

Appendix to A Data-Parsimonious Model for Long-Term Risk Assessments of West Nile Virus Spillover

Saman Hosseini^{1,*}, Lee W. Cohnstaedt², Matin Marjani¹, and Caterina Scoglio¹

¹Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS, USA

²Foreign Arthropod-Borne Animal Diseases Research Unit, USDA-ARS, Manhattan, KS, USA

*Shosseini@ksu.edu

Contents

-A	Study location	1
-A1	Orange County, California	1
-A2	Riverside County, California	2
-A3	Los Angeles County, California	2
-A4	Dallas County, Texas	2
-A5	Harris County, Texas	2
-A6	Duval County, Florida	2
-B	Compartmental model for WNV	3
-B1	Initial conditions of compartmental model	4
-B2	Coefficients and rates	4
-C	Predictive model of spillover onset	6
-D	Eco-Epi Poisson model of spillover severity	6
-E	Logarithmic Score	8
-F	Köppen climate classes	8
-G	Leave-one-year-out cross validation and retrospective forecast for spillover onset prediction	10
-G1	Leave-one-year-out cross validation	10
-G2	Retrospective forecast	10
-H	Leave-one-year-out cross validation and retrospective forecast for spillover severity prediction	10
-H1	Leave-one-year-out cross validation	10
-H2	Retrospective forecast	10
-I	Kernel Density Estimation	10
-I1	Scott's Rule of Thumb	11
-I2	Two-Dimensional KDE	12
-I3	Kernel Density Estimation for Surfaces	12
-J	Discussion	13

A. Study location

1) Orange County, California

Orange County, California, spans 792.9 square miles of land and is the 47th largest county in the state by area. The northwest of Orange County lies within the coastal plain of the Los Angeles Basin, while its southeast stretches to the lower elevations of the Santa Ana Mountains [1]. Most Orange County residents are concentrated in two broad coastal valleys in the basin: the Santa Ana Valley and the Saddleback Valley. Due to its compact size and the moderate influence of the nearby Pacific, Orange County and its entire land area fall within a single warm summer Mediterranean zone (Csa / Csb) on the Köppen map [2, 3]. Orange County has been recognized as a hotspot for WNV in California. The outbreak season usually begins in the summer and extends into the fall. The northern parts of the county experience higher WNV activity and report more cases. In 2014, Orange County had a record of 287 cases (198 neuroinvasive, 66 fever, 14 asymptomatic), and 8 deaths [4].

2) *Riverside County, California*

Riverside County spans from the suburbs of Los Angeles to the Arizona border. It includes diverse regions such as the Coachella Valley, which encompasses Palm Springs and other desert resort cities. The county spans both mountainous and desert terrain, with urban centers such as Riverside and Indio. It borders San Bernardino, Orange, San Diego, and Imperial counties and the state of Arizona to the east [5]. Riverside County has experienced a substantial number of cases of WNV, with the most severe outbreak occurring in 2015, during which 127 human cases were reported [6]. Riverside County traverses four Köppen zones: pockets of the warm summer Mediterranean (Csb) and hot summer Mediterranean (Csa) cling to its western foothills, a broad swath of hot semi-arid (BSh) covers the inland Perris–San Jacinto corridor, and the eastern two-thirds of the county, from the Coachella Valley to the Arizona border, lies in the stark hot desert regime (BWh). This sharp west-to-east transition produces some of the greatest intra-county temperature and moisture contrasts in southern California [2, 3, 7].

The marked severity of the activity of WNV in Riverside County is plausibly linked to ecological conditions in the Coachella Valley. Beyond the advantages that mild winters provide mosquito populations, periods of elevated temperature and drought intensify transmission dynamics in at least two ways. First, water scarcity drives both mosquitoes and avian hosts to converge at anthropogenic water sources, increasing their spatial overlap and thus elevating vector-host contact rates. Second, environmental stressors increase the susceptibility of juvenile birds, promoting greater viral amplification within avian communities. Together, these processes enhance viral circulation and substantially increase the risk of spillover to the human population [8–10].

A concise explanatory table of the Köppen-Geiger climate classification system is provided in Appendix F.

3) *Los Angeles County, California*

Los Angeles County lies along the Pacific coast of the United States. It is bordered by Ventura, Kern, San Bernardino, and Orange counties. The county includes coastal plains, the Los Angeles Basin, and mountain ranges. It is the most populous county in the United States [11]. The peak mosquito season in LA County usually lasts from June to November, and the county consistently reports a high incidence of human cases. The mean of the last 5 years shows a peak in September [12]. Los Angeles had the record for the highest number of WNV infections in both mosquitoes and humans [13]. Los Angeles County straddles at least four Köppen zones: the cool-summer Mediterranean (Csb) along the coast, the warmer Mediterranean (Csa) across the inland basin, hot semi-arid pockets (BSh) in the San Fernando and San Gabriel Valleys, and the cold-desert terrain (BWk) in the Antelope Valley [2, 3, 7].

4) *Dallas County, Texas*

Dallas County covers 873.2 square miles, ranking as the 177th largest county in the state by area. It shares boundaries with Kaufman, Ellis, Collin, Denton, Rockwall, and Tarrant counties [14]. Dallas County has repeatedly reported human cases of WNV and mosquito-positive pools, with some seasons marked by severe outbreaks. The 2012 epidemic was particularly significant, resulting in 173 neuroinvasive cases, 225 cases of West Nile fever, and 19 associated deaths [15]. More recently, in 2024 the county documented 37 human cases of WNV, including three fatalities, while mosquito surveillance identified 98 WNV-positive pools [16]. Dallas County has a humid subtropical climate (Cfa), with hot and humid summers, mild winters, and peak rainfall in spring and early fall [2, 3].

5) *Harris County, Texas*

Harris County covers 1,707.3 square miles, ranking as the 14th largest county in the state. It is bordered by Chambers, Galveston, Brazoria, Waller, Liberty, Montgomery, and Fort Bend counties [17]. This county falls into the humid subtropical climate zone (Cfa) under Köppen-Geiger. Summers are long, hot, and very humid due to the strong influence of the Gulf of Mexico, while winters are mild with rare freezes. Rainfall is high and fairly evenly distributed, with frequent sea-breeze thunderstorms in summer and occasional tropical cyclones [2, 3, 18]. The first human cases of WNV in Harris County were documented in 2002, signaling the establishment of the virus in the region. Since then, WNV activity has been detected annually, reflecting its endemic status within the county [19]. The most substantial outbreak occurred in 2014, during which 139 human cases, including two fatalities, were reported, highlighting the continuing public health importance of WNV transmission in Harris County [20].

6) *Duval County, Florida*

Duval County spans 762.6 square miles, making it the 28th largest county in the state by area, and shares borders with Baker, Clay, St. Johns and Nassau counties [21]. Located in northeast Florida along the Atlantic coast, it is coterminous with the consolidated city of Jacksonville and is bisected by the St. Johns River and its estuary. The low-lying coastal plain—with tidal marshes, riverine corridors, and urban–suburban neighborhoods—creates diverse mosquito habitats relevant to arboviral transmission. The climate is humid subtropical (Cfa) in the Köppen–Geiger classification, characterized by long, warm summers

and mild winters with a late-summer rainfall maximum [2, 3]. These conditions support an extended mosquito season and summer–fall West Nile virus activity.

B. Compartmental model for WNV

We have used a $S_H E_H I_H R_H - Egg_M A_M S_M E_M I_M - S_B E_B I_B R_B$ compartmental model to explain the mechanism of circulation of WNV between human, birds, and mosquito populations. In our model, the subscript H denotes the human population, B designates birds, and M refers to mosquitoes; within the mosquito classes, Egg_M represents the oviposition (egg) compartment, while A_M corresponds to the aquatic developmental stage, encompassing both larval and pupal phases. S, E, I , and R are compartments of susceptible, exposed, infectious (infected for human population) and recovered. We selected American Crow (*Corvus brachyrhynchos*) as the representative avian host and *Culex tarsalis* as the main mosquito vector. Our model has the following compartments.

Compartments of the Human Population:

- Susceptible
- Exposed
- Infected
- Recovered

Compartments of the Mosquito Population:

- Oviposition
- Aquatic
- Mosquitoes:
 - Susceptible Mosquitoes
 - Exposed Mosquitoes
 - Infectious Mosquitoes

Compartments of the Bird Population:

- Birds:
 - Susceptible Birds
 - Exposed Birds
 - Infectious Birds
 - Recovered Birds

Based on these compartments, we formulate the following system of differential equations. It was assumed that during the analysis window of the outbreak season, the demographic change is negligible. We have fixed population sizes for humans and birds during an outbreak season, while treating the mosquito population size as dynamically variable. In addition, mosquito population is assumed to be female and mosquitoes bite hosts in proportion to host abundance.

Model:

$$\dot{S}_H = \mu_H (S_H + E_H + I_H + R_H) - \beta_{M \rightarrow H} I_M S_H - \mu_H S_H$$

$$\dot{E}_H = \beta_{M \rightarrow H} I_M S_H - \delta_H E_H - \mu_H E_H$$

$$\dot{I}_H = \delta_H E_H - (\lambda_H + \mu_{WNV_H} + \mu_H) I_H$$

$$\dot{R}_H = \lambda_H I_H - \mu_H R_H$$

$$\dot{Egg}_M = \frac{EFGC}{LGC} (S_M + E_M + I_M) - \rho_E Egg_M - \mu_E Egg_M$$

$$\dot{A}_M = \rho_E Egg_M (1 - \frac{A_M}{K}) - \rho_A A_M - \mu_A A_M$$

$$\dot{S}_M = \rho_A A_M - \beta_{B \rightarrow M} I_B S_M - \mu_M S_M$$

$$\dot{E}_M = \beta_{B \rightarrow M} I_B S_M - \delta_M E_M - \mu_M E_M$$

$$\dot{I}_M = \delta_M E_M - \mu_M I_M$$

$$\dot{S}_B = \mu_B (S_B + E_B + I_B + R_B) - \beta_{M \rightarrow B} I_M S_B - \mu_B S_B$$

$$\dot{E}_B = \beta_{M \rightarrow B} I_M S_B - \delta_B E_B - \mu_B E_B$$

$$\dot{I}_B = \delta_B E_B - \lambda_B I_B - \mu_{WNV_B} I_B - \mu_B I_B$$

$$\dot{R}_B = \lambda_B I_B - \mu_B R_B$$

1) Initial conditions of compartmental model

The compartmental model was initialized with a constant human population of 300,000 individuals, all assumed to be susceptible ($E_H(0) = I_H(0) = R_H(0) = 0$). The mosquito population was seeded with one for the ovipositing compartment, one individual in the aquatic phase, 100 susceptible adults, a single exposed adult and a single infectious adult. The avian reservoir consisted of 40,000 birds, of which one was exposed, one was infectious, and one had recovered.

