Environ Health Perspect

DOI: 10.1289/EHP5899

Supplemental Material

Current and Projected Distributions of *Aedes aegypti* and *Ae. albopictus* in Canada and the U.S.

Salah Uddin Khan, Nicholas H. Ogden, Aamir A. Fazil, Philippe H. Gachon,
Guillaume U. Dueymes, Amy L. Greer, and Victoria Ng

Table of Contents

Figure S3. The relationship between the primary risk factors and the ecological niche of *Aedes albopictus* resulted from the model utilizing currently observed (2001-2016) climatic conditions. Results from the two primary contributing factors mean minimum daily temperature, number of days above 10°C are presented here. These results are from 120 bootstraps. The grey lines represent the predicted line for each bootstrap, the dashed black line represents the upper and lower boundary of the 95% confidence intervals, and the red continuous line represented the average of the bootstraps.

Table S3. Determining the presence/absence threshold cut-offs for the current (2001-2016 observed climatic data) and simulated climatic models (2006-2016). We considered the multiple recursive partitioning pathways (RCPs) leading to moderate (RCP 4.5) and high (RCP 8.5) greenhouse gas emission scenarios. The table shows the presence/absence threshold cut-offs determined by True Skill Statistics (*TSS*), threshold where maximum Kappa and the maximum percentages of the observations correctly classified (*MaxPCC*).

Table S4. Accuracy measures for the boosted regression tree models developed using simulated climatic data for the time period 2006-2016. The area under the receiver operating characteristic curve (AUC) are presented as accuracy measures.

Figure S4. The most influential variables identified in the *Aedes aegypti and Ae. albopictus* ecological niche models developed using simulated climatic data for the time period 2006-2016 and a single boosted regression trees model run. The figure represents the relationship between the most influential factors and *Aedes* ecological niche for each model and the black line represents a single iteration. The most influential covariates varied from one Regional Climatic Model (RCM) to another. This includes mean minimum temperature ("tmin", mean January temperature ("*tjan*"), mean daily temperature ("*avgtmp*") and average number of days ≥10° C ("ncdays10").

R codes and additional information's on boosted regression trees model

ADDITIONAL REFERENCE

Additional File- Excel Document

Table S1: A list of climatic and land cover data included in the boosted regression tree models. We also extracted identical variables for the following projected regional climatic models: CanRCM4-CanESM2, CRCM5-CanESM2, CRCM5-MPI-ESM-LR, and HIRHAM5-EC-EARTH. The vegetation index variable was not included in the list of projected climatic variables as it dropped out during the model simplification process.

Variable name	Unit	Source	Reference
Mean daily temperature*	°C	MOD11C2	(Wan et al. 2015)
Minimum daily temperature	°C	MOD11C2	(Wan et al. 2015)
Maximum daily temperature	°C	MOD11C2	(Wan et al. 2015)
Number of days ≥10°C	°C	MOD11C2	(Wan et al. 2015)
Number of days ≥20°C	°C	MOD11C2	(Wan et al. 2015)
Average total monthly precipitation	mm/month	MERRA2	((GMAO) 2015)
Enhanced vegetation index	-	MOD13C2	(Didan 2015)
Urban land cover	-	GRUMPv1	(Center for International Earth Science Information Network (CIESIN) et al. 2017)

Table S2. A brief outline of the Regional Climate Models (RCMs) in reference to the Coupled Global Climate Models (CGCMs). The simulations include three Regional Climate Models (RCMs) driven by three CGCMs under two representative concentration pathways (RCP4.5 and RCP8.5), and one RCM used two different boundary conditions from two CGCMs.

