Supplemental Information

Table of Contents

Results

SI-1: Model Validation

Table S1: Cross-validation results for overall, spatial, temporal (monthly and yearly), and null model performance of random forest predicting WNV presence/absence in *Cx. quinquefasciatus* pools. Overall results reflect predictive performance on randomly withheld observations. Blocked k-fold cross-validation on withheld years, months, and surveillance sites generated estimates of performance in novel temporal and spatial domains. Null model predictions assume WNV is either present (Null - All Present) or absent (Null - All Absent) in all samples.

	AUC	Sensitivity	Specificity	Accuracy	Balanced Acc.
Overall	0.88	0.82	0.80	0.80	0.81
Spatial	0.87	0.80	0.79	0.79	0.79
Monthly	0.87	0.77	0.80	0.79	0.79
Yearly	0.85	0.72	0.81	0.80	0.76
Null - All Absent	0.50	0.00	1.00	0.88	0.50
Null - All Present	0.50	1.00	0.00	0.12	0.50

SI-2: Predictor Importance and Marginal Effects of Temperature

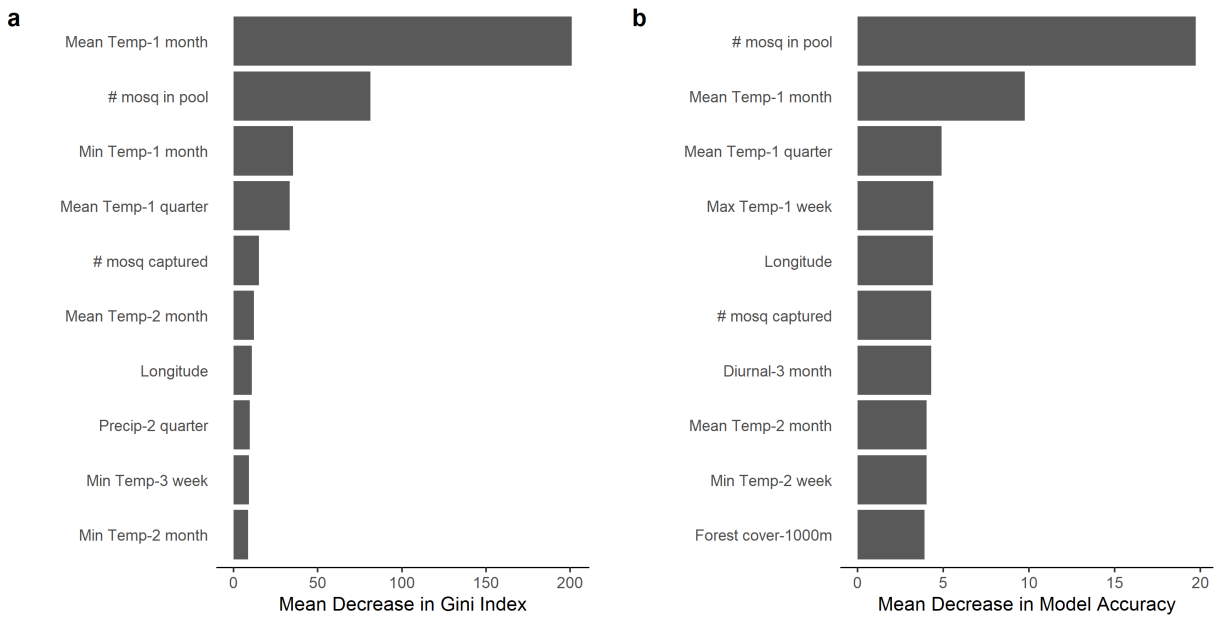

Figure S1. Random forest model variable importance of the ten best predictors according to (a) Gini index and (b) mean decrease in model accuracy. Variable names on y-axis include the temporal aggregation (week, month, quarter) and lag (1-3) or buffer radius of the predictor. Weekly, monthly and quarterly aggregations represent the average over 7, 30, and 90 days respectively.

