Decadal Weather Trends and Their Influence on West Nile Virus–Mosquito Occurrence and Public Health in Northern Italy:

A Distributed Lag Non-Linear Modeling Approach

Masterthesis

---

in the study program
'M.Sc. Applied Statistics'
(Prof. Dr. Thomas Kneib,
University of Göttingen)

by
Laura Böhme
laura.boehme@stud.uni-goettingen.de
24225289

at the
University of Göttingen
Department of Medical Statistics
Professorship for Computational Statistics
(Prof. Maike Hohberg)

in colaboration with
Karolinska Institutet Stockholm
Department of Global Public Health
Prime research group
(Assistant Prof. Elena Raffetti)

Submission date: 31/01/2026
Supervisors: Prof. Maike Hohberg, Assistant Prof. Elena Raffetti

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I'd like to express my thanks to my patient and supportive supervisors: Assistant Prof. Elena Raffetti, who made this project application possible, and invited me to a lot of new interesting meetings and gave me the opportunity to learn more and discover new interesting perspectives. Prof. Dr. Maike Hohberg, who was super supportive of starting this project, gave me valuable guidance, constructive feedback and supported me throughout this thesis.

I would also like to extend my special thanks to the PhD students Shivang Pandey, Manuel Marcellini O Nocentini, and Ricardo Berenguer Navarro and Markus Simon for their support of my project and great time in Stockholm.

To my parents and grandparents, I am very grateful for their constant encouragement, their intense support during the proof-reading phase, during my research trip to Stockholm and for standing by me during this whole exciting phase of my academic journey. I especially wish I could share this step with my other grandmother.

This whole journey would not have been possible without my friends, who supported me in countless ways throughout my time in Göttingen as well as in Stockholm.
To those who enrich my life and will do so in the chapters ahead, who made the different parts of this master's program feel light and inspiring, who supported me during the master and thesis process with encouragement and well-timed distractions.

Above all, to those who have always been there and continue to believe in me: thank you.

1 Introduction

The effects of climate change on human health are more and more perceptible. The global warming directly influences disease transmission dynamics through the geographical and seasonal spread of disease vectors. In the recent years, the West Nile Virus (WNV) has spread in multiple European countries, including Italy, causing hundreds of human infections. While the individual impacts of climate change on human health and mosquito distribution have been individually well-researched, research that links these elements remains limited.

The European Centre for Disease Prevention and Control (ECDC, 2025) released in August of 2025 a press statement stating that in Europe mosquito-borne diseases are getting more common and transmission timewindows are enlarging. WNV is getting detected in new areas every year and expected to rise in numbers in upcoming years. Therefore, intensifying research in the field of vector-borne diseases is crucial as this is becoming a major concern, especially without a vaccine solution (European Centre for Disease Prevention and Control (ECDC), 2021).
Various mosquito species act as vectors, with the widely distributed *Culex pipiens* mosquitoes in Europe (European Centre for Disease Prevention and Control (ECDC), 2021). Reported symptoms are e.g. headache, fever, rash and fatigue. In approximately 1% of the human cases, the nervous system can be affected and West Nile Neuroinvasive Disease (WNND) is caused (Brown et al., 2007; Nash et al., 2001; Orton et al., 2006; Zou et al., 2010).

This project was conducted within the Global Public Health Department, Karolinska Institute, Stockholm in collaboration with Assistant Prof. Elena Raffetti, with the aim of combining the climate and health perspective in the study of the WNV disease.

This thesis investigates the relationship between climate change, the distribution of mosquitoes infected with WNV and the effects on human health in Northern Italy over a 10-year period from May 2014 to October 2024. The region of interest is Emilia-Romagna. The effect of climate on the changes of mosquito populations and on public health, specifically WNND, is examined. A distributed lag non-linear modeling (DLNM) approach (Gasparrini et al., 2010) is employed, embedded within a generalized additive model (GAM), to examine complex exposure-response relationships (Zanobetti et al., 2000). This method accounts for time-delayed and non-linear predictor-response relationships. DLNMs use flexible, non-linear cross-basis functions to describe how predictor effects vary across exposure levels and lag periods, while the GAMs framework facilitates the inclusion of smooth temporal trends.

In existing literature, the delay between climate factors and the implication for mosquito occurrence and infection are widely discussed. Meta-analyses have identified a general pattern of temperature delay, above 6 weeks for dengue fever, above 7 weeks for malaria,

and even further delays for vector-borne diseases, highlighting the importance to account for delayed responses (Wang et al., 2021).