	Regional Climate Model			Coupled General Circulation Model		
No. simulation	Name	Responsible institution	Reference	Name	Responsible institution	Reference
1	Canadian Regional Climate Model (CRCM5)	Université du Québec à Montréal, Canada	Martynov et al. (2013), Šeparović et al. (2013)	Second generation Canadian Earth System Model (CanESM2)	Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment and Climate Change Canada (ECCC), Canada	http://climate-modelling.canada.ca/
2	Canadian Regional Climate Model (CRCM5)	Université du Québec à Montréal, Canada	Martynov et al. (2013), Šeparović et al. (2013)	Fifth version of Max Planck Institute Earth System Model (ECHAM5/MPI-M&MPI-ESM-LR)	Max Planck Institute for Meteorology, Germany	https://www. mpimet.mpg.de/
3	Canadian Regional Climate Model (CanRCM4)	CCCma of ECCC, Canada	Scinocca et al. (2016)	CanESM2	CCCma of ECCC, Canada	http://climate-modelling.canada.ca/
4	High Resolution Limited Area Model (HIRHAM5)	Danish Meteorological Institute, Denmark & Alfred Wegener Institute Foundation for Polar and Marine Research, Germany	Bøssing Christensen et al. (2007)	European Earth System Model (EC-EARTH;)	Irish Centre for High End Computing (ICHEC), Ireland	Hazeleger et al. (2010)

Figure S1. Reported distribution of *Ae. aegypti and Ae. albopictus* mosquitoes in Canada and the United States, 2001-2016. The orange dots represent the geographic location of mosquito occurrences referred in Excel table S1 and S2.

