

Effects of Climate and Built Environments on *Aedes aegypti*
distribution in the Western United States

Nathanial O'Dell

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Public Health

University of Washington

2026

Committee:

Joan Casey

Marissa Childs

Program Authorized to Offer Degree:

Epidemiology

©Copyright 2026

Nathanial O'Dell

University of Washington

Abstract

Effects of Climate and Built Environments on *Aedes aegypti*
distribution in the Western United States

Nathanial O'Dell

Chair of the Supervisory Committee:
Dr. Joan Casey
Department of Epidemiology

Since its establishment in California in 2013, the dengue and Zika vector *Aedes aegypti* (L.) has rapidly spread across the state, and as of 2025, has been found in 64% of counties with active vector surveillance. Previous studies modeling this vector's distribution in the United States typically neglect human-mediated dispersal effects that may be driving rapid, long distance movement of vectors that lead to introductions into new locations. To address this gap, we incorporated connectivity measures between trap locations into a shape-constrained additive model (SCAM). The model was trained on *Ae. aegypti* occurrence data from 2014–2025 from the California Vector-borne Disease Surveillance System (CalSurv) with connectivity measured by traffic flow from sites that collected *Ae. aegypti* in the prior year. We found that this connectivity measure, alongside fractional impervious surface and annual average minimum temperatures, had the strongest partial effect on *Ae. aegypti* occurrence throughout California. Model predictions for the current risk of *Ae. aegypti* invasion throughout California also indicated that there is significant risk within unsampled regions of southeast and northern California, while Oregon and Washington state were

predicted to be homogeneously low-risk. We document the significant increase in the range of *Ae. aegypti* throughout California over the past 12 years, and highlight the relevance of these results in guiding proactive and targeted mosquito trap placement and protecting the public from future *Aedes*-borne disease transmission in the western United States.

1 Introduction

2 *Aedes aegypti* (L.) is an important vector of dengue, Zika, and yellow fever. Originating in sub-
3 Saharan Africa, *Ae. aegypti* is hypothesized to have arrived to the Americas in the 15th century,
4 with the first confirmed outbreaks of Yellow Fever in this region occurring in both Havana, Cuba,
5 and Yucatan in 1648 [1]. Post-invasion, *Ae. aegypti* mosquitoes rapidly expanded throughout
6 South America and into eastern North America, bringing with them outbreaks of yellow fever and
7 occasionally dengue [2]. Following this period of exploding distribution, *Ae. aegypti* was at least
8 partially eradicated from many regions of the Americas and the Caribbean [2]. Today, after a
9 secondary introduction in the Americas, *Ae. aegypti* populations have been shown to be genetically
10 insecticide-resistant, which may be hampering mosquito control efforts and aiding in their continued
11 expansion throughout North America [3, 4].

12 These historical patterns—where *Ae. aegypti* not only invades a new geography but also success-
13 fully establishes, expands its range, and begins to transmit disease in rapid succession—reflect the
14 vector’s life history. Given optimal temperature and precipitation conditions, *Ae. aegypti* can thrive
15 in urban and suburban settings [5, 6]. *Ae. spp.* rely on small, often artificial, water containers as
16 sites for oviposition [7], which tend to be more abundant in areas populated by humans. Densely
17 populated regions also provide ample hosts for female mosquitoes to blood feed more frequently
18 than they otherwise may, which is particularly important in desert ecosystems that might otherwise
19 be inhospitable for *Ae. aegypti* [8].

20 Mosquito-borne disease incidence have been hypothesized to be increasing due to interacting effects
21 of climate change, socioeconomic and urbanicity factors [9]. Chikungunya is one such *Aedes*-borne
22 disease that is expected to see increased transmission in densely populated regions under future
23 climate scenarios [10]. Increasing urbanization via land development has also been found to predict
24 areas at risk of *Ae. aegypti* establishment and subsequent disease transmission, as measured by
25 urban growth rates in global models [4] and within field studies comparing abundance in urban
26 versus suburban/rural environments in Côte d’Ivoire [11] and Kenya [12].

27 *Ae. aegypti* is considered to be firmly established in portions of the US, being found primarily in

the south and southeastern states of the country. Beginning in 2013, however, *Ae. aegypti* have been consistently detected by mosquito control districts throughout California [13], and within 4 years expanded both north and deeper south. Immediately post-establishment, genetic differentiation between north and southern mosquito populations were consistent with multiple, independent invasion events rather than entirely contiguous spread [13]. In northern California cities that had introductions from 2019-2021, however, populations have been shown to largely share genetic ancestry with southern California clusters [14], suggesting that at least part of observed expansion was from outward expansion within California rather than repeated new introductions.

Since the first detection of breeding populations in California in 2013 in Fresno, Madera, and San Mateo Counties [13], as of December 2025, *Ae. aegypti* has been detected in nearly 50% of all counties in California and 64% of counties with active vector surveillance. Given that these mosquitoes are weak fliers with an estimated daily flight range of 100 m [15], the rapid, long-distance dispersal displayed in the 13 years following their establishment suggests passive dispersal mechanisms akin to hitchhiking on human transport that may be facilitating long-range vector movement. Consistent with their continued northward expansion, in 2024 and 2025 respectively, *Ae. aegypti* was found for the first time in Oregon (Jackson County) and Idaho (Canyon County).

This rapid range expansion have raised concerns among state vector control agencies regarding the possibility of future local transmission of *Aedes*-borne disease [16]. Although local *Aedes*-borne disease transmission within the continental United States continues to be rare, in 2024 there were a record 18 locally-acquired dengue cases recorded in California, 14 of which were in Los Angeles County neighborhoods with a high number of collected female *Ae. aegypti* mosquitoes (though all subsequent specimens were dengue-negative) [17, 18]. Further, *Ae. aegypti*-transmitted viruses have caused public health emergencies at both local and global scales including the 2015-2016 Zika outbreak [19], the 2023-2024 surge in dengue cases [20], and the 2024-2025 chikungunya outbreak [21].

Importantly, many invasion and disease outbreak events have been linked to trade-related human activity via maritime and air travel networks [22, 23]. Yet, with notable exceptions [4, 24], vector

modeling efforts typically do not include measures that directly describe human mobility in relation to dispersal dynamics. Instead, models focused on future distribution at spatial scales relevant to state vector surveillance systems typically treat *Ae. aegypti* presence as a function of climate and land cover, neglecting to consider how private and/or commercial vehicle movement patterns may be facilitating rapid, passive dispersal of vectors.

This study seeks to address this gap by incorporating human movement within California via vehicle traffic counts into a species distribution model (SDM) framework, linking regions of established *Ae. aegypti* presence to those at risk of invasion. To measure this, we first estimated the annual changes in *Ae. aegypti* range within California using concave hulls to understand the rate of expansion post-2013. Then, we fit a shape-constrained additive model (SCAM) to vector surveillance data from 2014-2025 to model the probability of presence at trap sites in California as a function of weather, landscape characteristics, and traffic flow from origin locations with known *Ae. aegypti* presence to understand drivers of year-to-year expansion. Finally, to identify potentially suitable regions with high likelihood of future invasion, we predicted presence probabilities over the entire state of CA as well as Oregon and Washington state.

Better understanding and characterizing the drivers of this species' range expansion is critical for informing targeted vector control efforts needed to prevent autochthonous disease transmission in newly invaded areas of the US. By disentangling human-mediated dispersal effects from weather and land cover types, we aim to better anticipate where *Ae. aegypti* will spread next in the western United States.

Methods

Software

Statistical analyses were performed in R V4.5.1 [25]. We used the `data.table` [26] package for all non-georeferenced data manipulation tasks. All spatial statistics were conducted with the `terra` package [27] using the EPSG:4087 coordinate reference system (CRS) to facilitate distance calcu-

lations between trap locations.

Study area

We obtained mosquito trapping data collected by local vector control programs from 2003-2025 California Vectorborne Disease Surveillance System (CalSurv) Data Request #000090). A total of 15 counties had no recorded collections (**Figure 1 & 2**), as mosquito control efforts in these suburban/rural regions may involve pesticide treatments but not mosquito trapping (personal communication, CalSurv Data Steward, March 2, 2026). Collection methodology and trapping effort were heterogeneous, and most traps (> 90%) that recorded *Ae. aegypti* occurrence were collected after a single trap night. During data entry, it was not standard practice to record all the species that were absent during collection events, and thus zero counts were rare.

For the purposes of statistical analysis, we limited data to years following the establishment of *Ae. aegypti* in California in 2013.

Data processing

Data were delivered in 5-year increments and systematically cleaned to standardize across datasets and handle inconsistencies in data entry. Data cleaning was performed using functions created for this project that are compiled into a single package hosted on a public GitHub repository (<https://github.com/nathanialaodell/skeetersweeper>). Briefly, dataset column and species (i.e. “*Ae aegypti*”) name representation was standardized. Locations with imprecise or missing coordinates were reverse geocoded using address information—in the event of simultaneously missing address and geospatial coordinate information, entries were removed (< 1% of the original surveillance data).

Species distribution models require “background” points in order to create estimates of environmental conditions that inhibit species presence. Prior to filtering the dataset to only include the species of interest, we defined absences as any trap collection dates where there was a record of collected mosquitoes sans *Ae. aegypti*. For probability modeling, trap location point data were

aggregated into 5 km cells with a binary classifier indicating whether *Ae. aegypti* was collected on any collection date in a given year.

Environmental covariate selection was informed by predictors consistently associated with *Ae. aegypti* presence across the US [4, 8, 28]. Yearly land cover (land cover type; percent impervious surfaces) and weather (annual average maximum and minimum temperature in Celsius; total annual precipitation in millimeters) variable rasters for 2013-2025 were downloaded at 30 m and 4 km resolution from the US Geological Service National Land Cover Database (NLCD) [29] and the Oregon State Parameter-elevation Regressions on Independent Slopes Model group (PRISM) [30] respectively. Rasters were then downsampled to 5 km to facilitate modeling and predictions. When downsampling, continuous variables were averaged and the modal value of aggregated cells was calculated for categorical variables. At the time of data extraction, the most current NLCD release was 2024; that year’s land cover information was carried over to 2025.

***aegypti*-sourced Annual Average Daily Traffic**

To more directly incorporate the effect of human mobility into model predictions, we used annual average daily traffic (AADT) values for road segments throughout state highway networks. Both state highway system shapefiles and yearly AADT were obtained from publicly accessible California Department of Transportation (Caltrans) [31, 32], Oregon [33] and Washington state [34] geoportals. AADT information was stored as point data with exact post mile identifiers. For bidirectional road segments, two AADT values were present: “back” (west or south bound) and “ahead” (east or north). Similarly, highway lines had an associated post mile range and traffic direction identifier, which were used when merging AADT to their respective road segment. Although AADT fluctuates year to year, the observed variation was relatively small (**Figure 3**) and thus we used a 12 year average (2013-2024; 2025 was not released at the time of analysis) for each road segment.

To calculate the path between any two trap locations we then turned this state highway system into a network using the `igraph` package [35, 36]. Formally, this network can be represented as a graph $G = (V, E)$ where nodes V correspond to road intersections or predefined (by Caltrans) breaks in

road segments, and edges E represent the road segments themselves [37]. Connections between pairs of nodes $v_1, v_2 \in V$ were weighted by the AADT value along the road segment connecting them. This graph was undirected such that any connection from v_1 to v_2 was reciprocated from v_2 to v_1 . In order to ensure proper vertex snapping (e.g. of on-ramps to highways), coordinates were rounded to the nearest 0.01 meter. Additionally, to ensure full connectivity was possible from and to every point in the road network, only the largest component was considered, which encapsulated over 98% of edges. Excluded edges included roads that were not connected to any other road in the main highway road network after coordinate rounding.

To create a measure of connectivity to potential sources of *Ae. aegypti*, we then calculated the sum of traffic flow from all known trap locations with observed *Ae. aegypti*. Specifically, we calculated the nearest road node to each trap location where *Ae. aegypti* was sampled in year j —called “origin” nodes $o \in V_j$ —and each trap location that was placed in year $j + 1$ —“destination” nodes $d \in V_{j+1}$ —using the k-nearest neighbor algorithm [38]. Then, for all pairs of (o, d) we calculated the shortest path $P(o, d)$ between them along the highway road network using Dijkstra’s algorithm. Each shortest path was made up of many individual road segments, each with their own AADT weight. To obtain an aggregated metric of traffic flow across each shortest path $P(o, d)$, the median value of weights was calculated and then to estimate connectivity for a potential destination node d , we summed over all paths from origin nodes to the given destination node. That is, for each $d \in V_{j+1}$, this “*aegypti*-sourced AADT” value can be expressed as

$$\textit{aegypti-sourced AADT} = \sum_{o_k \in V_j} (\text{Traffic flow}_{o_k \rightarrow d} = \textit{median}(\textit{aadt}(e))), \quad (1)$$

where V_j is the set of trap locations in year j with observed *Ae. aegypti* and $\textit{aadt}(e)$ is the AADT value for an edge $\{e \in E \mid e \in P(o_k, d)\}$. Prior to incorporating this metric as a predictor in our SDMs, we ran a two-sided Wilcoxon Rank Sum Test to compare the median “*aegypti*-sourced AADT” of trap locations with *Ae. aegypti* occurrence to those without.

To predict the current risk of invasion of *Ae. aegypti* in the western US, we then performed this calculation for every cell in 5 km grids covering California, Oregon, and Washington state. Each grid

cell in a respective state served as a destination node d . For California, origin nodes o were defined as California trap locations with *Ae. aegypti* presence in 2025. For Oregon and Washington state, the calculation was repeated for each year from 2021 to 2025, with nodes o defined as that year’s California presence locations. As variation in AADT was also low in Oregon and Washington state, road segment AADT values were averaged across 2021-2025 prior to the *aegypti*-sourced AADT calculation.

