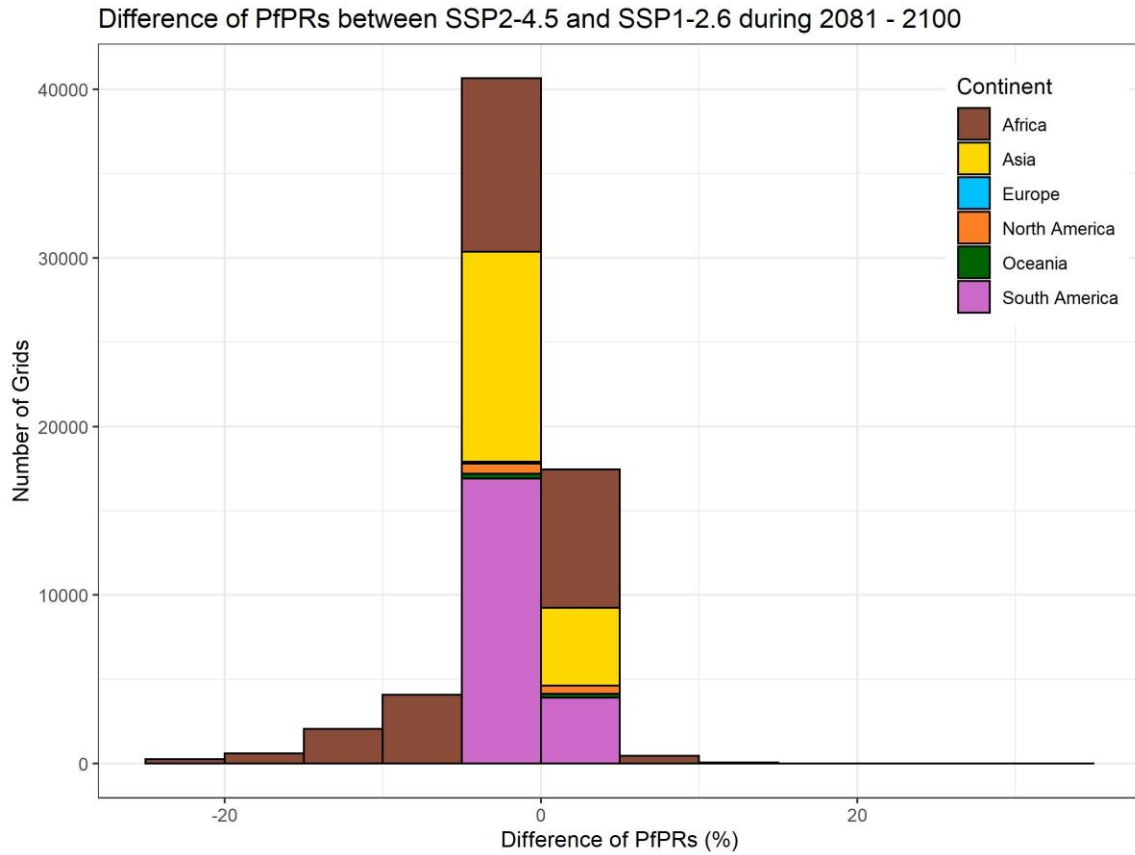


Figure S18: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP2-4.5 (2081 - 2100)

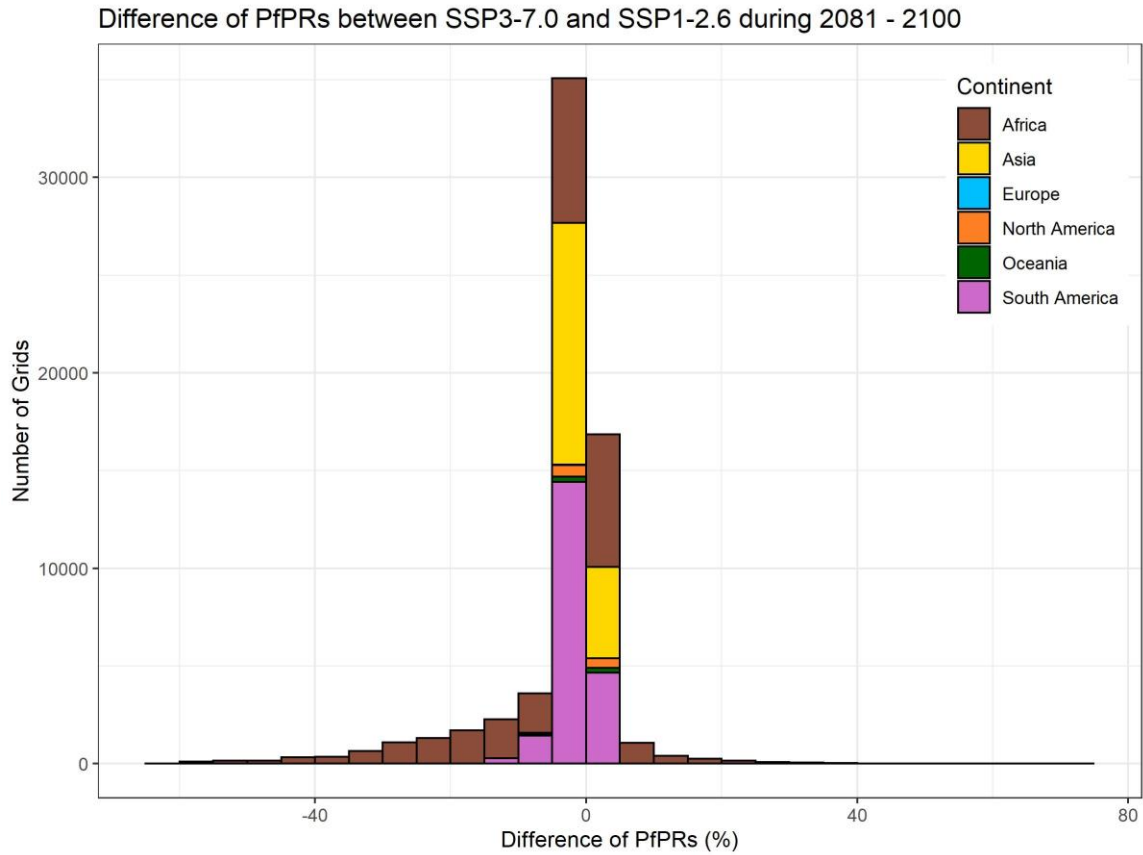


Figure S19: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP3-7.0 (2081 - 2100)