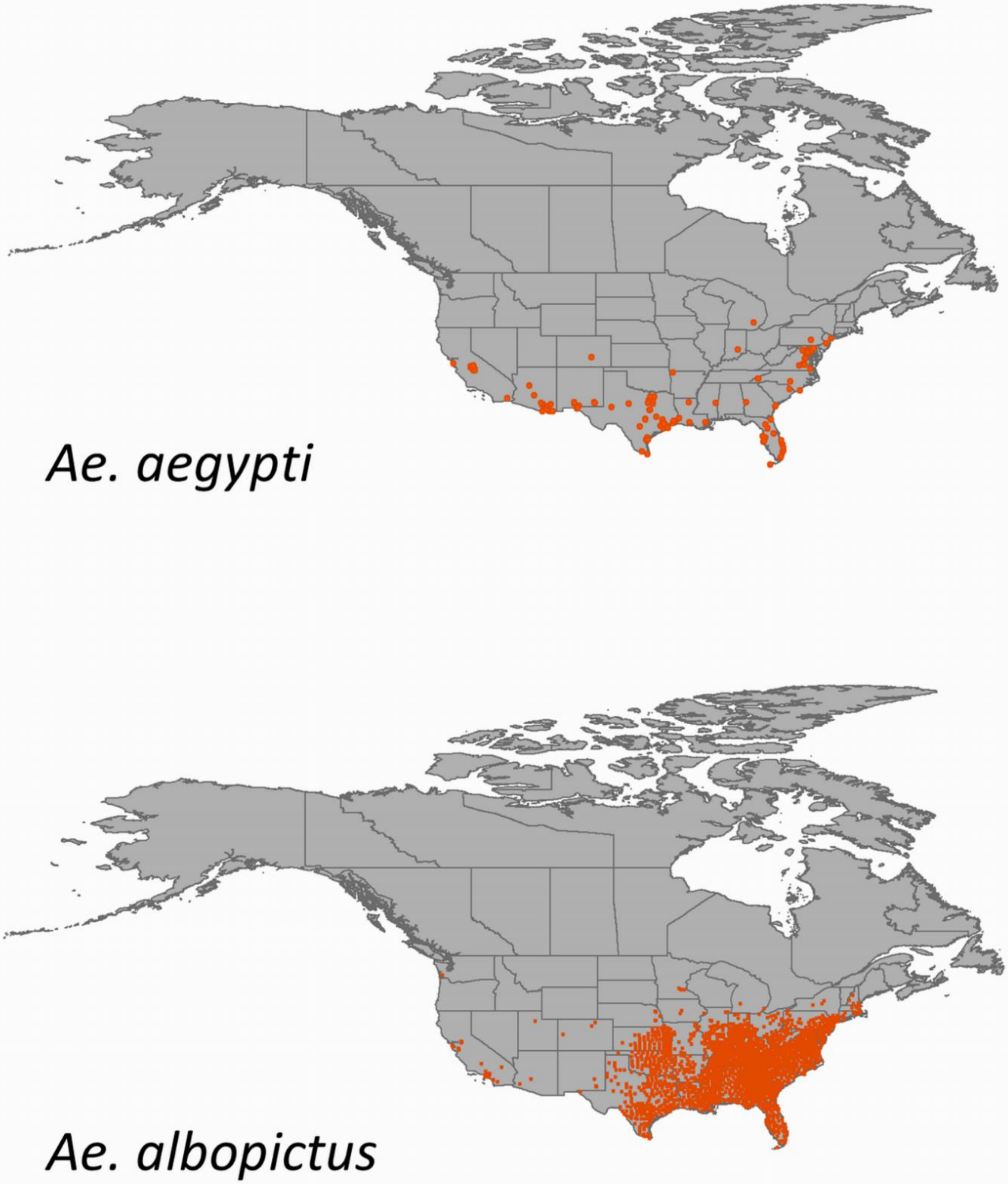


Figure S2. The relationship between the primary risk factors and the ecological niche of *Aedes aegypti* resulted from the model utilizing currently observed (2001-2016) climatic conditions. Results from the three primary contributing factors mean minimum daily temperature, mean maximum daily temperature and mean daily temperature in January are presented here. These results are from 120 bootstraps. The grey lines represent the predicted line for each bootstrap, the dashed black line represents the upper and lower boundary of the 95% confidence intervals, and the red continuous line represented the average of the bootstraps.