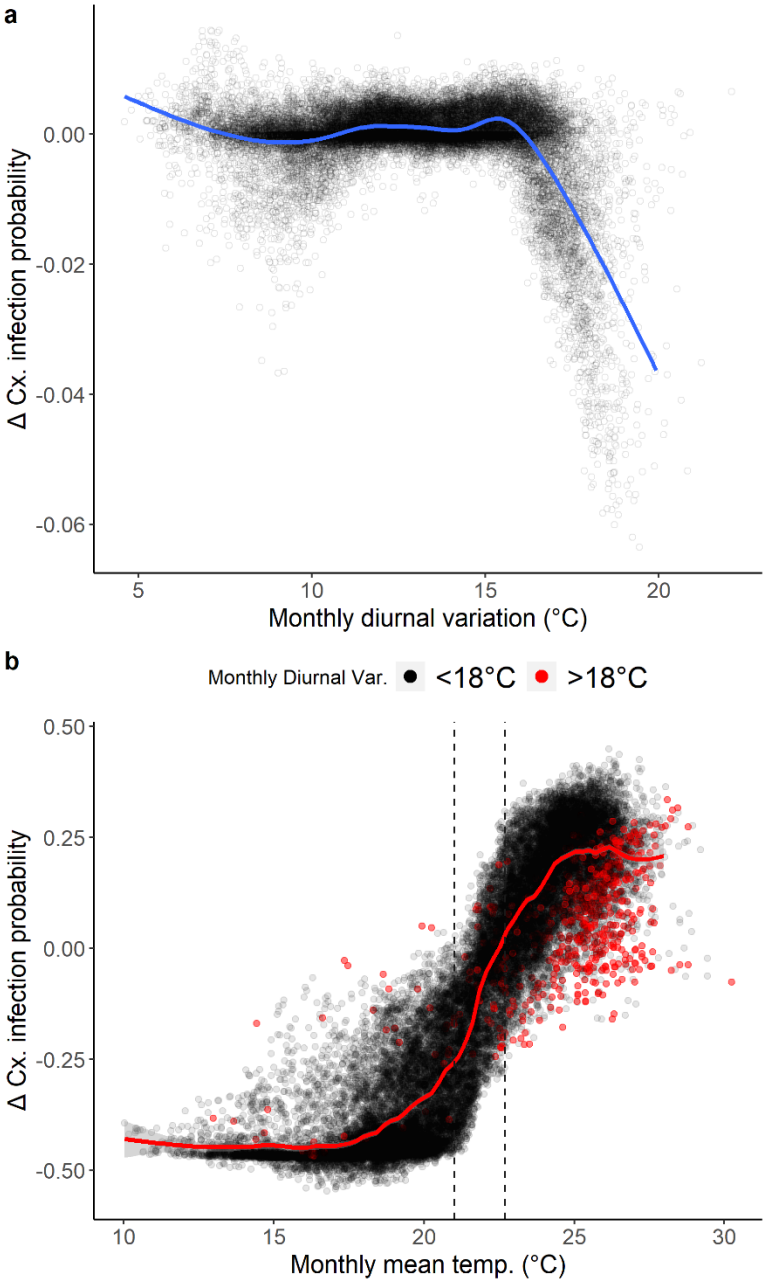

Figure S2: Marginal effects plots showing the marginal change in the predicted probability of WNV presence in *Cx. quinquefasciatus* pools (*Cx.* infection probability) associated with: (a) diurnal temperature variation (sum of effects of one-month lag and one-season lag); and (b) all temperature predictors in the random forest model (including all lags and daily, weekly, monthly, and quarterly aggregations). Observations in (b) in which the average diurnal variation in the month prior to sampling was greater than 18°C are highlighted in red. Vertical dashed lines denote the transitional range (21°C-22.7°C). Points in (a) and (b) represent all observations over the study period. The smoothed marginal effects curves were generated with a generalized additive model.

