1 # Supplementary appendix

2 # Supplement to: Global malaria infection risk from global

3 # warming

4

5 ## Authors

6 Chao Li[1], Shunsuke Managi*[1]

7

8 ## Affiliations

9 1 Urban Institute, Kyushu University, Japan

10 * Correspondent to: Shunsuke Managi, managi@doc.kyushu-u.ac.jp, Kyushu University

11 744 Motooka, Nishi-ku, Fukuoka 819-0395 Japan

12

13

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 $^{\circ}$ 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 ($PfPR_{2-10}$) 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 $PfPR_{2-10}$ 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

36   humidity, precipitation and wind speed. These variables are all monthly average values and

37   have converted into annual average values. However, the correlations among these climatic

38   variables are too high, which would lead to multicollinearity in the regressions. For

39   example, the correlation coefficient between annual average air pressure and temperature

40   is 0.745 (p-value < 0.1%). Additionally, another critical independent variable, vegetation

41   index, is significantly correlated with absolute humidity (correlation coefficient: 0.807, p-

42   value < 0.1%), precipitation (0.805, p-value < 0.1%) and wind speed (-0.716, p-value <

43   0.1%). Therefore, the only annual average temperature is kept among those five climatic

44   variables. Secondly, evidence shows that malaria transmission through *Anopheles gambiae*

45   peaks at 25°C [3]. Hence, the relationship between malaria infection rate and temperature is

46   non-linear and similar to the U-shape [4-6]. We take the temperature square into account.

47   Furthermore, extreme low and high temperatures are unsuitable for transmission vector

48   survival. In this way, the standard deviation of monthly average temperature in a specific

49   year would also affect the PfPR. In terms of climate, three variables, involving annual

50   average temperature, annual average temperature square, and standard deviation of

51   monthly average temperature in a specific year, are incorporated into the analysis.

52

53   ***Vegetation Index***

54   Local vegetation affects malaria vector population size, survival, biting, among

55   others [7-9]. Therefore, the local vegetation status should be considered in the analysis. The

56   normalized difference vegetation index (NDVI) is a widely used indicator to assess how

57   much the observed grid contains live green vegetation [10,11]. The NDVI ranges from -100%

58   (no live green vegetation) to 100% (full of live green vegetation). MODIS Terra and Aqua

59   satellites provide monthly NDVI data from 2000 to the present with a 0.05-arc-degree

60   spatial resolution. The MODIS vegetation index products used in the analysis are

61   MOD13C2 and MYD13C2 [12]. To link the NDVI data set to the temperature and $PfPR_{2\text{-}10}$

62   data set, we upscale the NDVI data set to 0.25 arc degree by the average method. Then,

63   these data are further converted into annual average values.

64

Population Density and GDP per Capita

66   The population density data are obtained from the WorldPop Project. The

67   WorldPop Project provides spatial grid data set of population count from 2000 to 2020

68   with a 1-km resolution [13,14]. Because the length of the latitudes decreases with the

69   increasing latitude, the conversion of population count from 1-km to 0.25-arc-degree

70   resolution is not perfect. At least, we cannot use the direct sum method. To solve this issue,

71   we regard the population count in every grid as the population density, and the unit of this

72   data set is regarded as capita per square kilometer (Cap/km^2). Here, we upscale the data

73   from 1-km to 0.25-arc-degree resolution by the average method. The data set eventually

74   illustrates the annual population density with a 0.25-arc-degree resolution.

75   GDP per capita is associated with development level and public health investment.

76   The higher development level and more public health investment, including a wide range

77   of international aids, are linked with a relatively lower malaria infection rate [1,15]. However,

78   the spatial grid data set about GDP per capita is unavailable. We employ the country-level

79   GDP per capita from the World Bank. The unit of GDP per capita is the current U.S. dollar.

80   We assume that the GDP per capita in every grid is the same in a country.

81

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. Simply speaking, the SSP1-2.6 scenario predicts that the

98   temperatures stabilize 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

104 monthly average temperature. The IPCC provides open access data based on over 30-model

105 stimulation with a 1-arc-degree resolution (https://interactive-atlas.ipcc.ch/).

106

107 *Grids*

108   We integrate the abovementioned data set to build a long panel spatial grid data set.

109 Because we only focus on the endemic areas, the grids are dropped, whose $PfPR_{2\text{-}10}$ is

110 always 0. Additionally, every grid is required to have at least two-year complete records.

111 Without enough records, the grid is also removed. Accordingly, 65,705 grids are kept. The

112 coverage of our study is over 40 million km^2. Our analysis includes temperature, NDVI,

113 population density, and GDP per capita according to statistical tests and data availability.

114 Previous studies indicate that climate change affects vector-borne diseases [18-20], so

115 temperature-related variables and NDVI are considered. Moreover, population density and

116 prevention also play a critical role in malaria transmission [1,21]. Although detailed

117 prevention investments are unavailable, we replace them with GDP per capita.

118

Methods

119

120 *Spatially Stationary Model*

121   At first, we assume that relationships between $PfPR_{2\text{-}10}$ and other selected

122 independent variables are the same in each grid. For example, the marginal effects of a 1%

123 increase in the NDVI in Africa and South America are the same, even though they are

124 thousands of kilometers away. Without considering the spatial contexts, those relationships

125 do not vary spatially [22,23]. To estimate the links among variables, we apply the panel

126   regression method. Because our data set is a panel data set, the hypotheses on the time-

127   fixed effects of each grid should be carefully tested. Three typical panel regression models

128   based on different time-fixed effect hypotheses, including fixed effects model (FEM),

129   random effects model (REM), and pooled ordinary least square (POLS), are widely used.

130   We perform several statistical tests to select the reasonable spatially stationary model.

131   Firstly, we execute the F test for individual effects [24,25]. The significant result indicates that

132   the time-fixed effects exist since the null hypothesis is that no time-fixed effects are needed

133   [25]. Hence, the POLS is rejected because this model assumes that no time-fixed effects exist.

134   Then, we implement the Hausman test to detect that either the REM or the FEM is

135   reasonable [26]. The null hypothesis of the Hausman test that the preferred model is the REM

136   is rejected due to the significant test result. Therefore, the FEM is preferred in the spatially

137   stationary analysis.

138   Here, the FEM is illustrated as follows:

$$PfPR_{it} = \boldsymbol{\beta} \boldsymbol{X}'_{it} + \alpha_i + \mu_{it} \tag{1}$$

139   where $PfPR_{it}$ denotes the annual PfPR$_{2\text{-}10}$ (ranging from 0 to 1) in grid i in t year, $\boldsymbol{X}_{it}$

140   denotes a matrix of independent variables, including annual average temperature (℃),

141   annual average temperature square (℃^{2}), the standard deviation of monthly average

142   temperature in t year (℃), NDVI, population density (Cap/km^2), and GDP per capita

143   (USD/Cap) in measurement i in t year, α_i denotes the time-fixed effects, μ_{it} denotes an

144   idiosyncratic error, and $\boldsymbol{\beta}$ is a vector of parameters to be estimated.

145

Spatially Non-stationary Model

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

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

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

168 where $MSPE(b)$ is the mean square prediction error when the bandwidth is b, m is the data

169 size, y_j is the dependent variable of the jth record, $\widehat{y_j(b)}$ is the predicted value of the jth

170 record when the bandwidth is b, and p is the number of the parameters in the analysis. In

171 the GWR, the basic version of the GWPR, in a way, previous studies generally assume that

172 **Equation 2** is a U-shape function without statistical or mathematical evidence [22,23,32]. This

173 assumption helps to reduce calculating time but also brings enormous risks to the analysis.

174   We apply the step increment selection without assuming that **Equation 2** is a U-

175 shape function. The bandwidth selection extent is from 0.25 to 20 arc degrees, and the

176 examined bandwidth increases 0.25 arc degree every time because the spatial resolution of

177 the grid data is 0.25 arc degree. The optimal bandwidth is 4.25 arc degrees (**Figure S2**).

178 The mean square prediction error reaches the minimum when the bandwidth equals 0.75,

179 but we reject it. According to the grid data set spatial resolution, every sub-dataset at most

180 has 25 grids, but some of them might only have one grid. If the sub-datasets have only one

181 grid, they do not fit FEM's statistical requirements. Moreover, some grids would not be

182 spatially linked with other grids, which leads to a not robust GWPR result. Therefore, the

183 optimal bandwidth, 4.25 arc degrees, is calibrated.