2) Coefficients and rates

In this model δ_H is the incubation rate, representing the rate with which individuals in the human population transit from the exposed phase to the infected compartment. In addition λ_H is the recovery rate of the human population, which is the rate of moving from the infected to the recovered compartment. Furthermore, μ_H , μ_E , μ_A , μ_M , and μ_B , denote the natural mortality rates for humans, the oviposition phase, the aquatic phase, mosquitoes and birds. In addition, μ_{WND_H} , μ_{WND_B} , denote the mortality rates caused by WND in the human and bird compartment. $\frac{EFGC}{LGC}$ represents the eggs per female per gonotrophic cycle (fecundity), indicating the number of eggs laid by each female mosquito in a single reproductive cycle. Also, ρ_E and ρ_A are the hatching rate in the oviposition phase and the larval development rate, respectively. δ_M represents the extrinsic incubation rate of the pathogen within the mosquito population. As is evident, the model includes the carrying capacity K for the aquatic phase of the mosquito population. We estimated K with a Kalman filter that assimilates observed human counts, so K is identified from human incidence data rather than entomological observations [22].

In our model, $\beta_{M \rightarrow H}$, $\beta_{M \rightarrow B}$ and $\beta_{B \rightarrow M}$ are transmission rates from mosquitoes to humans and birds and from birds to mosquitoes, respectively. The general form of the transmission rate coefficient is $\beta = \frac{BR \cdot p \cdot TE}{N_{host}}$, where BR is the biting rate, $p = p_M$ or p_H denotes the proportion of available hosts and TE is the transmission probability per bit (transmission efficacy) to humans, birds, or mosquitoes.

To reduce the complexity of the model, we assume that the transmission efficiencies of mosquito-to-human, mosquito-to-bird, and bird-to-mosquito transmission are identical, i.e. $TE_{MH} = TE_{BM} = TE_{MB}$.

So:

$$\beta_{M \rightarrow H} = ap_H b_{MH} \frac{1}{N_H} = BR \cdot \frac{N_H}{N_H + N_B} \cdot TE \cdot \frac{1}{N_H} = \frac{BR \cdot TE}{N_H + N_B}$$

$$\beta_{B \rightarrow M} = \beta_{M \rightarrow B} = ap_B b_{HM} \frac{1}{N_B} = BR \cdot \frac{N_B}{N_H + N_B} \cdot TE \cdot \frac{1}{N_B} = \frac{BR \cdot TE}{N_H + N_B}$$

#	Population	Symbol	Description	Type of Function	Reference
1	H	μ_H	Human mortality rate	Fixed value: 3.5×10^{-5}	CDC (2020) [23]
2	H	δ_H	Human incubation rate	Fixed value: 0.2	CDC (2020)[24]
3	H	λ_H	Human recovery rate	Fixed value: 0.1	CDC (2020) [24]
4	H	μ_{WNV_H}	Human mortality rate of WNV	Fixed value: 0.01	CDC (2020) [23]
5	M	$\frac{EFGC}{LGC}$	Fecundity	Quadratic function of temperature	Shocket et al. (2020) [25]
6	M	ρ_E	Hatching rate	Briere function of temperature	Heidecke et al. (2025)[26]
7	M	μ_E	Egg mortality rate	Function of temperature	Heidecke et al. (2025)[26]
8	M	ρ_A	Larva to adulthood rate	Function of temperature	Shocket et al. (2020) [25]
9	M	μ_A	Aquatic mortality rate	Function of temperature	Shocket et al. (2020) [25]
10	M	μ_M	Mosquitoes mortality rate	Linear function of temperature	Shocket et al. (2020) [25]
11	M	δ_M	Extrinsic incubation rate	Briere function of temperature	Shocket et al. (2020) [25]
12	M	TE	Transmission efficacy	Quadratic function of temperature	Shocket et al. (2020) [25]
13	M	BR	Biting rate	Briere function of temperature	Shocket et al. (2020) [25]
14	M	K	Carrying capacity of aquatic phase	Function of temperature	Evensen (2009) [22]
15	B	μ_B	Bird mortality rate	Fixed value 3.4×10^{-4}	Brault et al. (2004)[27] McLean et al. (2006) [28]
16	B	δ_B	Bird incubation rate	Fixed value 0.42	Brault et al. (2004)[27] McLean et al. (2006) [28]
17	B	λ_B	Recovery rate of birds	Fixed value: 0.005	Brault et al. (2004)[27] McLean et al. (2006) [28]
18	B	μ_{WNV_B}	Birds mortality rate of WNV	Fixed value: 0.2	Brault et al. (2004) [27] McLean et al. (2006) [28]

TABLE I: Rates and coefficients.

To reflect the high mortality of American crows infected with WNV while allowing rare survival, we set the recovery rate at a small value, $\lambda_B = 0.005 \text{ day}^{-1}$, implying a mean recovery time of $1/\lambda_B = 200$ days. Assuming a WNV-induced mortality rate of $\mu_{WNV,B} = 0.20 \text{ day}^{-1}$, the fatality rate of the case is:

$$\text{CFR} = \frac{\mu_{WNV,B}}{\mu_{WNV,B} + \lambda_B} = \frac{0.20}{0.20 + 0.005} \approx 0.976.$$

This yields a fatality rate of 97.6%, consistent with experimental studies showing near-total mortality in crows.

C. Predictive model of spillover onset

In this section, we explain how to establish predictive PDF for spillover onset. Before proceeding, we would like to illustrate the idea with a simple analogy. Consider the concept of a prediction interval in statistics, which specifies an interval that, with a given level of confidence, is expected to contain the true parameter value. In other words, the interval shows a meaningful region of the domain of a PDF (such as the normal PDF), in which the unknown parameter is most likely to lie, based on the associated probability mass. Here, we pursue a similar concept in a different way to derive a bivariate probability density function for the spillover onset, which identifies regions on the two-dimensional plane of (MoT) where the probability density of spillover onset is non-zero. After obtaining the PDF, we equivalently define the critical region for spillover onset on the (M, T) plane. By checking the daily (M, T) values against this region, we can classify each day as risky or not and, consequently, recover the within-year interval(s) during which an epidemic is expected to begin for the corresponding temporal regime.

As an example, to obtain the predictive PDF for 2024 for a given county, we first identified the timing of spillover onset for each year in the historical dataset (for example, 2006-2023). Then we ran the compartmental model for each calendar year using the observed daily temperatures of that year, producing a mosquito abundance curve that lasts the season $M_k(t)$, $k = 2006, \dots, 2023$. For each season in which spillover onset occurred at time (st) , we extracted the paired values (M, T) from the seven-day window $st - 13$ to $st - 7$ days before onset. This resulted in a set of paired values $\{(M_{st-i}, T_{st-i})\}_{i=7}^{13}$. All samples collected from these windows were then pooled in the training data set. We repeated this sampling process for each year in the historical record (2006-2023), building a comprehensive collection of $\{(M_{st_k-i}, T_{st_k-i})\}_{i=7}^{13}$, k is the index of the k 'th year in historical data and st_k is the onset of spillover time for in the k 'th year:

$$\text{Data sample for spillover onset} = \bigcup_{k=2006}^{2023} \{(M_{st_k-i}, T_{st_k-i})\}_{i=7}^{13}.$$

We assumed that the patient was infected approximately 14 days prior to the date the case was reported by the hospital to the relevant public health authority. In general, there is a temporal delay between the time of infection and the time of case reporting [29]; thus, this extra seven-day delay accounts for the delay associated with hospital reporting (reflecting a 7-day intrinsic incubation period for WNV plus an additional 7-day reporting delay) [30].

Next, we applied two-dimensional kernel density estimation (KDE) [31] directly to the data sample for spillover onset, producing a smooth, non-parametric estimate of the joint PDF. See Section -I for details, explanations, and examples of the kernel density estimation approach. This PDF delineates the regions of (M, T) s in the MoT plane that are most likely to correspond to the initial spillover event, as inferred from the historical disease dynamics reproduced by the compartmental model. In other words, within the MoT space (values of M on the horizontal axis and values of T on vertical axis), we identified the region(s) of (M, T) in which spillover is more likely to occur. Figure 1 presents a diagram that describes each step of this process. Figure 2 presents both the 2-D histogram and its fitted PDF for 2024.

D. Eco-Epi Poisson model of spillover severity

In this section, we construct the Eco-Epi rate surface $\lambda(M, T)$, which maps each combination of mosquito abundance M and normalized cumulative temperature T to the expected number of human infections. Our aim is to identify the regions in the MoT plane associated with different expected infection levels. Simply put, if (M_λ, T_λ) was observed when λ people were reported, then the neighborhood of (M_λ, T_λ) on the MoT plane is assigned a height of λ .