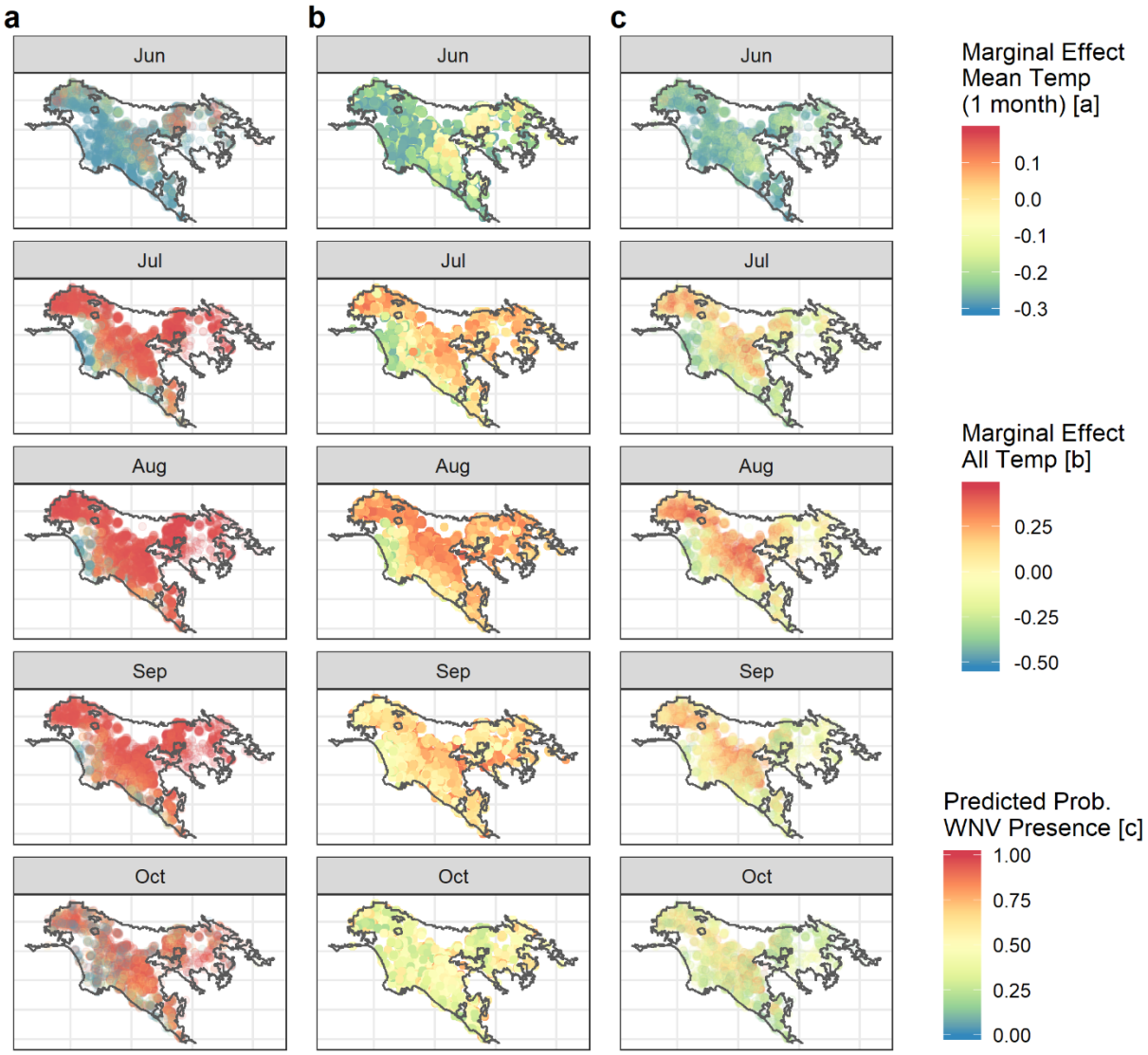

Figure S3: Geographic variability in the (a) marginal effects of mean temperature (one-month lag), (b) marginal effects of all temperature predictors including all lags and temporal aggregations, and (c) predicted probability of WNV presence in *Cx. quinquefasciatus* pools across the LA metro area. Points represent all observations at mosquito surveillance sites randomly jittered to limit point overlap.