184   The GWPR is performed after the optimal bandwidth has been calibrated. The

185 GWPR model is written as follows:

$$PfPR_{it} = \boldsymbol{\beta}_i \boldsymbol{X}'_{it} + \alpha_i + \epsilon_{it} \qquad (3)$$

186 where $\boldsymbol{\beta}_i$ denotes a vector of parameters in the regression with the sub-dataset taking the

187 grid i as the center. There are 65,705 sub-datasets, so the number of regressions on sub-

188   datasets is also 65,705, and every estimation is based on the sub-dataset's spatial weight

189   vector. For this reason, the parameters, $\boldsymbol{\beta}_i$, vary spatially. The spatial weight vector is

190   calculated as follows:

$$\boldsymbol{W}_i = \begin{cases} [1 - (\frac{\boldsymbol{d}}{b})^2]^2, d_k \in \boldsymbol{d} \; if \; d_k \leq \; b \\ \quad 0, \quad \quad d_k \notin \boldsymbol{d} \; if \; d_k > b \end{cases} \tag{4}$$

191   where $\boldsymbol{W}_i$ denotes a vector of spatial weights between the grid i and its neighbors, $\boldsymbol{d}$

192   represents a vector of distances between the grid i and its neighbors, b is the calibrated

193   optimal bandwidth, and d_k is the distance between the grid i and the grid k. Then, the

194   GWPR estimates the parameters, based on the spatial weight vectors. Because the time-

195   fixed effects term (α_i) in is unknown in **Equation 3** [24], we transform that equation as

196   follows:

$$PfPR_{it} - \overline{PfPR_\iota} = \boldsymbol{\beta}_i(\boldsymbol{X}_{it} - \overline{\boldsymbol{X}_\iota})' + (\epsilon_{it} - \overline{\epsilon_\iota}) \tag{5}$$

197   where $\overline{PfPR_\iota}$ denotes the mean of the PfPR$_{2\text{-}10}$ of the grid i, $\overline{\boldsymbol{X}_\iota}$ denotes a vector of means

198   of independent variables of the grid i, and $\overline{\epsilon_\iota}$ denotes the mean of the error of the grid i. To

199   simplify **Equation 5**, we define the matrix of transformed independent variables $\boldsymbol{x}_{it}$ and

200   the transformed dependent variable $pfpr_{it}$ as follows:

$$\boldsymbol{x}_{it} = \boldsymbol{X}_{it} - \overline{\boldsymbol{X}_\iota} \tag{6}$$

$$pfpr_{it} = PfPR_{it} - \overline{PfPR_\iota} \tag{7}$$

$$\sigma_{it} = \epsilon_{it} - \overline{\epsilon_\iota} \tag{8}$$

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

$$pfpr_{it} = \boldsymbol{\beta}_i \boldsymbol{x}'_{it} + \sigma_{it} \tag{9}$$

202   The parameter estimation could be displayed as follows:

$$\beta_i = [x_{it}^T W_i x_{it}]^{-1} x_{it}^T W_i pfpr_{it} \tag{10}$$

203

Statistical Indicators and 10-fold Cross Validation

205   Because we need to make the prediction based on the results of IPCC stimulations,

206   we must confirm the model's accuracy. Several statistical indicators, including R^2, root

207   mean square error (RMSE), mean absolute error (MAE), correlation coefficient (r) between

208   observed and predicted values, and regression coefficients between observed and predicted

209   values (intercept α and slope β), are applied to depict the accuracy of the model. R^2 is a

210   critical statistical indicator describing the goodness of fit, which is expressed as follows:

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

211   where n represents the observation number in the whole data set, $OPfPR_k$ represents the

212   kth record of the observed PfPR$_{2\text{-}10}$, $PPfPR_k$ represents the kth record of the predicted

213   PfPR$_{2\text{-}10}$, and $\overline{OPfPR}$ represents the mean of the observed PfPR$_{2\text{-}10}$. The RMSE is sensitive

214   to both systematic error and random error, expressed as follows:

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

215   The MAE is computed as follows:

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

216   In the analysis, the MAE is expected to be lower. Additionally, the regression coefficients

217   between observed and predicted values are estimated as follows:

$$OPfPR_k = \alpha + \beta PPfPR_k + \delta_k \tag{14}$$

218   where α is the intercept of the regression, the ideal value of α is 0, β is the slope, its ideal

219   value is 1, and δ_k is a random error term.

220   To check the reliability of the GWPR model, we perform the 10-fold cross

221   validation. In the 10-fold cross validation, the whole data set is randomly divided into ten

222   subsets. In every individual fold cross validation, nine subsets are used to train the GWPR

223   model, while the left one subset is employed to test the accuracy and ability of prediction.

224   All statistical indicators, involving R^2, RMSE, MAE, r, α, and β, are reported based on the

225   training and testing results.

226

227   ***Effects of Temperature Change on PfPR$_{2\text{-}10}$***

228   The effects of temperature change on PfPR$_{2\text{-}10}$ cannot directly be estimated because

229   other necessary variables, such as NDVI, population, and GDP per capita, cannot be

230   precisely predicted. The errors from the other variable predictions would drastically affect

231   the estimation of PfPR$_{2\text{-}10}$. However, if we assume that other variables are the same in the

232   various scenarios, we could extract the difference among the impacts of temperature

233   change in different periods. The difference between the two temperature scenarios is

234   calculated as follows:

$$\Delta PfPR_{it} = S2PfPR_{it} - S1PfPR_{it} \tag{15}$$

235   where $\Delta PfPR_{it}$ denotes the difference between the estimation of PfPR$_{2\text{-}10}$ with the

236   different temperature scenarios in the grid i during the period t, $S1PfPR_{it}$ denotes the

237   estimation of PfPR$_{2\text{-}10}$ based on the temperature scenario $S1$ in the grid i during the period

238   t, and $S2PfPR_{it}$ denotes the estimation of PfPR$_{2\text{-}10}$ based on the temperature scenario $S2$

239   in the grid i during the period t. We hope to obtain the impacts of the higher greenhouse

240   gas emission on PfPR$_{2\text{-}10}$, so the SSP1-2.6 scenario is always set as the temperature

241   scenario $S1$ and the temperature scenario $S2$ is one of the other three scenarios, including

242   SSP2-4.5, SSP3-7.0, and SSP5-8.5.

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

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

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

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

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

244   where $\boldsymbol{\beta}_{1i}$, $\boldsymbol{\beta}_{2i}$, $\boldsymbol{\beta}_{3i}$, $\boldsymbol{\beta}_{4i}$, $\boldsymbol{\beta}_{5i}$, and $\boldsymbol{\beta}_{6i}$ are parts of the estimated GWPR parameters $\boldsymbol{\beta}_i$

245   from **Equation 10**, $\boldsymbol{TeS1}_{it}$ denotes the stimulated annual average temperature under

246   SSP1-2.6 scenario in the grid i during the period t, $\boldsymbol{TSDS1}_{it}$ denotes the standard

247   deviation of stimulated monthly average temperature under SSP1-2.6 scenario in the grid

248   i during the period t, $\boldsymbol{TeS2}_{it}$ denotes the stimulated annual average temperature under

249   another scenario in the grid i during the period t, $\boldsymbol{TSDS2}_{it}$ denotes the standard deviation

250   of stimulated monthly average temperature under another scenario in the grid i during the

251   period t, and $\boldsymbol{NDVI}_{it}$, $\boldsymbol{POP}_{it}$, and $\boldsymbol{GDP}_{it}$ denote the NDVI, population density, and GDP

252   per capita in the grid i during the period t, respectively.

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

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

$$+ \boldsymbol{\beta}_{3i}(\boldsymbol{TSDS2}_{it} - \boldsymbol{TSDS1}_{it})'$$

254

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^2s of the training and testing data sets are over 94%, which hints at the model's stability. Besides, the RMSEs, MEAs, αs, βs, and *r*s are close to their ideal values, further proposing the high reliability of the estimated impacts of temperature change on PfPR$_{2\text{-}10}$.

Figure S3-S5 illustrate the difference between the estimation of PfPR$_{2\text{-}10}$ with the different temperature scenarios during 2041 – 2060, 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 PfPR$_{2\text{-}10}$ 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 of temperature change between SSP2-4.5 and SSP1-2.6 is tiny, the impacts of climate change are not potent, but they appear 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 PfPR$_{2\text{-}10}$ range from -1.68% to 2.66%,

14

278   and the grid number of positive impacts outstrips the number of grids with negative effects.

279   **Figure S8**'s distribution is similar to **Figure S7**'s. Moreover, we compare the geographical

280   distributions of average $PfPR_{2-10}$ and temperature change's impacts. Notably, CD, CF, CM,

281   BF, and NG are the worst hit by malaria in the past 20 years, compared with other countries.

282   Worse, they might face the most harmful impacts of climate change on the malaria

283   infection rates during 2021 - 2040.

284   With further increase in temperature during 2041-2060, the distribution and

285   intention of negative impacts shift. **Figure S9-S11** show the difference between SSP2.4-5

286   and SSP1-2.6 ($\Delta PfPR_{2041-2060}^{SSP245-SSP126}$), the difference between SSP3.7-0 and SSP1-2.6

287   ($\Delta PfPR_{2041-2060}^{SSP370-SSP126}$), and the difference between SSP5.8-5 and SSP1-2.6

288   ($\Delta PfPR_{2041-2060}^{SSP585-SSP126}$) during 2041 – 2060. $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$ becomes minus in most

289   areas of CD, CF, CM, BF, and NG. The positive impacts of the increase in temperature on

290    $PfPR_{2-10}$ also rise, according to $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$ and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$ because in

291   SSP3.7-0 and SSP5.8-5, the predicted temperatures are higher than the scenario SSP2.4-5.

292   The negative impacts mainly appear in Gabon (GA), Angola (AO), and Zambia (ZM),

293   which are at the boundaries of high malaria-affected regions. The possible result of climate

294   change, especially in Sub-Saharan Africa, is that the increase in temperature would spur

295   the high-risk area to spread to the higher latitude regions and coastlines. Moreover, the

296   original high-risk areas are no longer suitable for human living, owing to extreme

297   temperatures. **Figure   S12-S14**   demonstrate   the   statistical   distribution   of

298    $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$, $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$, and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$, respectively.