For each year, environmental covariates and *aegypti*-sourced AADT were combined into a single multi-layer raster at 5 km resolution. Equivalent grids were created for Oregon and Washington state.

Calculation of the “*aegypti*-sourced AADT” metric was completed on Hyak, UW’s high performance computing cluster. This resource was funded by the UW student technology fee.

Yearly range estimates

We estimated of annual changes in the land area encompassed by *Ae. aegypti* using a combination of spatial clustering methods and concave hull fitting.

Many spatial clustering methods assume clusters with convex, spherical shapes, and require users to know and specify the number of clusters that should be present in a system *a priori*; as is the case in k -means clustering [39]. For data with considerable noise, which often describes geospatial datasets, density-based clustering methods can be a powerful, alternative clustering method that relaxes these assumptions. One common density-based algorithms is density-based spatial clustering of applications with noise (DBSCAN) [40]. This algorithm is implemented in R using the `dbscan` package.

DBSCAN obtains density estimates of points in a dataset, $p \in D$, by using the concept of an ϵ -neighborhood. The ϵ -neighborhood, $N_\epsilon(p)$, of a point p is the set of points within a specified radius ϵ around p . For a given year, this neighborhood is

$$N_\epsilon(p) = \{q \in D_j \mid d(p, q) < \epsilon\} \quad (2)$$

180 where $d(p, q)$ is the distance in kilometers between two trap locations $q, p \in D_j$ with *Ae. aegypti*
 181 presence in year j , and $\epsilon \in \mathbb{R}^+$.

182 Within the DBSCAN algorithm it is also necessary to specify a minimum points threshold, $minPt \in$
 183 $\mathbb{Z}^+$. This parameter is used to classify points in a dataset into core, border, and noise points, where
 184 core points form the center of a cluster, border points are the points on the edge of a cluster,
 185 and noise points are points that are not part of a cluster. To form contiguous dense regions from
 186 points, we must also define the notions of reachability and connectedness.

187 A point $q \in D$ is **directly** density-reachable from a point $p \in D$ if, and only if p is a core point
 188 (there are $\geq minPts$ from p within the radius ϵ) and the distance from p to q is $\leq \epsilon$. More formally,

- 189 1. $|N_\epsilon(p)| \geq minPts$
- 190 2. $q \in N_\epsilon(p)$

191 Additionally, p is considered density-reachable from q if there exists a set of points $(p_1, p_2, \dots, p_n)$
 192 with $q = p_1$ and $p = p_n$ such that p_{i+1} is directly density-reachable from $p_i \forall i \in \{1, 2, \dots, n-1\}$.

193 Finally, we define a density-connected point. A point $p \in D$ is density-connected to a point $q \in D$
 194 if $\exists o \in D$ such that both p and q are density-reachable from o .

195 A density-based cluster $C \neq \{\emptyset\}$ is a subset of D satisfying:

- 196 1. If $p \in C$ and q is density-reachable from p , then $q \in C$.
- 197 2. $\forall p, q \in C$, p is density-connected to q .

198 Clusters were formed for each year separately, then concave hulls were fit to each cluster present
 199 in a given year. A concave hull is a polygon which encapsulates all the points of a specified input
 200 (a DBSCAN cluster). To construct each hull, a Delaunay Triangulation was first computed over a
 201 cluster's points, which is a method of connecting a set of 2D points to form triangles such that no
 202 point falls inside the circumscribed circle of any triangle (a circle that passes through all vertices
 203 of a triangle whose center is located at the intersection of the perpendicular bisector of the sides)
 204 [41]. The longest outer edges of this triangulation were then progressively removed until the ratio

of the lengths of the longest to shortest boundary edges reached a target threshold, which we set to 0.15.

A *minPts* parameter that is too low in combination with a high ϵ can lead to unreasonably large clusters that encapsulate outliers and regions where no trapping occurs, while the inverse can lead to overly conservative estimates. In an attempt to disentangle spatial clustering that is a consequence of trap placement bias from meaningful clusters of vector presence, several combinations of *minPts* and ϵ values were tested.

A tight concave hull algorithm was used to create polygons. Following the fitting of the hulls, the land area contained within each shape was summed to obtain a yearly estimate of the range of *Ae. aegypti* in California.

Model fitting and calibration

We use logistic regressions to fit the probability of *Ae. aegypti* occurrence as a function of land use, meteorology, and importation risk. Logistic regression models inherently assume linear dependence of the log-odds of an outcome on predictors. For environmental variables like temperature and precipitation in particular, such an assumption broadly fails with respect to *Ae. aegypti*. This has been shown in a variety of modeling contexts including adult mosquito abundance in Arizona [42] and Mexico [43]. Field and laboratory results reinforce these aforementioned findings, with evidence that *Ae. aegypti* egg development's dependence on temperatures between 25-35 °C can be estimated by a gamma distribution [44], and that tolerance to fluctuations in temperatures is related to the geographical location that they are developing in [45].

As such, modeling frameworks that are too inflexible to allow for non-linear relationships may not be able to capture the full extent of environmental signals. Thus, to model and predict the invasion risk of *Ae. aegypti* throughout the western US, we fit and compared two sets of models: a generalized additive model (GAM) [46] and, in an effort to create more biologically relevant smoothing terms, a shape-constrained additive model (SCAM) [47], using the `mgcv` and `scam` packages.

GAMs are generalized linear models (GLM) that include a sum of smooth functions on predictors

[48] which allow for flexible specification of the dependence of a response on a set of predictors.

Consider a GLM with link function $g(\mu)$ and p predictors. We express this GLM as

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3)$$

With GAMs, we can fit smooth functions on any number of terms within equation (3). For simplicity, we consider a model with $p = 3$ predictors, two of which are smooth terms, f_1 , f_2

$$g(\mu) = \beta_0 + f_1(x_1) + f_2(x_2) + \beta_1 x_3 \quad (4)$$

In order to estimate f , it is necessary to represent it in such a way that equation (3) becomes a linear model. This is done by choosing a basis function, which defines the space of functions of which f is an element. If $b_j(x)$ is the j th basis function, then f has the representation

$$f(x) = \sum_{j=1}^k b_j(x) \beta_j \quad (5)$$

for some values of the unknown parameters, β_j . Substituting (5) into (4)

$$g(\mu) = \beta_0 + \sum_{j=1}^k b_j(x_1) \beta_j^{(1)} + \sum_{j=1}^k b_j(x_2) \beta_j^{(2)} + \beta_1 x_3 \quad (6)$$

(a linear model).

The choice of the basis dimension size k of f is considered arbitrary as it only sets an upper limit to the flexibility of a term, and as such increases in size usually solely come with a trade-off in computational efficiency of estimation [49]. However, it is possible for biologically unrealistic smooths to be estimated if this upper limit is not restricted. A brief sensitivity analysis was conducted on the effect of increasing k for smooths for minimum temperature maximum temperature, and cumulative precipitation.

Within a GAM framework there are a number of methods to estimate f in the form of *smoothing splines* [49]. For weather predictors, we used a first derivative penalty on a quadratic B-spline, and a thin plate regression spline with an extension of the smoothing penalty on the entire smooth for impervious surface percentage and *aegypti*-sourced AADT variables.

SCAMs extend this framework by imposing monotonicity or convexity constraints on individual
smooths via shape constrained P-splines [47], allowing prior ecological knowledge to be considered
explicitly in the model structure. For example, suppose we want to constrain the estimate of x_1 in
equation (6) to be monotone increasing. Then $\beta^{(1)}$ would be reparameterized as

$$\beta^{(1)} = \sum \exp(D\gamma^{(1)}) \quad (7)$$

where Σ is the lower triangular cumulative sum matrix with $\Sigma_{ij} = 1$ if $(i \geq j)$, D is a second-order
difference matrix, and $\gamma^{(1)}$ are unconstrained parameters. It also follows that equation (6) could
be rewritten as

$$g(\mu) = \beta_0 + \sum_{j=1}^k b_j(x_1) \left[\sum \exp(D\gamma^{(1)}) \right] + \sum_{j=1}^k b_j(x_2) \beta_j^{(2)} + \beta_1 x_3 \quad (8)$$

Since $\exp(\cdot) > 0$, the cumulative sum guarantees $\beta_j^{(1)} \geq \beta_{j-1}^{(1)}$ for all j (see [47] for a complete list
of possible constraints). We chose monotone increasing for minimum temperatures and concave for
maximum temperatures as constraints in the model. This was motivated by previous studies exam-
ining how temperatures below 16 °C and above 36 °C decreased egg laying behavior significantly
[50], the positive effect of increasing temperatures between 20 °C and 30 °C have on *Ae. aegypti*
development [51], as well as temperature’s effect on female bloodmeal behaviors [52]. Due to a lack
of clarity as to how precipitation at the extreme ends of its distribution may impact *Ae. aegypti*
occurrence, we left this term unconstrained.

The SCAM is hereafter referred to as our “constrained” model; conversely, the GAM is “uncon-
strained”. Formally, the constrained model is expressed as

$$\begin{aligned} \text{logit}(p) = & \beta_0 + \beta_1 \cdot \text{Year} + \beta_3 \cdot \text{LndCov} + f_{1,\uparrow}(t_{\min}) + \\ & f_{2,\cap}(t_{\max}) + f_3(\text{ppt}) + f_4(\text{FctImp}) + f_5(\text{AADT}) \end{aligned}$$

and for the unconstrained model,

$$\begin{aligned} \text{logit}(p) = & \beta_0 + \beta_1 \cdot \text{Year} + \beta_3 \cdot \text{LndCov} + f_{1,bs}(t_{\min}) + \\ & f_{2,bs}(t_{\max}) + f_{3,bs}(\text{ppt}) + f_4(\text{FctImp}) + f_5(\text{AADT}) \end{aligned}$$

where $f_{\uparrow}$ denotes a monotone increasing P-spline; $f_{\cap}$ a concave P-spline, f a thin plate spline with shrinkage and selection; and f_{bs} a B-spline with a derivative penalty ($m = (3, 1)$), with 4 knots: the outer two encompass the extrapolation range of the out of sample data, while the inner two encompass the interpolation range of the in sample data.

Model comparisons

Non-linear regression models and SDMs at large are typically poor at extrapolation [53]. Although SCAMs can create smooths that are constant at the tails, values that fall far outside the range of the training data run the risk of leading to wildly inaccurate predictions.

To compare how our models performed with unseen data, the dataset was divided in a spatial train/test set, with a 4,250,000 meter northing threshold. This value was chosen after a brief comparison of possible thresholds, with the intent being to choose a value that provided a degree of separation of weather variable distributions and ample sample size for modeling testing. Trap locations north of this value were used to test out of sample model performance prior to extrapolation of results into Oregon and Washington state. Out of sample accuracy of the constrained and unconstrained models were compared using both the area under curve (AUC), to evaluate the models' ability to correctly rank positive observations over negative ones, [54] and Brier score [55], to measure the actual percent accuracy of modeled *Ae. aegypti* occurrence. The latter was calculated as

$$\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \quad (9)$$

where p_i is the i th predicted probability and o_i is the i th observed value that takes on values $(0, 1)$.

Results from this procedure were then used to choose a “preferred” model. Model diagnostics, including residuals independence checks, were done using simulated residuals (i.e. bootstrapping) in the DHARMA package. The Akaike Information Criterion (AIC) [56] of nested versions of this model were compared for the purposes of feature selection. For Oregon and Washington state predictions, weather variables were clamped to the minimum and maximum values of the corresponding variable distribution in California.

Results

Summary statistics

There were a total of 872,047 unique trapping events across 79,976 trap sites from 2003-2025. After subsetting to only years after *Ae. aegypti* established in California (2013), we were left with 693,073 collection events across 75,328 sites—of these, 20.29% events had *Ae. aegypti* presence. Trapping effort varied greatly (**Figure 5**), with a median number of collections of 2 (IQR 3) collections per site from 2013-2025.

A correlation matrix for continuous model predictors (all predictors sans land cover type) indicated that temperature and precipitation variables were strongly correlated (**Figure 6**). However, because correlation coefficients were less than a 0.7 cutoff, no variables were removed for collinearity.

Excluding 2013-2014, Wilcoxon Rank Sum Test results showed significant differences in the median “*aegypti*-sourced AADT” metric in sites with *Ae. aegypti* compared to those without (**Table 1**).

Concave hull estimates

Final parameters for DBSCAN fitting were chosen by comparing land area estimates against trends in the annual number of traps with occurrence, with $minPts = 20$ and cluster radius $\epsilon = 17.5$ km providing the closest agreement (**Figure 1, Figure 4**).

Annual changes in estimated *Ae. aegypti* range can be found in **Table 2**. Initial expansion was nonlinear but gradual relative to increases estimated post-2017; from 2017-2025 the average annual occupied land area increased by $\sim 2,500$ km² each year. *Ae. aegypti* was estimated to have expanded its range in all years except from 2023 to 2024, when the cumulative hull area decreased by 452.23 km² (**Figure 7**). From 2022-2023 the estimated range expanded by over 5,500 km², which was the largest increase across all years.

The number of cluster fit by the DBSCAN algorithm also grew nonlinearly over time, with the peak number of clusters fit being 13 in 2022 and 2025 (**Figure 8**).

Model selection and training

To ensure biological plausibility without fitting splines that were too flexible, basis dimension sizes of 6, 5, and 6 were selected for the final unconstrained model's minimum temperature, maximum temperature, and cumulative precipitation functions (**Figures 9, 10, 11**).