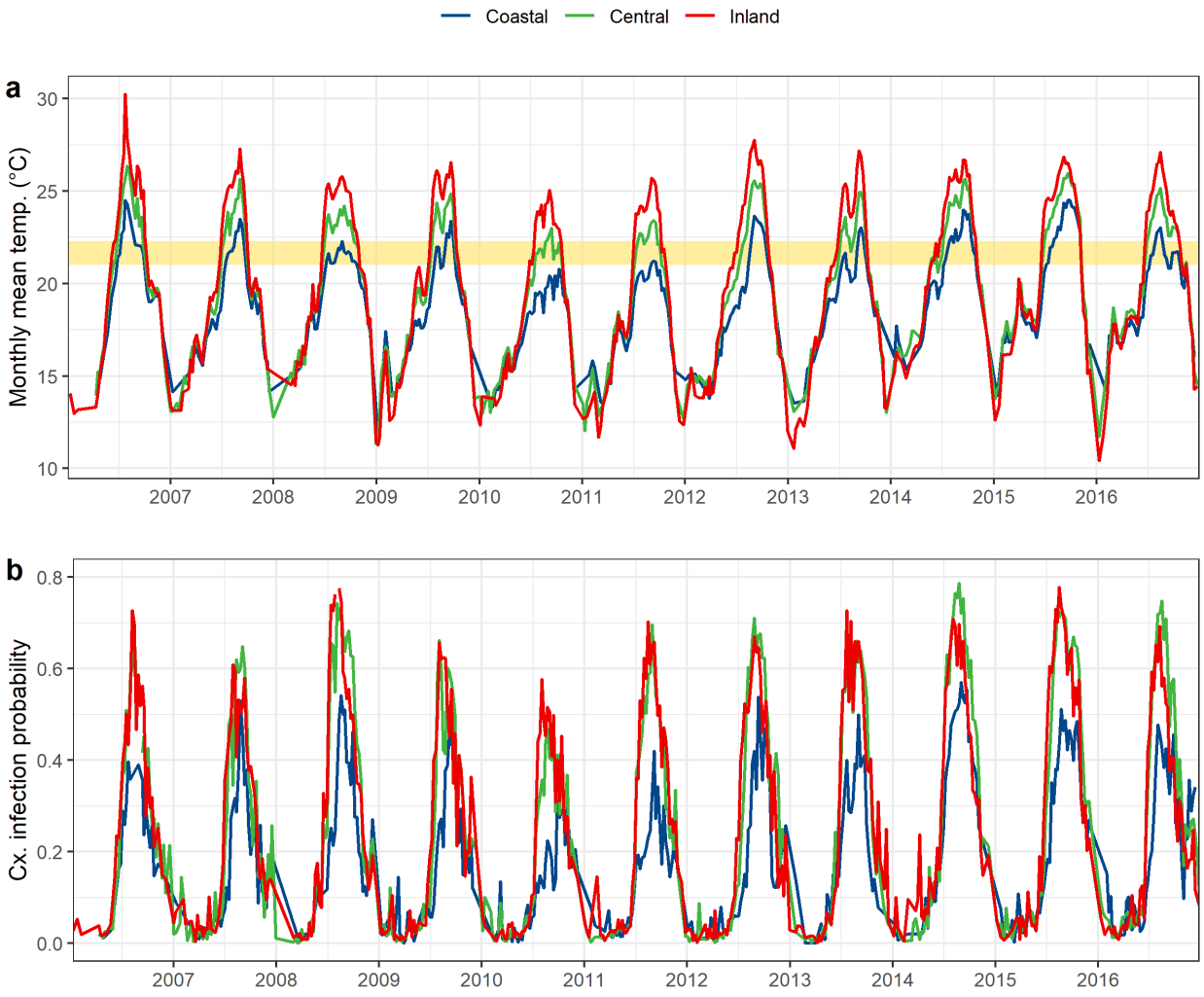

Figure S4: Weekly trends in (a) monthly mean temperature and (b) *Cx.* infection probability from 2006-2016. The yellow box in (a) marks transitional temperatures (21.0 - 22.7°C) between the inhibitory and favorable temperature ranges.

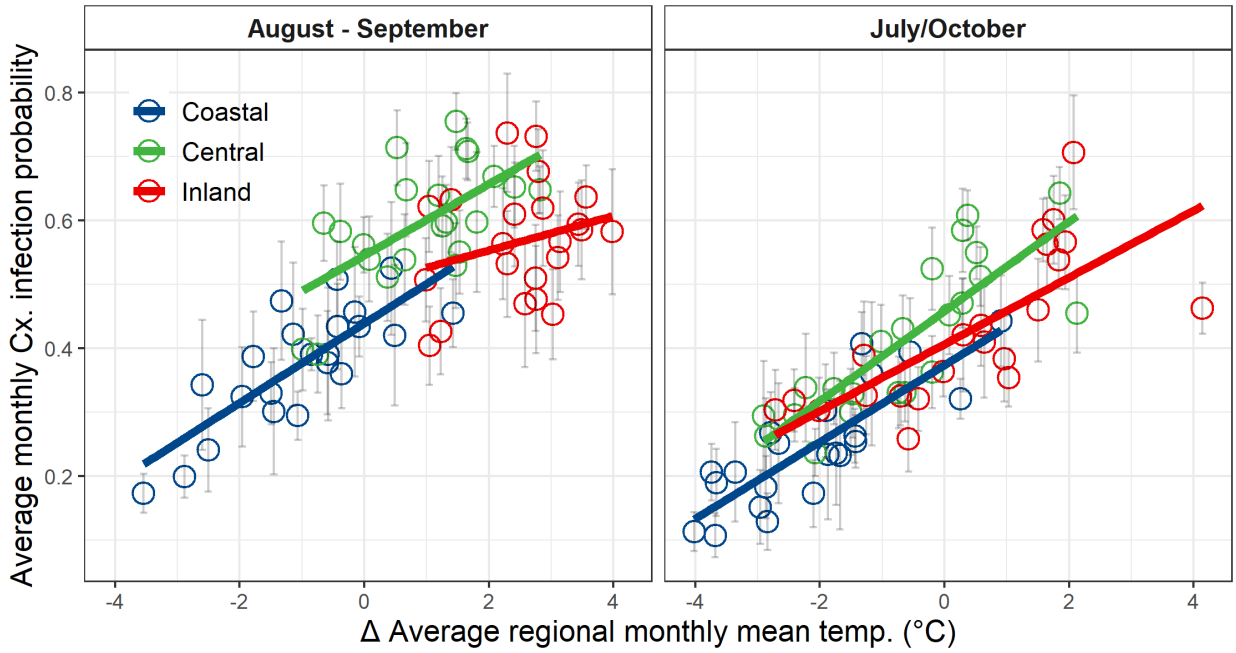

Figure S5: The fitted linear relationships between the difference in monthly mean temperature compared to the regional average and the average monthly *Cx.* infection probability for each metropolitan LA climate zone in August and September (left) and July and October (right). Slopes estimated separately by zone (color).