299   Compared with differences among the scenarios during 2021- 2040, the values of

300    $\Delta PfPR_{2041-2060}^{SSP245-SSP126}$, $\Delta PfPR_{2041-2060}^{SSP370-SSP126}$, and $\Delta PfPR_{2041-2060}^{SSP585-SSP126}$ are absolutely

301   larger. Moreover, the impacts of the increase in temperature are positive in most grids of

302   Africa, Asia, and South America.

303   By the end of the 21st century, the four scenarios end in different temperature

304   predictions. **Figure S15-S17** demonstrate the difference between SSP2.4-5 and SSP1-2.6

305   ($\Delta PfPR_{2081-2100}^{SSP245-SSP126}$), the difference between SSP3.7-0 and SSP1-2.6

306   ($\Delta PfPR_{2081-2100}^{SSP370-SSP126}$), and the difference between SSP5.8-5 and SSP1-2.6

307   ($\Delta PfPR_{2081-2100}^{SSP585-SSP126}$) during 2081 – 2100. Because the difference of temperatures of the

308   four scenarios becomes more significant, the impacts of temperature change on PfPR$_{2-10}$

309   are intensified. The spatial distributions of $\Delta PfPR_{2081-2100}^{SSP245-SSP126}$, $\Delta PfPR_{2081-2100}^{SSP370-SSP126}$,

310   and $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$ are similar to the situation of $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$, but the

311   values of the difference become larger during 2080 – 2100. **Figure S18-S20** propose the

312   statistical distribution of $\Delta PfPR_{2081-2100}^{SSP245-SSP126}$, $\Delta PfPR_{2081-2100}^{SSP370-SSP126}$, and

313    $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$. In fact, $\Delta PfPR_{2081-2100}^{SSP585-SSP126}$ ranges -77.20% to 101.54%.

314   **Figure S21** illustrates the shapes of the relationship between annual average

315   temperature and PfPR$_{2-10}$ in each grid. The basic shapes of the relationship are two types,

316   specifically, U-shape and inverted U-shape. The coefficient of temperature square term

317   decides the basic shape of the relationship. Additionally, together with the coefficient of

318   temperature term, the symmetry lines of the relationship of each grid can be pinned down.

319   In **Figure S21**, we set the interested temperature range from 0°C to 50°C. If the symmetry

320   lines of the relationships of some grids do not belong to the interested range, we consider

321   those relationships are either monotonically decreasing or monotonically increasing in the

322   range, mathematically. Therefore, the other four shapes are derived from the basic types,

323   including increase based on inverted U-shape (Increase Inverted U), decrease based on U-

324 shape (Decrease U), decrease based on inverted U-shape (Decrease Inverted U), and

325 increase based on U-shape (Increase U). Previous studies generally put forward that the

326 relationship in national or regional levels is inverted U-shape, and the temperature impacts

327 would peak at a specific temperature. However, the relationship at the grid level is

328 obviously more complicated and various. Hence, we must treat the impacts of climate

329 change carefully. Although we accept the assumption that the relationship between

330 temperature and malaria infection rates is non-linear, we did not force to set the malaria-

331 transmission peaking temperature to 25℃ or other values, as the previous studies

332 mentioned [4,6]. In our results, most grids' transmission peaking temperatures are not 25℃.

333 We employ the annual average temperature to probe the relationship, whereas previous

334 studies use laboratory environments and real-time or shorter-term temperatures. As we

335 have mentioned before, even though the annual average temperature is 25℃, how many

336 days and hours are at that temperature remains unknown. For this reason, in many grids,

337 the transmission peaking temperatures are not 25℃. The shapes in each grid are even

338 different, though either U-shape or inverted U-shape. Furthermore, real-world

339 environments are significantly more complex than laboratory environments. The ignored

340 factors by the laboratory environments are also related to the malaria-transmission peaking

341 temperature. These ignored variables lead to the difference in the transmission peaking

342 temperatures between our result and previous studies.

343

Figure:

344

345
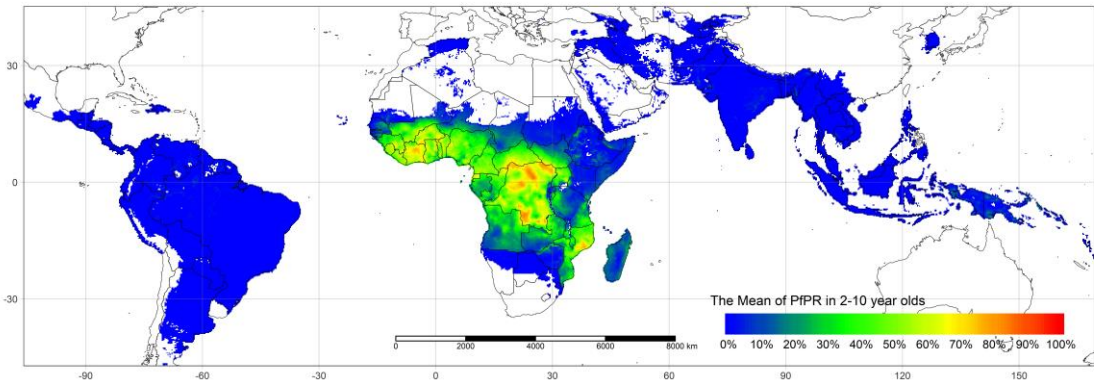
346   **Figure S1: The mean of PfPR$_{2\text{-}10}$ from 2000 to 2019**

347   (Note: if the PfPR$_{2\text{-}10}$ in a grid is always 0 from 2000 to 2019, the grids are blank)
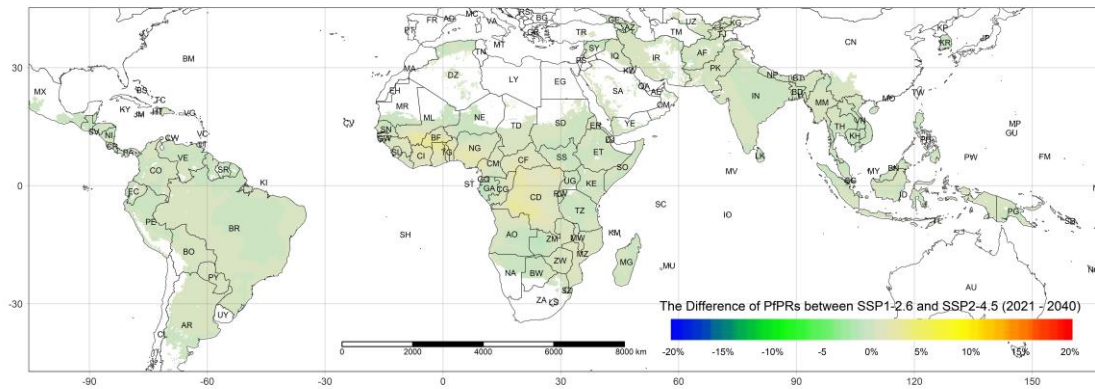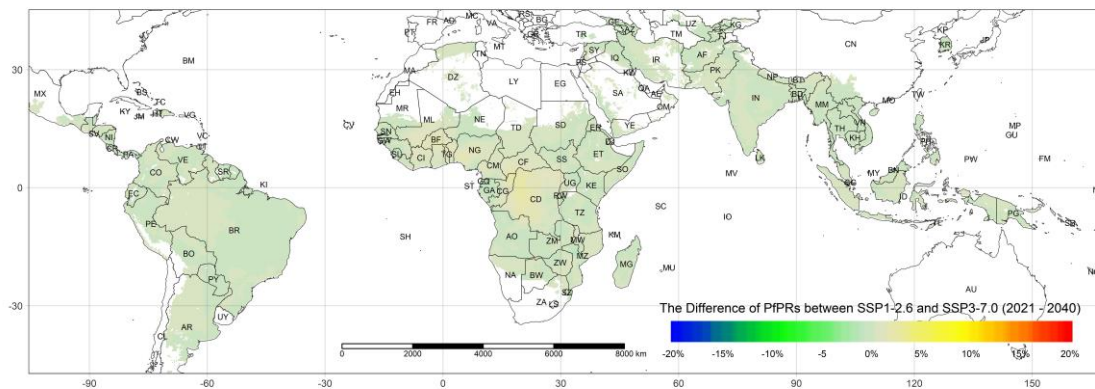
348