Both models were able to predict spatial out-of-sample North California presence and absence with high accuracy, with AUC and Brier scores of 0.935 and 0.091; and 0.939 and 0.090 for the constrained and unconstrained models, respectively (**Table 3**). Although the unconstrained model performed slightly better than the constrained in out of sample validation tests, we chose the constrained model as our preferred model due to the smooth functions' fit at the tail of weather variable distributions, and because performance differences were very small (**Figure 12**). In nested model comparisons, the model incorporating all predictors had the lowest AIC and was used for final predictions (**Table 4**).

The estimated partial effects of the full preferred model are shown in **Figure 13**.

Minimum temperatures' partial effect was constant until $\sim 7.5^{\circ}\text{C}$, and increased steeply at temperatures above $\sim 7.5^{\circ}\text{C}$. Maximum temperatures had a positive relationship that peaked near $25\text{--}28^{\circ}\text{C}$. Cumulative annual precipitation had a positive association up to 1250 mm, although confidence intervals at higher precipitation values (greater than 1000 mm) were wide given the limited observational data in this range. Increasing *aegypti*-sourced AADT and fractional impervious surface were both associated with higher probability of *Aedes aegypti* occurrence, with the greatest increases at low levels. The magnitude of effect varied for the two variables, with *aegypti*-sourced AADT peaking at nearly a 4 log-odds partial effect.

Aside from the year indicator, land cover type was the only variable to not be fit with a smooth function. Relative to open water, cultivated cropland (OR = 1.49, $p = 0.012$) and developed land covers (OR = 1.25–1.28, $p > 0.05$) were positively associated with presence, while evergreen forest (OR = 0.35, $p = 0.003$) and barren land (OR = 0.41, $p = 0.013$) were negatively associated.

Modeled *Aedes aegypti* invasion risk in the Pacific Northwest

Our model predicts the current invasion risk of *Ae. aegypti* throughout California to be high in multiple regions where they are not currently found or are only found in a limited number of 5 km grid cells (**Figure 14**).

Southeastern California, specifically Imperial, Riverside, and the eastern edges of San Bernardino County are predicted to have major areas of land where conditions are suitable for *Ae. aegypti* invasion. Roughly 21.10% of the land area of California was predicted to have at least a 50% risk of invasion, with the majority those locations being located in the southern counties of Los Angeles, Orange, San Diego, Kern, San Bernardino, and Riverside. These counties also experience average maximum temperatures in the range of 22.5-33.3°C. Areas with high invasion risk correspond to the conditions associated with higher probability of *Ae. aegypti* occurrence: Los Angeles, San Bernardino, and San Diego County have large percentages of built environments and had the highest *aegypti*-sourced AADT values in the state.

At the edges of northern California predicted risk was homogeneously low, with most locations estimated to have < 1% risk—these regions are mostly classified as evergreen forests or shrub/scrubs and have relatively low road connectivity to occupied sites. The magnitude of these risk estimates are reflected in both Oregon and Washington state predictions—neither had any areas with estimates $\geq 0.57\%$ (**Figures 15, 16**). Both states have similar environmental and road connectivity conditions to areas of northern California where predictions were similarly low.

Discussion

In this study, we used mosquito trap collections from 2014-2025 from the California Vectorborne Disease Surveillance System (CalSurv) to model and predict the current risk of *Aedes aegypti* invasion throughout the Pacific Northwest. We observed a significant increase in the observed range of *Ae. aegypti* post-establishment in 2013.

Estimated speeds of *Ae. aegypti* expansion were far outside what would be expected if dispersal

were to occur through mosquito flight alone (~ 100 m daily) [15], with calculated year to year changes in occupied land area being as large as 5,5000 km² from 2022 to 2023 (**Table 2**). This burst in range may be a result of abnormally wet winter conditions in 2022 and 2023 throughout California [57] creating a large number of breeding habitats in regions where conditions are not necessarily conducive for full establishment.

In our modeling analysis, the preferred model had a high prediction accuracy in out of sample testing (AUC = 0.935 (95% CI: 0.925-0.945), Brier score = 0.091 (95% CI: 0.086-0.96)). Although all variables in our model were associated with *Ae. aegypti* occurrence, *aegypti*-sourced AADT, fractional impervious surface, and minimum temperatures had the strongest partial effects, with estimates near or above 2 log-odds at the right tail of their distributions.

By 2025, the average *aegypti*-sourced AADT value in California was 45.83e⁶ (**Figure 17**)—values above this average were where the partial effect of this covariate was strongest in our model. Los Angeles, Orange, San Diego, and San Bernardino Counties encompassed the most extreme traffic flow values, though the Central Valley, Shasta, Siskiyou, Mendocino, and Coastal California had large land masses with traffic flow above the average (**Figure 17**). Although fractional impervious surface and *aegypti*-sourced AADT might be expected to be collinear, the Pearson coefficient for these two predictors was 0.17, indicating that they capture meaningfully different environmental signals. This is, however, at least in part due to the fact that *aegypti*-sourced AADT variates from one year to the next because of changes in mosquito distribution. Comparatively, land cover measures tend to change at a lower rate.

Ae. aegypti thrives in densely populated and/or urbanized environments throughout the United States, as seen in Arizona [8] and throughout regions of Texas that have experienced recent increases in urban development [58, 59]. Within our model, impervious surface do not drive the observed occurrence patterns.

Further illustrating the importance of climate constraints are the model extrapolations into Oregon and Washington, where we estimate that there is little risk of *Ae. aegypti* invasion. This is consistent even in urban centers with high population density such as King County, Washington, and

Multnomah County, Oregon that theoretically would have ideal conditions for the urban dwelling vector [11, 60]. Low minimum and maximum temperatures may be serving as key environmental barriers for the species, as the maximum observed average minimum and maximum temperatures in both states were below 10°C and 23°C, respectively; collection dates with temperatures that did not exceed a respective value had $\sim 13\%$ and 14% occurrence.

The low predicted risk in Oregon is of particular interest, as it is contradicted by recent trap collections by the Jackson County Vector Control District. *Ae. aegypti* has been detected twice in Jackson County, Oregon: the first in Talent, Oregon, in 2024, and the second ~ 9 km away in Ashland, Oregon, in 2025 [61]. Whether these represent independent introductions or persistence of a single population between years is still unclear. Jackson County’s mean annual minimum temperature ($\sim 3^\circ\text{C}$) falls well below the range where our model estimates meaningful presence probability.

If these detections do represent repeated introductions rather than persistence, it would imply a high importation rate into the region and it’s possible that traffic flow, particularly along the Interstate 5 corridor, represents a quantifiable pathway for *Ae. aegypti* dispersal from established populations. Both the partial effect of *aegypti*-sourced AADT and estimates of current invasion risk therefore have important implications for vector control strategies. Areas where *Ae. aegypti* is not currently found but are known tourist and economic centers may benefit from an expansion of proactive surveillance efforts to ensure timely detection.

One question raised by our results are the implications of high predicted risk in regions with low population density, as is the case in most of San Bernardino County. The vast majority of traps in this county—where *Ae. aegypti* has been detected annually since at least 2017—are located in the southwest corner of its geopolitical boundary. Nonetheless, we estimate that large swaths of land east of these traps present conditions that are suitable for *Ae. aegypti* invasion. The mean *aegypti*-sourced AADT throughout the county is $74e^6$ and is made up mostly of shrub/scrub land cover. Observed temperatures are also suitable for occurrence, with an average minimum and maximum temperature of 13°C and 26°C. Future research should seek to overlap these low-trapping effort

regions with high predicted invasion risk and high population density in order to identify areas where surveillance expansion would be most beneficial.

Our prediction map estimates more extensive invasion risk in California than previous studies that have estimated suitability for the vector throughout the US [62, 4, 63]. Overall, our model predicted high levels of risk in the east and southeast regions of the state, as well as unsampled areas of the Central Valley. This may be related to the inclusion of more recent records of *Ae. aegypti* occurrence, as these studies included public records of *Ae. aegypti* occurrence through 2016 at most.

There are limitations to this work that must be acknowledged. First, our model was intentionally sparse to focus on the effects of climate and importation risk, and did not include variables that have been shown in previous studies to impact species occurrence, such as population density, seasonal weather conditions, humidity, elevation, and vegetation level. Population density in particular has been shown to be an important measure for urban areas, as increases in density can result in more opportunities for female mosquitoes to bloodfeed—increasing the risk of oviposition and disease transmission [8, 64, 65]. Our model also did not include importation effects from neighboring states, instead considering dispersal within California to be a closed system despite the fact that *Ae. aegypti* has been established in Arizona since its reintroduction to the state in 1998 in Tucson [66], and found regularly in Las Vegas, Nevada, since 2017 [67].

Nevertheless, the present study is an important step toward understanding the dispersal dynamics of *Ae. aegypti* in the western US. Our approach demonstrates the utility of transportation-informed dispersal modeling for identifying areas of elevated invasion risk and can help guide targeted surveillance and vector control efforts, particularly in regions where trapping data is sparse. Future modeling efforts should incorporate a more robust *aegypti*-sourced AADT measure that allows for movement in and out of the study area in a saturated model that contains additional sociodemographic and weather covariates.

Conclusions

This study documents the significant increase in the range of *Ae. aegypti* throughout California over the past 12 years, and highlights the importance of measures that quantify the risk of human-mediated dispersal in invasion risk estimates. To the best of our knowledge, this study is the first to primarily model *Ae. aegypti* occurrence and expansion in the western US, with an attempt to extrapolate results into the Pacific Northwest—a region where the species is not considered to be established. Our findings emphasize the importance of temperatures as a limiting factor to *Ae. aegypti* expansion into the Pacific Northwest and traffic flow as a facilitator of long-range vector movement.

We also highlight the need for an increase in sampling effort throughout California; especially in the northern extents of the state where *Ae. aegypti* has encroached in the past 5 years. Integration of long-term surveillance data—and the expansion of its spatial coverage—will be important to effectively track dispersal and *Aedes*-borne disease transmission as the species continues expand into portions of the western US that were previously considered to be unsuitable. While we do not directly model locally transmitted mosquito-borne disease case data, descriptions of vector occurrence provide a foundation to motivate an expansion of proactive vector surveillance, especially in more rural regions of northern California that do not currently report trap collections to CalSurv. Such efforts are imperative to ensure timely insecticide treatment and protect the public from future disease transmission.

Acknowledgments

I would like to first thank my committee members, Dr. Marissa Childs and Dr. Joan Casey, for having as much patience as they did during this project; for being so willing to advise a project that is (at least somewhat) outside of their usual research purview; and for lending their expertise and advice throughout the entire process. I also thank the state and local vector control districts throughout California that were willing to let us use their data through CalSurv for the purposes of my master of public health thesis.

472 Finally, I thank the many people in my life who have been supportive over the past two years, and
473 especially the last few months of this process. To my parents; my siblings Lydia, Dakotah, and
474 Kaylin; Nana; uncle Bobby; Tåsi; and to Jesus, Eric, and Benji, who all might as well be family:
475 none of my accomplishments while at the University of Washington would have been achievable
476 without your support.

References

1. McNeill W. Plagues and peoples. Anchor, 1977. Chap. 5
2. McNeill JR. Mosquito empires: ecology and war in the Greater Caribbean, 1620–1914. Cambridge University Press, 2010
3. Crawford JE, Balcazar D, Redmond S, Rose NH, Youd HA, Lucas ER, Ali RSM, Alnazawi A, Badolo A, Chen CH, et al. 1206 genomes reveal origin and movement of *Aedes aegypti* driving increased dengue risk. *Science* 2025; 389:eads3732
4. Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, Yi D, Johnson K, Earl L, Marczak LB, Shirude S, Davis Weaver N, Bisanzio D, Perkins TA, Lai S, Lu X, Jones P, Coelho GE, Carvalho RG, Van Bortel W, Marsboom C, Hendrickx G, Schaffner F, Moore CG, Nax HH, Bengtsson L, Wetter E, Tatem AJ, Brownstein JS, Smith DL, Lambrechts L, Cauchemez S, Linard C, Faria NR, Pybus OG, Scott TW, Liu Q, Yu H, Wint GRW, Hay SI, and Golding N. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus* SI. *Nature Microbiology*. 2019 Mar 4; 4:854–63. DOI: 10.1038/s41564-019-0376-y. Available from: <https://www.nature.com/articles/s41564-019-0376-y> [Accessed on: 2025 Oct 9]
5. Reinhold J, Lazzari C, and Lahondère C. Effects of the Environmental Temperature on *Aedes aegypti* and *Aedes albopictus* Mosquitoes: A Review. *Insects*. 2018 Nov 6; 9:158. DOI: 10.3390/insects9040158. Available from: <https://www.mdpi.com/2075-4450/9/4/158> [Accessed on: 2026 Feb 7]
6. Misslin R and Daudé É. An environmental suitability index based on the ecological constraints of *Aedes aegypti*, vector of dengue. *Revue Internationale de Géomatique*. 2017 Oct; 27. Ed. by Emsellem K, Josselin D, and Moreno Sierra D:481–501. DOI: 10.3166/rig.2017.00044. Available from: <https://rig.revuesonline.com/10.3166/rig.2017.00044> [Accessed on: 2026 Feb 7]
7. Christophers SR. *Aedes aegypti*: the yellow fever mosquito. CUP Archive, 1960
8. Coalson JE, Richard DM, Hayden MH, Townsend J, Damian D, Smith K, Monaghan A, and Ernst KC. *Aedes aegypti* abundance in urban neighborhoods of Maricopa County, Arizona, is