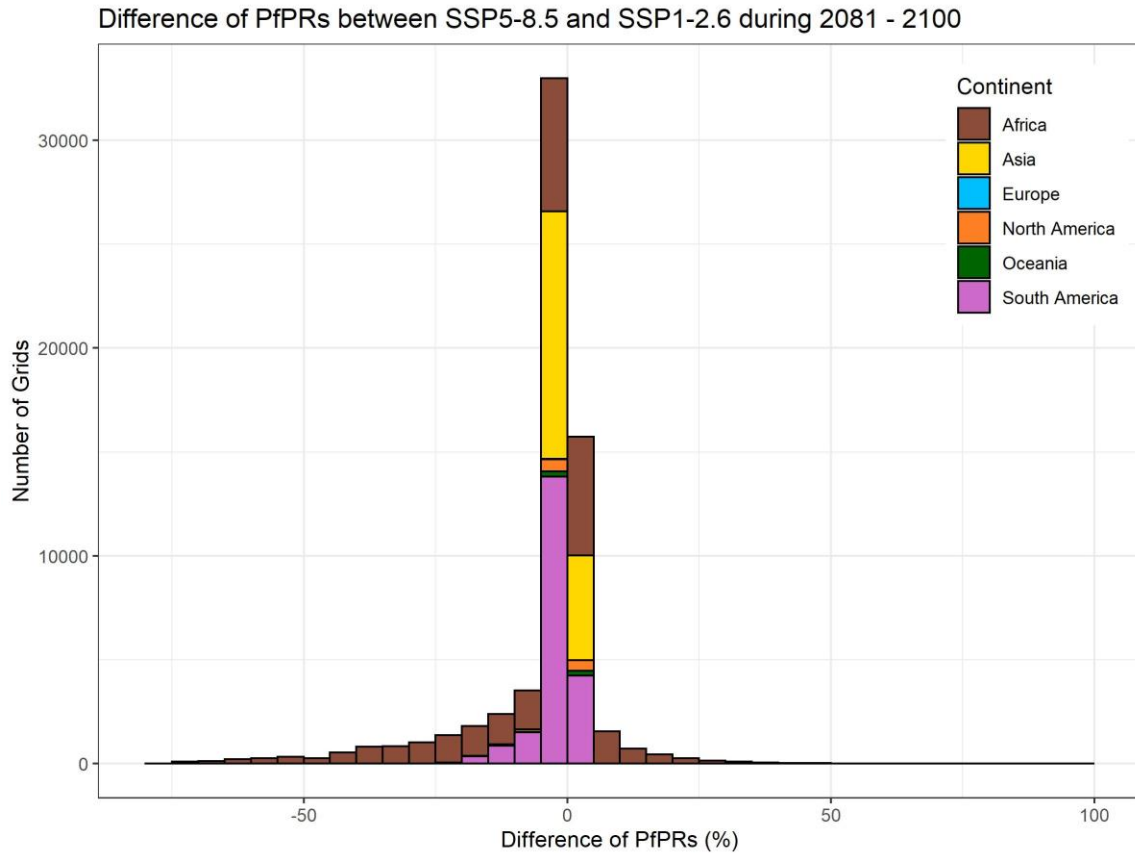


Figure S20: The Distribution of the Difference of PfPR₂₋₁₀ between SSP1-2.6 and SSP5-8.5 (2081 - 2100)

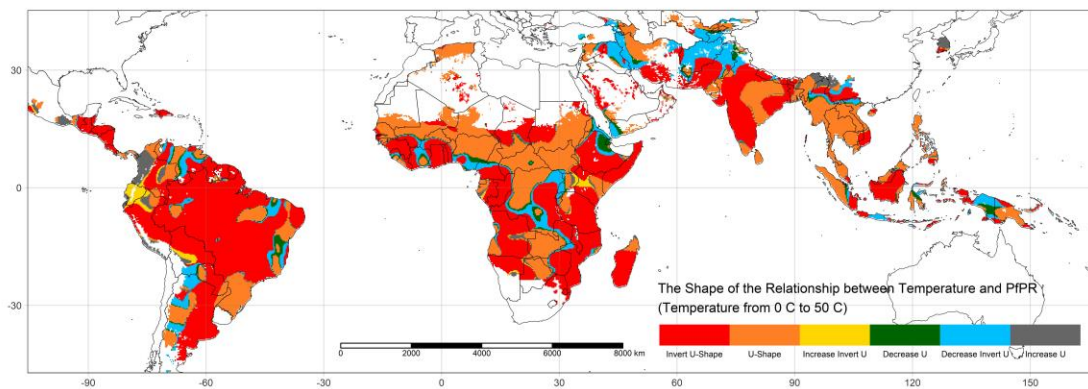


Figure S21: The Shape of Relationship between Temperature and PfPR₂₋₁₀

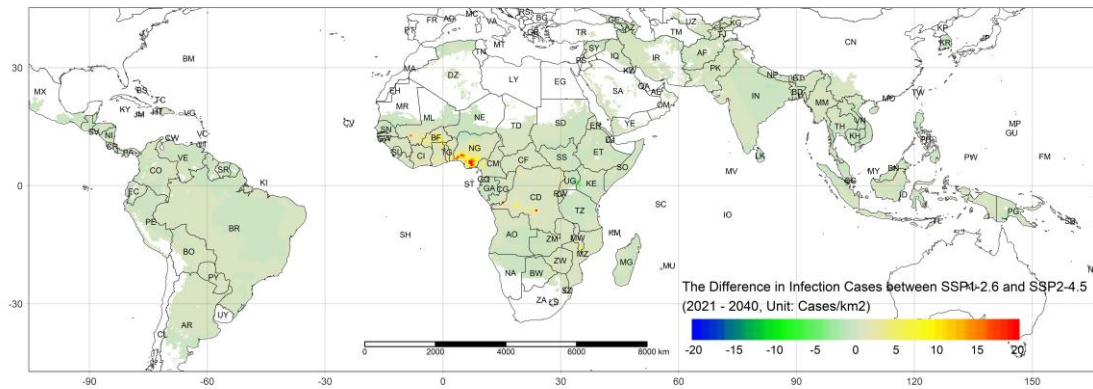


Figure S22: The Difference in Infection Cases between SSP1-2.6 and SSP2-4.5 during 2021 - 2040

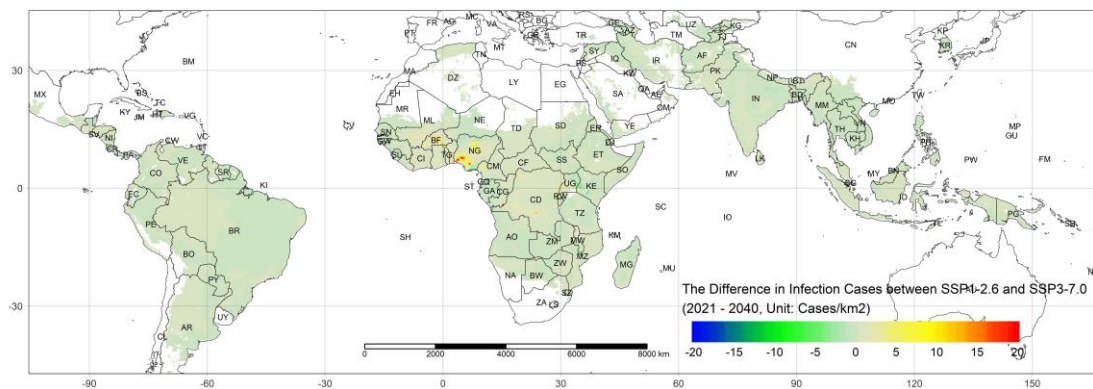


Figure S23: The Difference in Infection Cases between SSP1-2.6 and SSP3-7.0 during 2021 - 2040

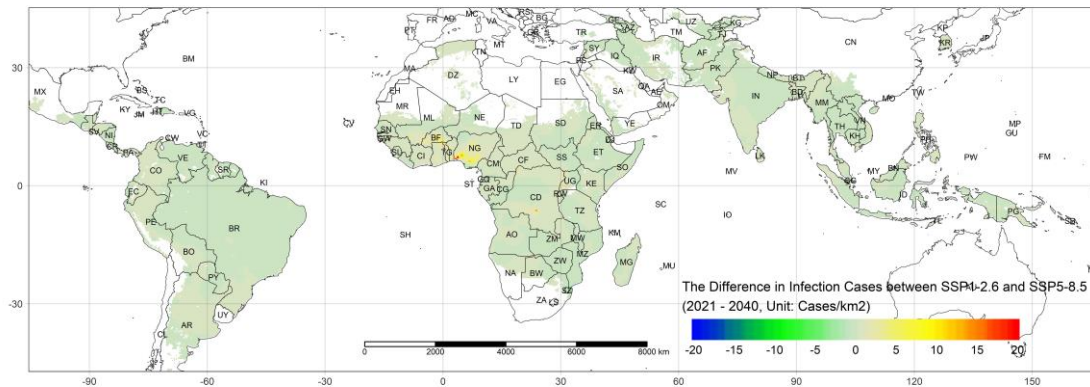


Figure S24: The Difference in Infection Cases between SSP1-2.6 and SSP5-8.5 during 2021 - 2040