349
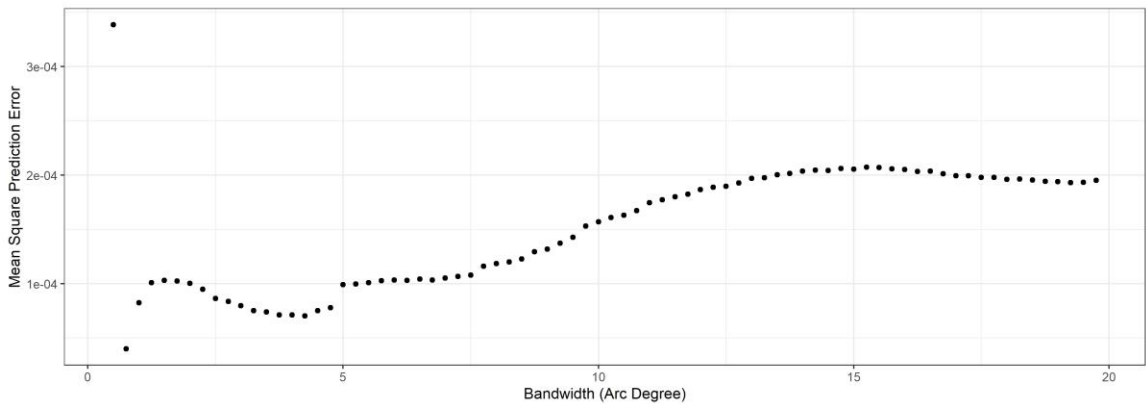
350   **Figure S2: Step Increment Bandwidth Selection**

351

Figure S3: The Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP2-4.5 (2041 - 2060)

353

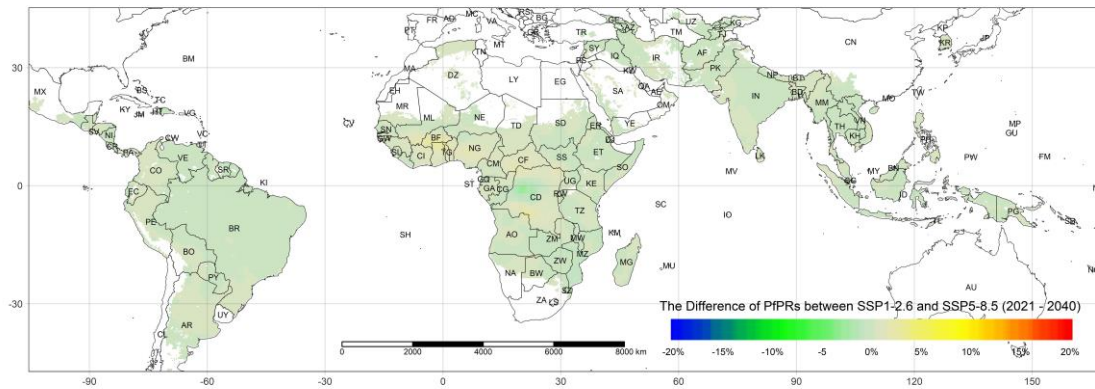**Figure S4: The Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP3-7.0 (2041 - 2060)**

355

356   **Figure S5: The Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP5-8.5 (2041 - 2060)**

357

358

359
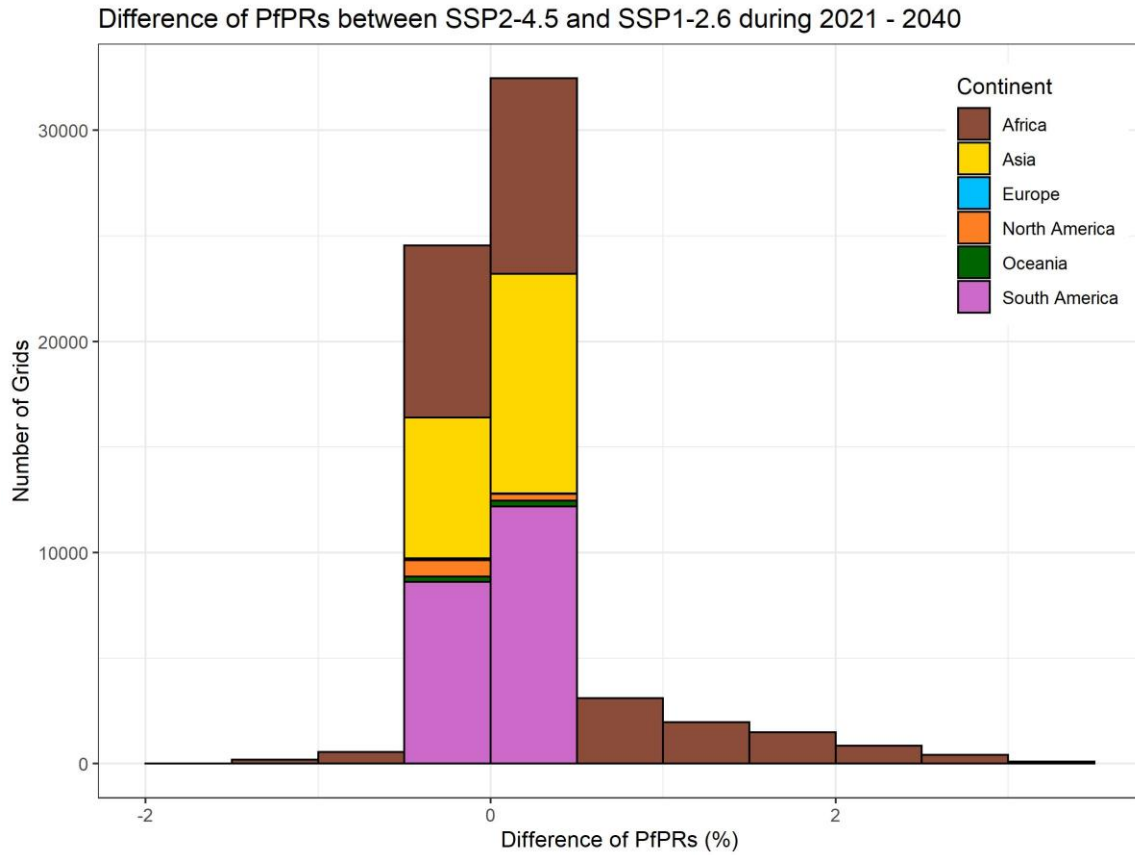
Figure S6: The Distribution of the Difference of $PfPR_{2-10}$ between SSP1-2.6 and SSP2-4.5 (2041 - 2060)

Difference of PfPRs between SSP3-7.0 and SSP1-2.6 during 2021 - 2040
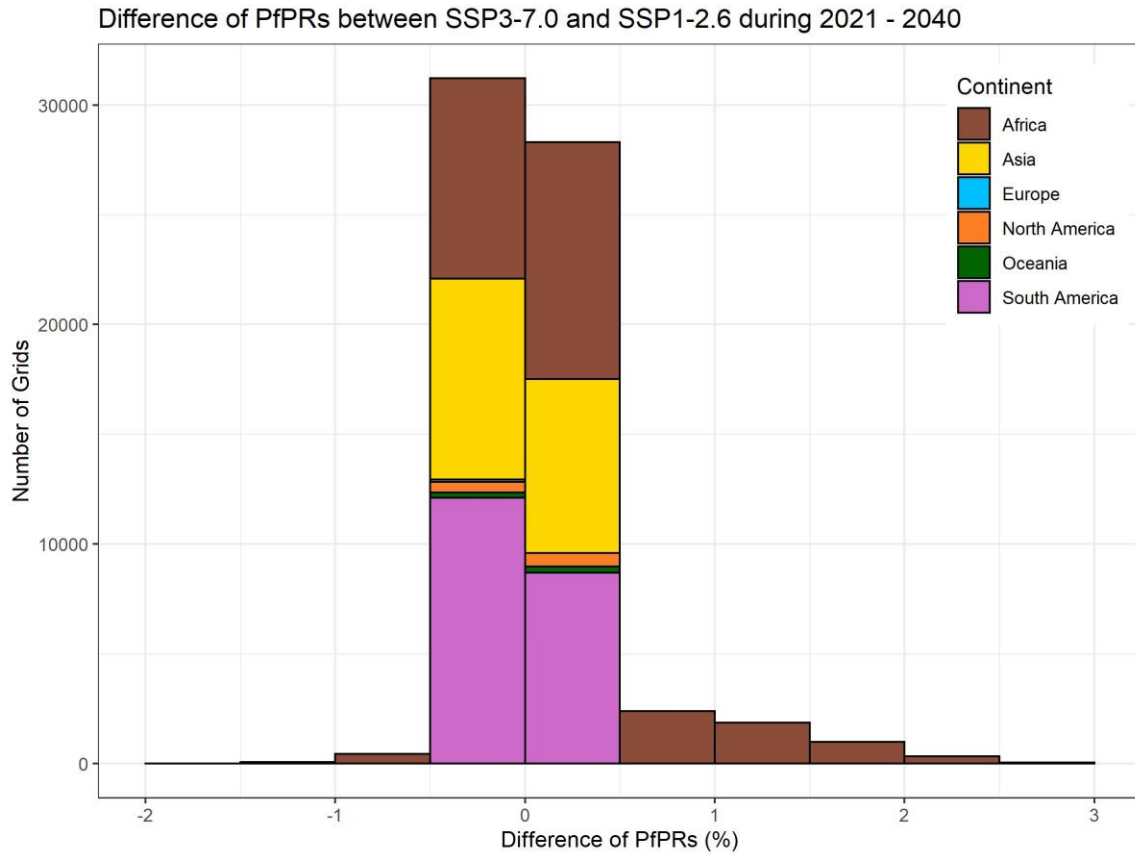
Figure S7: The Distribution of the Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and SSP3-7.0 (2041 - 2060)