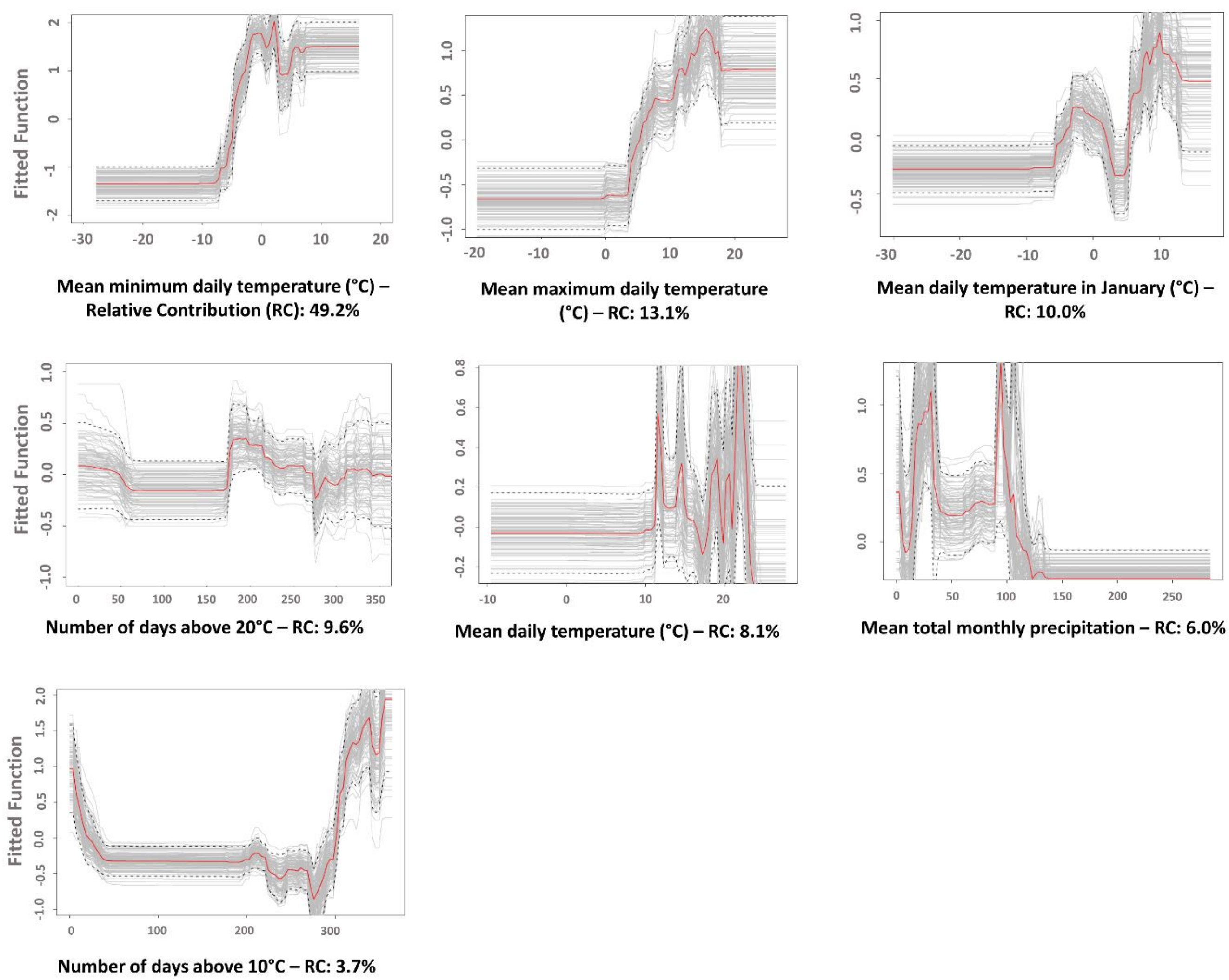


Figure S3. The relationship between the primary risk factors and the ecological niche of *Aedes albopictus* resulted from the model utilizing currently observed (2001-2016) climatic conditions. Results from the two primary contributing factors mean minimum daily temperature, number of days above 10°C are presented here. These results are from 120 bootstraps. The grey lines represent the predicted line for each bootstrap, the dashed black line represents the upper and lower boundary of the 95% confidence intervals, and the red continuous line represented the average of the bootstraps.