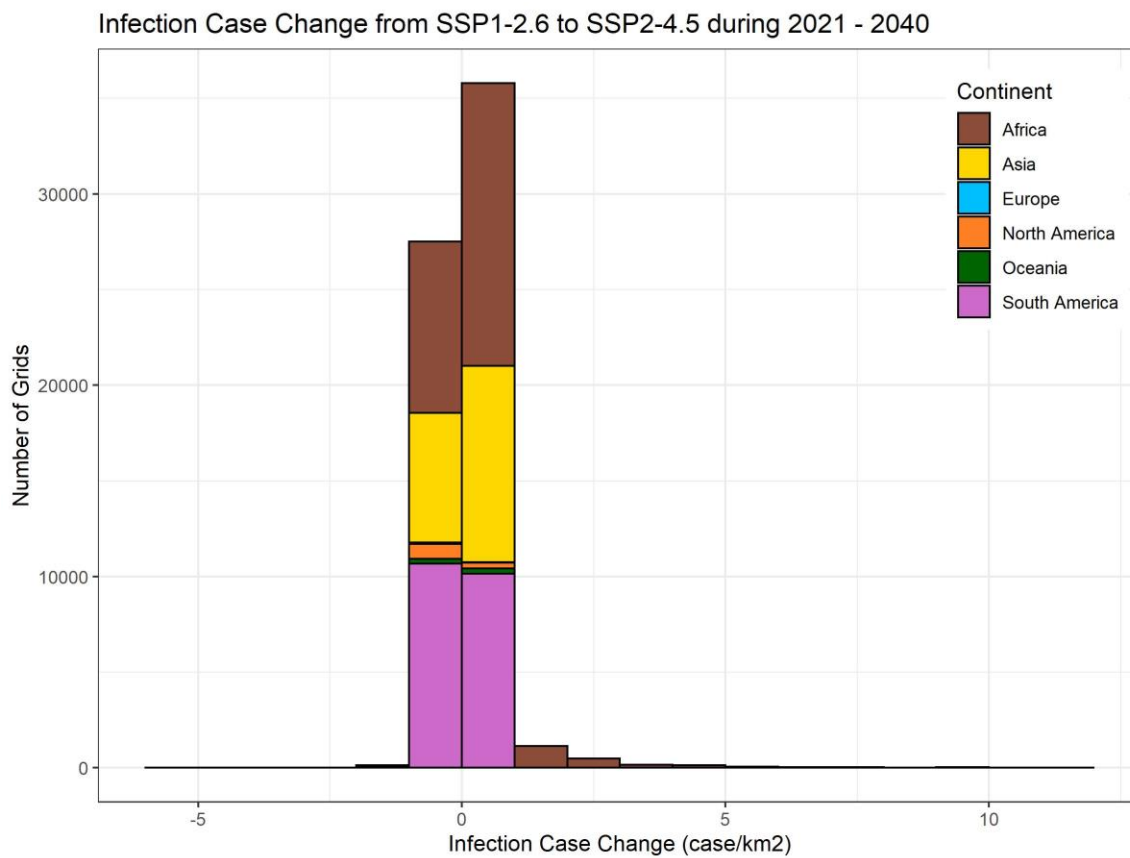
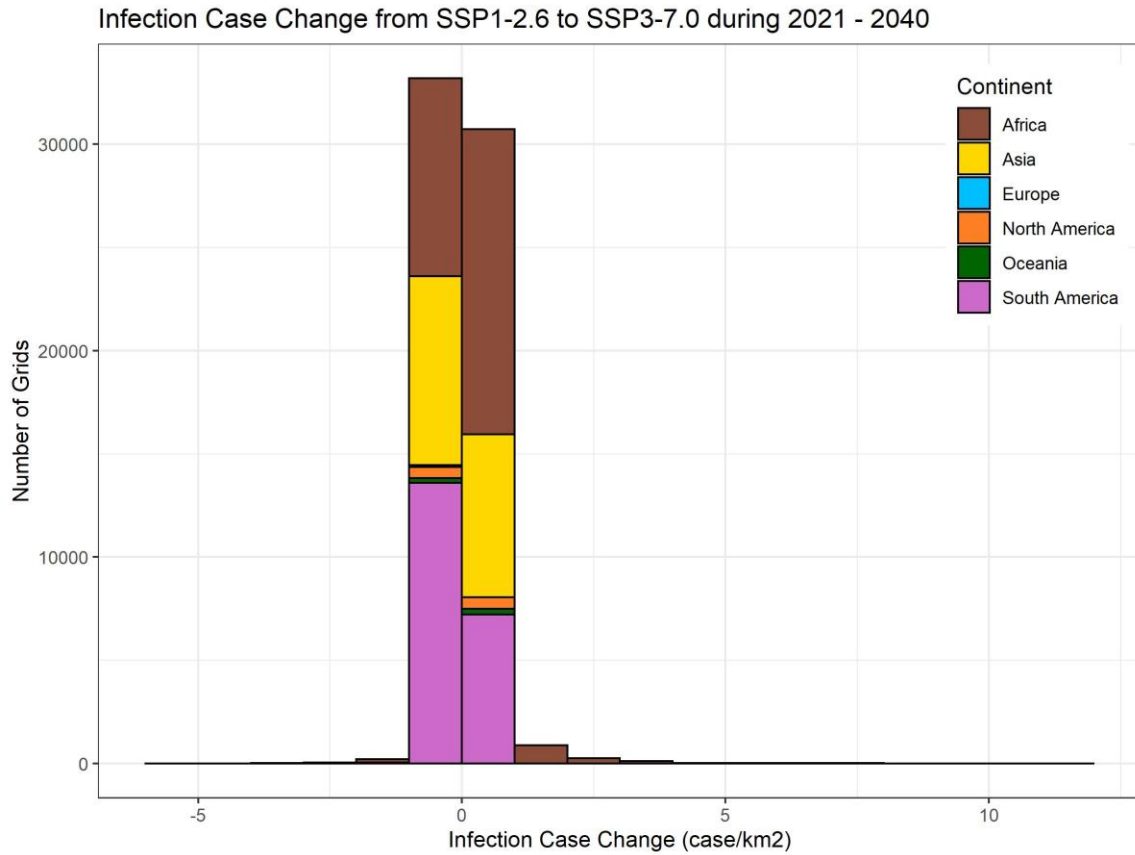


Figure S25: Infection Case Change from SSP1-2.6 to SSP2-4.5 during 2021 – 2040



487

488 **Figure S26: Infection Case Change from SSP1-2.6 to SSP3-7.0 during 2021 – 2040**