Difference of PfPRs between SSP5-8.5 and SSP1-2.6 during 2021 - 2040

366
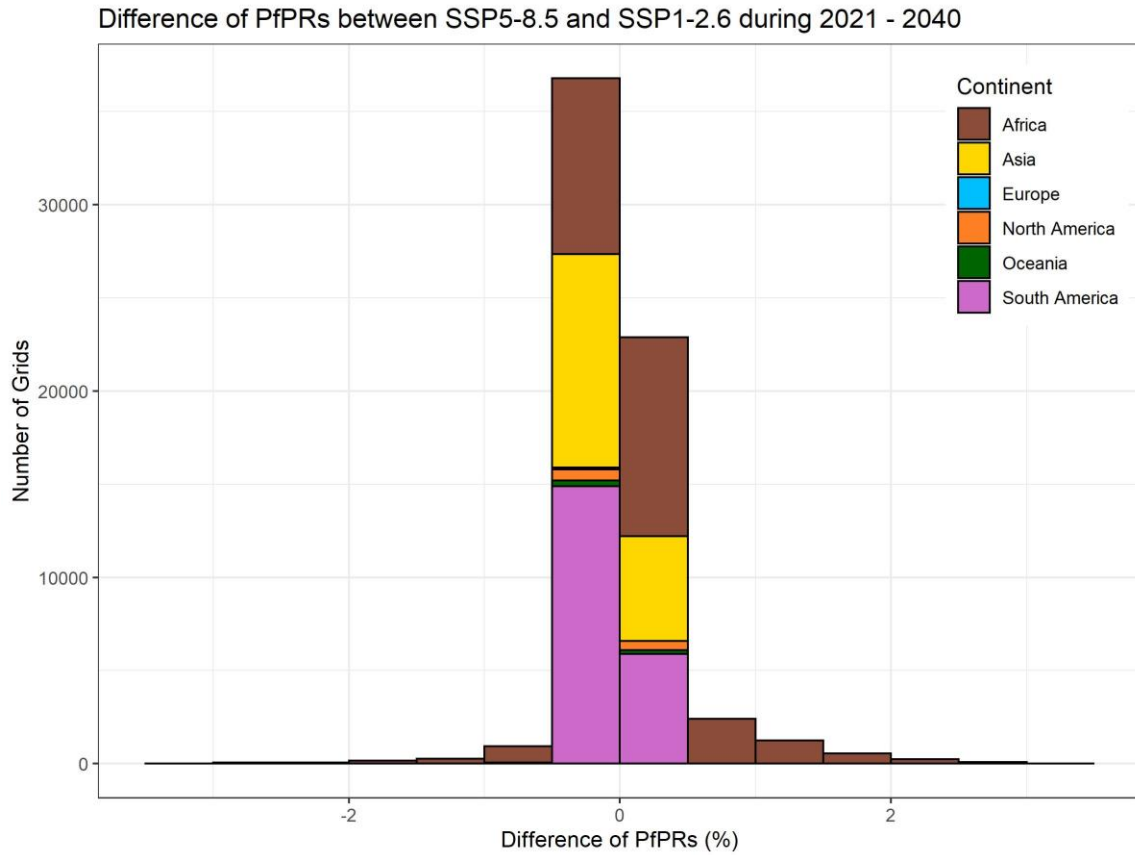
Figure S8: The Distribution of the Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and

SSP5-8.5 (2041 - 2060)

367

368

369

370

371
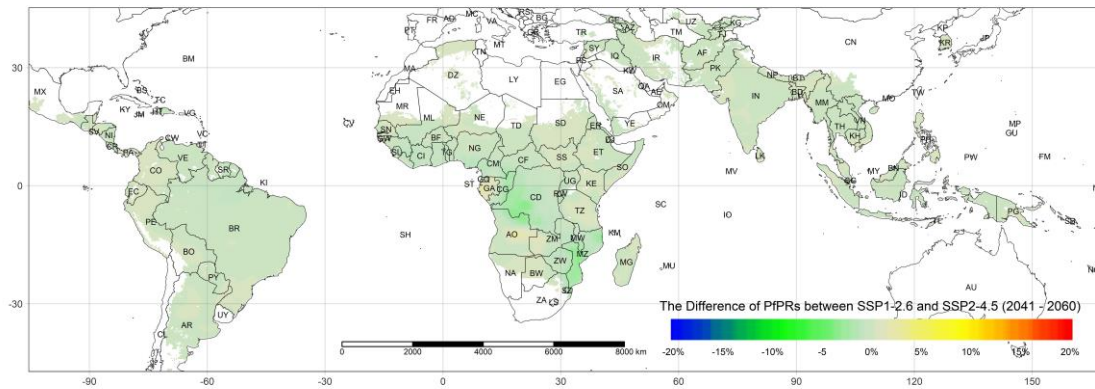
Figure S9: The Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and SSP2-4.5 (2061 - 2080)

373
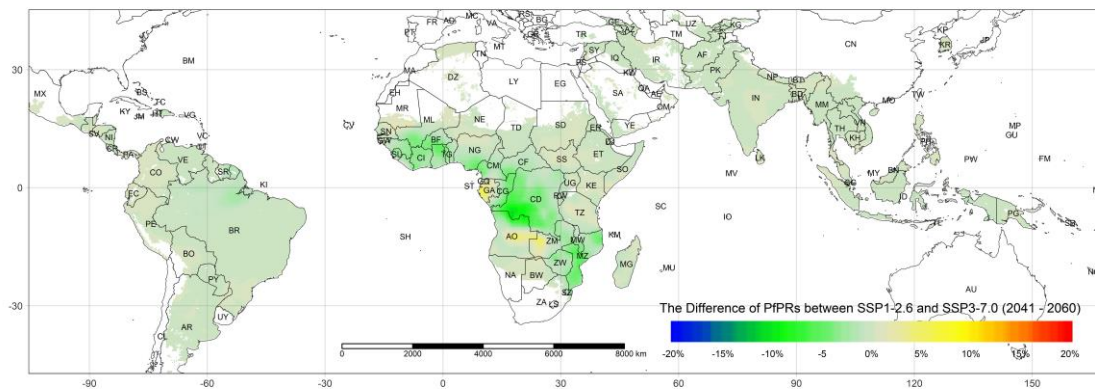
Figure S10: The Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and SSP3-7.0 (2061 - 2080)

The Difference of PfPRs between SSP1-2.6 and SSP5-8.5 (2041 - 2060)

375
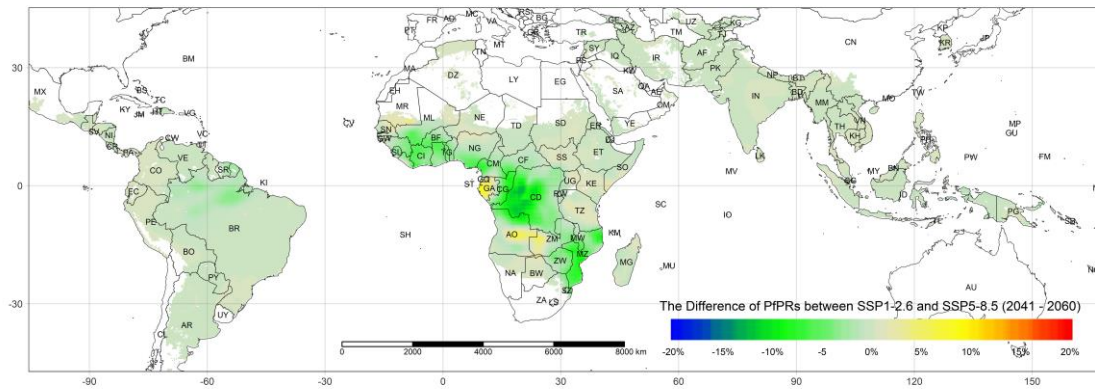
Figure S11: The Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and SSP5-8.5 (2061 - 2080)