Building on the above idea, we use historical human case data to derive the rate function $\lambda(M, T)$ in three steps:

- 1) **Data aggregation and weighting.** For each epidemiological week with at least one reported human case (spillover times st_{jk} , $k = 2006, \dots, 2023$), we collected data $(M$ and $T)$ in a 7-day window, 14 to 7 days before the time of reported cases from the curve of mosquito abundance $M_k(t)$ and normalized cumulative temperature. Each pair receives a weight $\omega_{st_{jk}} = N_{st_{jk}}$, where $N_{st_{jk}}$ is the number of cases in week st_{jk} ,

$$\text{weighted sample} = \bigcup_{k=2006}^{2023} \bigcup_{j_k} \left\{ (M_{st_{jk}-i}, T_{st_{jk}-i}, \omega_{st_{jk}}) \right\}_{i=7}^{13}.$$

This weighting amplifies the influence of environmental conditions associated with larger outbreaks. As described in Section -C, the sampling window, from 14 to 7 days prior to spillover onset, accounts for two key delays: the intrinsic incubation period required for symptoms to appear (approximately 7 days, during which data are needed), and the reporting delay from hospitals to public health agencies (an additional 7 days, which should be considered for the starting point for sampling, but is not itself sampled). This window was chosen to ensure biological plausibility and consistency with the observed reporting patterns [29].

- 2) **Obtaining non homogeneous Poisson surface rate by KDE** The weighted sample is smoothed with a kernel density estimator (KDE) to obtain a two-dimensional PDF, $\hat{f}(M, T)$. To convert this PDF into a surface that is numerically consistent with the empirical weights, we rescale it by the factor $\omega_{\max}/f_{\max}$, where $f_{\max} = \max_{M, T} \hat{f}(M, T)$, $\omega_{\max} = \max_i \omega_i$. The resulting intensity surface is therefore $\lambda(M, T) = \frac{\omega_{\max}}{f_{\max}} \hat{f}(M, T)$. See Section -I3 for further details.

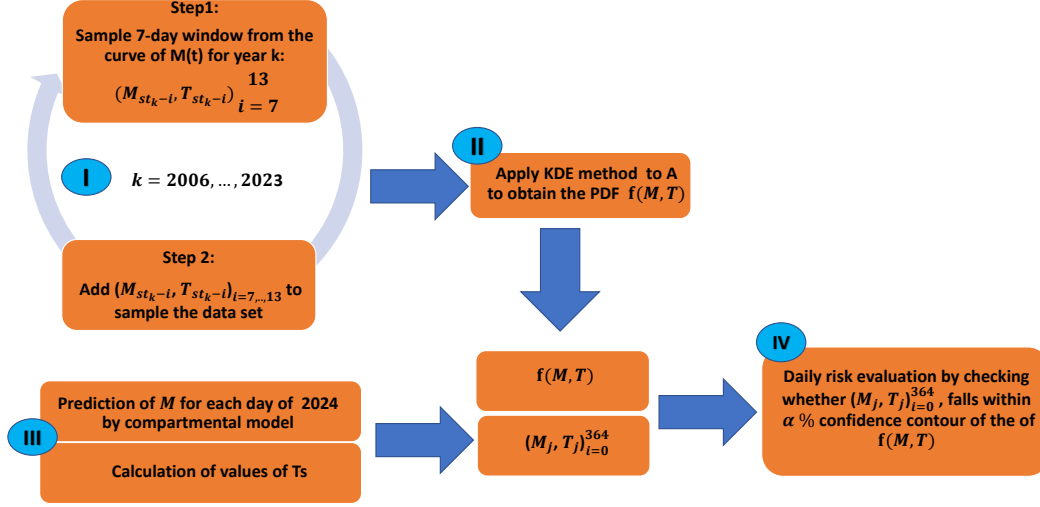


Fig. 1: For constructing the PDF to predict spillover onset for 2024, the following steps were performed, I–for each historical year $k \in \{2006, \dots, 2023\}$, paired samples $(M_{st_k-i}, T_{st_k-i})_{i=7}^{13}$ were collected from the mosquito abundance curve $M_k(t)$ and the temperature record. Specifically, M_{st_k-i} were taken from $M(t)$ and T_{st_k-i} from the temperature series for the interval 14 to 7 days preceding the reported onset date. II- These collected pairs across all historical years $((M_{st_k-i}, T_{st_k-i})_{i=7}^{13})$ were combined to generate a dataset $(\bigcup_{k=2006}^{2023} \{(M_{st_k-i}, T_{st_k-i})_{i=7}^{13}\})$, from which a continuous PDF was derived via kernel density estimation (KDE). III- For the target prediction year (2024), the compartmental model was run using the average temperatures from 2019-2023, yielding predicted $(M_j, T_j)_{j=0}^{364}$ pairs. IV- Using the derived predictive PDF and the forecasted $(M_j, T_j)_{j=0}^{364}$ coordinates, daily spillover risk was then assessed by checking which one of $(M_j, T_j)_{j=0}^{364}$ coordinates fall inside the contour of α percent confidence.

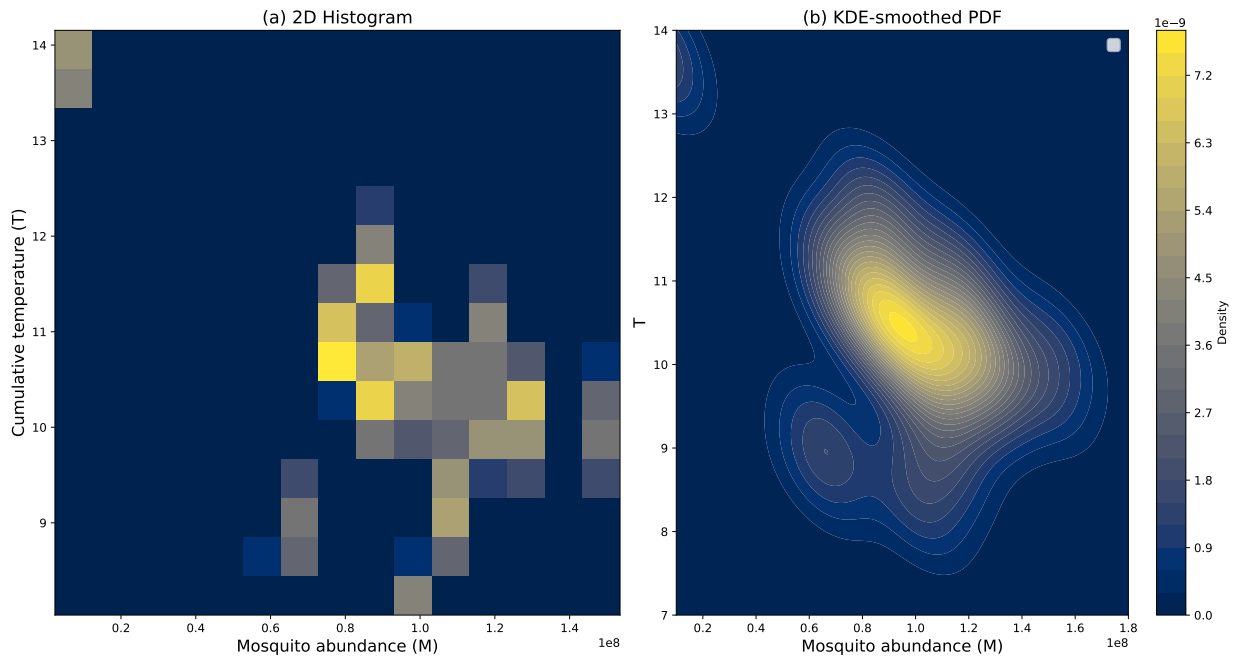


Fig. 2: Panel (a) illustrates the 2D histogram of mosquito abundance (M) and normalized cumulative temperature (T) at the timing of WNV spillover onset to the human population, based on data from 2006 to 2023. Panel (b) presents the corresponding smoothed probability density function estimated using a kernel density estimation (KDE) method.

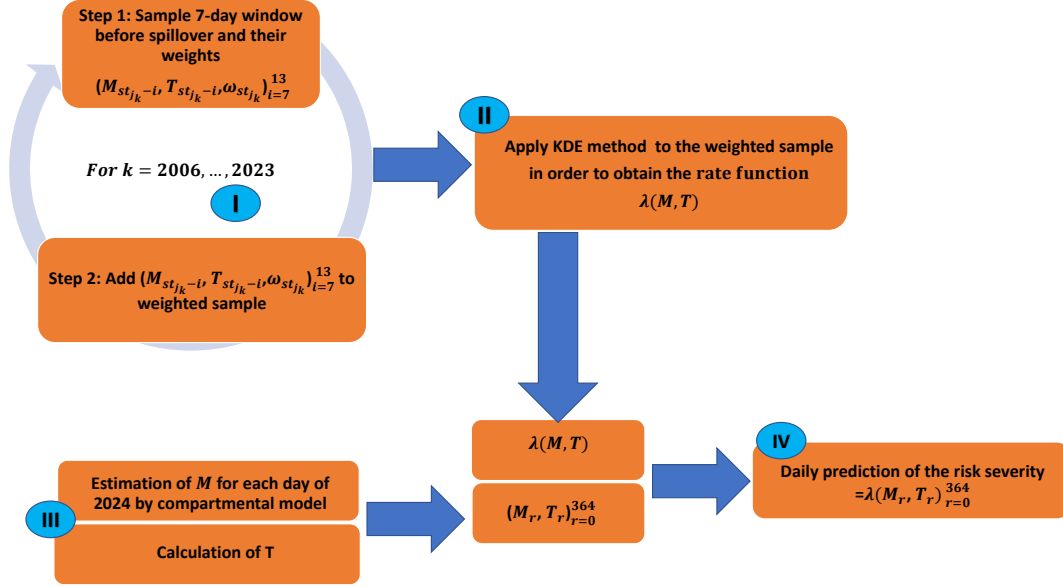


Fig. 3: Schematic representation of obtaining the Poisson rate function $\lambda(M, T)$ for WNV spillover severity for 2024 using the historical data from 2006 to 2023. I- The workflow begins with sampling of mosquito abundance from the curve of $M_k(t)$, real temperature of year (to estimate T), and weights for each week with reported cases. This process should be done for all historical years $k = 2006, \dots, 2023$. II- Using the weighted sample of step (I) and KDE methods we find a surface for rate of Poisson process as function of (M, T) . III- The compartmental model should be run for the year 2024 to predict values of (M_r, T_r) for the entire season. IV – Using the rate function obtained in step (II) and the predicted values of (M_r, T_r) from step (III), the severity of the risk for each day can be obtained as the rate of a Poisson process.