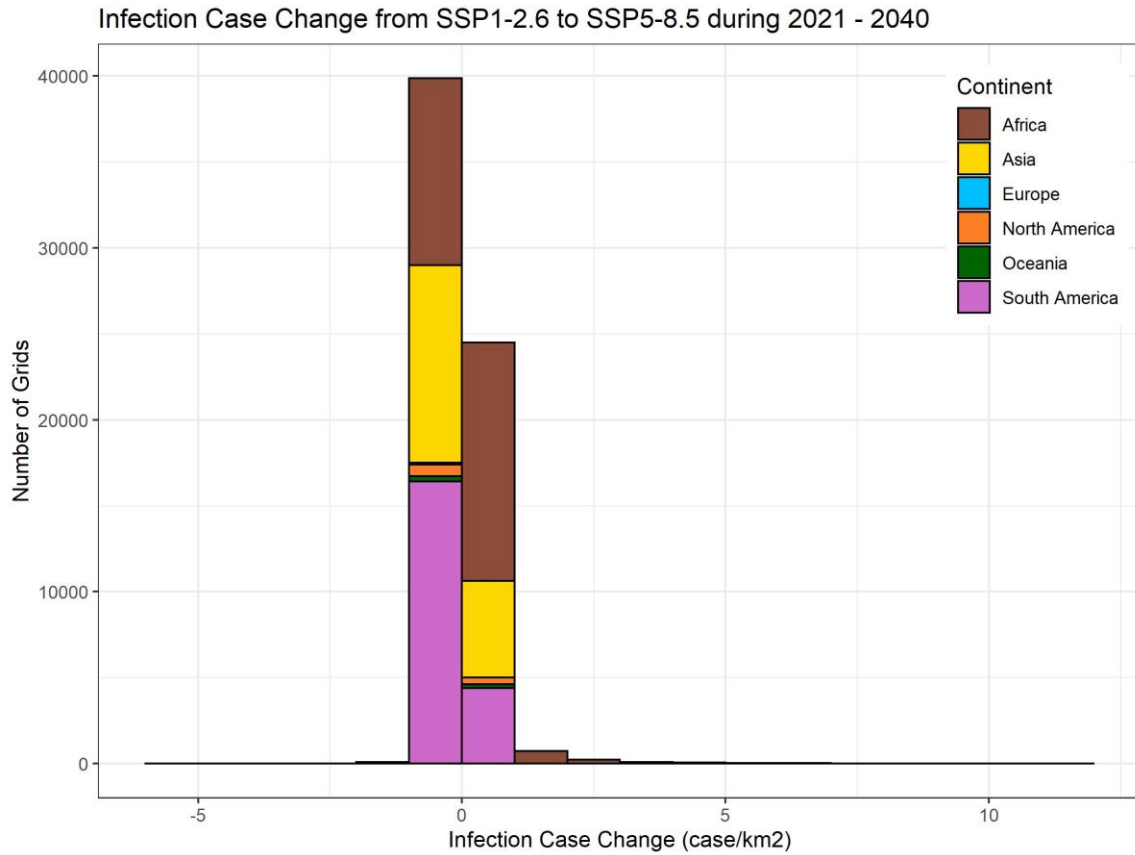


Figure S27: Infection Case Change from SSP1-2.6 to SSP5-8.5 during 2021 – 2040

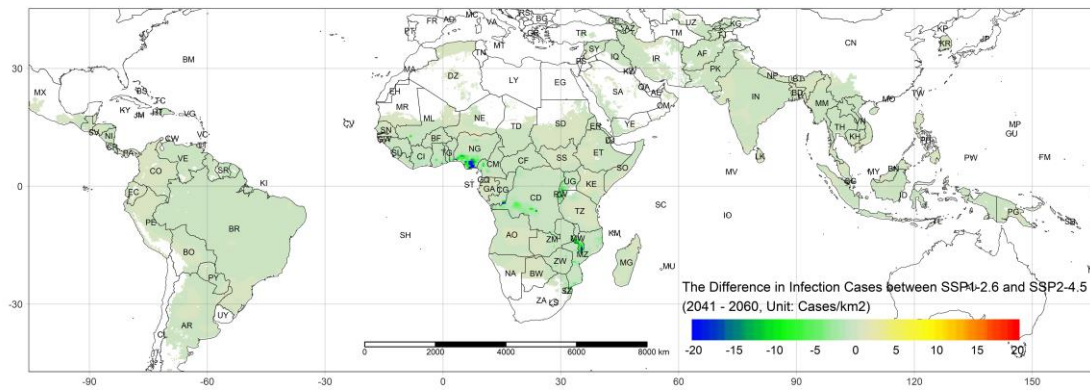


Figure S28: The Difference in Infection Cases between SSP1-2.6 and SSP2-4.5 during 2041 - 2060

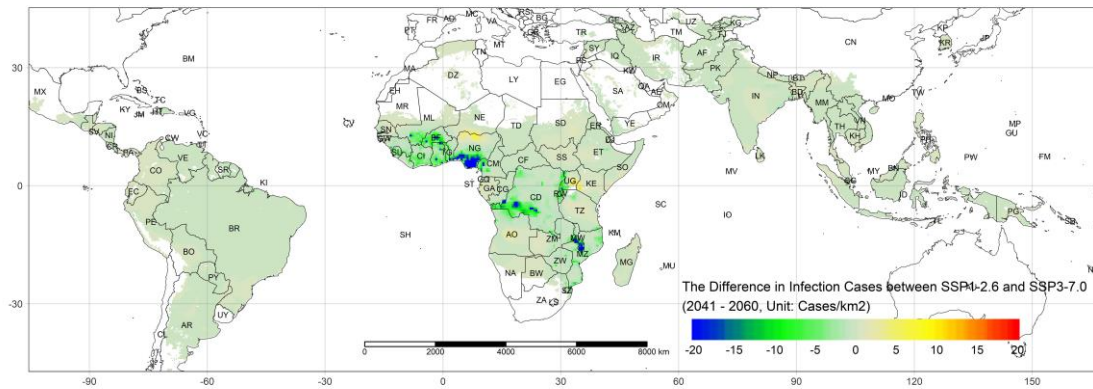


Figure S29: The Difference in Infection Cases between SSP1-2.6 and SSP3-7.0 during 2041 - 2060

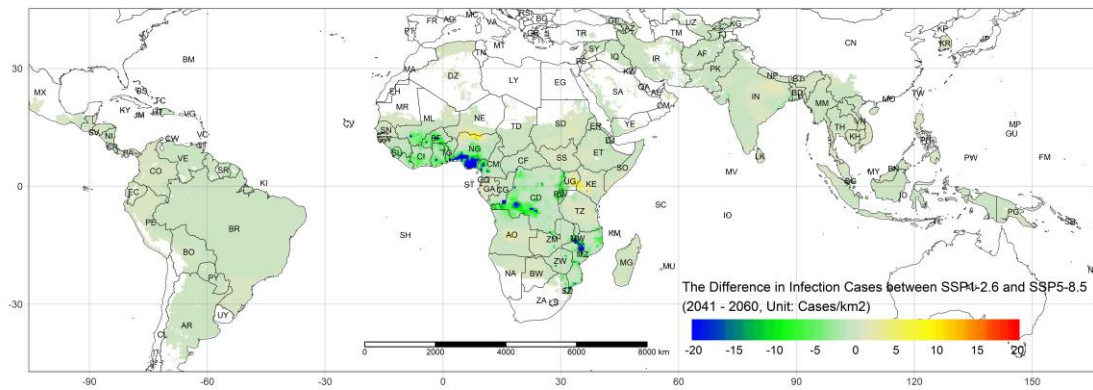


Figure S30: The Difference in Infection Cases between SSP1-2.6 and SSP5-8.5 during 2041 - 2060