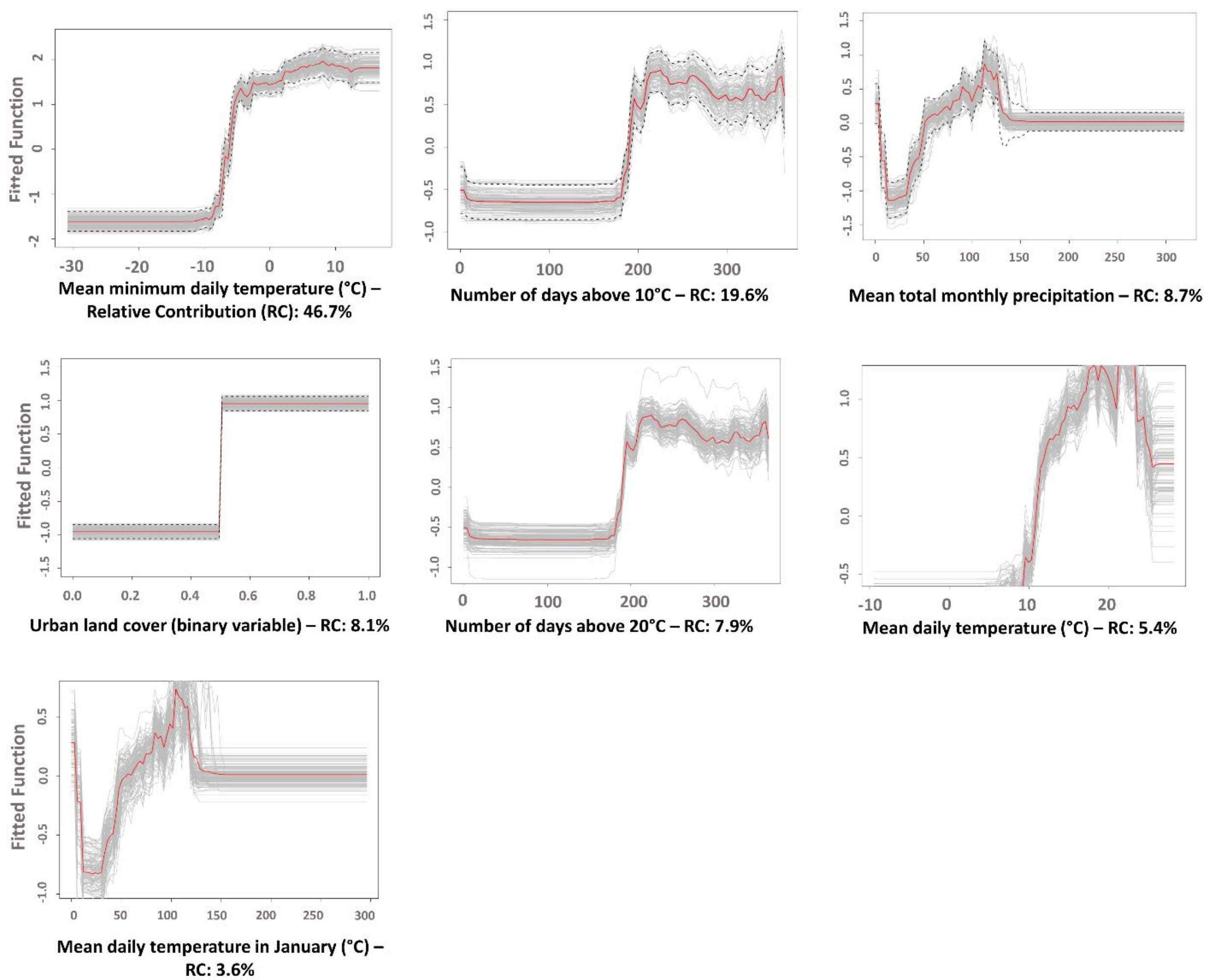


Table S3: Determining the presence/absence threshold cut-offs for the current (2001-2016 observed climatic data) and simulated climatic models (2006-2016). We considered the multiple recursive partitioning pathways (RCPs) leading to moderate (RCP 4.5) and high (RCP 8.5) greenhouse gas emission scenarios. The table shows the presence/absence threshold cut-offs determined by True Skill Statistics (*TSS*), threshold where maximum Kappa and the maximum percentages of the observations correctly classified (*MaxPCC*).

Optimal Threshold for *Aedes* Presence/Absence Niche					
Model period	**Species**	**Model RCPs**	**TSS**	**Kappa**	**MaxPCC**
Observed climate (2001-2016)					
	Ae. albopictus	N/A	0.80	0.80	0.90
	A. aegypti	N/A	0.69	0.69	0.84
Simulated climate (2006-2016)					
	Ae. albopictus	4.5*	0.65	0.65	0.65
	A. aegypti	4.5*	0.53	0.53	0.53