Figure 3 summarizes the estimation workflow, while the surface derived $\lambda(M, T)$ in Figure 4 shows the surface rate of the Eco-Epi model.

E. Logarithmic Score

In the logarithmic scoring rule, the predictive probability density is mapped onto a set of epidemiologically motivated bins, that is, the single value 0; the ranges $[1, 5]$, $[6, 10], \dots, [46, 50]$; the broader intervals $[50, 100]$, $[101, 150], \dots, [201, 250]$; and an out-of-range category. For a given forecast, let P denote the probability mass assigned to the bin that contains the observed case count; the score is then calculated as $-\log P$. It is important to note that for this metric, the highest score is achieved when the observation falls within the bin of highest density. If an observation falls outside the meaningful domain of the PDF or receives a score below -10 , we assign a default score of -10 and flag it as an outlier prediction.

F. Köppen climate classes

Below is a concise overview of the Köppen climate classifications used in this study to analyze the results [3].

TABLE II: Brief descriptions of Köppen climate classes occurring across the three study counties

Köppen code	Name	Key characteristics
Csa	Hot-summer Mediterranean	Dry summers; warmest-month mean $\geq 22^\circ\text{C}$; mild, wet winters.
Csb	Warm-summer Mediterranean	Dry summers; warmest-month mean $< 22^\circ\text{C}$ but $\geq 10^\circ\text{C}$ for ≥ 4 months; mild, wet winters.
BSh	Hot semi-arid	Annual evapotranspiration $>$ precipitation; hot summers; brief wet season, typically in winter or late summer.
BWk	Cold desert	Arid; mean annual $T < 18^\circ\text{C}$; large diurnal range; sparse precipitation year-round.
BWh	Hot desert	Arid; mean annual $T \geq 18^\circ\text{C}$; extremely hot summers; minimal, erratic rainfall.

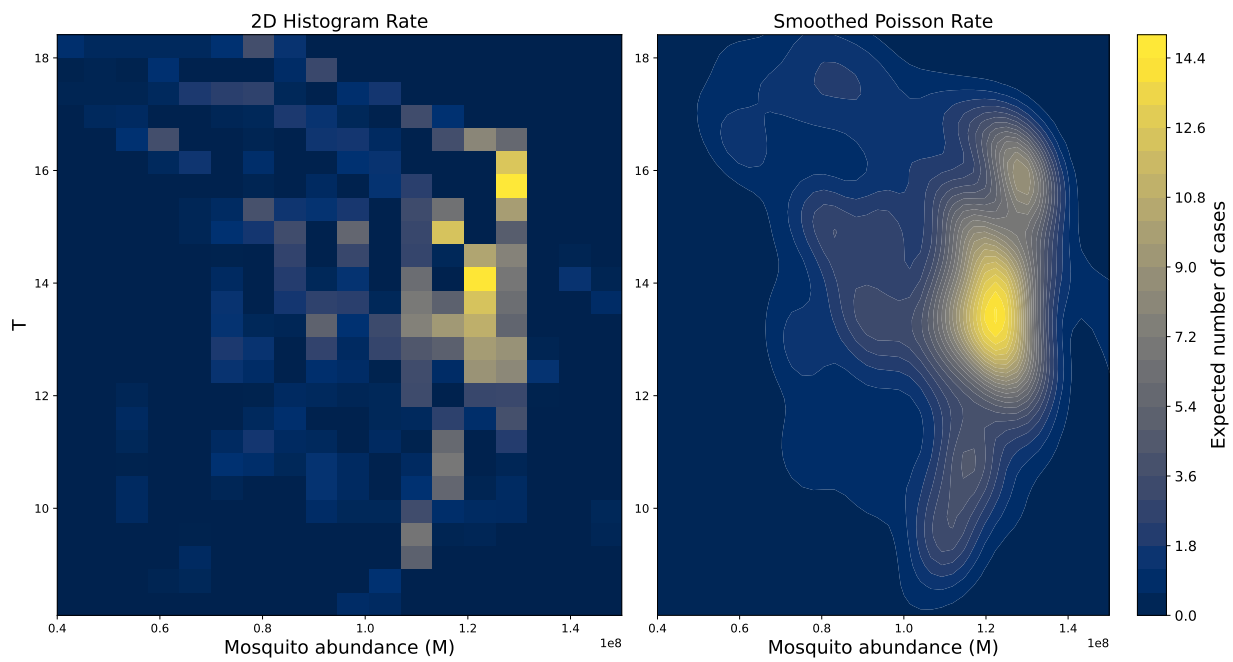


Fig. 4: Left: Two-dimensional histogram of the estimated Poisson rate function for Orange County, California, based on historical data from 2006 to 2023. Each cell represents the estimated rate as a function of mosquito abundance (M) and normalized cumulative temperature (T). Right: Smoothed surface of the rate function obtained by applying Kernel Density Estimation to the collected data.

G. Leave-one-year-out cross validation and retrospective forecast for spillover onset prediction

1) Leave-one-year-out cross validation

For a given target year about which we wanted to make a prediction, the data of that year (number of cases) were withheld, the predictive PDF is derived from the remaining years (as we explained in Section -C), the spillover onset prediction interval for the excluded year was generated by the predictive PDF. The performance of the model was then evaluated by determining whether the real onset of spillover for that year fell within the predicted interval and other metrics defined in the paper. And, as explained before, the compartmental model was run using a proxy temperature series obtained by a combination of temperature data from the year withheld and the five preceding years. We used daily temperatures from the withheld year up to the start of the WNV in human population season (we used the temperature of withholding year to day number 140); for the remainder of that year, the model was driven by the five-year climatological mean temperature calculated from the immediately preceding years.

2) Retrospective forecast

Our target years for the retrospective forecast were the last two years for each county. For example, for 2023, we used human case data up to 2022 to construct the predictive PDF (as described in Section -C) and applied it to forecast the spillover onset period in 2023. For 2024, we incorporated data until 2023 to establish the predictive PDF and subsequently predicted the spillover onset period for 2024. Similarly to the leave-one-year-out cross-validation, the compartmental model was run using a proxy temperature series constructed by combining temperature data from the withheld year with that of the five preceding years. We note that because Riverside County did not report cases in 2022, data through 2021 and 2023 were used to forecast 2023 and 2024, respectively. For Duval County, Florida, we conducted retrospective forecasts for 2019 and 2024 because no data was available for 2020-2023.

H. Leave-one-year-out cross validation and retrospective forecast for spillover severity prediction

1) Leave-one-year-out cross validation

For leave-one-year-out cross validation, we excluded the data for a given year and obtained the surface rate function of the Eco-Epi Poisson model, using the remaining years as explained in Section -D. The fitted surface rate was then applied to predict the severity of spillover for the omitted year, allowing us to evaluate its predictive performance by checking the results with reported cases and defined metrics such as logarithmic score, AVE and NRMSE. For the temperature input for the compartmental model to predict M and T , as described before, we used the combination of the temperature of the excluded year and the average temperature of five years prior to the excluded year.

2) Retrospective forecast

Similarly to spillover onset, our target years for the retrospective forecast were the last two years for each county. For example, for 2023, we used human case data up to 2022 to construct the surface rate of the Exo-Epi Poisson model (as described in Section -D) and applied it to forecast the severity of the outbreak in 2023. For 2024, we incorporated data until 2023 to establish the surface rate and subsequently predicted the spillover severity period for 2024. Then, as explained, the compartmental model was run using a proxy temperature series constructed by combining temperature data from the withheld year with those of the five preceding years. For Riverside County, we used data to 2021 to forecast 2023. Similarly to spillover onset, for Duval County, Florida, we performed retrospective forecasts for 2019 and 2024 because no data were available for 2020-2023.

I. Kernel Density Estimation

Kernel density estimation (KDE) is a non-parametric method to fit a probability density function (PDF) to observed data without assuming a parametric family (e.g. normal, gamma). Given a sample $x_1, \dots, x_n$, KDE places a smooth kernel on each observation and averages their contributions to produce a continuous estimate $\hat{f}$ that integrates into one. This approach is particularly useful when the underlying distribution may be multimodal or skewed, or when we wish to visualize/highlight highest-density regions and contours. Various KDE formulations exist, but all approximate the unknown density by summing bell-shaped kernels centered on the data points. In this work, we employ a Gaussian kernel for its smooth, bell-shaped profile. Formally, the uni-variate KDE with kernel K and bandwidth h is:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{n h} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

in which x_i , n , and h are sample of observations, is the sample size, and bandwidth, respectively. In this estimation, the critical parameter is the width of the kernel - or bandwidth $-h$, which determines how widely each kernel spreads its mass.

1) Scott's Rule of Thumb

Scott's rule of thumb provides a practical, data-driven way to choose the bandwidth h without knowing the true underlying density f . In theory, optimal bandwidth minimizes the asymptotic mean integrated squared error (AMISE), which depends on the optimum bandwidth.

$$R(f'') = \int (f''(x))^2 dx,$$

the integrated squared second derivative of f , a measure of its “roughness.” Since f is unknown, Scott instead assumes f is approximately normal and replaces $R(f'')$ with its exact value under a Gaussian distribution. For a uni-variate Gaussian kernel

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2},$$

this substitution, together with known kernel constants, yields the simple formula

$$h_{\text{Scott}} = \hat{\sigma} n^{-1/5},$$

where n is the number of data points and $\hat{\sigma}$ is their standard deviation of the sample. This choice automatically widens h when data are sparse (large $n^{-1/5}$) and narrows when data are concentrated (small $\hat{\sigma}$), striking a balance between over- and under-smoothing.