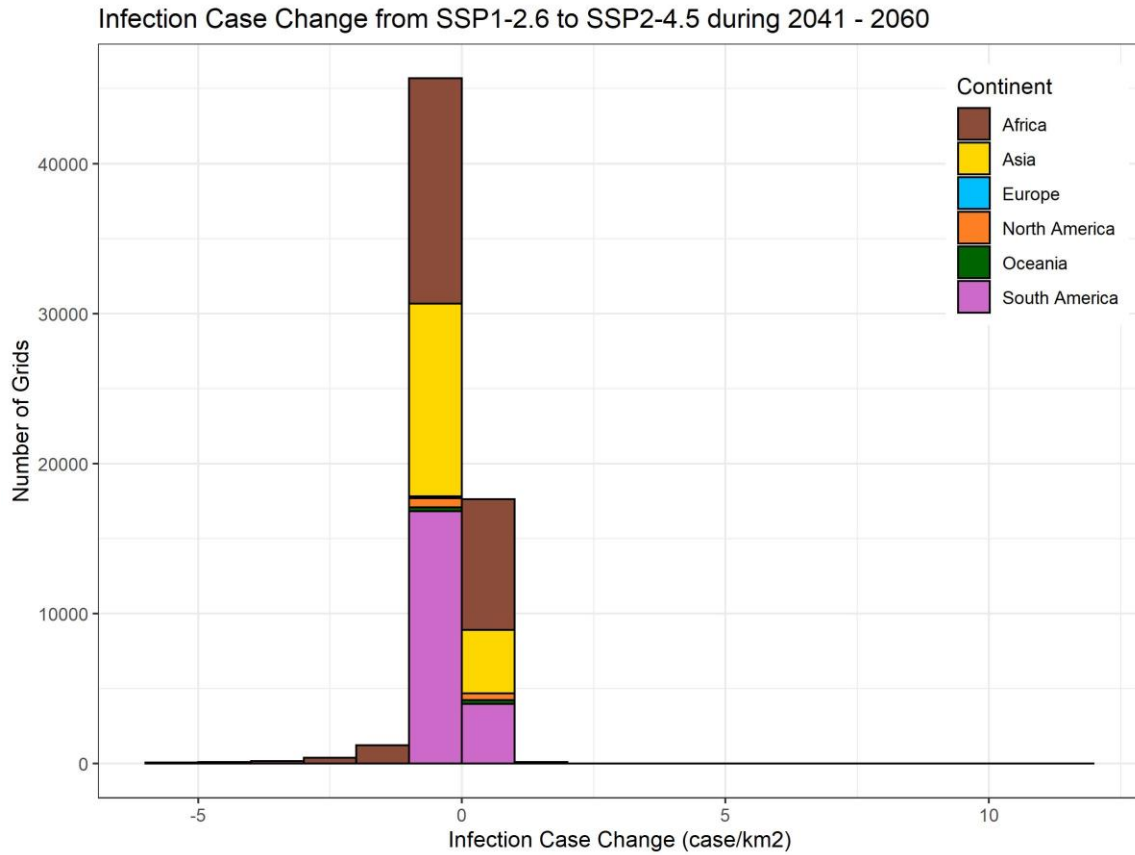


Figure S31: Infection Case Change from SSP1-2.6 to SSP2-4.5 during 2041 – 2060

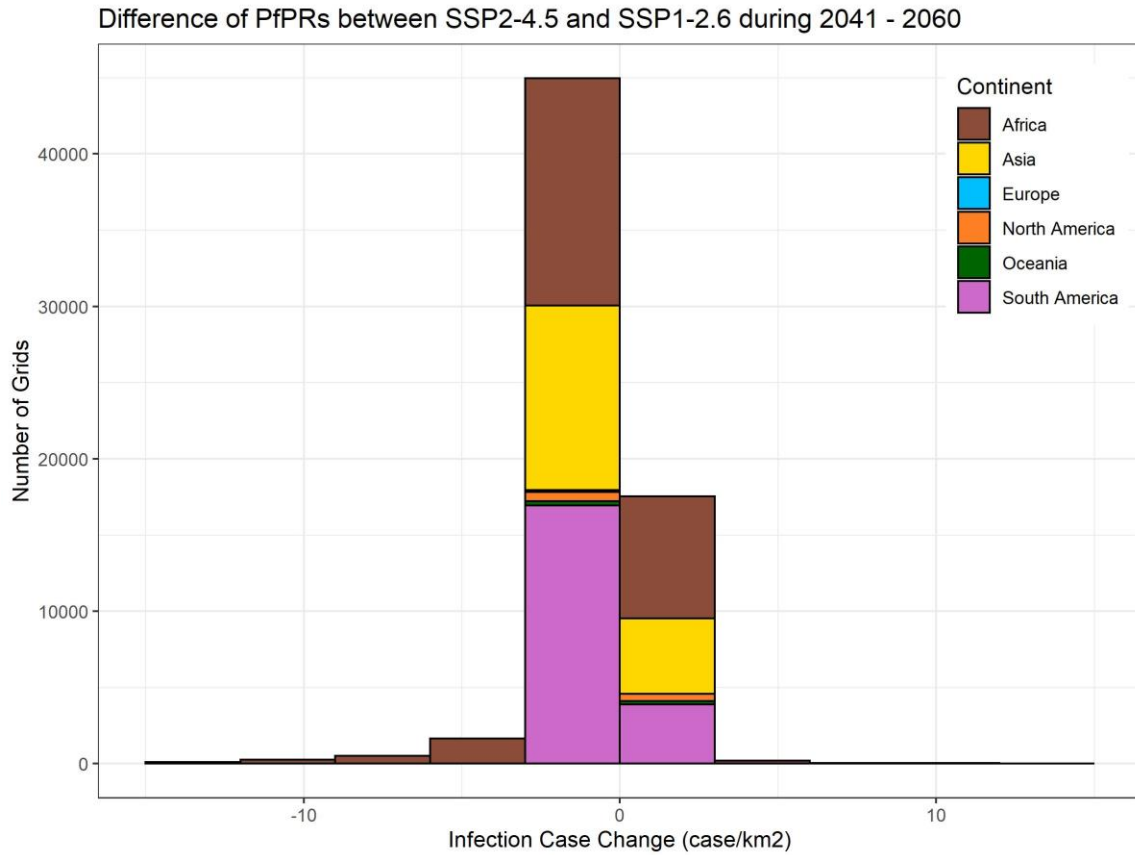


Figure S32: Infection Case Change from SSP1-2.6 to SSP3-7.0 during 2041 – 2060

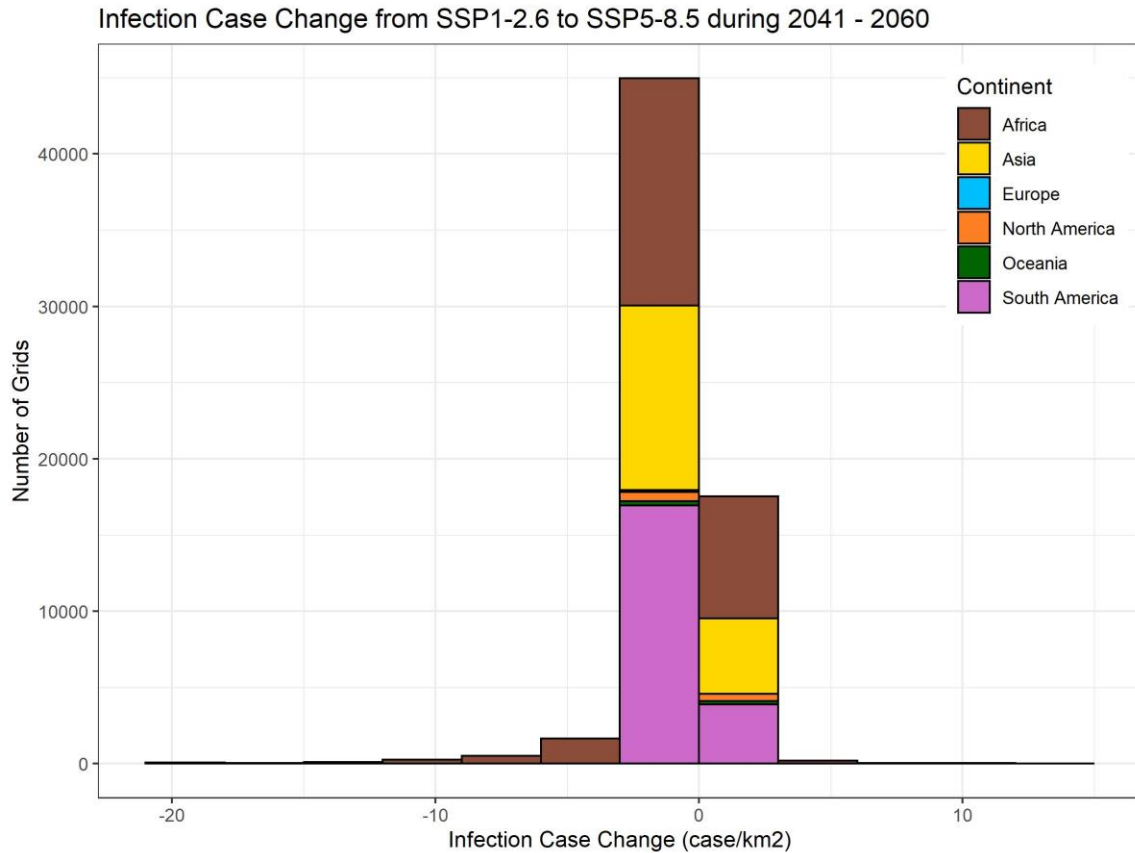