SI-3: Future Climate Projections

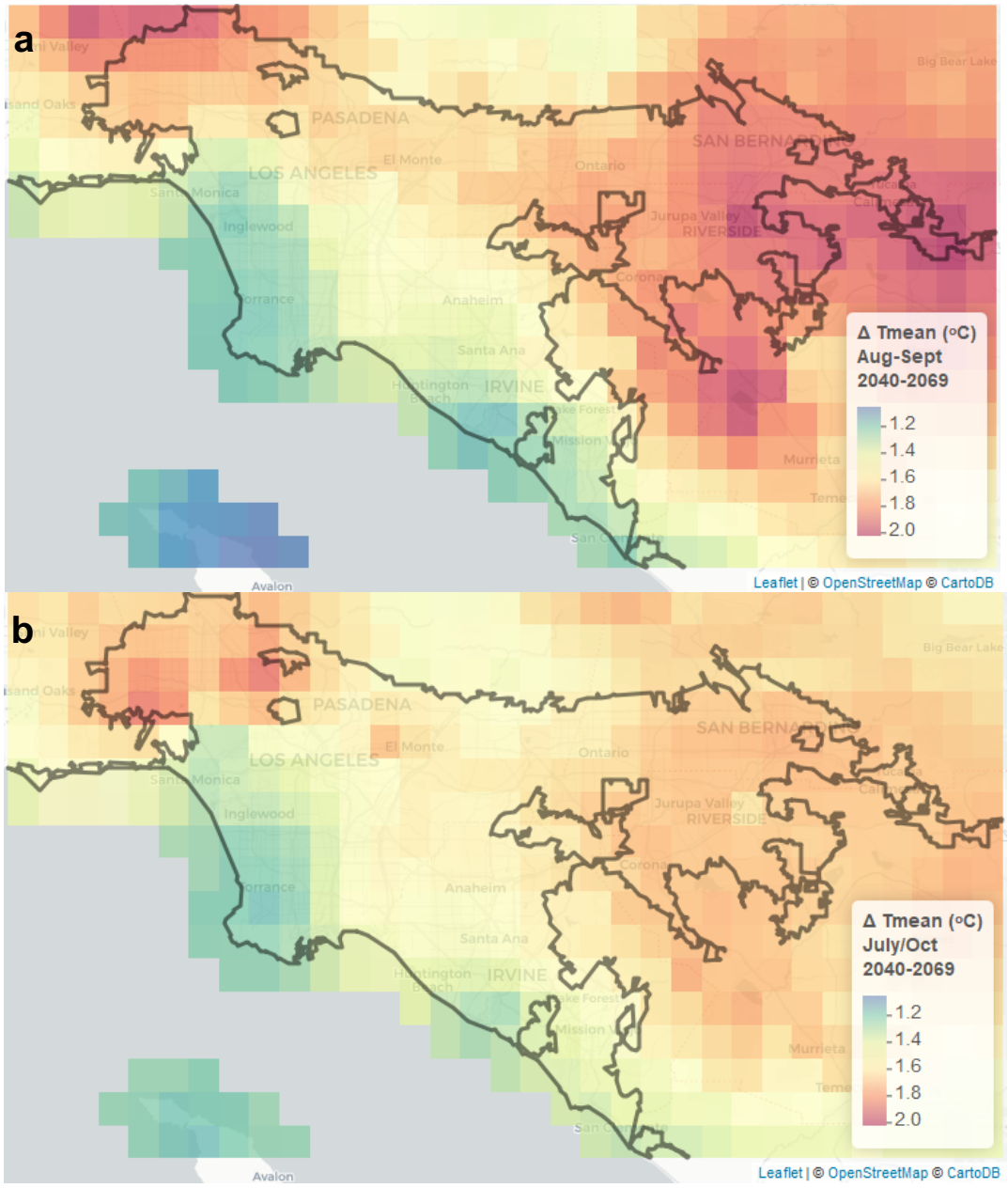

Figure S6: Projected mid-century (2040 to 2069) changes in mean temperature in a) August and September and b) July and October derived from an ensemble of four global climate projections of daily minimum and maximum temperature following Representative Concentration Pathway 4.5. The four models used (HadGEM2-ES, CNRM-CM5, CanESM2, and MIROC5) were selected by California's Climate Action Team Research Working Group to represent a range of possible simulations (17). Global models were downscaled to ~6 km resolution using the "Localized Constructed Analogs" statistical technique (18). Projected changes represent the

difference between mean temperature at mid-century (2040 to 2069) and during a 30-year period centered on the study period (1996-2025).