In higher dimensions ($d > 1$), the same idea applies: one uses the sample covariance matrix $\hat{\Sigma}$ and scales it by $n^{-2/(d+4)}$. Thus, the multivariate bandwidth matrix is

$$\mathbf{H}_{\text{Scott}} = n^{-2/(d+4)} \hat{\Sigma},$$

or, if a single scalar is preferred,

$$h_{\text{Scott}} = \hat{\sigma} n^{-1/(d+4)},$$

with $\hat{\sigma}$ a representative scale (e.g., the average marginal standard deviation). Scott's rule therefore provides a self-contained recipe - no cross-validation or pilot estimates required - that adapts seamlessly to sample size, data dimension and variability. Once $\hat{f}$ is fitted to h_{Scott} , you obtain a smooth density estimate whose overall shape faithfully mirrors your data and is ready for any subsequent processing, such as rescaling to match empirical counts [31].

Figure 5 clarifies how the resulting PDF is for a given sample, we can also notice that in the Scott form of KDE the resulting PDF does not depend on the number of the bins of the histogram.

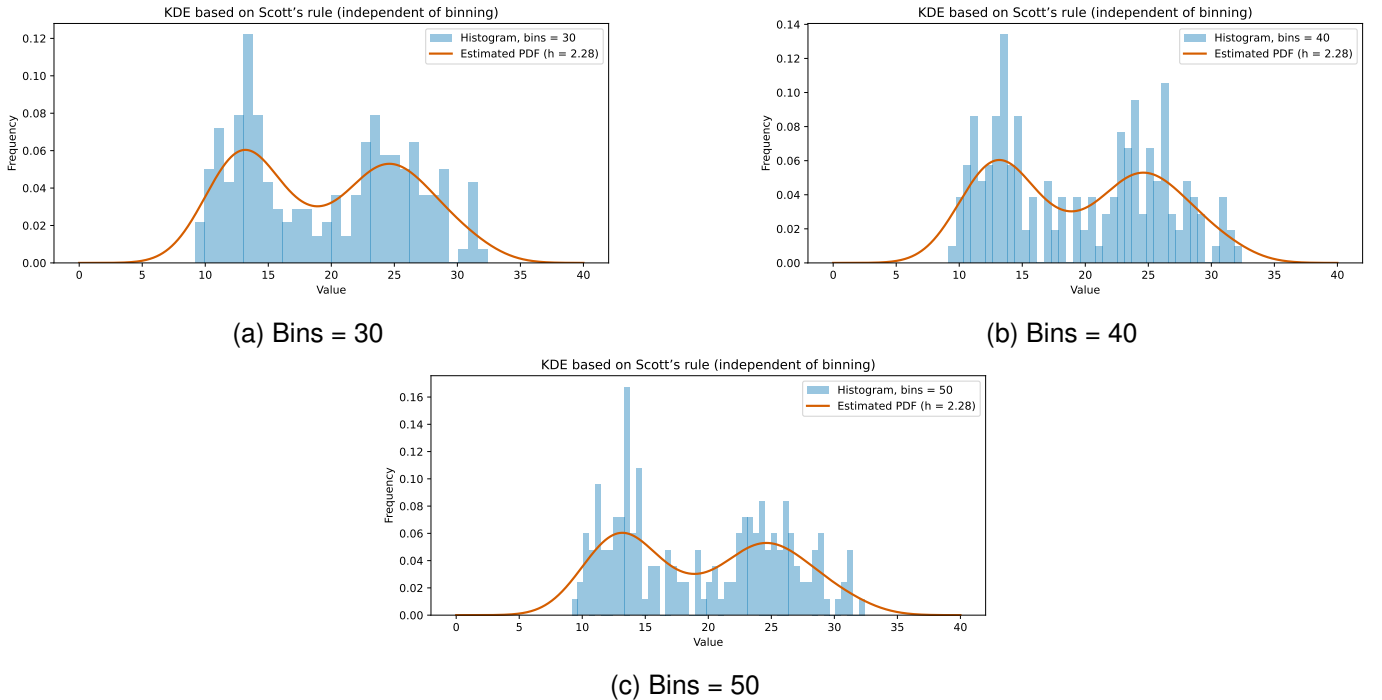


Fig. 5: KDE estimates PDF and different histogram bin counts.

2) Two-Dimensional KDE

When the data are bivariate, $\{(x_i, y_i)\}_{i=1}^n$, the KDE generalizes naturally to two dimensions. Using an isotropic Gaussian kernel, the estimate at (x, y) is

$$\hat{f}_{\mathbf{H}}(x, y) = \frac{1}{n \sqrt{|\mathbf{H}|}} \sum_{i=1}^n \exp\left(-\frac{1}{2} \begin{bmatrix} x - x_i \\ y - y_i \end{bmatrix}^T \mathbf{H}^{-1} \begin{bmatrix} x - x_i \\ y - y_i \end{bmatrix}\right),$$

where $\mathbf{H}$ is the 2×2 bandwidth matrix. A convenient choice is the scalar matrix $\mathbf{H} = h^2 \mathbf{I}$, which yields

$$\hat{f}_h(x, y) = \frac{1}{n h^2} \sum_{i=1}^n \frac{1}{2\pi} \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2h^2}\right).$$

Scott's rule of thumb extends to d dimensions by setting

$$\mathbf{H}_{\text{Scott}} = n^{-2/(d+4)} \hat{\Sigma},$$

so in two dimensions ($d = 2$) one has

$$\mathbf{H}_{\text{Scott}} = n^{-1/3} \hat{\Sigma} \implies h_{\text{Scott}} = n^{-1/6} \hat{\sigma}$$

if a scalar bandwidth h is preferred. This yields a smooth surface over the (x, y) plane whose height at each point reflects the local concentration of observations, independent of any histogram binning 6d.

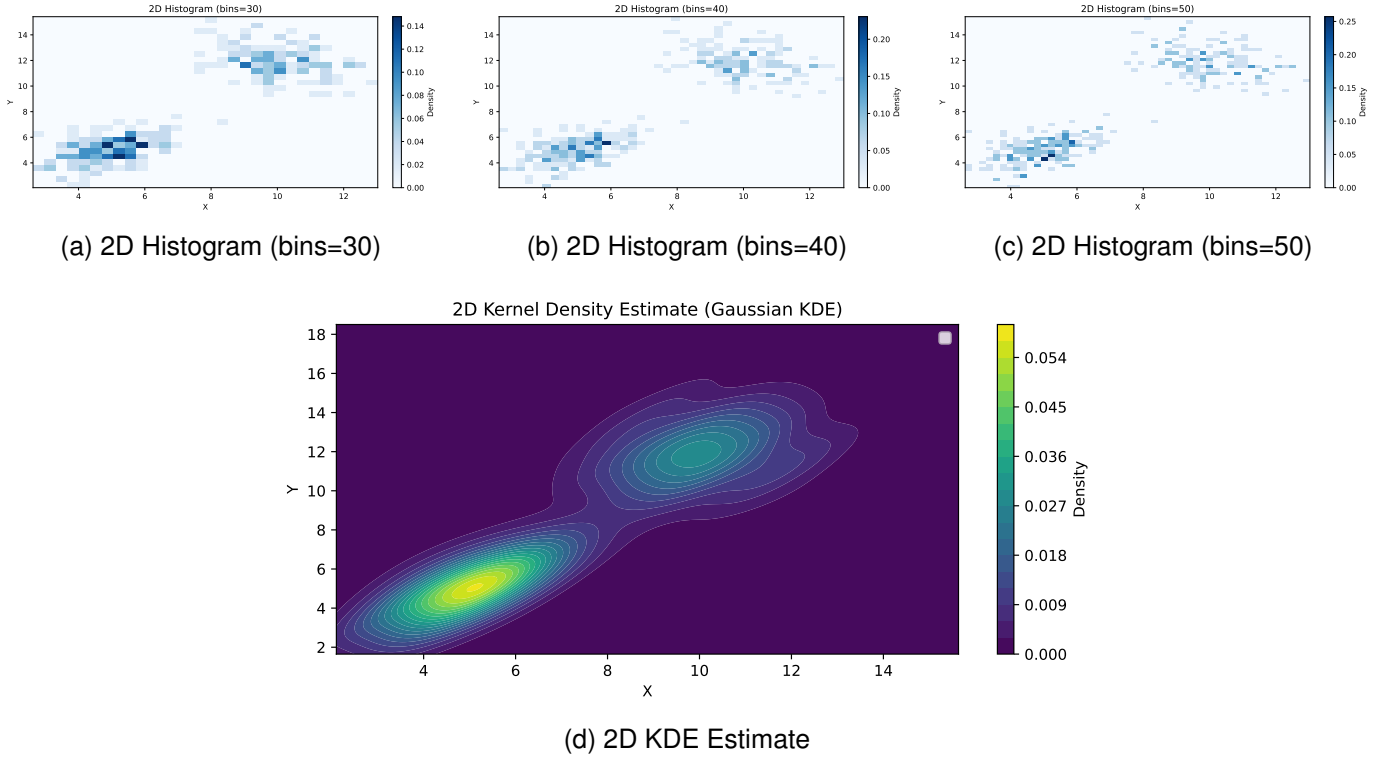


Fig. 6: (a–c) Density histograms with increasing bin counts, and (d) the corresponding smoothed 2D kernel density estimate.

3) Kernel Density Estimation for Surfaces

Consider a bivariate sample $\{(x_i, y_i)\}_{i=1}^n$ in which repeated coordinates are interpreted as integer weights $\omega_i \in \mathbb{N}$. Our goal is to construct a smooth *surface* whose height approximates these weights while preserving the spatial smoothness afforded by a kernel estimator.

a) *Step 1: fit the weighted PDF.*

Using Scott's bandwidth rule, we obtain the weighted KDE $\hat{f}_h(x, y)$ that integrates to 1.