Figure S33: Infection Case Change from SSP1-2.6 to SSP5-8.5 during 2041 – 2060

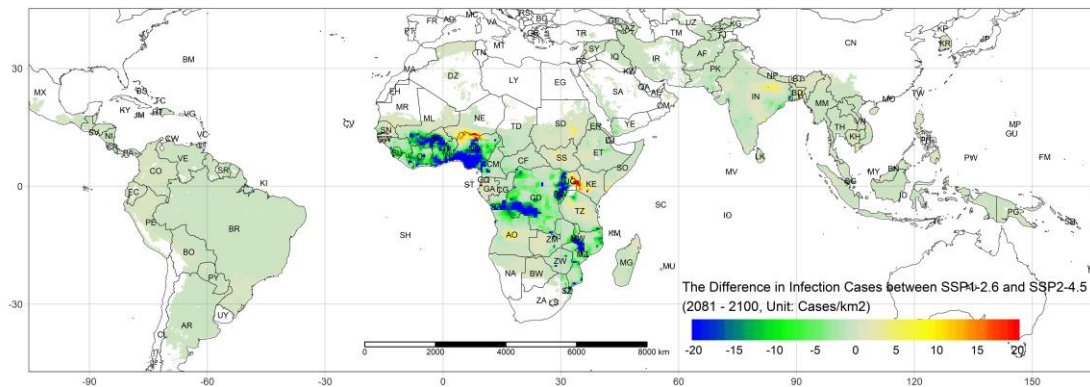
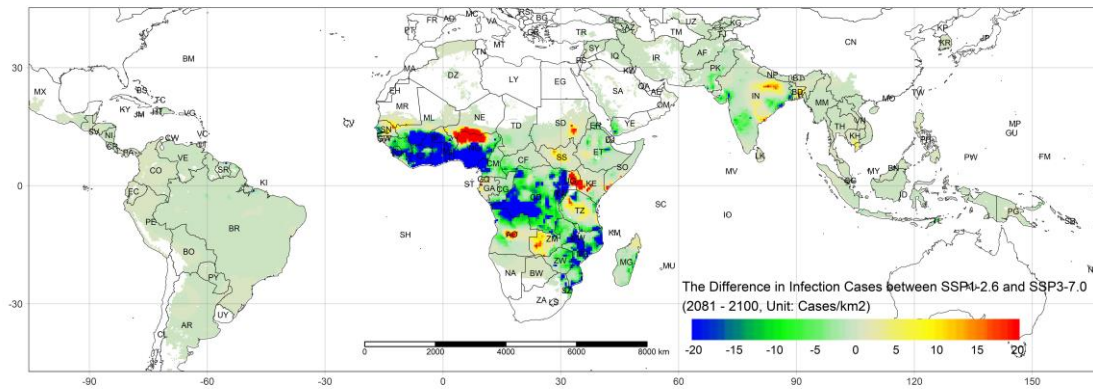
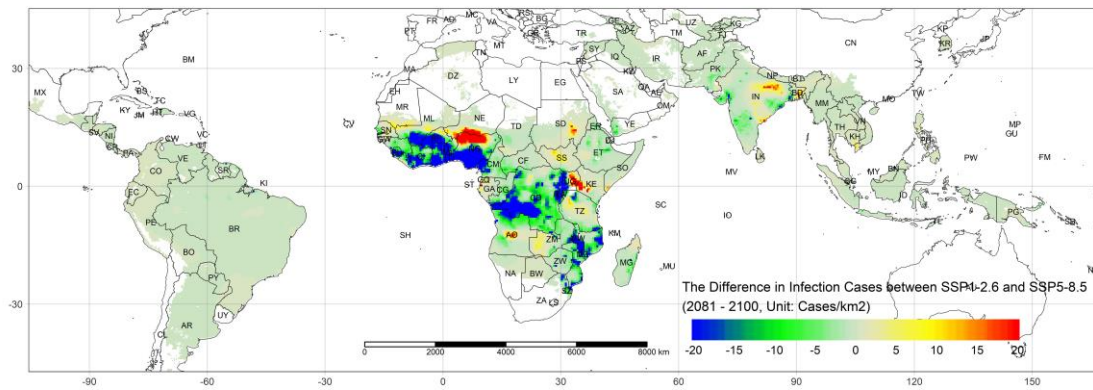


Figure S34: The Difference in Infection Cases between SSP1-2.6 and SSP2-4.5 during 2081 - 2100