- linked to increasing socioeconomic status and tree cover. *Parasites & Vectors*. 2023 Oct 8; 16:351. DOI: 10.1186/s13071-023-05966-z. Available from: <https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-023-05966-z> [Accessed on: 2025 Dec 9]
9. Franklins LHV, Jones KE, Redding DW, and Abubakar I. The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*. 2019 Sep; 19:e302–e312. DOI: 10.1016/S1473-3099(19)30161-6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1473309919301616> [Accessed on: 2025 Dec 12]
 10. Tjaden NB, Suk JE, Fischer D, Thomas SM, Beierkuhnlein C, and Semenza JC. Modelling the effects of global climate change on Chikungunya transmission in the 21st century. *Scientific reports* 2017; 7:3813
 11. Zahouli JB, Koudou BG, Müller P, Malone D, Tano Y, and Utzinger J. Urbanization is a main driver for the larval ecology of *Aedes* mosquitoes in arbovirus-endemic settings in south-eastern Côte d'Ivoire. *PLoS neglected tropical diseases* 2017; 11:e0005751
 12. Ndenga BA, Mutuku FM, Ngugi HN, Mbakaya JO, Aswani P, Musunzaji PS, Vulule J, Mukoko D, Kitron U, and LaBeaud AD. Characteristics of *Aedes aegypti* adult mosquitoes in rural and urban areas of western and coastal Kenya. *PloS one* 2017; 12:e0189971
 13. Pless E, Gloria-Soria A, Evans BR, Kramer V, Bolling BG, Tabachnick WJ, and Powell JR. Multiple introductions of the dengue vector, *Aedes aegypti*, into California. *PLOS Neglected Tropical Diseases*. 2017 Aug 10; 11. Ed. by Scarpino SV:e0005718. DOI: 10.1371/journal.pntd.0005718. Available from: <https://dx.plos.org/10.1371/journal.pntd.0005718> [Accessed on: 2026 Feb 6]
 14. Campos M, Lee Y, Brisco K, Crepeau M, Cornel AJ, and Lanzaro GC. Genetics of range expansion and admixture of *Aedes aegypti* populations in California. *BMC Genomics*. 2025 Dec 19; 26:1135. DOI: 10.1186/s12864-025-12443-7. Available from: <https://link.springer.com/10.1186/s12864-025-12443-7> [Accessed on: 2026 Feb 14]
 15. Moore TC and Brown HE. Estimating *Aedes aegypti* (Diptera: Culicidae) Flight Distance: Meta-Data Analysis. *Journal of Medical Entomology*. 2022 Jul 13; 59. Ed. by Alto B:1164–70. DOI:

- 10.1093/jme/tjac070. Available from: <https://academic.oup.com/jme/article/59/4/1164/6596097> [Accessed on: 2026 Mar 5]
- 16.Porse CC, Kramer V, Yoshimizu MH, Metzger M, Hu R, Padgett K, and Vugia DJ. Public health response to *Aedes aegypti* and *Ae. albopictus* mosquitoes invading California, USA. *Emerging infectious diseases* 2015; 21:1827
- 17.CDPH Weekly Update on Number of Dengue Cases in California. California Department of Public Health Division of Communicable Disease Control, 2025 Dec 23. Available from: <https://westnile.ca.gov/pdfs/WeeklyUpdateDengueInfectionsCA.pdf> [Accessed on: 2026 Mar 8]
- 18.Vaughan AM, Park C, Ngo VP, Contreras ZA, Lee JJ, Danza P, Haddix M, Moir O, Green N, Brown M, Burleson T, Marutani A, Nicholas A, Hallum T, Vetrone S, Ortiz L, Fernandez G, El-Tobgy E, Escobar J, and Gandela TN. Investigation of and Response to Autochthonous Dengue, Los Angeles County, California, USA, August–November 2024. *Emerging Infectious Diseases* 2026 May; 32. DOI: <https://doi.org/10.3201/eid3205.251812>. Available from: https://wwwnc.cdc.gov/eid/article/32/5/25-1812_article
- 19.Heukelbach J, Alencar CH, Kelvin AA, De Oliveira WK, and Góes Cavalcanti LP de. Zika virus outbreak in Brazil. *The Journal of Infection in Developing Countries* 2016; 10:116–20
- 20.Cremades R, Sandoval-Pinto E, Ortega-Prieto AM, Jimenez-Guardeño JM, Pérez-Gómez HR, Lona Reyes JC, Sierra-Díaz E, and Regla-Nava JA. Epidemiological Characteristics of Dengue Disease in Mexico (2014–2025): A Descriptive Analysis of a Hyperendemic Country. *Pathogens* 2026; 15:190
- 21.Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, Carcelen AC, Ott CT, Sheffield JS, Ferguson NM, et al. Assessing the global threat from Zika virus. *Science* 2016; 353:aaf8160
- 22.Bennett KL, Gómez Martínez C, Almanza A, Rovira JR, McMillan WO, Enriquez V, Barraza E, Diaz M, Sanchez-Galan JE, Whiteman A, et al. High infestation of invasive *Aedes* mosquitoes in used tires along the local transport network of Panama. *Parasites & vectors* 2019; 12:264

23. Guagliardo SAJ, Lee Y, Pierce AA, Wong J, Chu YY, Morrison AC, Astete H, Brosi B, Vazquez-Prokopec G, Scott TW, et al. The genetic structure of *Aedes aegypti* populations is driven by boat traffic in the Peruvian Amazon. *PLoS neglected tropical diseases* 2019; 13:e0007552
24. Eritja R, Palmer JR, Roiz D, Sanpera-Calbet I, and Bartumeus F. Direct evidence of adult *Aedes albopictus* dispersal by car. *Scientific reports* 2017; 7:14399
25. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2025. Available from: <https://www.R-project.org/>
26. Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, Schwendinger B, and Krylov I. data.table: Extension of ‘data.frame’. R package version 1.18.4. 2026. DOI: 10.32614/CRAN.package.data.table. Available from: <https://CRAN.R-project.org/package=data.table>
27. Hijmans RJ. terra: Spatial Data Analysis. R package version 1.8-70. 2025. DOI: 10.32614/CRAN.package.terra. Available from: <https://CRAN.R-project.org/package=terra>
28. Liu-Helmersson J, Brännström Å, Sewe MO, Semenza JC, and Rocklöv J. Estimating Past, Present, and Future Trends in the Global Distribution and Abundance of the Arbovirus Vector *Aedes aegypti* Under Climate Change Scenarios. *Frontiers in Public Health*. 2019 Jun 12; 7:148. DOI: 10.3389/fpubh.2019.00148. Available from: <https://www.frontiersin.org/article/10.3389/fpubh.2019.00148/full> [Accessed on: 2026 Jan 26]
29. U.S. Geological Survey (USGS). Annual NLCD Collection 1 Science Products. U.S. Geological Survey data release. 2024. DOI: 10.5066/P94UXNTS. Available from: <https://doi.org/10.5066/P94UXNTS> [Accessed on: 2025 Dec 30]
30. PRISM Climate Group. PRISM Climate Data. Available from: <https://prism.oregonstate.edu> [Accessed on: 2025 Dec 30]
31. Transportation (Caltrans) CD of. State Highway Network Lines. California State Geoportal. 2025 Aug 11. Available from: https://gis.data.ca.gov/datasets/77f2d7ba94e040a78bfbe36feb6279da_0/about
32. Transportation (Caltrans) CD of. Traffic Volumes: Annual Average Daily Traffic (AADT). Traffic Census Program. 2025. Available from: <https://dot.ca.gov/programs/traffic-operations/census>

33. Transportation (ODOT) OD of. Traffic Flow (AADT). ODOT TransGIS. 2025-10-01. Available from: <https://gis.odot.state.or.us/transgis/>
34. Transportation (WSDOT) WSD of. Traffic Sections, Washington State Department of Transportation. Washington State Geospatial Open Data. 2024-05-16. Available from: <https://geo.wa.gov/datasets/WSDOT:wsdot-traffic-sections-aadt-current/about>
35. Csardi G and Nepusz T. The igraph software package for complex network research. *InterJournal* 2006; *Complex Systems*:1695. Available from: <https://igraph.org>
36. Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, and Müller K. igraph: Network Analysis and Visualization in R. R package version 2.1.1. 2026. DOI: 10.5281/zenodo.7682609. Available from: <https://CRAN.R-project.org/package=igraph>
37. Diestel R. Graph theory. eng. Fifth edition. Graduate Texts in Mathematics, 173. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017. Available from: <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2533092>
38. Cover T and Hart P. Nearest neighbor pattern classification. *IEEE transactions on information theory* 1967; 13:21–7
39. Hartigan JA and Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* 1979; 28:100–8
40. Ester M, Kriegel HP, Sander J, Xu X, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd*. Vol. 96. 34. 1996 :226–31
41. Lee DT and Schachter BJ. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences* 1980; 9:219–42
42. Newman EA, Feng X, Onland JD, Walker KR, Young S, Smith K, Townsend J, Damian D, and Ernst K. Defining the roles of local precipitation and anthropogenic water sources in driving the abundance of *Aedes aegypti*, an emerging disease vector in urban, arid landscapes. *Scientific Reports*. 2024 Jan 24; 14:2058. DOI: 10.1038/s41598-023-50346-3. Available from: <https://www.nature.com/articles/s41598-023-50346-3> [Accessed on: 2026 Feb 9]
43. Lubinda J, Treviño C. JA, Walsh MR, Moore AJ, Hanafi-Bojd AA, Akgun S, Zhao B, Barro AS, Begum MM, Jamal H, Angulo-Molina A, and Haque U. Environmental suitability for

- Aedes aegypti* and *Aedes albopictus* and the spatial distribution of major arboviral infections in Mexico. *Parasite Epidemiology and Control*. 2019 Aug; 6:e00116. DOI: 10.1016/j.parepi.2019.e00116. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2405673119300790> [Accessed on: 2026 May 1]
- 44.Li M, Guo R, Ding W, and Ma J. Temperature dependent developmental time for the larva stage of *Aedes aegypti*. *Mathematical Biosciences and Engineering* 2022; 19:4396–406
 - 45.De Majo MS, Zanotti G, Campos RE, and Fischer S. Effects of constant and fluctuating low temperatures on the development of *Aedes aegypti* (Diptera: Culicidae) from a temperate region. *Journal of Medical Entomology* 2019; 56:1661–8
 - 46.Buja A, Hastie T, and Tibshirani R. Linear smoothers and additive models. *The Annals of Statistics* 1989 :453–510
 - 47.Pya N and Wood SN. Shape constrained additive models. *Statistics and Computing*. 2015 May; 25:543–59. DOI: 10.1007/s11222-013-9448-7. Available from: <http://link.springer.com/10.1007/s11222-013-9448-7> [Accessed on: 2026 May 1]
 - 48.Hastie T and Tibshirani R. Generalized additive models. *Statistical science* 1986; 1:297–310
 - 49.Wood SN. Generalized additive models: an introduction with R. chapman and hall/CRC, 2017. Chap. 4-5
 - 50.Grech MG, Sartor PD, Almirón WR, and Luduena-Almeida FF. Effect of temperature on life history traits during immature development of *Aedes aegypti* and *Culex quinquefasciatus* (Diptera: Culicidae) from Córdoba city, Argentina. *Acta tropica* 2015; 146:1–6
 - 51.Couret J, Dotson E, and Benedict MQ. Temperature, larval diet, and density effects on development rate and survival of *Aedes aegypti* (Diptera: Culicidae). *PloS one* 2014; 9:e87468
 - 52.Carrington LB, Armijos MV, Lambrechts L, Barker CM, and Scott TW. Effects of fluctuating daily temperatures at critical thermal extremes on *Aedes aegypti* life-history traits. *PloS one* 2013; 8:e58824
 - 53.Velazco SJE, Rose MB, De Marco P, Regan HM, and Franklin J. How far can I extrapolate my species distribution model? Exploring shape, a novel method. *Ecography*. 2024 Mar;

- 2024:e06992. DOI: 10.1111/ecog.06992. Available from: <https://nsojournals.onlinelibrary.wiley.com/doi/10.1111/ecog.06992> [Accessed on: 2026 May 1]
- 54.Fawcett T. An introduction to ROC analysis. *Pattern recognition letters* 2006; 27:861–74
- 55.Glenn WB et al. Verification of forecasts expressed in terms of probability. *Monthly weather review* 1950; 78:1–3
- 56.Akaike H. Factor analysis and AIC. *Psychometrika* 1987; 52:317–32
- 57.Arellano CJ, Rouzegari N, Dao V, Zadeh MB, Shearer EJ, Gorrooh VA, Nguyen P, Hsu KL, Sorooshiana S, et al. Beyond expectations: investigating anomalous 2022–23 winter weather conditions and water resources impacts in California. *American Meteorological Society* 2026
- 58.Myer MH, Fizer CM, Mcpherson KR, Neale AC, Pilant AN, Rodriguez A, Whung PY, and Johnston JM. Mapping *Aedes aegypti* (Diptera: Culicidae) and *Aedes albopictus* vector mosquito distribution in Brownsville, TX. *Journal of medical entomology* 2020; 57:231–40
- 59.O'Dell N, Bolling BG, Dacko N, Carr JT, Hambrick B, Chaves LF, and McMillan JR. Identifying environmental drivers of *Aedes aegypti* and *Aedes albopictus* abundance in the Dallas-Fort Worth metroplex using Random Forest modeling. *Journal of Medical Entomology* 2025; 62:789–99
- 60.Wilke ABB, Vasquez C, Carvajal A, Medina J, Chase C, Cardenas G, Mutebi JP, Petrie WD, and Beier JC. Proliferation of *Aedes aegypti* in urban environments mediated by the availability of key aquatic habitats. *Scientific reports* 2020; 10:12925
- 61.District JCVC. Public Health Advisory: *Aedes aegypti* Mosquito. 2025 Aug. Available from: <https://jcvcd.gov/mosquito/aedes-aegypti/>
- 62.Capinha C, Rocha J, and Sousa CA. Macroclimate determines the global range limit of *Aedes aegypti*. *EcoHealth* 2014; 11:420–8
- 63.Campbell LP, Luther C, Moo-Llanes D, Ramsey JM, Danis-Lozano R, and Peterson AT. Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2015; 370
- 64.Holeva-Eklund WM, Young SJ, Will J, Busser N, Townsend J, and Hepp CM. Species distribution modeling of *Aedes aegypti* in Maricopa County, Arizona from 2014 to 2020. *Frontiers in Environmental Science* 2022; 10:1001190