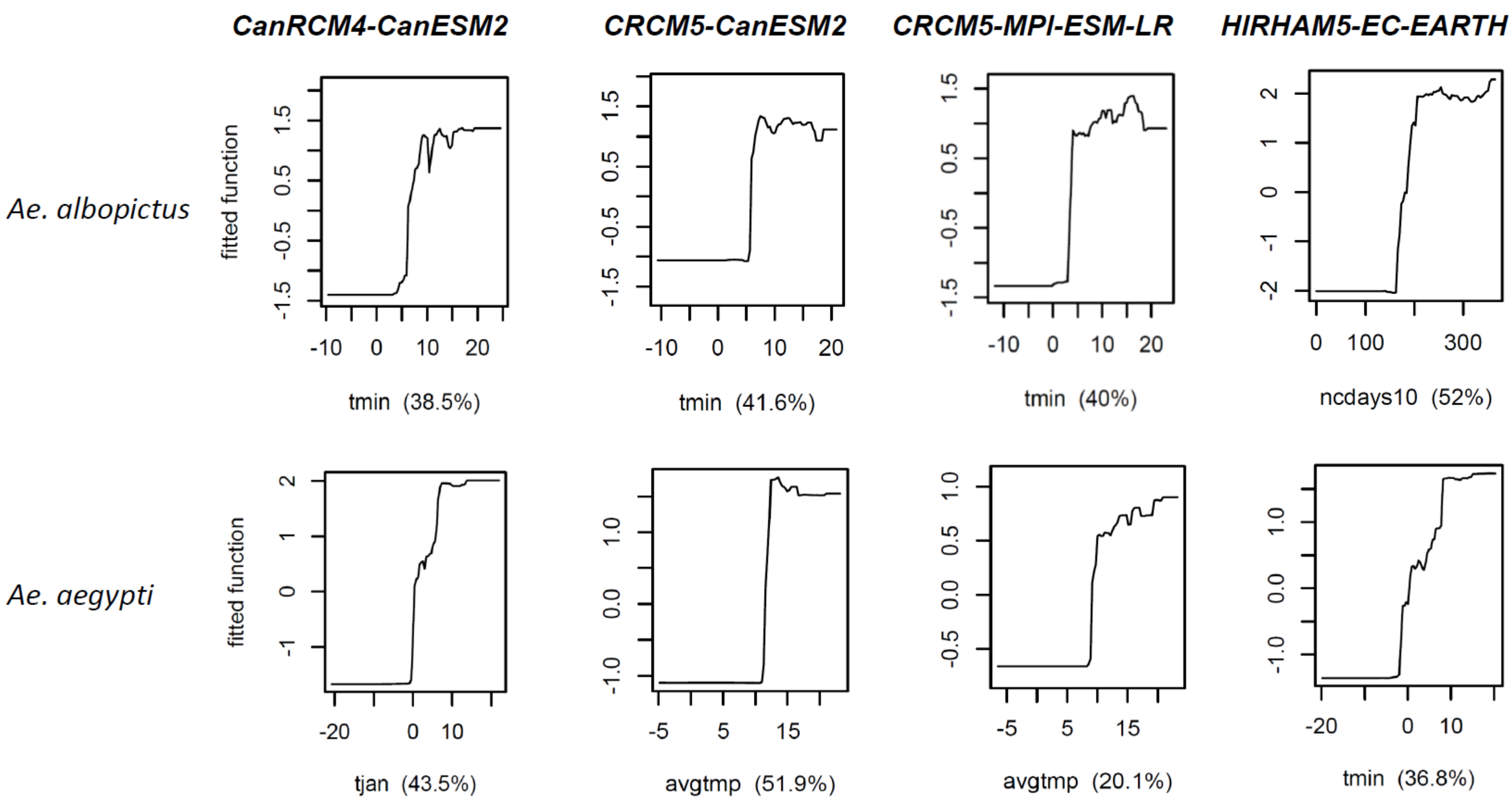
*Greenhouse gas concentration are be similar for this short time window close to the current period and both RCPs (4.5 and 8.5) will have similar outcome.

Table S4: Accuracy measures for the boosted regression tree models developed using simulated climatic data for the time period 2006-2016. The area under the receiver operating characteristic curve (AUC) are presented as accuracy measures.

Projected climatic model	**Boosted Regression Tree Model Accuracy**	
	Aedes albopictus (AUC*)	***Ae. Aegypti (AUC*)***
CanRCM4-CanESM2	0.962	0.976
CRCM5-CanESM2	0.965	0.934
CRCM5-MPI-ESM-LR	0.964	0.949
HIRHAM5-EC-EARTH	0.961	0.980

*The AUC values are rounded to three decimal digits.

Figure S4. The most influential variables identified in the *Aedes aegypti and Ae. albopictus* ecological niche models developed using simulated climatic data for the time period 2006-2016 and a single boosted regression trees model run. The figure represents the relationship between the most influential factors and *Aedes* ecological niche for each model and the black line represents a single iteration. The most influential covariates varied from one Regional Climatic Model (RCM) to another. This includes mean minimum temperature ("tmin", mean January temperature ("*tjan*"), mean daily temperature ("avgtmp") and average number of days ≥10º C ("ncdays10").