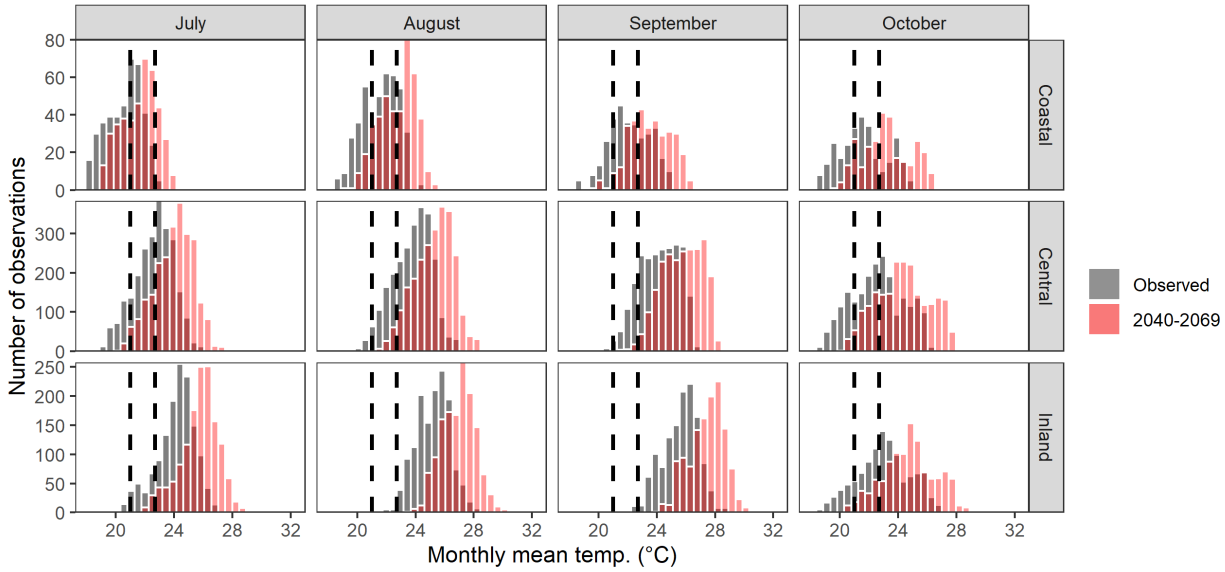

Figure S7: Histograms of observed monthly mean temperature conditions (grey bars) and the expected distribution if observed temperatures increased by mid-century (2040 to 2069) according to conservative Representative Concentration Pathway 4.5 projections (red bars). Separate histograms are provided for each climate zone (rows) from July through October (columns). Vertical dashed lines identify the bounds of the transitional temperature range (21.0 - 22.7°C). The four models used (HadGEM2-ES, CNRM-CM5, CanESM2, and MIROC5) were selected by California's Climate Action Team Research Working Group to represent a range of possible simulations (17). Global models were downscaled to ~6 km resolution using the "Localized Constructed Analogs" statistical technique (18) and averaged over each climate zone.

Methods

SI-4: Mosquito Surveillance Data and Human WNV Cases

CalSurv data were acquired via a data request for mosquito traps and pools, which provide information on trapping efforts to survey mosquito abundance and efforts to detect WNV within these mosquitos, respectively. We included data from several vector control agencies in

the Los Angeles metropolitan area (LA metro), including the Greater LA County Vector Control District (N=15,971 trap nights), West Valley Mosquito and Vector Control District (N=1,393 trap nights), LA County West Vector and Vector-borne Disease Control District (N=1,081 trap nights), San Bernardino Mosquito and Vector Control Program (N=291 trap nights), and the Long Beach Vector Control Program (N= 148 trap nights). These data were subsetted to include only adult female *Cx. quinquefasciatus* surveillance records. Further information on the reason for excluding *Cx. tarsalis* from our analysis is available in the main text. Additionally, several other nationally/globally important WNV vectors, *Cx. pipiens* and *Cx. restuans,* were not detected in our study area.

To account for potential differences in trapping and processing methodology, we included dummy variables for each agency and for observations with an unreported agency (N=17,577 trap nights). A dummy predictor variable was also included for trap type to account for differences in the abundance of female *Cx. quinquefasciatus* captured and differences in the probability detecting WNV in pools among the two trap types (1, 2). Trap data were collected from January-December, although the trapping frequency differed between surveillance sites (N = 928 sites) and ranged from 1-441 observations (mean=39.2) over the 11 year study period. The location and operational period of surveillance sites were determined based on the specific priorities of the vector control agencies and differed between years. We only included observations that employed RT-PCR assays to test for WNV in mosquito pools, because it is considered the gold standard for detecting WNV (3).

WNV presence/absence in female *Cx. quinquefasciatus* pools was the outcome variable in the random forest model. Maximum likelihood estimates of daily vector infection rates (MLE-IR)—a common alternative outcome—could not be calculated from our dataset comprising many single pools typically collected on one trap-night (>75% of observations had only one pool); sampling at individual traps was also too infrequent to generate temporally aggregated MLE-IR estimates at individual traps across the study area. Likewise, estimates of minimum infection rate (MIR), in which only one mosquito from each infected pool is assumed to be infected, contained significant outliers in cases where pools size was very small and offered no substantive advantage over presence/absence estimates since pool size was included as a predictor in the model.

The geographic location of human West Nile non-neuroinvasive and neuroinvasive disease cases was attributed to the census tract in which the individual resided. Though exposure to WNV may not have occurred in an individual's census tract of residence, it is likely to be a reasonable approximation of the location of exposure given that *Culex* vectors are most active at dawn and dusk when individuals are likely to be in closer proximity to their home. WNV cases in California are identified through clinical and laboratory criteria reported to CDPH, which subsequently confirms that they meet the Council of State and Territorial Epidemiologists case definition (4). We imputed missing onset dates (n=11, 0.4% of cases) by estimating them to be four days prior to known hospitalization dates, based on the average difference between onset and hospitalization in cases where both dates were available. In general, onset dates can be unreliable because they may depend on individual patients remembering when symptoms began, but for the purposes of our analysis we attributed cases to months rather than specific days.