**Figure S35: The Difference in Infection Cases between SSP1-2.6 and SSP3-7.0
during 2081 - 2100**



**Figure S36: The Difference in Infection Cases between SSP1-2.6 and SSP5-8.5
during 2081 - 2100**

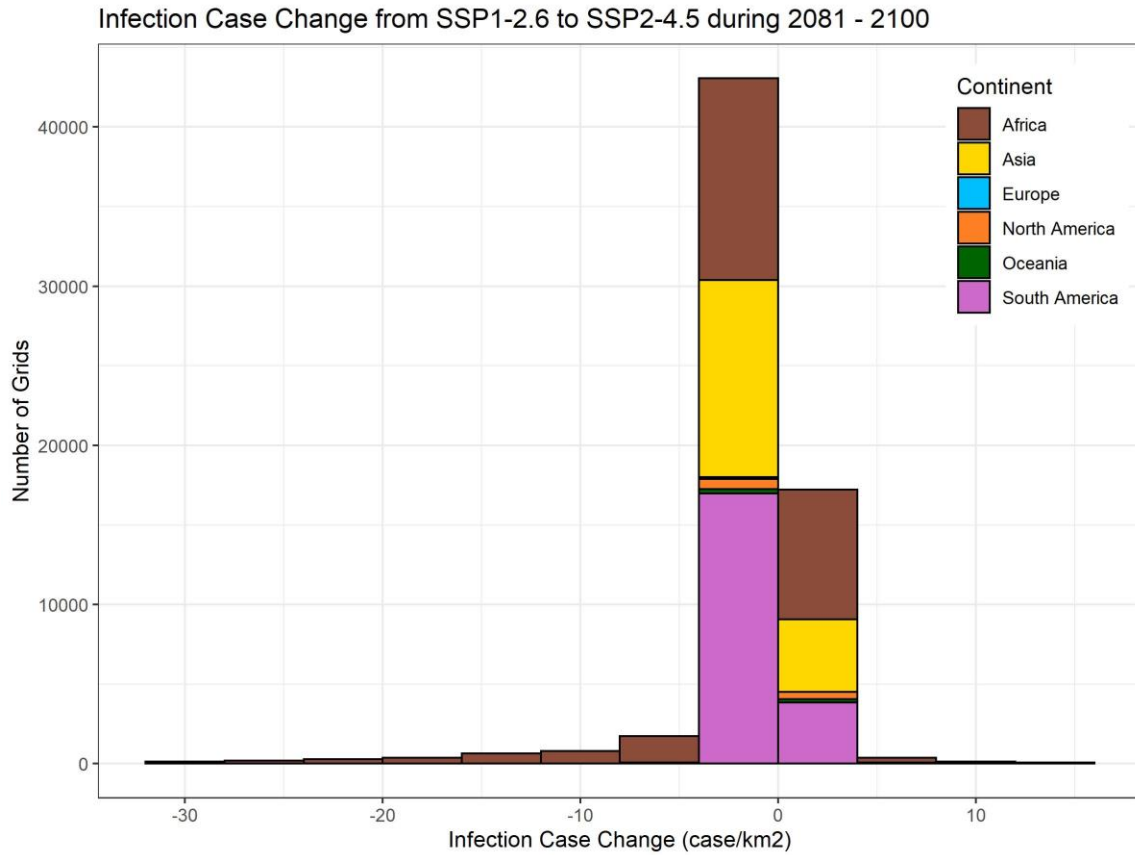


Figure S37: Infection Case Change from SSP1-2.6 to SSP2-4.5 during 2081 – 2100

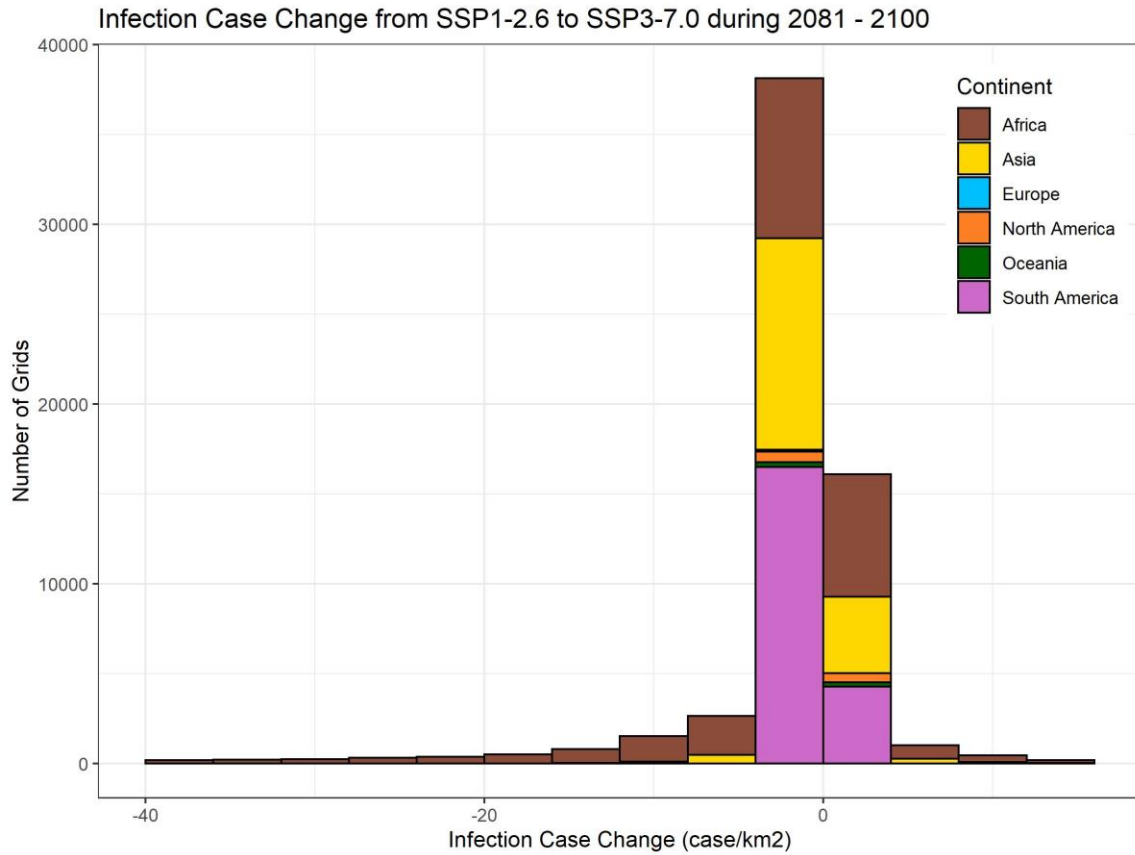


Figure S38: Infection Case Change from SSP1-2.6 to SSP3-7.0 during 2081 – 2100

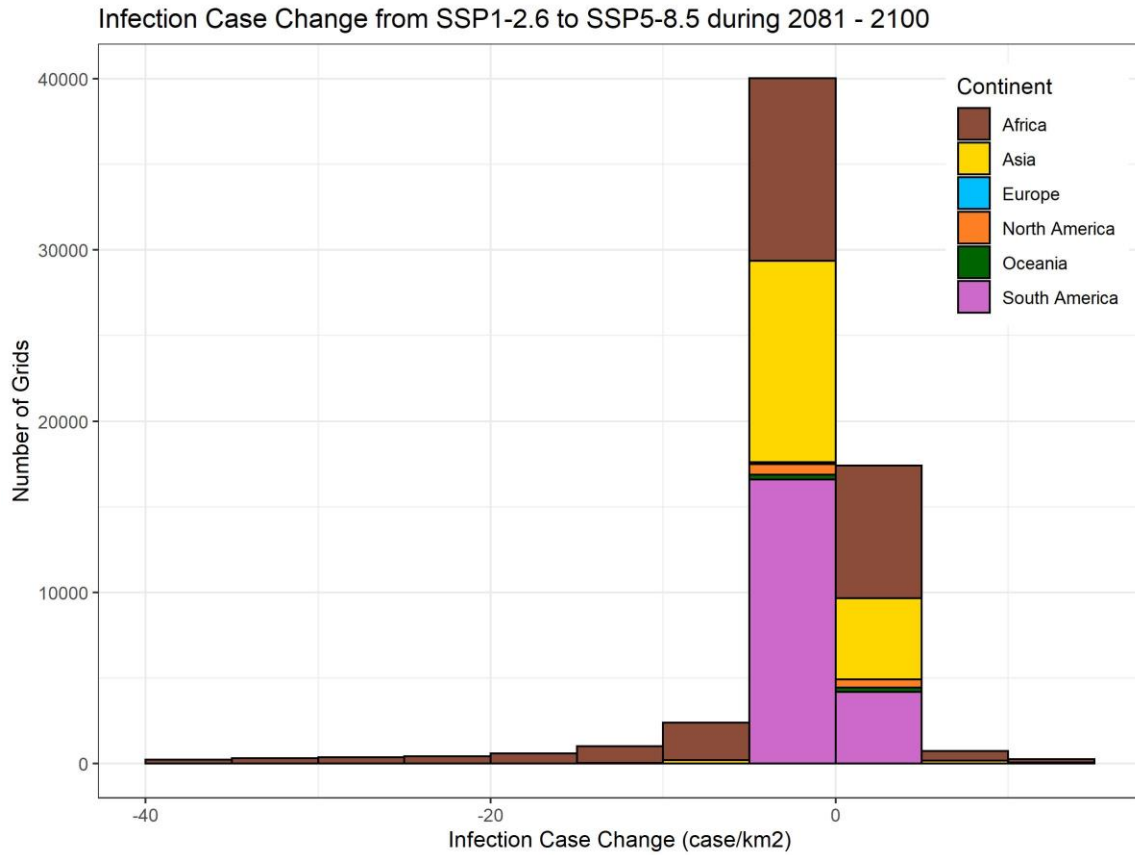


Figure S39: Infection Case Change from SSP1-2.6 to SSP5-8.5 during 2081 – 2100

Table S1: Statistical Indicators Summary of 10-Fold Cross Validation

Statistical Indicator of Testing								Statistical Indicator of Training							
	N	R ²	RMSE(%)	MAE(%)	α	β	r	N	R ²	RMSE(%)	MAE(%)	α	β	r	
1	1158867	94.75%	4.45%	1.94%	0.006	0.944	0.973	128763	94.67%	4.49%	1.95%	0.006	0.944	0.973	
2	1158867	94.74%	4.45%	1.94%	0.006	0.944	0.973	128763	94.78%	4.44%	1.93%	0.006	0.943	0.974	
3	1158867	94.75%	4.45%	1.94%	0.006	0.943	0.973	128763	94.72%	4.44%	1.94%	0.006	0.944	0.973	
4	1158867	94.75%	4.45%	1.94%	0.006	0.944	0.973	128763	94.76%	4.46%	1.94%	0.006	0.944	0.973	
5	1158867	94.75%	4.45%	1.94%	0.006	0.943	0.973	128763	94.69%	4.46%	1.94%	0.006	0.945	0.973	
6	1158867	94.74%	4.45%	1.94%	0.006	0.943	0.973	128763	94.75%	4.45%	1.94%	0.006	0.945	0.973	
7	1158867	94.75%	4.45%	1.94%	0.006	0.944	0.973	128763	94.65%	4.48%	1.94%	0.006	0.942	0.973	
8	1158867	94.74%	4.45%	1.94%	0.006	0.944	0.973	128763	94.82%	4.45%	1.95%	0.006	0.943	0.974	
9	1158867	94.75%	4.45%	1.94%	0.006	0.943	0.973	128763	94.73%	4.46%	1.93%	0.006	0.945	0.973	
10	1158867	94.74%	4.45%	1.94%	0.006	0.944	0.973	128763	94.75%	4.45%	1.94%	0.006	0.942	0.973	

Note: The Ideal value of β is 1, and the ideal value of α is 0. When the values of statistical indicators are close between training results and testing result, the predictions of model are reliable.

525 **Reference:**

- 526 1. Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on *Plasmodium*
527 *falciparum* in Africa between 2000 and 2015. *Nature* 2015; **526**(7572): 207-11.
- 528 2. Rodell M, Houser PR, Jambor U, et al. The Global Land Data Assimilation System.
529 *Bulletin of the American Meteorological Society* 2004; **85**(3): 381-94.
- 530 3. Mordecai EA, Ryan SJ, Caldwell JM, Shah MM, Labeaud AD. Climate change
531 could shift disease burden from malaria to arboviruses in Africa. *The Lancet Planetary*
532 *Health* 2020; **4**(9): e416-e23.
- 533 4. Mordecai EA, Paaijmans KP, Johnson LR, et al. Optimal temperature for malaria
534 transmission is dramatically lower than previously predicted. *Ecology Letters* 2013; **16**(1):
535 22-30.
- 536 5. Mordecai EA, Caldwell JM, Grossman MK, et al. Thermal biology of mosquito-
537 borne disease. *Ecology Letters* 2019; **22**(10): 1690-708.
- 538 6. Johnson LR, Ben-Horin T, Lafferty KD, et al. Understanding uncertainty in
539 temperature effects on vector-borne disease: a Bayesian approach. *Ecology* 2015; **96**(1):
540 203-13.
- 541 7. Lourenço PM, Sousa CA, Seixas J, Lopes P, Novo MT, Almeida APG. *Anopheles*
542 *atroparvus* density modeling using MODIS NDVI in a former malarious area in Portugal.
543 *Journal of Vector Ecology* 2011; **36**(2): 279-91.
- 544 8. Reiner RC, Geary M, Atkinson PM, Smith DL, Gething PW. Seasonality of
545 *Plasmodium falciparum* transmission: a systematic review. *Malaria Journal* 2015; **14**(1).
- 546 9. Bruguera S, Fernández-Martínez B, Martínez-De La Puente J, et al. Environmental