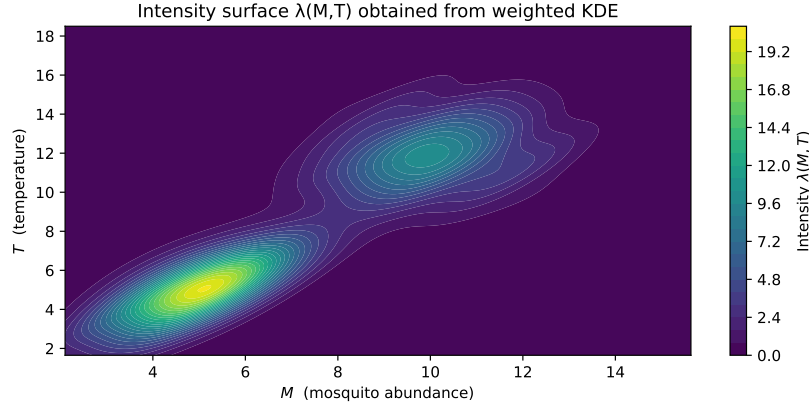


Fig. 7: Weight-consistent surface obtained from the weighted KDE after rescaling so that the peak equals the maximum empirical weight.

b) Step 2: rescale the PDF to a weight-consistent surface.

Let $f_{\max} = \max_{x,y} \hat{f}_h(x,y)$ and $\omega_{\max} = \max_i \omega_i$ denote the highest kernel-density value and the largest empirical weight, respectively. Multiplying the PDF by the factor $c = \omega_{\max}/f_{\max}$ yields

$$\lambda(x,y) = \frac{\omega_{\max}}{f_{\max}} \hat{f}_h(x,y),$$

so that $\lambda(x_i, y_i) = \omega_{\max}$ at the global mode and, more generally, the surface height at any location is proportional to the local weighted concentration of observations. Consequently, $\lambda(x,y)$ is both smooth (because it inherits the structure of the kernel) and numerically consistent with the empirical weights (Figure 7).

J. Discussion

Duval County

Florida’s climate permits mosquito activity in all months. In the humid subtropical setting of Duval County, arboviruses such as WNV can therefore circulate beyond the traditional summer–fall peak. Seroconversion analyses of sentinel chickens from northeast Florida show that both the WNV and the eastern equine encephalitis virus (EEEV) are detected outside of the peak season, indicating a non-zero transmission risk year round [32, 33]. Consistent with this, our Eco-Epi model yields an early-year risk signal that intensifies from summer to fall and identifies early-onset cases in 2015 and 2019 (Supplementary X). This daily resolution interpretation is consistent with recent WNV hazard maps derived from sentinel chicken and meteorological data, providing pre-season and in-season guidance at monthly resolution [34].

Furthermore, in Duval County the Eco-Epi model predicted the annual mean risk in years without reported human cases (2017, 2020–2023) at levels comparable to those in low incidence years (e.g. 2016, 2018, 2019, 2024; 0–3 reported cases; Supplementary XI). Florida’s Department of Health records of WNV activity, such as positive dead birds in multiple years during this period, support the presence of underlying transmission activity despite low reported human case counts. Additionally, for 2020–2022, pandemic-era effects on health care search, testing, and surveillance, as well as broader non-pharmaceutical interventions, may have altered the observed burden of WNV in Duval County [35, 36].

Furthermore, during the sampling step for the establishment of the surface rate of the Eco-Epi model, we noticed that the highest spillover rates occur in (23–29 °C). This pattern is consistent with laboratory findings on effective temperature ranges for mosquito transmission [25, 37–40]. The Eco-Epi spillover rate surface for Duval County shows that the highest spillover values center at intermediate temperatures $T \approx (10\text{--}18 \text{ }^{\circ}\text{C})$ on the normalized cumulative temperature scale, with markedly lower rates at both lower and higher values of T (Supplementary III).

Throughout 2010–2019 and 2024, the Eco-Epi rate surfaces of each year $\lambda(M, T)$ and the daily risk maps show a consistent mechanism. Seasonal (M, T) trajectories typically enter the high-risk ridge from late July onward, producing a principal risk window around the day of year (DOY) (210–310) with occasional spring / early winter shoulders. The early spillovers in 2015 and 2019 were also predicted by daily maps (Supplementary X). Additionally, the Eco-Epi model outperformed both baselines: Annual logarithmic scores were consistently less negative and the NRMSE was markedly lower than the null and negative binomial models in every evaluated year (largest gains in 2011–2012, 2015 and 2018; Supplementary IV).

Taken together, these results show that our parsimonious Eco-Epi framework captures the magnitude of the WNV spillover in Duval County and consistently outperforms simple baselines. The resulting daily risk signal is actionable: It complements routine surveillance and can guide preseason preparedness and in-season control during the late July to early November window.

Dallas County

The Eco-Epi model showed an accurate prediction in Dallas County in both cross-validation of the one-year-leave-out cross-validation and retrospective forecast. In 100% years the NRMSE model was less than 1, and showed a great difference from the performance of the NB and the null model (Supplementary IV).

The surface of the Eco-Epi Poisson rate for Dallas County shows a clear bimodal pattern. Consequently, over time, the temporal risk maps of the model identify two seasonal maxima: a primary summer peak around DOY 196–231 (epidemiologic weeks 28–33) and a secondary early fall rise around DOY 255–289 (weeks 37–41) (Supplementary X). This pattern aligns with Dallas County Health and Human Services (DCHHS) field surveillance summaries, where the average abundance of *Culex quinquefasciatus* shows a dominant midsummer maximum with a recurrent early fall rebound in several seasons [41].

Furthermore, during the sampling step used to obtain the surface rate of the Eco-Epi model, we observed that the highest spillover values cluster under relatively warm conditions (24–32 °C). This slightly higher temperature window, compared with (23–27 °C), is plausibly explained by habitat compression (oasis/island effect) during heat and drought, in which mosquitoes and avian hosts co-aggregate at scarce water sources, increasing contact rates.

This interpretation aligns with the analyses of the Coachella Valley in California linking hydro meteorological anomalies with an elevated risk of WNV [8] and the broader evidence that drought intensifies the transmission of WNV and that higher temperatures increase *Culex* infection metrics [42, 43]. Although Dallas lacks a desert valley setting as Riverside County's Coachella Valley, functionally similar 'oasis' conditions occur along the Trinity River floodplain (Great Trinity Forest and the Chain of Wetlands) and in urban lakes, riparian corridors, and storm-water basins. These water-replete patches within a hotter, impervious urban matrix can concentrate avian hosts and *Culex quinquefasciatus*, increasing contact rates during heat/drought [44].

For Dallas County, we hypothesize that thermal forcing contributes to the observed bimodality: supra-optimal summer temperatures attenuate adult mosquito activity and survival, with activity resuming as conditions cool in early fall. This mechanism is encoded in our temperature-dependent rate model and is consistent with the timing of the two seasonal maxima indicated by surveillance. Taken together, the bimodal ridge in $\lambda(M, T)$ and the seasonal signal in these two periods indicate sustained hazard through late summer with renewed risk following early fall rainfall, even when total abundance is below the midsummer maximum.

In general, in Dallas County, the Eco-Epi framework provides an interpretable daily hazard signal. Specifically, our model recovers the bimodal mosquito pattern without using any mosquito surveillance data for input or calibration; the mosquito component is inferred from the model, yet it reproduces the same two-period signal observed in surveillance. Consequently, this parsimony and agreement with independent observations support its use as a decision aid. Consequently, given the two predictable windows, vector control programs can stage two targeted surges: (i) a pre-peak operation beginning 2 to 3 weeks before weeks 28 to 33 (source reduction, larviciding in high-productivity basins and intensified gravid trapping) and (ii) a second surge in weeks 37 to 41 after late summer heat breaks and rainfall.

Harris County

In Harris County, performance is strong for both metrics: The Eco-Epi model consistently outperforms the baselines of the null and negative binomials by logarithmic score and achieves a low error with $\text{NRMSE} \leq 0.5$ in 12 of 14 years.

Furthermore, the sampling process of the compartmental model indicates that the highest spillover values cluster under relatively warm conditions (23–29 °C). The lower temperature ceiling for spillover in Harris County than in Dallas County (29 °C vs. 32 °C) is consistent with Harris's wetter and more humid Gulf-influenced climate, which can dampen late summer amplification at the hottest temperatures. The Eco-Epi rate surface indicates that Harris County experiences a broader spillover-conducive temperature window (T) than Dallas County on the normalized cumulative temperature scale. (e.g., lower / upper bounds in Harris vs. Dallas: ~ 4 –19.45°C vs. 7–21.4°C) (Supplementary III). This pattern is consistent with regional climatology: Houston's proximity to the Gulf of Mexico moderates temperatures, mild winters, and damped summer extremes, so Harris County remains warmer on average, while inland Dallas exhibits a more continental thermal regime, with greater seasonal amplitude and colder winter minima, and 1991–2020 climate normals indicate that it is cooler on average [45]. Mechanically, a moderated regime maintains daily conditions within the unimodal temperature-sensitive zone for vector survival, extrinsic incubation, and biting rate for longer periods of the season, thus widening the period during which spillover can occur [25, 38, 39, 46].

Throughout 2011–2024, our one-year-out cross-validation predictions for Harris County place the peak season in late summer to early fall: a recurrent high-risk window of approximately DOY 230–305 (mid-August to early November), with a median peak near DOY ~ 270 (late September) and some years shifting later into October–November. For 2014, the prediction of the Eco-Epi model indicates a maximum risk window that spans DOY 250–300 (approximately early September to late October) (Supplementary X). During this severe outbreak, the predicted severity of the risk in time was broadly consistent with the field study: mosquito pool positives were first detected in week 23 (DOY 157), peaked in week 30 (DOY 201–207), and spillover peaked in week 33 (mid-August; DOY ~ 221 –228), with elevated activity continuing through week 42 (mid-October; DOY ~ 277 –894), consistent with local guidance that risk is highest from summer through fall and often peaks in August to September [47].