SI-5: Climate and Land Cover Data

Daily values for all gridded temporal predictors (temperature [°C], precipitation [mm], and drought status [mm]) from 2005-2016 were determined by extracting the value of the raster cell containing the coordinates of each mosquito surveillance site (raster package in R version 3.5.1). Average percent forest cover, average percent impervious cover, average elevation (m), and total palustrine and riverine wetland area (m^2) were estimated within radial buffers (10m, 100m, and 1000m) surrounding mosquito surveillance sites (raster package in R version 3.5.1). Daily temperature, and precipitation were aggregated and lagged to generate several additional predictors, including average daily (1, 2, 3 day lags), average weekly (1-7, 8-14, 15-21 day lag), average monthly (1-30, 31-60, 61-90 day lags), and average quarterly (1-90, 91-180, 181-270 day lags) values during the period that preceded the date of collection at a mosquito surveillance trap. Drought/total column soil moisture was calculated at average weekly (1-7, 8-14, 15-21 day lags), monthly (1-30, 31-60, 61-90 day lags) and quarterly (1-90, 91-180, 181-270 days lags) aggregations. Large quarterly lags of temperature, precipitation and drought were included in order to account for the effects of climate conditions in the months prior to the WNV transmission season, including the preceding winter and fall. We did not resample the gridded datasets because they were never directly compared in the analysis and any adjustments to the resolution would have introduced error or sacrificed information.

Climate and land cover predictor variables were derived from weather stations, aerial photography, and satellite monitoring data:

1. Temperature: Gridded daily minimum and maximum temperature (°C) at 800m resolution were acquired from TopoWX (Data Version 2016.1; 5). We also derived daily gridded mean temperature ((maximum+minimum)/2) and diurnal variation (maximum-minimum) at 800m resolution using this dataset.
2. Precipitation: Gridded daily total precipitation (mm) at 4km resolution were acquired from PRISM (6).
3. Drought: We quantified drought using two related metrics: daily total column soil moisture (mm) and daily anomalies in total column soil moisture at 1/16 degree (~6km) spatial resolution. Soil moisture in three soil strata, spanning the surface to ~1.5m deep (~0-10cm, ~10-50cm, ~50-150cm), were estimated using the Variable Infiltration Capacity (VIC) hydrologic model (7). Baseline total column soil moisture for calculating total column anomalies was estimated at each mosquito surveillance site as the average total column soil moisture for each day of the year (1-365) across the entire study period (2006-2016), VIC-based soil moisture simulations compare well with standard drought indices like Palmer Drought Severity Index, and are available at higher spatial and temporal resolution (8, 9).
4. Wetland Cover: Palustrine and riverine wetland cover were acquired from the U.S. Fish and Wildlife Service National Wetlands Inventory (NWI), which created vectorized outlines of wetlands using at least 1:40,000 scale (1-meter ground resolution) aerial imagery captured in 2005 or 2006 (10).

5. Forest Cover: Gridded 30m spatial resolution percent canopy cover during peak growing season (circa 2010) was acquired from the United States Geological Survey (USGS) 2010 Global Tree Canopy Cover Landsat-based dataset (11).
6. Impervious Cover: Gridded 30m resolution percent impervious surface cover (circa 2011) was acquired from the USGS 2011 Landsat-based National Land Cover Database Percent Developed Imperviousness layer (12).
7. Elevation: Gridded 10m resolution digital elevation data (based on digitized contour maps and, where available, other sources like aircraft-based lidar) were acquired from the USGS National Elevation Dataset (13).

Table S3: Predictor variables included in random forest model.

Model Predictors	Spatial Resolution	Temporal Resolution	Data Source
Climate (3 lags at daily, weekly, monthly, quarterly aggregations)			
Mean Temperature	800m	Daily	TopoWX
Minimum Temperature	800m	Daily	TopoWX
Maximum Temperature	800m	Daily	TopoWX
Diurnal Variation	800m	Daily	TopoWX
Total Precipitation	4km	Daily	PRISM
Drought (total column soil moisture/total column anomaly)	~6km	Daily	Variable Infiltration Capacity (VIC) Model
Land-cover (quantified within 10, 100, 1000m buffers)			
Palustrine and Riverine Wetland	Vectorized	—	National Wetlands Inventory (NWI)
Canals and ditches	Vectorized	—	National Hydrography Dataset (NHD)
Impervious surfaces	30m	—	National Land Cover Dataset (NLCD)
Canopy cover	30m	—	USGS Global Tree Canopy Cover
Elevation	10m	—	National Elevation Dataset (NED)
Non-Environmental Predictors			
# Collected *Cx. quinquefasciatus*	—	1 trap-night	California Vectorborne Disease Surveillance System
# Pooled *Cx. quinquefasciatus*	—	1 trap-night	California Vectorborne Disease Surveillance System
Trap Latitude	—	—	—
Trap Longitude	—	—	—
Year	—	—	—
Month	—	—	—
Week of Year	—	—	—
Vector Control Agency	—	—	California Vectorborne Disease Surveillance System
Trap Type	—	—	California Vectorborne Disease Surveillance System

SI-6: Model Tuning

A grid search designed to optimize several random forest hyperparameters, including the bootstrap sample size and the appropriate level of undersampling to correct class imbalance, was conducted on a dataset consisting of a random subsample of 20% of observations. Models were built using ensembles of 500 trees, which was sufficient for classification accuracy to stabilize in both classes. Out-of-bag (OOB) cross validation (i.e., predictions on observations not included in bootstrap samples employed in tree construction) was used to select the hyperparameter set that maximized the balanced accuracy of model predictions (14). The best performing hyperparameter set included a 50/50 class balance and a bootstrap sample size equal to 25% of the observations in the minority class. All predictor variables were included as candidates at each split (maximum mtry) to improve interpretability of marginal effects.