65. Padmanabha H, Durham D, Correa F, Diuk-Wasser M, and Galvani A. The interactive roles of *Aedes aegypti* super-production and human density in dengue transmission. *PLOS Neglected Tropical Diseases* 2012
66. Fink T, Hau B, Baird B, Palmer S, Kaplan S, Ramberg F, Mead D, and Hagedorn H. *Aedes aegypti* in Tucson, Arizona. *Emerging Infectious Diseases* 1998; 4:703
67. Lozano S, Raman V, and Connelly CR. The expansion of the yellow fever mosquito (*Aedes aegypti*) and evidence of its establishment in the Las Vegas metropolitan area, Nevada, United States of America. *Journal of medical entomology* 2025; 62:1611–5

Figures and tables

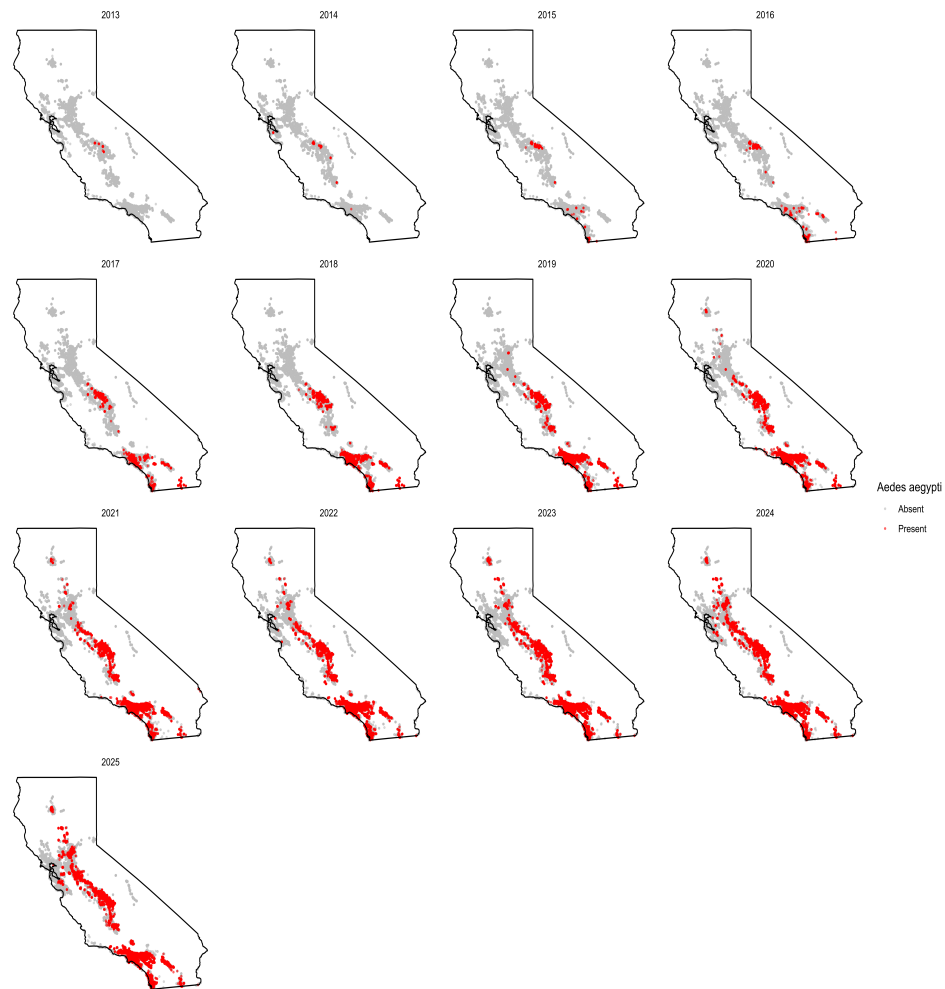


Figure 1: *Aedes aegypti* presence and absence in California from 2013 to 2025. White regions are areas where no sampling occurred, while point indicate trap locations with *Ae. aegypti* presences (red) or absences (grey).

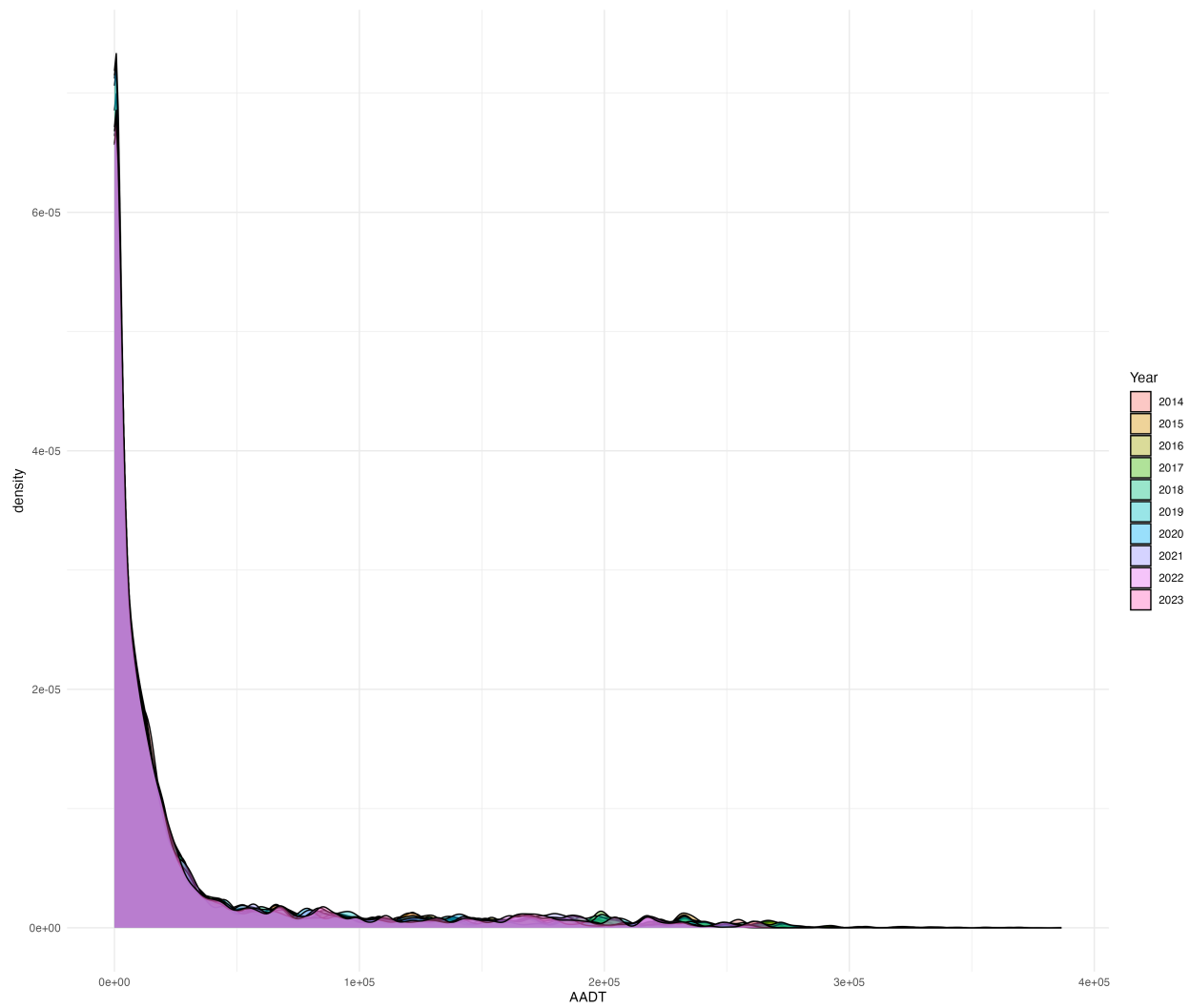


Figure 3: **Distribution of average annual daily traffic (AADT) for road segments in the California State Highway system.** Overlapping distributions show the values for each year. Values are scaled by $1e^6$

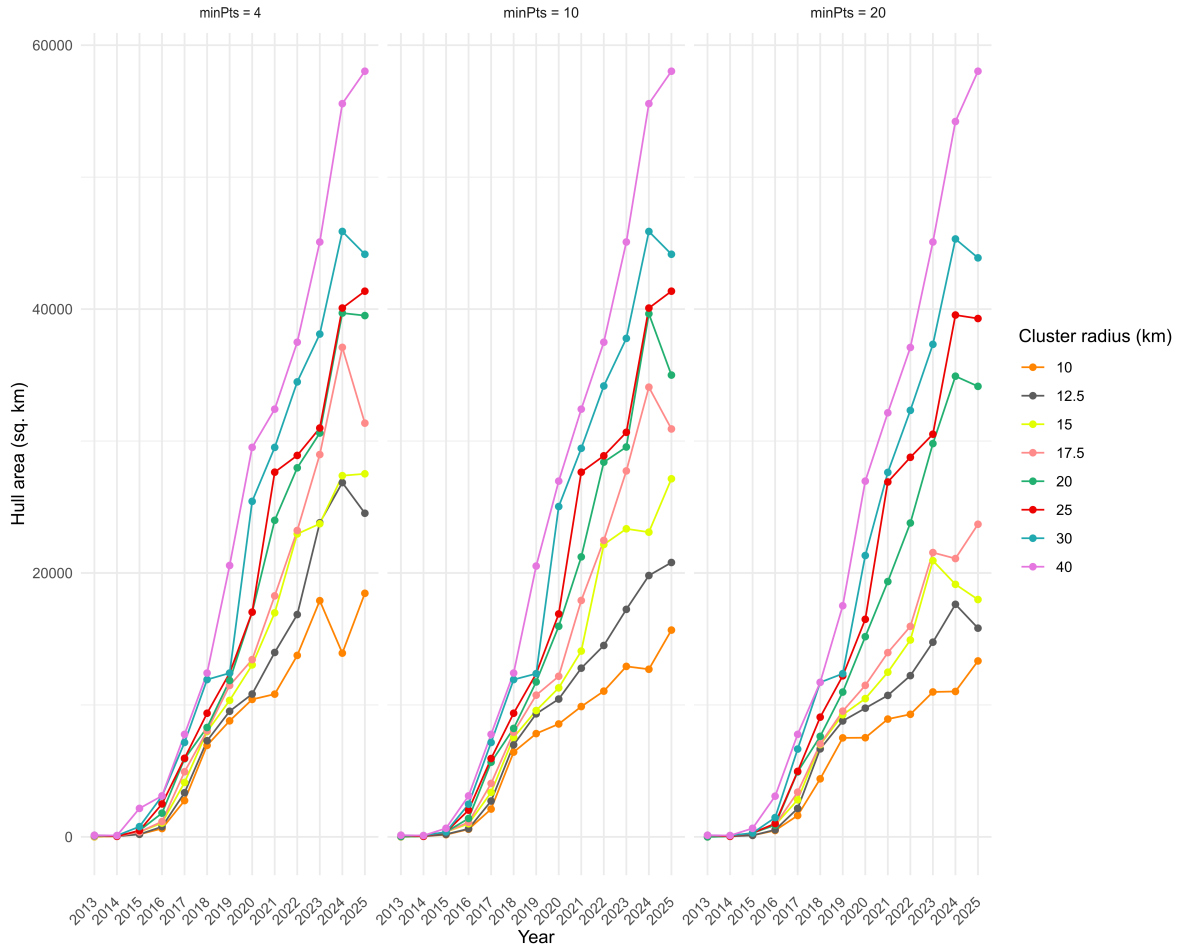


Figure 4: **Sensitivity analysis results for DBSCAN $minPts$ and ϵ parameters.** Individual panels correspond to tested $minPts$ values (4, 10, and 20 points) and colors represent unique ϵ values.