547 drivers, climate change and emergent diseases transmitted by mosquitoes and their vectors
548 in southern Europe: A systematic review. *Environ Res* 2020; **191**: 110038.
- 549 10. Amadi JA, Olago DO, Ong'Amo GO, et al. Sensitivity of vegetation to climate
550 variability and its implications for malaria risk in Baringo, Kenya. *PLoS One* 2018; **13**(7):
551 e0199357.
- 552 11. Midekisa A, Senay G, Henebry GM, Semuniguse P, Wimberly MC. Remote
553 sensing-based time series models for malaria early warning in the highlands of Ethiopia.
554 *Malaria Journal* 2012; **11**(1): 165.
- 555 12. Didan K, Munoz AB, Solano R, Huete A, University of Arizona: Vegetation Index
556 Phenology Lab. MODIS vegetation index user's guide (MOD13 series). 2015.
- 557 13. Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ. High-
558 resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015,
559 and 2020. *Scientific Data* 2015; **2**(1): 150045.
- 560 14. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population Distribution,
561 Settlement Patterns and Accessibility across Africa in 2010. *PLoS One* 2012; **7**(2): e31743.
- 562 15. World Health Organization. World malaria report 2021, 2021.
- 563 16. IPCC. Climate change 2022: impacts, adaptation and vulnerability, 2022.
- 564 17. IPCC. Climate change 2014 synthesis report. Geneva, Switzerland: IPCC, 2014.
- 565 18. Rogers DJ, Randolph SE. Climate Change and Vector-Borne Diseases. In: Hay SI,
566 Graham A, Rogers DJ, eds. *Advances in Parasitology*: Academic Press; 2006: 345-81.

19. Altizer S, Ostfeld Richard S, Johnson Pieter TJ, Kutz S, Harvell CD. Climate Change and Infectious Diseases: From Evidence to a Predictive Framework. *Science* 2013; **341**(6145): 514-9.
20. Buchwald AG, Hayden MH, Dadzie SK, Paull SH, Carlton EJ. Aedes-borne disease outbreaks in West Africa: A call for enhanced surveillance. *Acta Tropica* 2020; **209**: 105468.
21. Nyaruaba R, Mwaliko C, Mwau M, Mousa S, Wei H. Arboviruses in the East African Community partner states: a review of medically important mosquito-borne Arboviruses. *Pathogens and Global Health* 2019; **113**(5): 209-28.
22. Brunson C, Fotheringham S, Charlton M. Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician)* 1998; **47**(3): 431-43.
23. Fotheringham A, Brunson C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships: John Wiley & Sons; 2002.
24. Croissant Y, Mollo G. Panel Data Econometrics in R: The plm Package. *J Stat Softw* 2008; **27**(2).
25. Breusch TS, Pagan AR. The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies* 1980; **47**(1): 239.
26. Kang S. A note on the equivalence of specification tests in the two-factor multivariate variance components model. *Journal of Econometrics* 1985; **28**(2): 193-203.
27. Giesen C, Roche J, Redondo-Bravo L, et al. The impact of climate change on mosquito-borne diseases in Africa. *Pathogens and Global Health* 2020; **114**(6): 287-301.
28. Kulkarni MA, Duguay C, Ost K. Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: a scoping review of reviews. *Globalization and Health* 2022; **18**(1).
29. Brunson C, Fotheringham AS, Charlton ME. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 2010; **28**(4): 281-98.
30. Fotheringham AS, Oshan TM. Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems* 2016; **18**(4): 303-29.
31. Beenstock M, Felsenstein D. The econometric analysis of non-stationary spatial panel data: Springer; 2019.
32. Gollini I, Lu B, Charlton M, Brunson C, Harris P. GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *J Stat Softw* 2015; **63**(17).