The remaining 80% of observations were used to evaluate model performance and assess the importance and marginal effects of predictors. OOB validation was used to calculate overall performance of the model, including measures of AUC (area under the receiver operating characteristic curve), sensitivity, specificity and accuracy. Blocked k-fold cross validation was used to estimate the dependence of model predictions on temporal, spatial, and repeated measure structures (15). These error estimates helped to determine the generalizability of model predictions, or the anticipated error on novel predictions in spatial or temporal domains.

To assess temporal prediction error, we generated separate cross-validation folds, each representing all observations collected in one of the years or months of mosquito collection. We trained a model on all years or months except one, then predicted WNV presence/absence for observations in the hold-out fold. Each year or month was held out once and the hold-out AUC, sensitivity, specificity, and accuracy were calculated (15, 16).

To estimate combined spatial and repeated measures prediction error, each surveillance site with at least 10 independent observations (enough to generate robust error estimates) and all sites within a 10km radius of that site were iteratively withheld from a training model. Over the 11 years, there were 456 surveillance sites with at least 10 independent observations. Predictions were then generated on the hold-out observations and the AUC, sensitivity, specificity, and accuracy were calculated. This method ensured that predictions on the hold-out sites were not informed by observations from that site (repeated measures) or nearby sites (spatial autocorrelation). The removal of sites within a 10km radius was a conservative measure to ensure that spatial dependencies between nearby sites did not inform predictions.

A consequence of the rarity of WNV detection in pools (12% positive, 88% negative) and the desire to achieve balanced sensitivity and specificity is the model's tendency to overpredict the presence of WNV; 80% accuracy for both presence and absence leads to many more false positives than false negatives (18% and 2% of all observations were false positive and false negatives, respectively).

Supplemental References

1. Williams GM, Gingrich JB (2007) Comparison of light traps, gravid traps, and resting boxes for West Nile virus surveillance. *J Vector Ecol* 32(2):285–292.
2. L'Ambert G, Ferré J-B, Schaffner F, Fontenille D (2012) Comparison of different trapping methods for surveillance of mosquito vectors of West Nile virus in Rhône Delta, France. *J Vector Ecol* 37(2):269–275.
3. Kesavaraju B, et al. (2012) Evaluation of a rapid analyte measurement platform for West Nile virus detection based on United States mosquito control programs. *Am J Trop Med Hyg* 87(2):359–363.
4. West Nile virus in California: Guidelines for human testing and surveillance. http://westnile.ca.gov/website/publications/2006%20WNV%20Guidelines_rev0706.pdf (2006).
5. Oyler JW, Ballantyne A, Jencso K, Sweet M, Running SW (2015) Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Int J Climatol* 35(9):2258–2279.
6. Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG (2008) Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J Appl Meteorol Climatol* 47(2):475–497.
7. Variable Infiltration Capacity (VIC) Macroscale Hydrology Model Available at: http://www.hydro.washington.edu/Lettenmaier/Models/VIC/index-old.shtml [Accessed April 18, 2019].
8. Sheffield J, Goteti G, Wen F, Wood EF (2004) A simulated soil moisture based drought analysis for the United States. *J Geophys Res Atmospheres* 109(D24).
9. Sheffield J, Wood EF (2008) Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J Clim* 21(3):432–458.
10. National Wetland Inventory Available at: https://www.fws.gov/wetlands/ [Accessed April 18, 2019].
11. USGS Global Tree Cover Canopy Database Available at: https://web.archive.org/web/20180205193014/https://landcover.usgs.gov/glc/TreeCoverDescriptionAndDownloads.php [Accessed April 18, 2019].
12. NLCD 2006 Percent Developed Imperviousness (2011 Edition) (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium Available at: https://www.mrlc.gov/data/nlcd-2006-percent-developed-imperviousness-2011-edition-conus [Accessed April 18, 2019].
13. National Elevation Dataset (NED) Available at: https://gdg.sc.egov.usda.gov/Catalog/ProductDescription/NED.html [Accessed April 18, 2019].
14. Velez DR, et al. (2007) A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genet Epidemiol* 31(4):306–315.
15. Roberts DR, et al. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8):913–929.
16. Forman G, Scholz M (2010) Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *ACM SIGKDD Explor Newsl* 12(1):49–57.
17. Pierce DW, Kalansky JF, Cayan DR. Climate, drought, and sea level rise scenarios for the fourth California climate assessment. California's Fourth Climate Change Assessment, California Energy Commission. Publication number: CNRA-CEC-2018-006. 2018.
18. Pierce DW, Cayan DR, Thrasher BL (2014) Statistical Downscaling Using Localized Constructed Analogs (LOCA). *J Hydrometeorol* 15(6):2558–2585.