377

378

379

Difference of PfPRs between SSP2-4.5 and SSP1-2.6 during 2041 - 2060

380
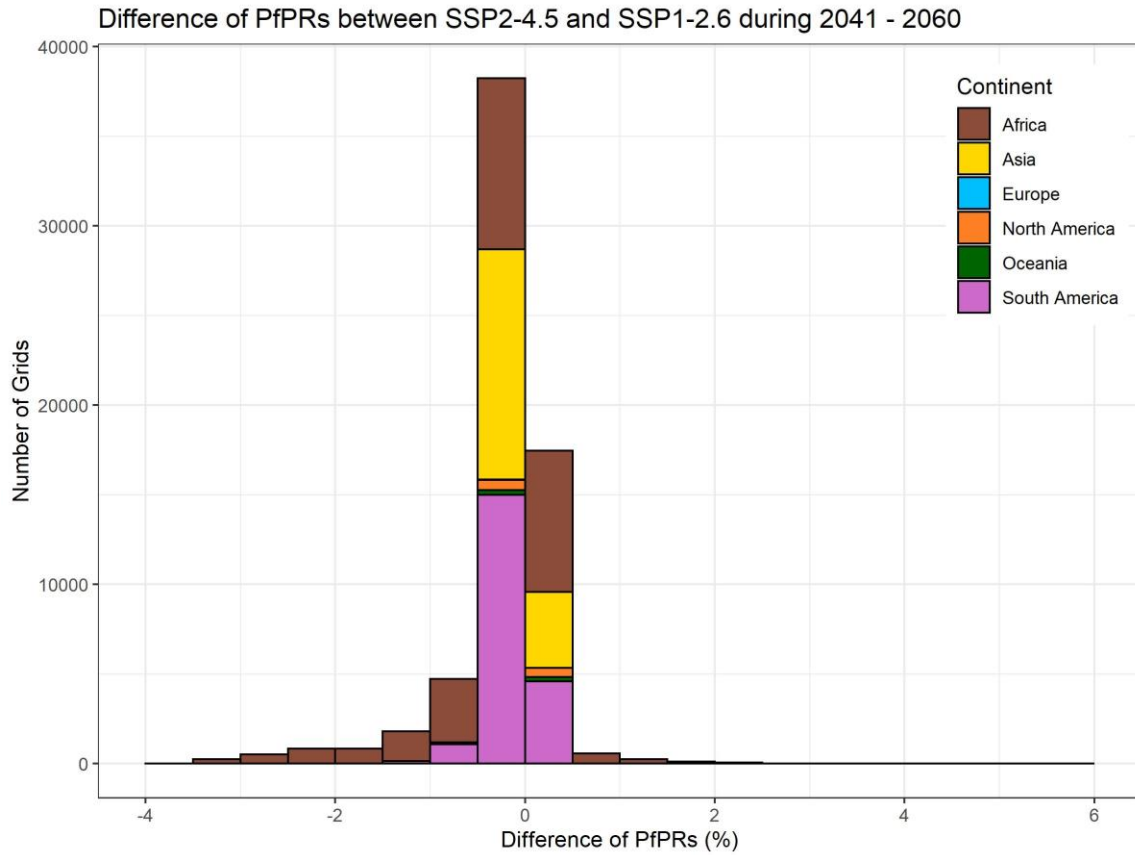
Figure S12: The Distribution of the Difference of PfPR$_{2-10}$ between SSP1-2.6 and

SSP2-4.5 (2061 - 2080)

Difference of PfPRs between SSP3-7.0 and SSP1-2.6 during 2041 - 2060

383
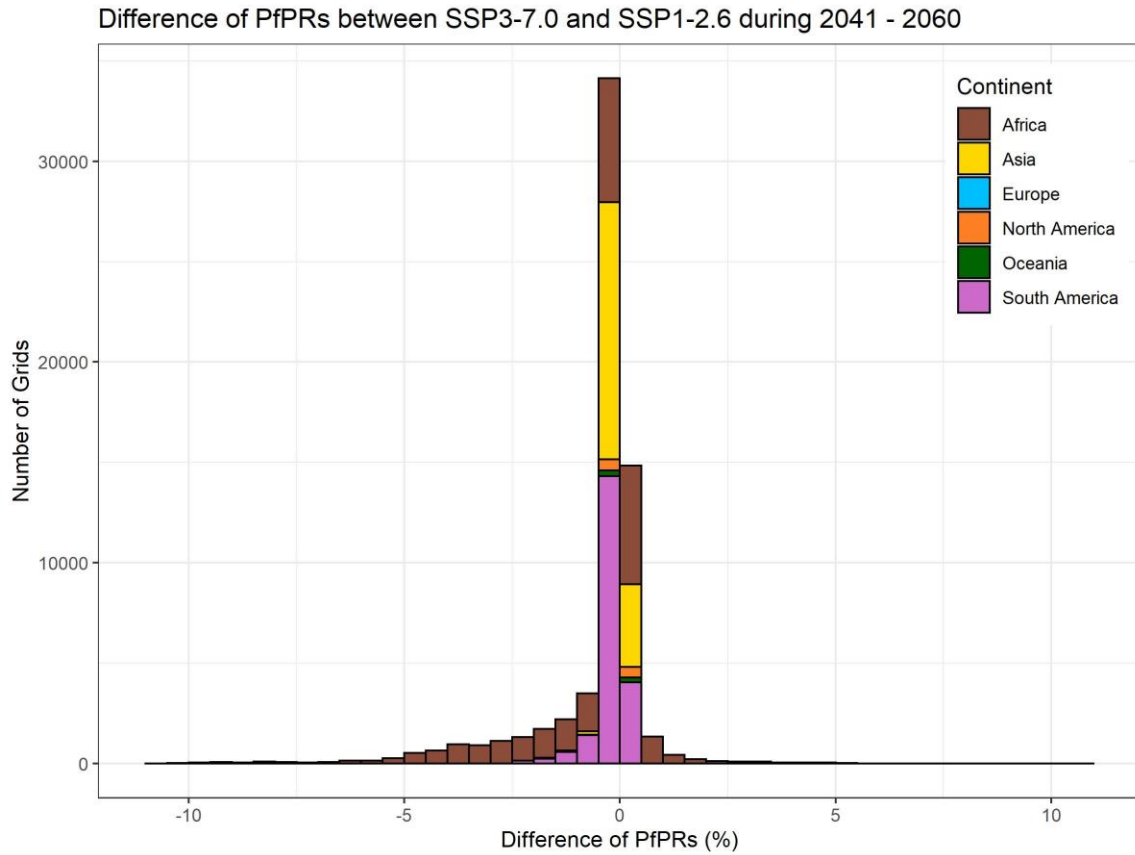
Figure S13: The Distribution of the Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and

384

SSP3-7.0 (2061 - 2080)

385

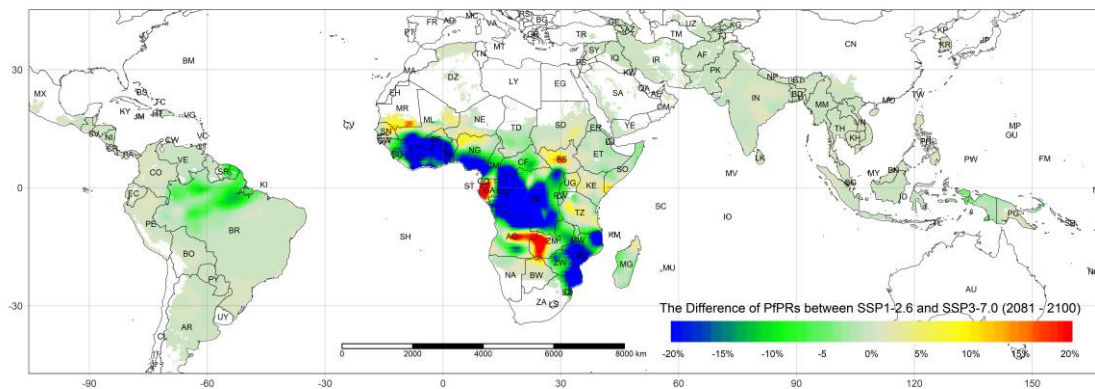Difference of PfPRs between SSP5-8.5 and SSP1-2.6 during 2041 - 2060

386

Figure S14: The Distribution of the Difference of PfPR$_{2-10}$ between SSP1-2.6 and

388 **SSP5-8.5 (2061 - 2080)**

389

390

The Difference of PfPRs between SSP1-2.6 and SSP2-4.5 (2081 - 2100)

391

Figure S15: The Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP2-4.5 (2081 - 2100)

The Difference of PfPRs between SSP1-2.6 and SSP3-7.0 (2081 - 2100)