S9. R codes and additional information's on boosted regression trees model

R codes for the boosted regression trees model

```
brt.model.name<- gbm.step(data=training.data.sebset,
                          gbm.x = c(3:4, 6:9, 11), # The predictor variables in dataset
                          gbm.y = 2, # The outcome variables in dataset
                          family = "bernoulli",
                          tree.complexity = 8, #defines the complexity of individual trees
                          learning.rate = 0.01, #defines the weight applied to the individual trees
                          bag.fraction = 0.5, # the proportion of observations used in selecting variables
                          n.folds=10) # number of folds
```

Introduction of some degree of randomness into a boosted regression trees model often increases its accuracy, run-time and reduces overfitting (*Friedman 2002*). This process also introduces variance in the fitted values and predictive outcomes between each model run (*Elith et al. 2008*). In this model, stochasticity was introduced through a 'bag function' that specified the proportion of data to be randomly selected for each step.

R-codes for model simplification

```
Simplified.brt.model<- gbm.simplify(model.name, n.folds=10, n.drop="auto",
                        prev.stratify = TRUE, eval.data = NULL, plot = TRUE)
```

A detailed description of the model simplification theories and procedure are illustrated in the 'dismo' package available for R (Hijmans et al. 2017). In brief, the model simplification is done through dropping the least-informative predictors from the model, followed by re-fitting, and successively repeating the process until some stopping criteria is achieved (Elith et al. 2008). The simplifications begin with primary cross-validation (CV) and then initializes the assessment procedure for removing predictors using k-fold CV; the process progressively simplifies the model fit to each fold, and utilizes the CV error to assess the number of variables to be removed from the primary model without affective the predictive performance (Elith et al. 2008).

S10. ADDITIONAL REFERENCE

(GMAO) GMaAO. 2015. Merra-2 tavgm_2d_flx_nx: 2d,monthly mean,time-averaged,single-level,assimilation,surface flux diagnostics v5.12.4, greenbelt, md, USA, goddard earth sciences data and information services center (ges disc), accessed: 15 march 2017, 10.5067/0jrlvl8yv2y4.

Bøssing Christensen O, Drews M, Hesselbjerg Christensen J, Dethloff K, Ketelsen K, Hebestadt I, et al. 2007. The hirham regional climate model. Version 5 (beta). (Technical Report). Denmark:Danish Meteorological Institute.

Center for International Earth Science Information Network (CIESIN), Columbia University, CUNY Institute for Demographic Research (CIDR), International Food Policy Research Institute (IFPRI), The World Bank, (CIAT) CIdAT. 2017. Global rural-urban mapping project, version 1 (grumpv1): Urban extent polygons, revision 01. Palisades, ny: Nasa socioeconomic data and applications center (sedac). Http://dx.Doi.Org/10.7927/h4z31wkf. Accessed on: 27 march 2017.

Didan K. 2015. Mod13c2 modis/terra vegetation indices monthly l3 global 0.05deg cmg v006, cmg 0.05 deg monthly evi. Nasa eosdis lp daac. Doi: 10.5067/modis/mod13c2.006.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. J Anim Ecol 77:802-813.

Hazeleger W, Severijns C, Semmler T, Ştefănescu S, Yang S, Wang X, et al. 2010. Ec-earth: A seamless earth-system prediction approach in action. Bulletin of the American Meteorological Society 91:1357-1364.

Hijmans RJ, Phillips S, Leathwick J, Elith J, Hijmans MRJ. 2017. Package 'dismo'. Circles 9:1-68.

Martynov A, Laprise R, Sushama L, Winger K, Šeparović L, Dugas B. 2013. Reanalysis-driven climate simulation over cordex north america domain using the canadian regional climate model, version 5: Model performance evaluation. Climate Dynamics 41:2973-3005.

Scinocca J, Kharin V, Jiao Y, Qian M, Lazare M, Solheim L, et al. 2016. Coordinated global and regional climate modeling. Journal of Climate 29:17-35.

Šeparović L, Alexandru A, Laprise R, Martynov A, Sushama L, Winger K, et al. 2013. Present climate and climate change over north america as simulated by the fifth-generation canadian regional climate model. Climate Dynamics 41:3167-3201.

Wan Z, S. Hook, Hulley. G. 2015. Mod11c2 modis/terra land surface temperature/emissivity 8-day l3 global 0.05deg cmg v006.