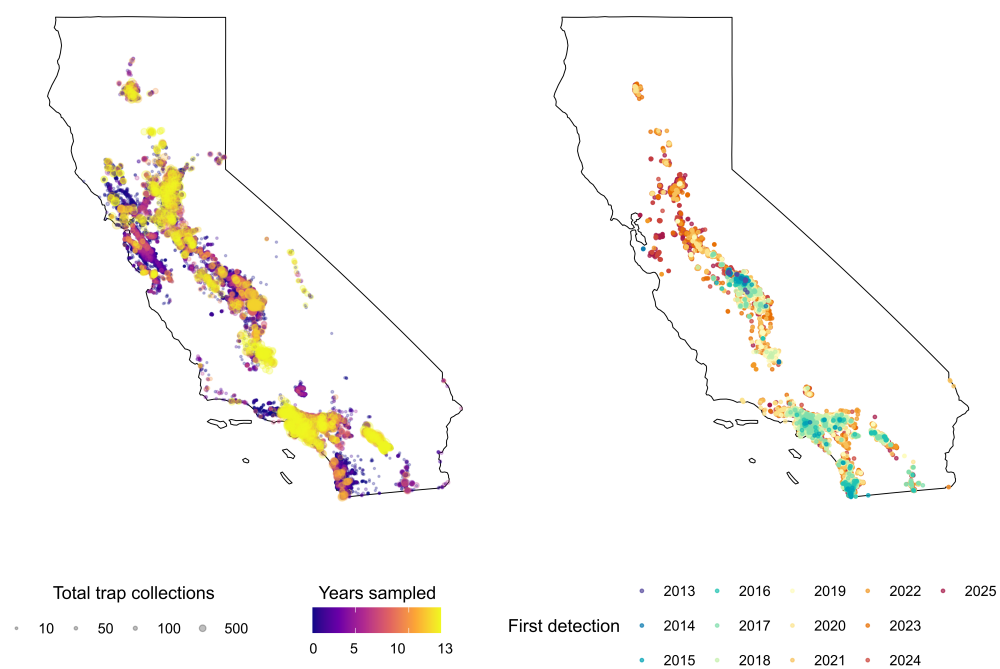


Figure 5: **Trapping effort in California from 2013-2025.** “First detection” corresponds to the year that *Aedes aegypti* was first detected at a trap location.

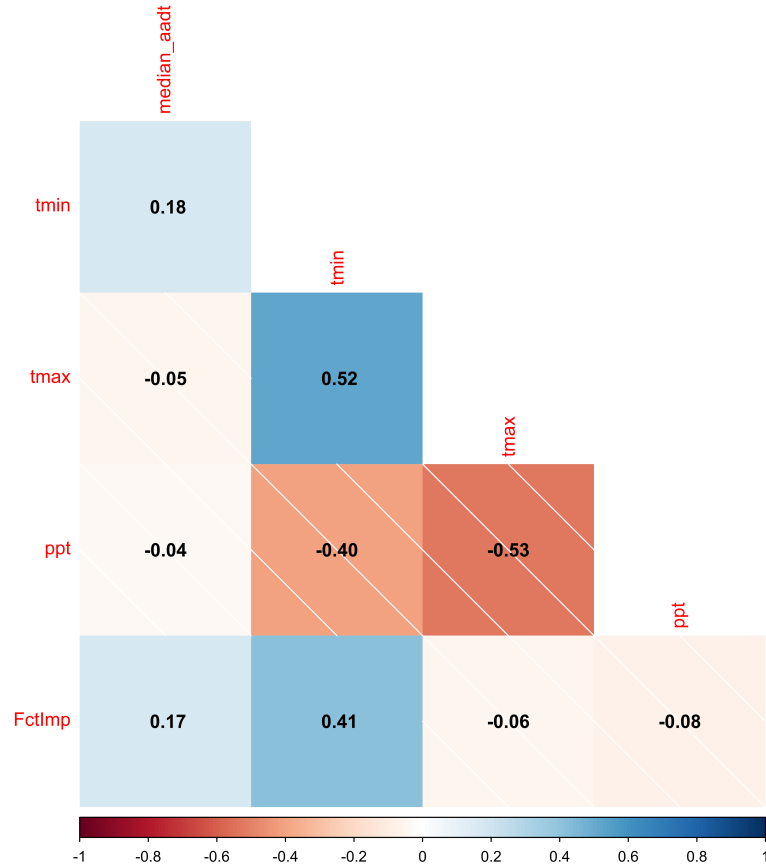


Figure 6: **Collinearity tests for model predictors.** Correlation matrix plot for continuous predictors in the “preferred” model (*median_aadt* = *aegypti*-sourced AADT; *tmin* = annual average minimum temperature; *tmax* = annual average maximum temperature; *ppt* = annual cumulative precipitation; *FctImp* = fractional impervious surface).

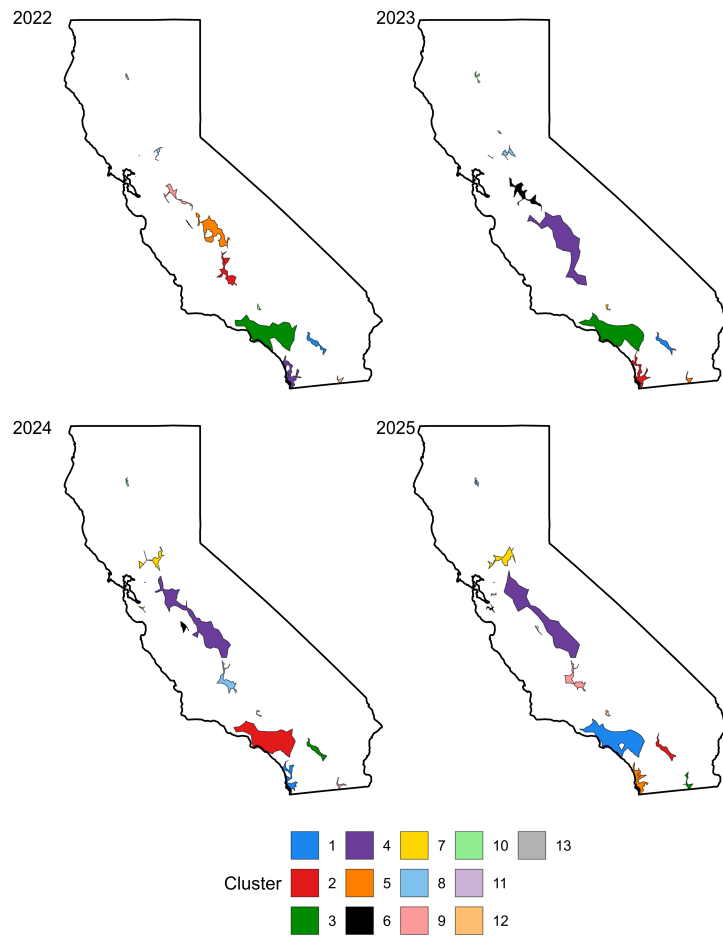


Figure 7: **Spatial clusters and convex hulls over time.** Spatial cluster formations for **A)** 2022 through 2025 *Aedes aegypti* presence. Colors indicate cluster assignment.

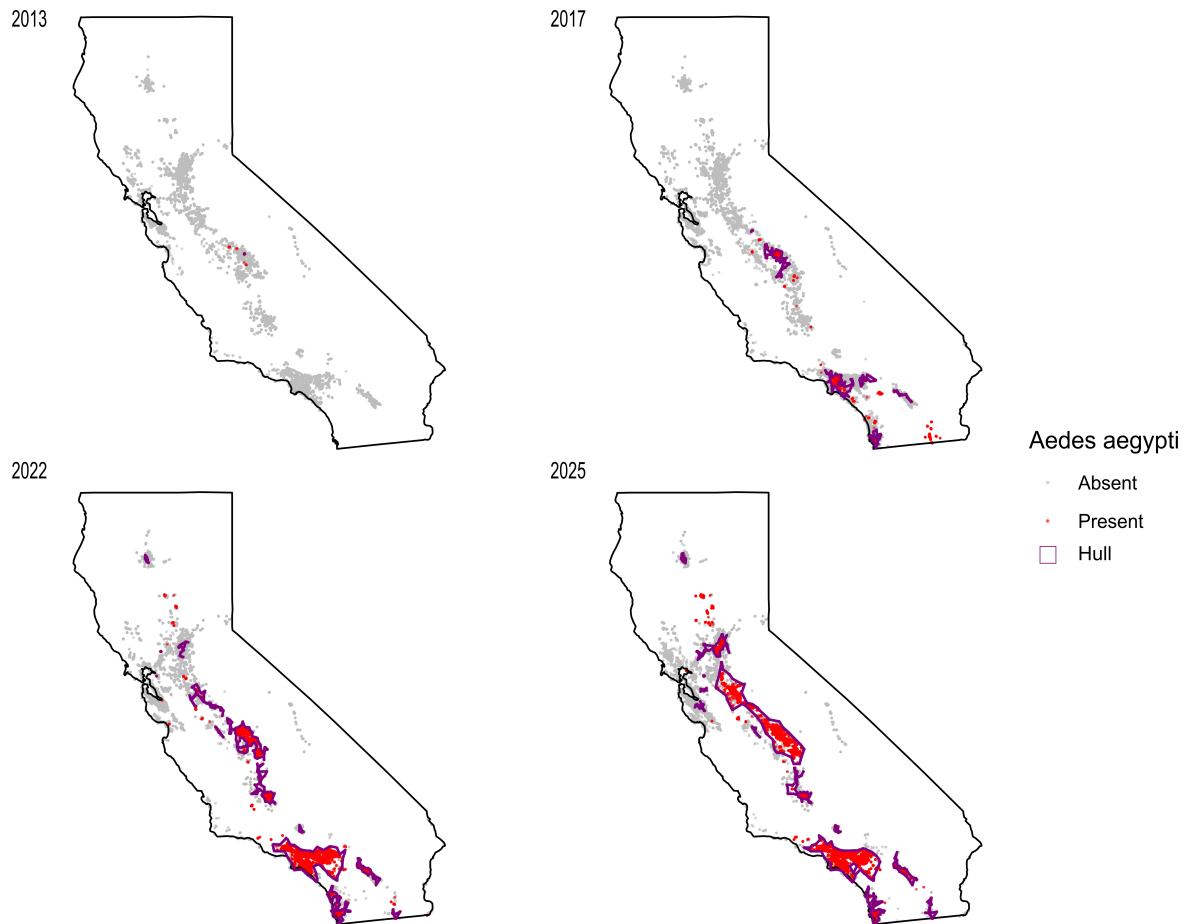


Figure 8: **Range shift estimates.** Fitted hulls for 2013, 2017, 2022, and 2025 *Aedes aegypti* presence using a minimum points parameter of 20 and cluster radius of 17.5 km.

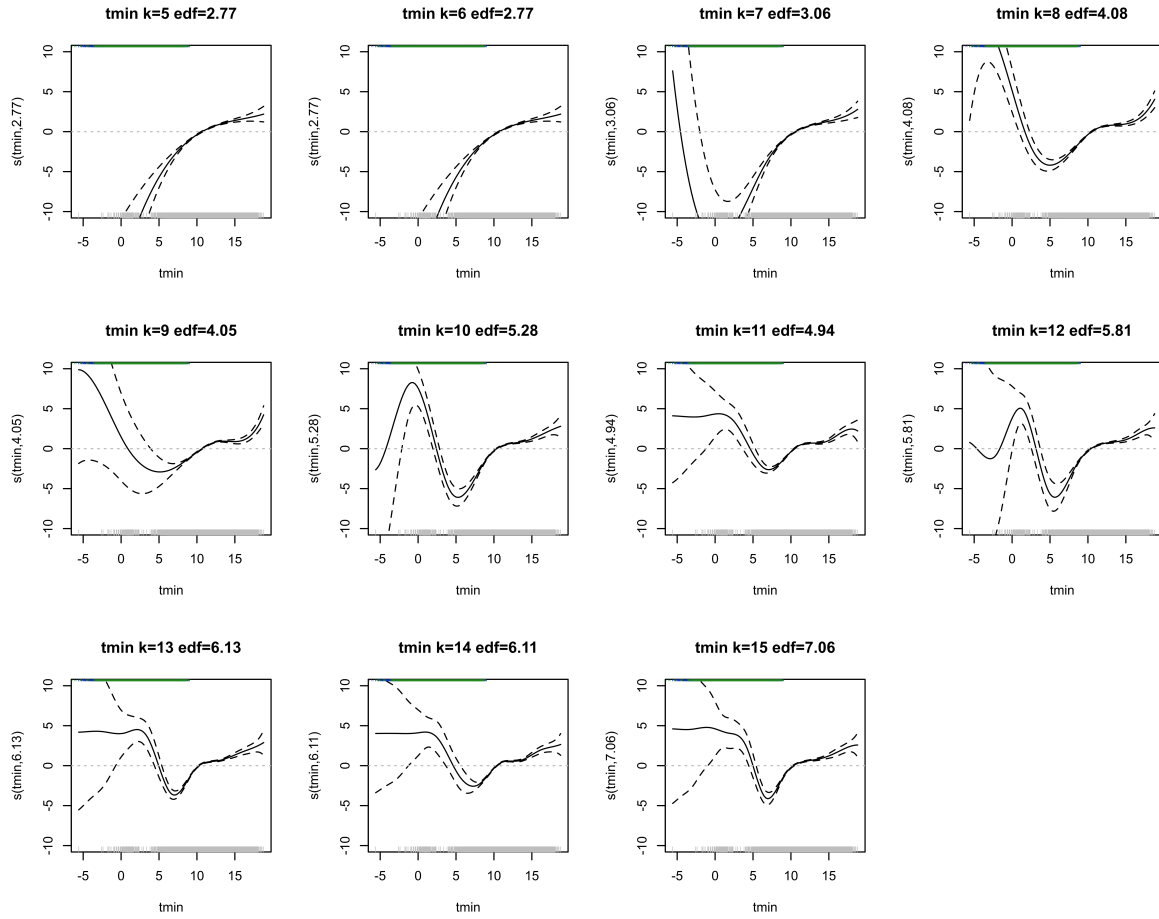


Figure 9: **Minimum temperature k-basis testing.** Partial effect plots of minimum temperatures (“ t_{min} ”) on *Aedes aegypti* occurrence in California, faceted by basis dimension (k) values. The gray-colored rug plot corresponds to the distribution of the variable for California, while green and blue are for Oregon and Washington state, respectively. edf = effective degrees of freedom