393

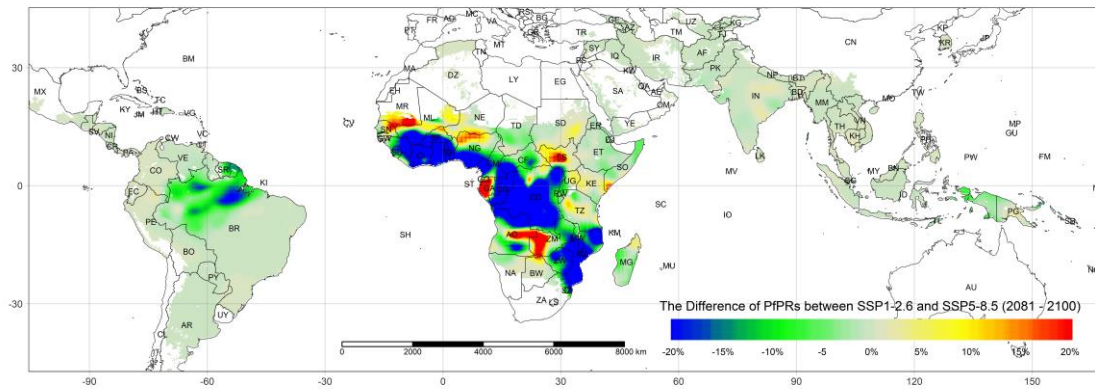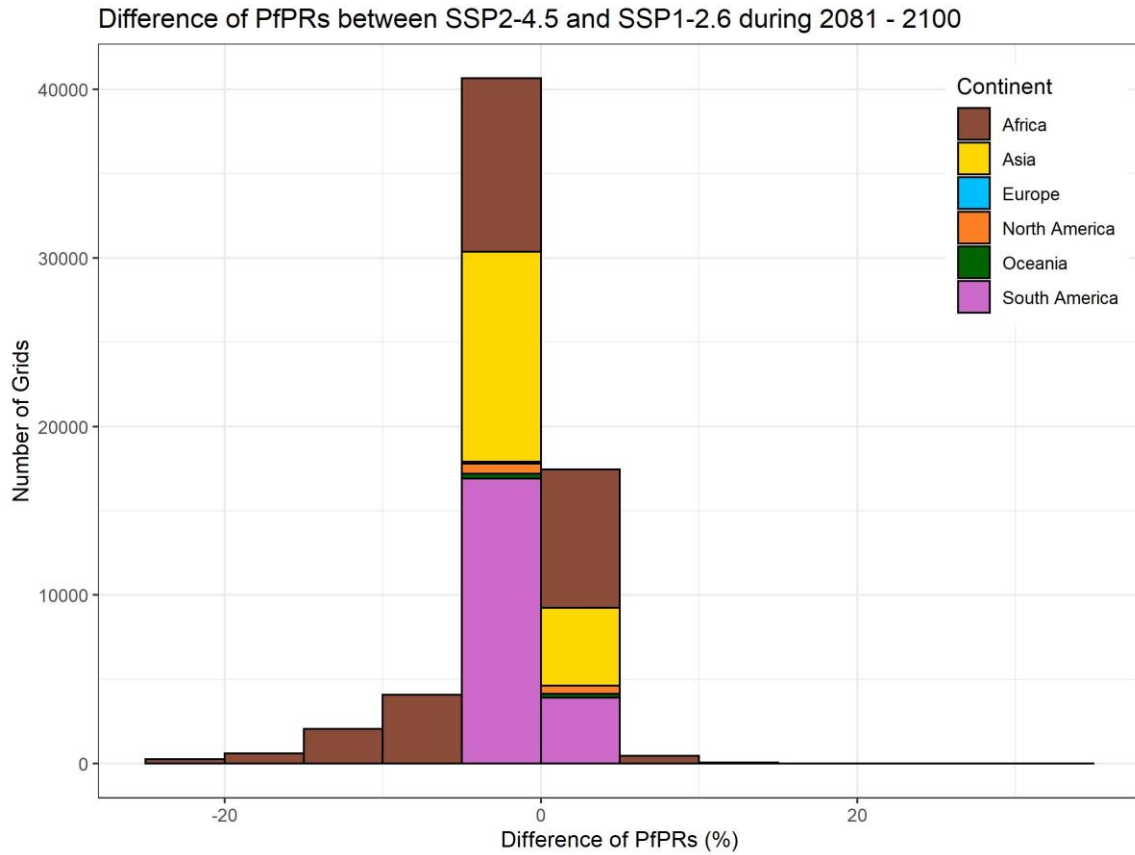**Figure S16: The Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP3-7.0 (2081 - 2100)**

29

395

Figure S17: The Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and SSP5-8.5 (2081 - 2100)

397

398

399

Difference of PfPRs between SSP2-4.5 and SSP1-2.6 during 2081 - 2100

Figure S18: The Distribution of the Difference of PfPR$_{2-10}$ between SSP1-2.6 and SSP2-4.5 (2081 - 2100)

400

401

402

Difference of PfPRs between SSP3-7.0 and SSP1-2.6 during 2081 - 2100

403

Figure S19: The Distribution of the Difference of PfPR$_{2\text{-}10}$ between SSP1-2.6 and

SSP3-7.0 (2081 - 2100)

Figure S20: The Distribution of the Difference of PfPR$_{2-10}$ between SSP1-2.6 and
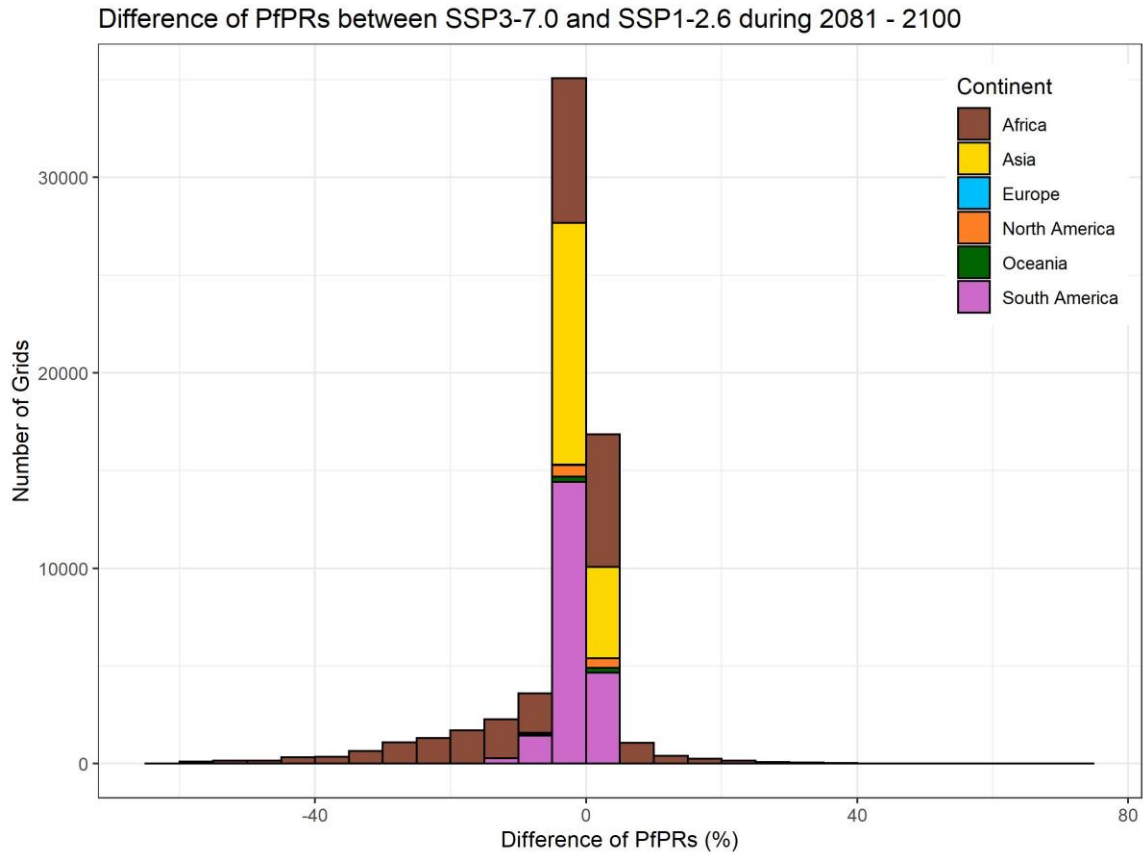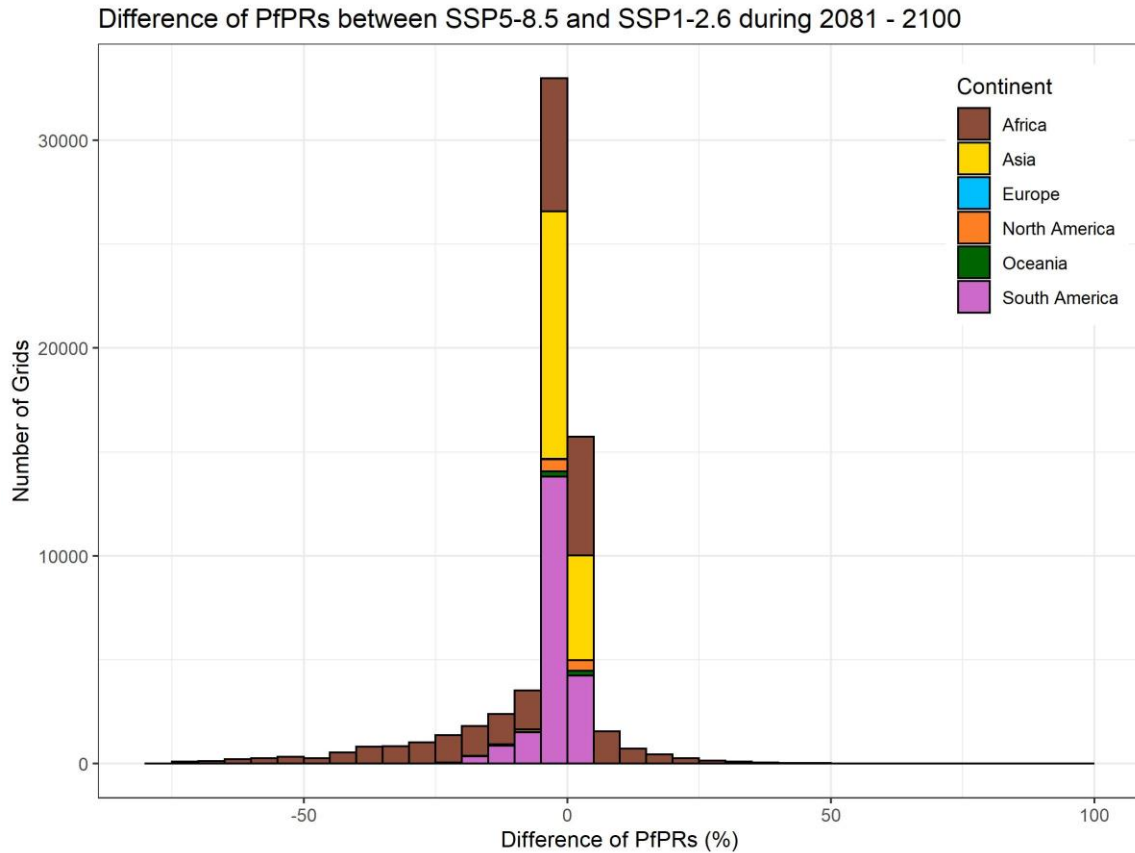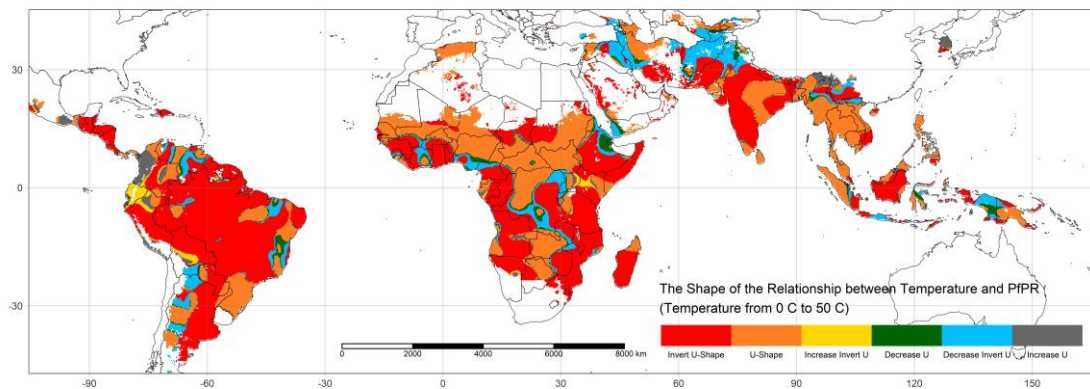
SSP5-8.5 (2081 - 2100)