Our model suggests that Harris County's longer temperature-suitable period (for spillover) warrants starting routine surveillance and control earlier (late spring) and maintaining it through late fall; this aligns with the CDC's guidance to maintain season-long WNV surveillance and with local public health descriptions of the mosquito season as summer–fall [48].

Orange County

In Orange County, 16 out of 19 years in one-year-out cross-validation have NRMSE ≤ 1 and close to 0.5. Furthermore, the logarithmic score results show that the performance of the Eco-Epi model in severe years like 2014 and 2015 is much better than the null and NB models, and even based on the NRMSE the prediction error for these years is at the same level as for other years, without a significantly higher error (Supplementary IV).

The peak of the surface of the Eco-Epi Poisson rate for Orange County occurs in the range of normalized cumulative temperature ($10\text{--}16^\circ\text{C}$) T , moderate temperature compared to the entire domain, which is ($8\text{--}18^\circ\text{C}$). The Eco-Epi rate of the Poisson model for Orange County in 2006–2024 reflects a repeatable structure within seasons that is largely invariant over years. The modeled risk rises abruptly in late July, concentrates during late summer–early fall (about DOY 240–290), and then decays through October. This timing is consistent with county surveillance: Orange County Health Care Agency reports that most human WNV infections occur in August to September with substantial activity in July and October, and the Orange County Mosquito and Vector Control District response plan shows peak *Culex* abundance from July through October [49, 50].

The rate of the Eco-Epi model indicates that inter-annual variation is governed by (i) how quickly the mosquito abundance M climbs to high rate region ($1.1\text{--}1.5 \times 10^8$) and remains there, (ii) whether temperature T persists in the warm seasonal band (approximately $T \in (11 - 14^\circ\text{C})$).

Years 2008, 2012, 2014, 2015 and 2020 also exhibit unusually late activity extending toward DOY 330–360 which logarithmic score and Eco-Epi Poisson distributions show that the model captured this late activity well (Supplementary V).

Although the onset changes by approximately 3–4 weeks from year to year, the modal risk window remains anchored in the late summer, indicating that the timing of spillover is governed by seasonal covariates and that sustained late season vigilance is consistently warranted.

Riverside County

Throughout 2006–2024, the Eco-Epi rate surface $\lambda(M, T)$ for Riverside County shows a consistent high- λ lobe at combinations of higher mosquito abundance (M) and approximately ($8\text{--}17^\circ\text{C}$) window for normalized cumulative temperatures (T). Years with larger outbreaks are those in which the observed seasonal trajectory ($M(t), T(t)$) spends longer time within or passes repeatedly through this lobe during late summer and early autumn.

This mechanism is visible especially in 2015: the trajectory tracks the core of the high- λ region, and the day-of-year panels show a broad favorable window (roughly day 230–300) accompanied by sustained weekly case counts.

Forecasts for lower activity years (e.g., 2009, 2011, 2016, 2021) exhibit brief or late-season intersections with the lobe and correspondingly short, low-intensity windows in the bottom panels. Operationally, the maps imply a recurrent phenology. The predicted risk on average begins in late August (day $\sim 220\text{--}240$), peaks during September (day $\sim 250\text{--}290$), and decreases in late October (day ≥ 300) and is consistent with field studies [51] (Supplementary X).

Furthermore, analyzing the surface of the Eco-Epi rate, the inter-annual variation is governed by (i) how quickly M rises, (ii) how warm the season remains through the early fall, and (iii) how long the trajectory (M, T) dwells inside the high- λ lobe confirmed by [52, 53]. Independent field modeling from the Coachella Valley may support this persistence mechanism. Ward et al. developed an environmentally informed forecast for the mosquito infection rate (MIR) and showed that a cool, dry winter followed by a warm, wet spring and a relatively cool summer predicts a higher MIR in space and time [8]. Mechanically interpreted, prolonged residence of environmental drivers in conducive states, i.e. longer stretches when temperature-moisture conditions remain favorable, raises MIR, which is conceptually aligned with our dwell time in the high- λ region (longer time inside the favorable band means higher risk).

The Eco-Epi model achieves higher annual summed log-scores than both the null and the NB baselines in nearly every year, including large seasons (e.g., 2015 Eco-Epi markedly outperforms both baselines) and moderate ones (e.g., 2007, 2008, 2013, 2016, 2018–2020, 2024). NRMSE is consistently smallest for Eco-Epi, with the clearest gains in severe outbreak years (e.g., 2015), indicating improved accuracy and sharper timing of the seasonal peak. Taken together, the results support the view that explicitly linking transmission risk to joint ecological drivers (M, T) captures both the late-summer timing and the year-to-year amplitude of Riverside County spillover.

Los Angeles County

The surface of the Eco-Epi rate $\lambda(M, T)$ for Los Angeles County shows that the peak of the spillover rate occurred in a higher range of mosquito abundance than in Orange County and Riverside County. Also in the low abundance range of M practically in the interval of $((1 - 1.3) \times 10^8)$ it shows a noticeable spillover rate (Supplementary III), which counts for late spillovers in Los Angeles.

Throughout the years, Eco-Epi surfaces show a recurring high-risk lobe with a higher abundance of mosquitoes (M) and warm seasonal temperatures ($T \in (12 - 16^\circ\text{C})$). Seasons with the highest predicted risk are those in which the observed trajectory ($M(t), T(t)$) enters the lobe early and remains there the longest during late summer and early fall, most clearly

in 2008, 2012 to 2015, and 2016 to 2017, which also show broad September to October windows with occasional tails into November.

However, the skill of the model in Los Angeles is only modest. The annual log score and NRMSE summaries indicate that Eco-Epi improves on null and NB on average but is not consistently superior. A plausible explanation is the contrast in climatic heterogeneity and ecological diversity between the three counties of California. Los Angeles spans coastal to desert gradients (with elevation and urban heat-island effects), broadening before onset T and diversifying habitats; consequently, a single county-wide (M, T) trajectory can blur the forecast. Operationally, fit the model on the subcounty scale (e.g., coastal versus interior or vector control districts) to tighten the onset windows and improve calibration.

REFERENCES

- [1] U.S. Census Bureau, “Orange County, California,” https://data.census.gov/profile/Orange_County,_California?g=050XX00US06059, 2025, profile page for Orange County, California; accessed 2025-09-12.
- [2] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, “Present and future Köppen-geiger climate classification maps at 1-km resolution,” *Scientific data*, vol. 5, no. 1, pp. 1–12, 2018.
- [3] C. Beck, J. Grieser, M. Kottek, F. Rubel, and B. Rudolf, “Characterizing global climate change by means of Köppen climate classification,” 2006.
- [4] Orange County Health Care Agency, “2014 west nile virus activity report,” <https://ochealthinfo.com/sites/hca/files/import/data/files/102025.pdf>, 2015, county of Orange, Health Care Agency, Epidemiology and Assessment; accessed 2025-06-05.
- [5] U.S. Census Bureau, “Riverside County, California,” https://data.census.gov/profile/Riverside_County,_California?g=050XX00US06065, 2025, profile page for Riverside County, California; accessed 2025-09-12.
- [6] California Department of Public Health, “West Nile Virus Annual Reports,” <https://westnile.ca.gov/>, 2024, accessed 2025-09-12.
- [7] J. E. Kesseli, “The climates of California according to the Köppen classification,” *Geographical Review*, pp. 476–480, 1942.
- [8] M. J. Ward, M. Sorek-Hamer, J. A. Henke, E. Little, A. Patel, J. Shaman, K. Vemuri, and N. B. DeFelice, “A spatially resolved and environmentally informed forecast model of West Nile virus in Coachella Valley, California,” *GeoHealth*, vol. 7, no. 12, p. e2023GH000855, 2023.
- [9] J. Shaman, J. F. Day, and M. Stieglitz, “Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida,” *Journal of medical entomology*, vol. 42, no. 2, pp. 134–141, 2005.
- [10] W. J. Landesman, B. F. Allan, R. B. Langerhans, T. M. Knight, and J. M. Chase, “Inter-annual associations between precipitation and human incidence of West Nile virus in the United States,” *Vector-Borne and Zoonotic Diseases*, vol. 7, no. 3, pp. 337–343, 2007.
- [11] U.S. Census Bureau, “Los Angeles County, California,” https://data.census.gov/profile/Los_Angeles_County,_California?g=050XX00US06037, 2025, profile page for Los Angeles County, California; accessed 2025-09-12.
- [12] Los Angeles County Department of Public Health, “West nile virus surveillance data,” <http://publichealth.lacounty.gov/acd/WNVData.htm>, 2025, accessed 2025-06-05.
- [13] H. Liu and Q. Weng, “Environmental factors and risk areas of West Nile virus in southern California, 2007–2009,” *Environmental Modeling & Assessment*, vol. 17, pp. 441–452, 2012.
- [14] U.S. Census Bureau, “Dallas County, Texas,” https://data.census.gov/profile/Dallas_County,_Texas?g=050XX00US48113, 2025, profile page for Dallas County, Texas; accessed 2025-09-12.
- [15] W. M. Chung, C. M. Buseman, S. N. Joyner, S. M. Hughes, T. B. Fomby, J. P. Luby, and R. W. Haley, “The 2012 West Nile encephalitis epidemic in Dallas, Texas,” *Jama*, vol. 310, no. 3, 2013.
- [16] Dallas County Health and Human Services, “Arbovirus Surveillance Reports,” <https://www.dallascounty.org/departments/dchhs/data-reports/arbovirus-surveillance.php>, 2025, accessed 2025-09-12.
- [17] U.S. Census Bureau, “Harris County, Texas,” https://data.census.gov/profile/Harris_County,_Texas?g=050XX00US48201, 2025, profile page for Harris County, Texas; accessed 2025-09-12.
- [18] M. D. Statkiewicz, R. Talbot, and B. Rappenglueck, “Changes in precipitation patterns in Houston, Texas,” *Environmental Advances*, vol. 5, p. 100073, 2021.
- [19] K. M. Lillibridge, R. Parsons, Y. Randle, A. P. T. Da Rosa, H. Guzman, M. Siirin, T. Wuithiranyagool, C. Hailey, S. Higgs, A. A. Bala *et al.*, “The 2002 introduction of West Nile virus into Harris County, Texas, an area historically endemic for St. Louis encephalitis,” *The American journal of tropical medicine and hygiene*, vol. 70, no. 6, pp. 676–681, 2004.
- [20] Texas Department of State Health Services, “Texas Department of State Health Services,” <https://www.dshs.texas.gov/mosquito-borne-diseases/west-nile-virus>, 2025, accessed 2025-09-12.
- [21] U.S. Census Bureau, “Duval County, Florida,” https://data.census.gov/profile/Duval_County,_Florida?g=050XX00US12031, 2025, profile page for Duval County, Florida; accessed 2025-09-12.
- [22] G. Evensen, “The ensemble kalman filter for combined state and parameter estimation,” *IEEE Control Systems Magazine*, vol. 29, no. 3, pp. 83–104, 2009.