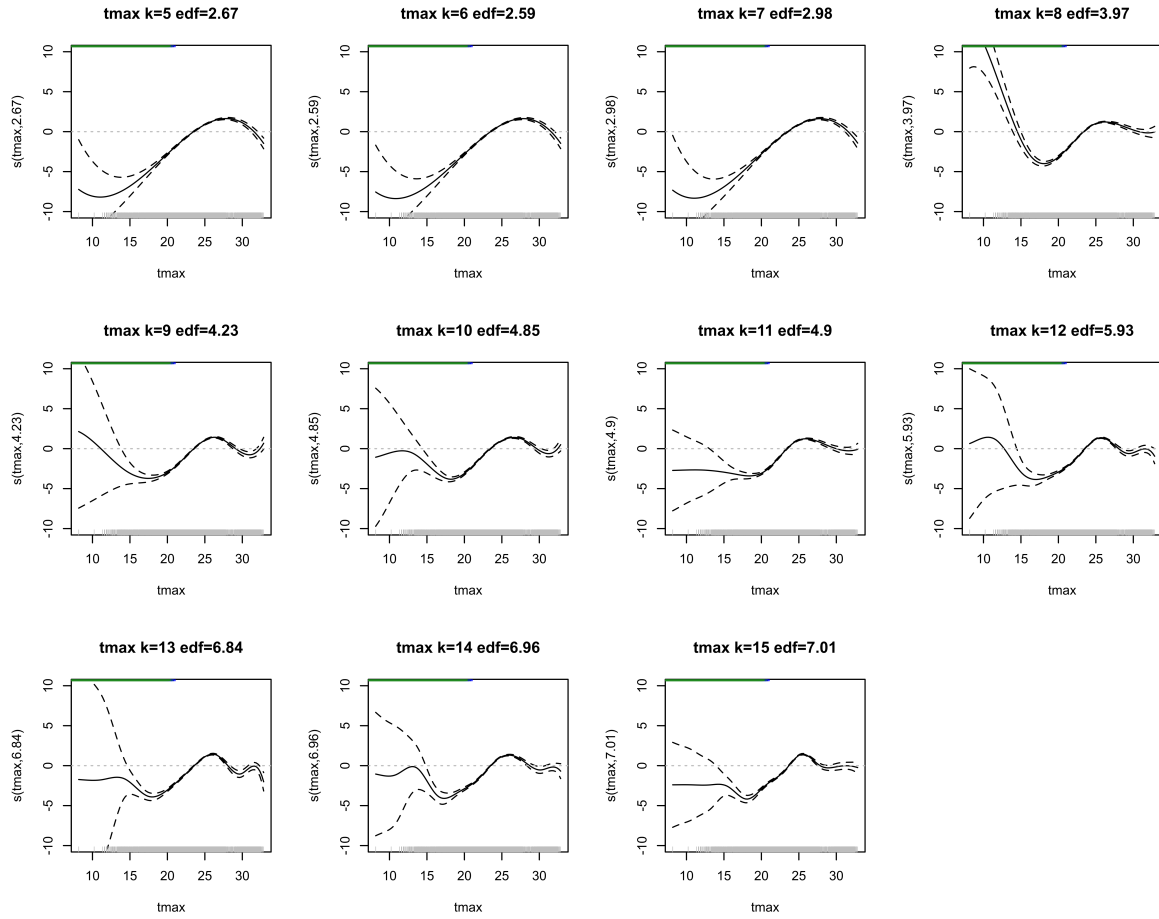


Figure 10: **Maximum temperature k-basis testing.** Partial effect plots of maximum temperatures (“ t_{max} ”) on *Aedes aegypti* occurrence in southern California, faceted by basis dimension (k) values. The gray-colored rug plot corresponds to the distribution of the variable for California, while green and blue are for Oregon and Washington state, respectively.

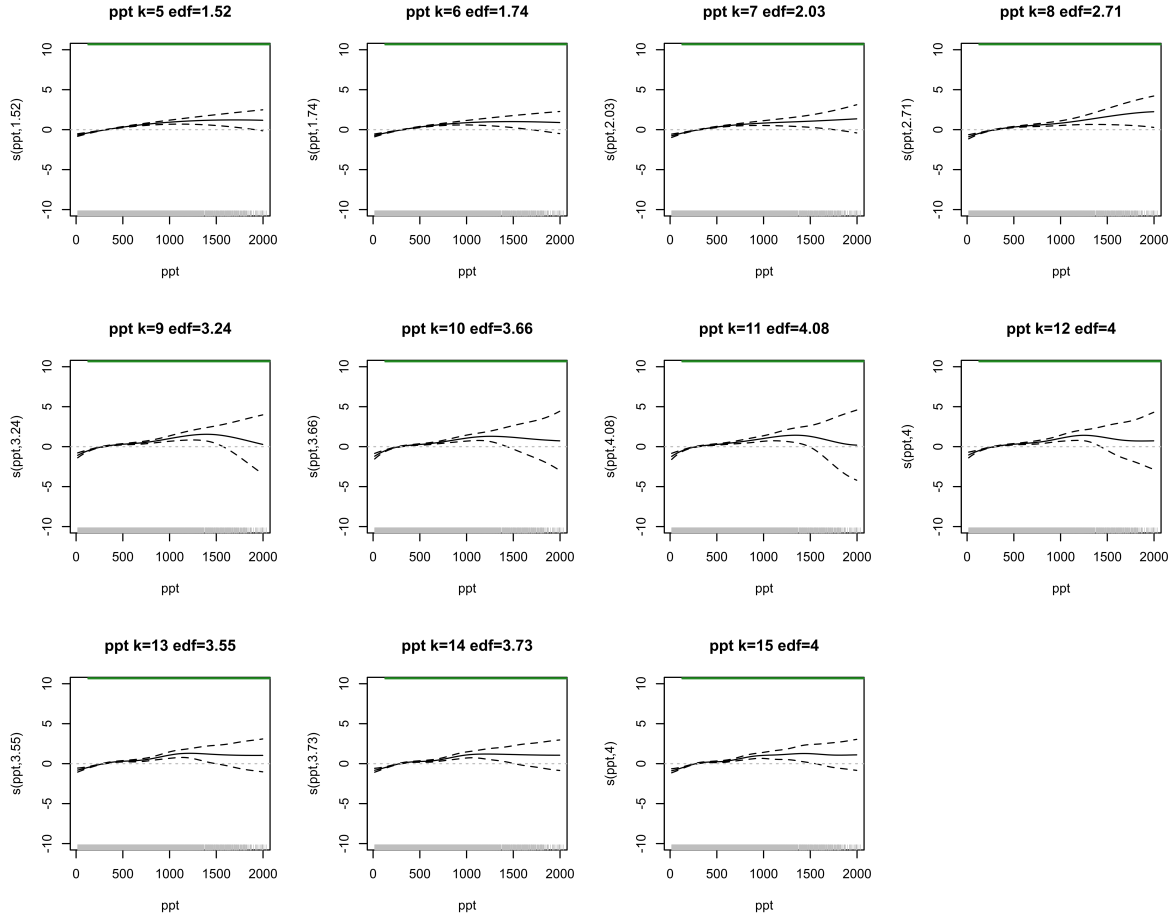


Figure 11: **Precipitation k-basis testing.** Partial effect plots of cumulative precipitation (“ppt”) on *Aedes aegypti* occurrence in southern California, faceted by basis dimension (k) values. The gray-colored rug plot corresponds to the distribution of the variable for California, while green and blue are for Oregon and Washington state, respectively.

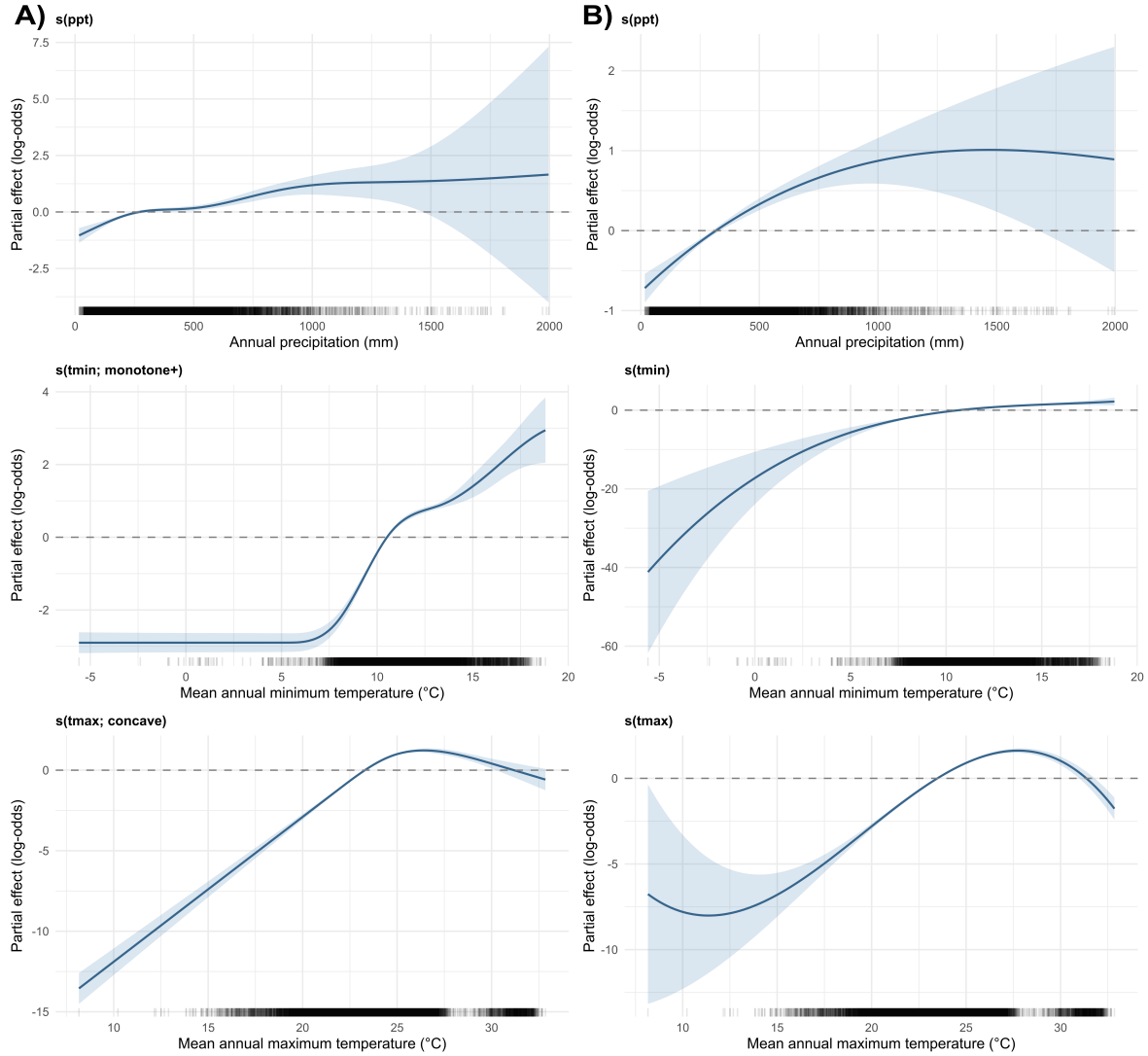


Figure 12: **Model partial effect plots compared.** Comparison of **A)** constrained and **B)** unconstrained smooth terms. Blue envelopes represent standard errors. Rug plots at the bottom show the distribution of values for California.

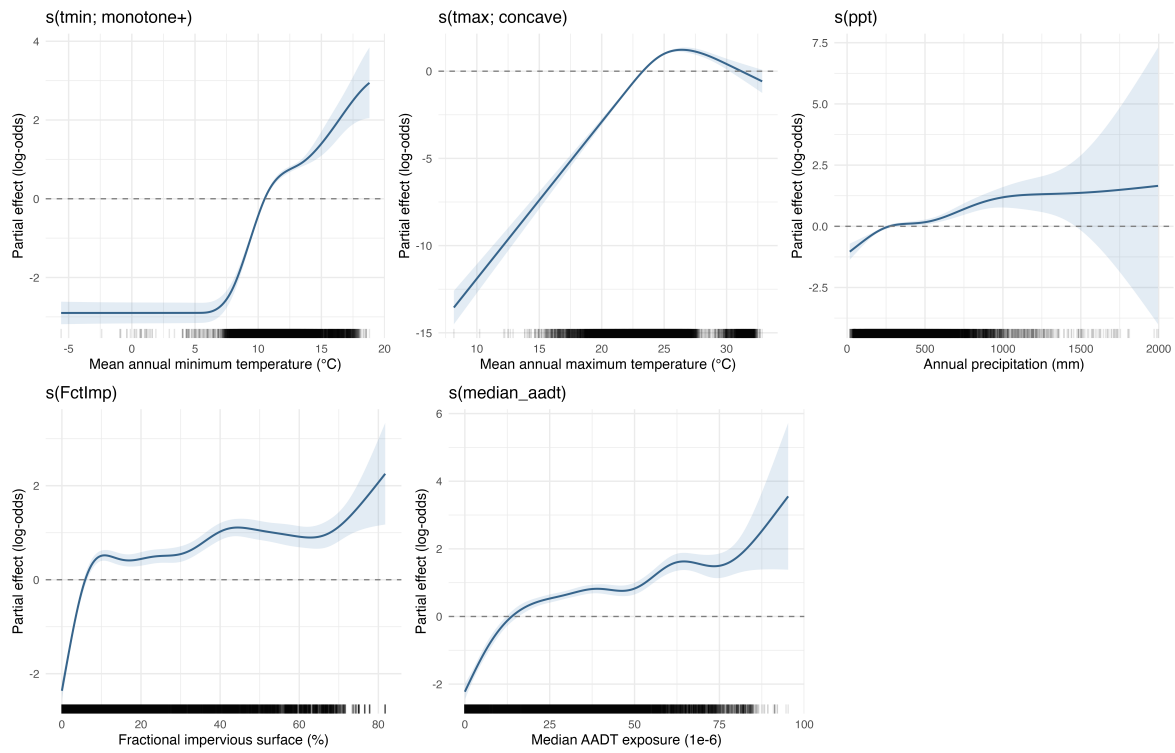


Figure 13: **Model partial effect plots.** “Preferred” model partial effects plots, for each variable fit with a smooth function. Blue envelopes represent standard errors. Rug plots at the bottom show the distribution of values for California.