Figure S21: The Shape of Relationship between Temperature and PfPR$_{2-10}$

Table:

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.

Reference:

1.   Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* 2015; **526**(7572): 207-11.

2.   Rodell M, Houser PR, Jambor U, et al. The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society* 2004; **85**(3): 381-94.

3.   Mordecai EA, Ryan SJ, Caldwell JM, Shah MM, Labeaud AD. Climate change could shift disease burden from malaria to arboviruses in Africa. *The Lancet Planetary Health* 2020; **4**(9): e416-e23.

4.   Mordecai EA, Paaijmans KP, Johnson LR, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters* 2013; **16**(1): 22-30.

5.   Mordecai EA, Caldwell JM, Grossman MK, et al. Thermal biology of mosquito-borne disease. *Ecology Letters* 2019; **22**(10): 1690-708.

6.   Johnson LR, Ben-Horin T, Lafferty KD, et al. Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach. *Ecology* 2015; **96**(1): 203-13.

7.   Lourenço PM, Sousa CA, Seixas J, Lopes P, Novo MT, Almeida APG. Anopheles atroparvus density modeling using MODIS NDVI in a former malarious area in Portugal. *Journal of Vector Ecology* 2011; **36**(2): 279-91.

8.   Reiner RC, Geary M, Atkinson PM, Smith DL, Gething PW. Seasonality of Plasmodium falciparum transmission: a systematic review. *Malaria Journal* 2015; **14**(1).

9.   Brugueras S, Fernández-Martínez B, Martínez-De La Puente J, et al. Environmental drivers, climate change and emergent diseases transmitted by mosquitoes and their vectors in southern Europe: A systematic review. *Environ Res* 2020; **191**: 110038.

10.   Amadi JA, Olago DO, Ong'Amo GO, et al. Sensitivity of vegetation to climate variability and its implications for malaria risk in Baringo, Kenya. *PLoS One* 2018; **13**(7): e0199357.

11.   Midekisa A, Senay G, Henebry GM, Semuniguse P, Wimberly MC. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malaria Journal* 2012; **11**(1): 165.

12.   Didan K, Munoz AB, Solano R, Huete A, University of Arizona: Vegetation Index Phenology Lab. MODIS vegetation index user's guide (MOD13 series). 2015.

13.   Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data* 2015; **2**(1): 150045.

14.   Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS One* 2012; **7**(2): e31743.

15.   World Health Organization. World malaria report 2021, 2021.

16.   IPCC. Climate change 2022: impacts, adaptation and vulnerability, 2022.

17.   IPCC. Climate change 2014 synthesis report. Geneva, Switzerland: IPCC, 2014.

18.   Rogers DJ, Randolph SE. Climate Change and Vector-Borne Diseases. In: Hay SI, Graham A, Rogers DJ, eds. Advances in Parasitology: Academic Press; 2006: 345-81.

455   19.   Altizer S, Ostfeld Richard S, Johnson Pieter TJ, Kutz S, Harvell CD. Climate
456 Change and Infectious Diseases: From Evidence to a Predictive Framework. *Science* 2013;
457 **341**(6145): 514-9.
458   20.   Buchwald AG, Hayden MH, Dadzie SK, Paull SH, Carlton EJ. Aedes-borne disease
459 outbreaks in West Africa: A call for enhanced surveillance. *Acta Tropica* 2020; **209**:
460 105468.
461   21.   Nyaruaba R, Mwaliko C, Mwau M, Mousa S, Wei H. Arboviruses in the East
462 African Community partner states: a review of medically important mosquito-borne
463 Arboviruses. *Pathogens and Global Health* 2019; **113**(5): 209-28.
464   22.   Brunsdon C, Fotheringham S, Charlton M. Geographically Weighted Regression.
465 *Journal of the Royal Statistical Society: Series D (The Statistician)* 1998; **47**(3): 431-43.
466   23.   Fotheringham A, Brunsdon C, Charlton M. Geographically Weighted Regression:
467 The Analysis of Spatially Varying Relationships: John Wiley & Sons; 2002.
468   24.   Croissant Y, Millo G. Panel Data Econometrics in R: The plm Package. *J Stat Softw*
469 2008; **27**(2).
470   25.   Breusch TS, Pagan AR. The Lagrange Multiplier Test and its Applications to
471 Model Specification in Econometrics. *The Review of Economic Studies* 1980; **47**(1): 239.
472   26.   Kang S. A note on the equivalence of specification tests in the two-factor
473 multivariate variance components model. *Journal of Econometrics* 1985; **28**(2): 193-203.
474   27.   Giesen C, Roche J, Redondo-Bravo L, et al. The impact of climate change on
475 mosquito-borne diseases in Africa. *Pathogens and Global Health* 2020; **114**(6): 287-301.
476   28.   Kulkarni MA, Duguay C, Ost K. Charting the evidence for climate change impacts
477 on the global spread of malaria and dengue and adaptive responses: a scoping review of
478 reviews. *Globalization and Health* 2022; **18**(1).
479   29.   Brunsdon C, Fotheringham AS, Charlton ME. Geographically Weighted
480 Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 2010;
481 **28**(4): 281-98.
482   30.   Fotheringham AS, Oshan TM. Geographically weighted regression and
483 multicollinearity: dispelling the myth. *Journal of Geographical Systems* 2016; **18**(4): 303-
484 29.
485   31.   Beenstock M, Felsenstein D. The econometric analysis of non-stationary spatial
486 panel data: Springer; 2019.
487   32.   Gollini I, Lu B, Charlton M, Brunsdon C, Harris P. GWmodel: An R package for
488 exploring spatial heterogeneity using geographically weighted models. *J Stat Softw* 2015;
489 **63**(17).

490