- [23] (2024) Births. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics. [Online]. Available: <https://www.cdc.gov/nchs/fastats/births.htm>
- [24] (2024) Clinical signs and symptoms. Centers for Disease Control and Prevention (CDC). [Online]. Available: <https://www.cdc.gov/west-nile-virus/hcp/clinical-signs/index.html>
- [25] M. S. Shocket, A. B. Verwillow, M. G. Numazu, H. Slamani, J. M. Cohen, F. El Moustaid, J. Rohr, L. R. Johnson, and E. A. Mordecai, “Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23 C and 26 C,” *Elife*, vol. 9, p. e58511, 2020.
- [26] J. Heidecke, J. Wallin, P. Fransson, P. Singh, H. Sjödin, P. C. Stiles, M. Treskova, and J. Rocklöv, “Uncovering temperature sensitivity of west nile virus transmission: Novel computational approaches to mosquito-pathogen trait responses,” *PLOS Computational Biology*, vol. 21, no. 3, p. e1012866, 2025.
- [27] A. C. Brault, S. A. Langevin, R. A. Bowen, N. A. Panella, B. J. Biggerstaff, B. R. Miller, and N. Komar, “Differential virulence of West Nile strains for American crows,” *Emerging infectious diseases*, vol. 10, no. 12, p. 2161, 2004.
- [28] R. G. McLean, “West Nile virus in North American birds,” *Ornithological Monographs*, pp. 44–64, 2006.
- [29] N. B. DeFelice, R. Birger, N. DeFelice, A. Gagner, S. R. Campbell, C. Romano, M. Santoriello, J. Henke, J. Wittie, B. Cole *et al.*, “Modeling and surveillance of reporting delays of mosquitoes and humans infected with West Nile virus and associations with accuracy of West Nile virus forecasts,” *JAMA network open*, vol. 2, no. 4, pp. e193 175–e193 175, 2019.
- [30] Centers for Disease Control and Prevention, “Clinical signs and symptoms — west nile virus,” https://www.cdc.gov/west-nile-virus/hcp/clinical-signs/index.html?utm_source=chatgpt.com, 2025, accessed 2025-10-02.
- [31] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.
- [32] L. P. Campbell, R. P. Guralnick, B. V. Giordano, M. F. Sallam, A. M. Bauer, Y. Tavares, J. M. Allen, C. Efstathion, S. Bartlett, R. Wishard *et al.*, “Spatiotemporal modeling of zoonotic arbovirus transmission in northeastern Florida using sentinel chicken surveillance and earth observation data,” *Remote Sensing*, vol. 14, no. 14, p. 3388, 2022.
- [33] A. B. Wilke, C. Vasquez, J. Medina, A. Carvajal, W. Petrie, and J. C. Beier, “Community composition and year-round abundance of vector species of mosquitoes make Miami-Dade County, Florida a receptive gateway for arbovirus entry to the United States,” *Scientific reports*, vol. 9, no. 1, p. 8732, 2019.
- [34] J. A. Baecher, V. Akshay, R. P. Guralnick, A. M. Bauer, Y. N. Tavares, Y. Sánchez, J. T. Thorson, and L. P. Campbell, “Toward ecological forecasting of west nile virus in florida: Insights from two decades of sentinel chicken surveillance,” *Science of the Total Environment*, vol. 1000, p. 180308, 2025.
- [35] R. A. Soto, “West Nile virus and other domestic nationally notifiable arboviral diseases—United States, 2020,” *MMWR. Morbidity and Mortality Weekly Report*, vol. 71, 2022.
- [36] X. Guo, C. Ma, L. Wang, N. Zhao, S. Liu, and W. Xu, “The impact of COVID-19 continuous containment and mitigation strategy on the epidemic of vector-borne diseases in China,” *Parasites & vectors*, vol. 15, no. 1, p. 78, 2022.
- [37] E. A. Mordecai, J. M. Caldwell, M. K. Grossman, C. A. Lippi, L. R. Johnson, M. Neira, J. R. Rohr, S. J. Ryan, V. Savage, M. S. Shocket *et al.*, “Thermal biology of mosquito-borne disease,” *Ecology letters*, vol. 22, no. 10, pp. 1690–1708, 2019.
- [38] A. T. Ciota, A. C. Matarachiero, A. M. Kilpatrick, and L. D. Kramer, “The effect of temperature on life history traits of *Culex* mosquitoes,” *Journal of medical entomology*, vol. 51, no. 1, pp. 55–62, 2014.
- [39] W. K. Reisen, Y. Fang, and V. M. Martinez, “Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera: Culicidae),” *Journal of medical entomology*, vol. 43, no. 2, pp. 309–317, 2006.
- [40] D. J. Dohm, M. L. O’Guinn, and M. J. Turell, “Effect of environmental temperature on the ability of *Culex pipiens* (Diptera: Culicidae) to transmit West Nile virus,” *Journal of medical entomology*, vol. 39, no. 1, pp. 221–225, 2002.
- [41] Dallas County Health and Human Services, “Arbovirus Surveillance Report,” <https://www.dallascounty.org/departments/dchhs/data-reports/arbovirus-surveillance.php>, 2024, weekly arbovirus surveillance reports and dashboard for Dallas County, TX.
- [42] S. H. Paull, D. E. Horton, M. Ashfaq, D. Rastogi, L. D. Kramer, N. S. Diffenbaugh, and A. M. Kilpatrick, “Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 284, no. 1848, p. 20162078, 2017.
- [43] P. P. Marra, S. Griffing, C. Caffrey, M. A. Kilpatrick, R. McLean, C. Brand, E. Saito, A. P. Dupuis, L. Kramer, and R. Novak, “West Nile virus and wildlife,” *Bioscience*, vol. 54, no. 5, pp. 393–402, 2004.
- [44] K. C. Poh, M. C. Medeiros, and G. L. Hamer, “Landscape and demographic determinants of *Culex* infection with west nile virus during the 2012 epidemic in dallas county, tx,” *Spatial and Spatio-Temporal Epidemiology*, vol. 33, p. 100336, 2020.
- [45] National Weather Service, “National Weather Service,” <https://www.weather.gov/>, accessed: 2025-09-26.
- [46] L. Rueda, K. Patel, R. Axtell, and R. Stinner, “Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae),” *Journal of medical entomology*, vol. 27, no. 5, pp. 892–898, 1990.
- [47] D. Martinez, K. O. Murray, M. Reyna, R. R. Arafat, R. Gorena, U. A. Shah, and M. Debboun, “West Nile virus outbreak in Houston and Harris county, Texas, USA, 2014,” *Emerging infectious diseases*, vol. 23, no. 8, p. 1372, 2017.

- [48] Centers for Disease Control and Prevention, “West Nile Virus: Surveillance and control guidelines,” U.S. Centers for Disease Control and Prevention, Atlanta, GA, Tech. Rep. CS 342983-A, Aug. 2024, accessible web version: <https://www.cdc.gov/west-nile-virus/php/surveillance-and-control-guidelines/index.html>. [Online]. Available: https://www.cdc.gov/west-nile-virus/media/pdfs/2024/08/WestNileVirus-SurveillanceControlGuidelines_508-h.pdf
- [49] Orange County Health Care Agency, Communicable Disease Control, “West Nile virus, Orange County 2011–2018,” <https://ochealthinfo.com/sites/hca/files/import/data/files/102025.pdf>, Santa Ana, CA, 2019, county summary of WNV activity; seasonal peak in Aug–Sep with substantial activity in Jul and Oct (see Fig. 2).
- [50] Orange County Mosquito and Vector Control District, “Emergency response plan,” Orange County Mosquito and Vector Control District, Garden Grove, CA, Tech. Rep., 2015, includes average monthly *Culex quinquefasciatus* abundance and human WNV onset timing used for seasonal context. [Online]. Available: https://www.ocvector.org/files/adb1334da/OCMVCDMERGENCY2015_8_20_15FINAL.pdf
- [51] R. E. Snyder, T. Feiszli, L. Foss, S. Messenger, Y. Fang, C. M. Barker, W. K. Reisen, D. J. Vugia, K. A. Padgett, and V. L. Kramer, “West Nile virus in California, 2003–2018: A persistent threat,” *PLoS neglected tropical diseases*, vol. 14, no. 11, p. e0008841, 2020.
- [52] A. M. Kilpatrick and W. J. Pape, “Predicting human West Nile virus infections with mosquito surveillance data,” *American Journal of Epidemiology*, vol. 178, no. 5, pp. 829–835, 2013.
- [53] C. M. Barker, B. F. Eldridge, and W. K. Reisen, “Seasonal abundance of *Culex tarsalis* and *Culex pipiens* complex mosquitoes (Diptera: Culicidae) in California,” *Journal of medical entomology*, vol. 47, no. 5, pp. 759–768, 2010.