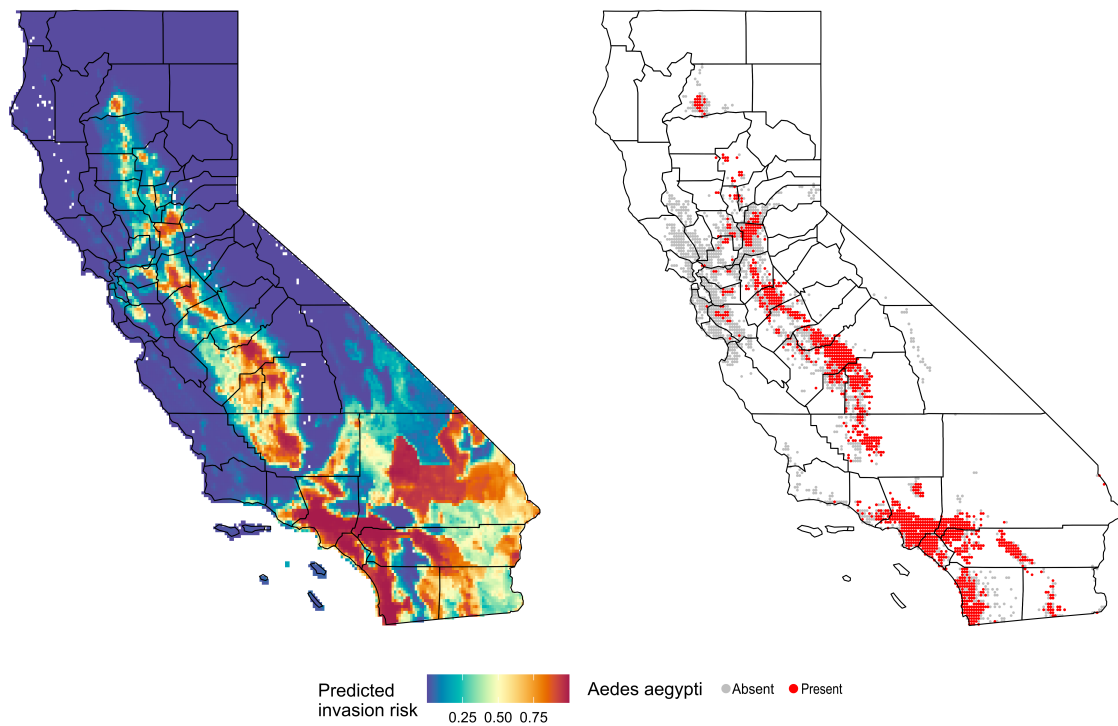


Figure 14: **California *Aedes aegypti* invasion risk.** Predicted current invasion risk in California (left panel) and 2025 records of *Aedes aegypti* occurrence in CalSurv data (right panel). Occurrence and absence records for 2025 are shown on the 5km prediction grid, where a cell is defined as having *Ae. aegypti* present if any trap in that cell-year found *Ae. aegypti*.

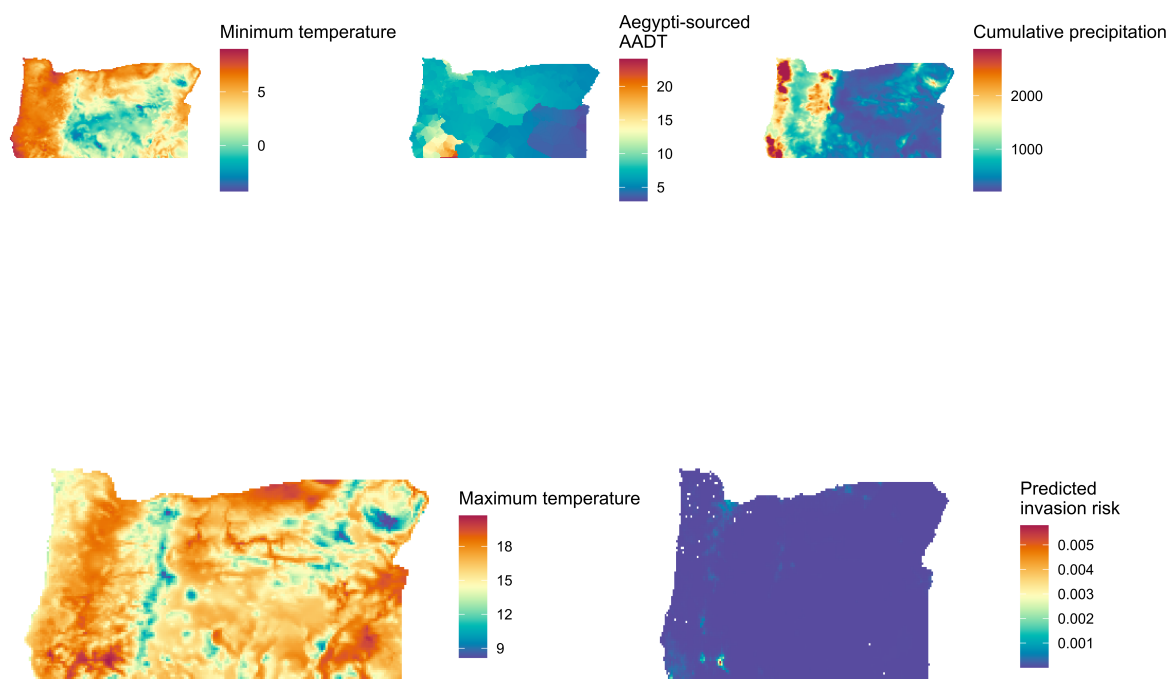


Figure 15: **Oregon *Aedes aegypti* invasion risk.** Panels 1-4 correspond to the 5-year average for minimum temperature, *aegypti*-sourced AADT, cumulative precipitation, and maximum temperature. Panel 5 is the final predicted invasion risk.

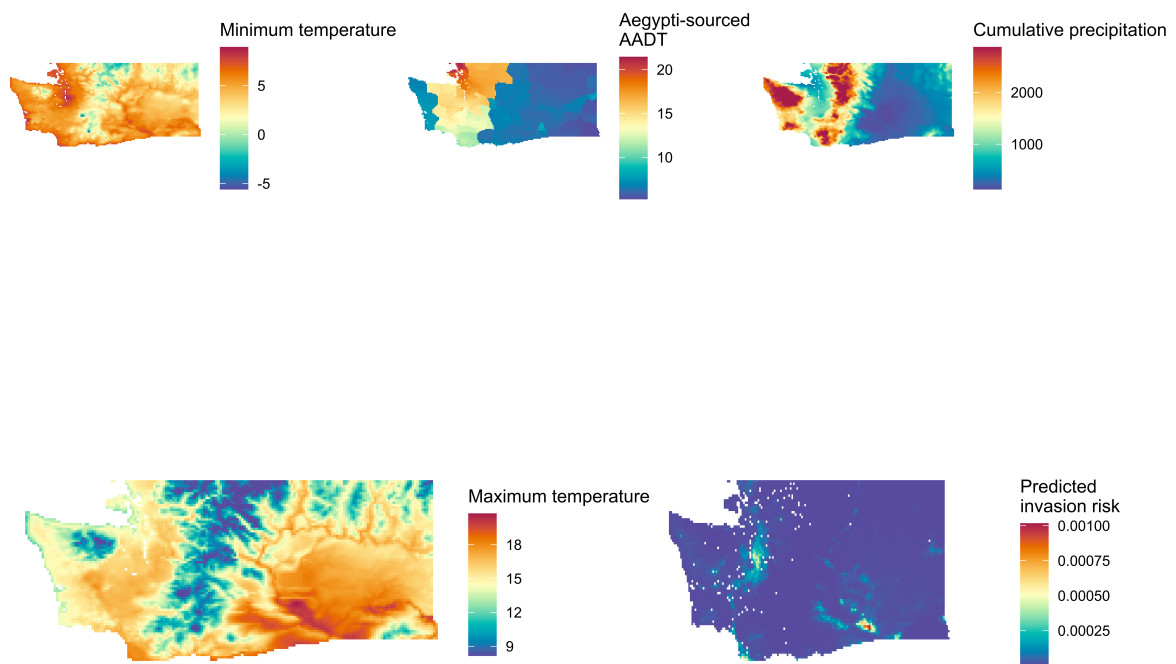


Figure 16: **Washington state *Aedes aegypti* invasion risk.** Panels 1-4 correspond to the 5-year average for minimum temperature, *aegypti*-sourced AADT, cumulative precipitation, and maximum temperature. Panel 5 is the final predicted invasion risk.

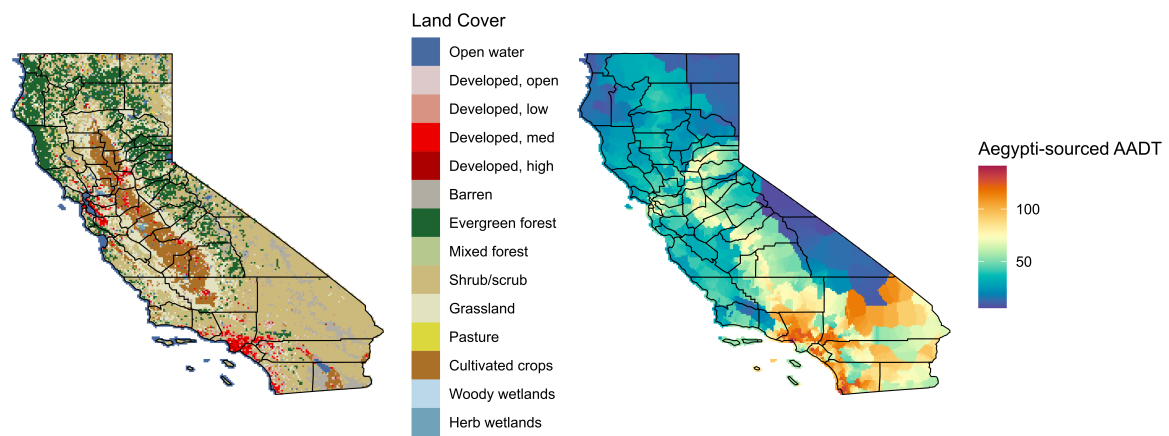


Figure 17: **California landscape.** Land cover types and *aegypti*-sourced AADT values throughout California as of 2025.

Table 1: Two-tailed Wilcoxon Rank Sum Test results for median *aegypti*-sourced AADT values from 2014-2025.

Year	Wilcoxon-statistic	p^1
2014	11,066.0	3.97e^{-01}
2015	45,721.0	1.16e^{-04}
2016	107,676.5	9.20e^{-14}
2017	242,692.0	2.28e^{-23}
2018	378,948.0	2.28e^{-23}
2019	469,303.5	7.19e^{-53}
2020	502,228.0	1.01e^{-59}
2021	552,188.5	6.61e^{-58}
2022	581,991.0	1.88e^{-66}
2023	720,278.0	4.06e^{-86}
2024	676,463.0	2.46e^{-81}
2025	732,828.5	6.18e^{-102}

¹ Based off a two-sided test comparing the median *aegypti*-sourced AADT value in occupied and unoccupied sites for a given year.

Table 2: Annual estimated land area occupied by *Aedes aegypti* using concave hulls fitted to DBSCAN clusters.

Year	Hull area (sq. km)	Δ Area ¹	# of Clusters
2013	0.98	–	1
2014	43.51	42.54	2
2015	292.58	249.07	1
2016	897.46	604.88	4
2017	3,393.33	2,495.87	6
2018	7,042.99	3,649.66	7
2019	9,529.50	2,486.50	9
2020	11,485.47	1,955.98	8
2021	13,966.45	2,480.97	11
2022	15,955.64	1,989.19	13
2023	21,548.11	5,592.48	11
2024	21,095.88	-452.23	12
2025	23,694.84	2,598.96	13

¹Change in cumulative area of concave hulls from previous year.

Table 3: Out-of-sample predictive performance on northern California holdout data. The constrained model used monotone increasing P-splines for minimum temperature and concave P-splines for maximum temperature.

Model	AUC (95% CI)	Brier score (95% CI ¹)
Constrained	0.935 (0.925–0.945)	0.091 (0.086 - 0.096)
Unconstrained	0.939 (0.930–0.949)	0.090 (0.085 - 0.094)

¹1000 sample bootstrap.

Table 4: Nested model comparison on out of sample California data. The full model includes all environmental covariates and traffic flow; the environment-only model excludes *Ae. aegypti*-sourced AADT; the traffic-only model includes only AADT; the null model is intercept-only.

Model	AIC	AUC	Brier score (95% CI)	Δ AIC
Full	12,041.97	0.935	0.091 (0.086 - 0.096)	—
Environmental	12,617.59	0.943	0.099 (0.094 - 0.105)	575.63
Traffic	16,558.96	0.699	0.111 (0.106 - 0.116)	4,517.00
Null	21,547.61	0.500	0.148 (0.146 - 0.151)	9,505.65