Previous studies looked at infections rates of WNV depending on delayed temperature and precipitation (Cotar et al., 2016) with zero-inflated binomial models. Since the introduction of DLNMs, they have been applied in environmental epidemiology, particularly to assess the effects of temperature and air pollution on mortality (Schwartz, 2000; Armstrong, 2006; Braga et al., 2001; Curriero et al., 2002). These studies have demonstrated the suitability of DLNMs for capturing delayed and cumulative effects that are not properly addressed by other traditional regression approaches.

By embedding the DLNM within a GAM, this work extends the generalized additive distributed linear model (GADLM) (Zanobetti et al., 2000) framework to a generalized additive distributed lag non-linear model (GADLNM), thereby incorporating non-linearity into the embedding.

Building on the existing literature, this project integrates climatic variables with vector-borne disease dynamics, especially with focus on climate change. Regions of Southern Europe, where these diseases are increasingly emerging as a health concern, are particularly interesting. The analysis jointly evaluates climate-driven patterns of mosquito occurrence and associated human health outcomes applying the DLNM framework. In addition, the study presents a detailed description of the model selection process, the included lag decision and basis specifications.

2 West Nil Virus Disease

West Nile Virus (WNV) has spread in various European countries, including Italy, causing hundreds of animal and human infections. It's an important public health concern with no vaccine solution but only supportive care (European Centre for Disease Prevention and Control (ECDC), 2021). The WNV is a flavivirus that was first isolated in Uganda in 1937 (Smithburn et al., 1940). Now, it is spread in Europe and North America.

Through bird-mosquito transmission that mostly involves *Culex pipiens* mosquitoes, a cycle is maintained (Zeller & Schuffenecker, 2004; Garrigós et al., 2023). Outbreaks in humans and horses, which are hosts in the transmission cycle, are connected with the lineage I and II of WNV. Lineage I is found in Africa, India, and Australia and this lineage is also responsible for the outbreaks in Europe and the Mediterranean basin. Lineage II is only detected in sub-Saharan Africa. *Culex pipiens* follow a four-stage life cycle: egg, larva, pupa and adult. The development from egg to adult typically takes about 7 to 10 days. Eggs are laid in stagnant water, such as ponds or rain tanks, and hatch within 48 hours. Both larvae and pupae live in the water. After approximately 5 days in the larval stage and then further $2-3$ days as pupae, adult mosquitoes emerge and become capable of flying (Centers for Disease Control and Prevention (CDC), 2022; SurveillanceMoustiques, 2026). Adult mosquitoes generally live for 2 to 4 weeks. Adult females feed on the blood of humans and animals, in order to get the nutrients that are required for egg production and subsequent laying new eggs (Centers for Disease Control and Prevention (CDC), 2022).

WNV is transmitted to humans through the bite of infected mosquitoes. Infected birds support the grow in numbers (ECDC, 2024), as mosquitoes are infected by them. While mosquito-borne transmission is the most common route, WNV can also be transmitted through blood donations or transplantation of cells, tissues or organs from infected donors. During pregnancy and breast-feeding the disease can be transmitted as well (Nanni Costa et al., 2011).

Human infections are asymptomatic for 80%. About 20% develop symptoms, including: headache, fever, myalgia, vomiting, rash, fatigue and eye pain. Acute WNV fever is defined as at least one of the four following criteria fever, enzephalitis, meningitis or rash (Robert-Koch-Institut, 2023). Approximately, in 1% of the human cases, the nervous system can be affected and West Nil neuroinvasive disease (WNND) is caused. Among people with WNND the mortality rate is up to 17% (Brown et al., 2007; Nash et al., 2001; Orton et al., 2006; Zou et al., 2010). The number of cases increased over the recent years and has been reported in an expanding number of regions, including new affected regions every year. The ECDC (2025) states that the WNND cases are overrepresented as they have severe symptoms and are likelier getting diagnosed than asymptomatic patients.

3 Project Workflow

The process of this project is mirrored in the structure of this thesis.

At first, a literature research in the field of WNV disease was conducted (Section 2). Section 4 introduces the data that reflect the three pillars of this project: Mosquito Surveillance, Weather Trend and Health Data. Further descriptive details on the datasets can then be found in the Section 6.1.1, 6.1.2 and 6.1.3.

The methodology is outlined in Section 5. The foundation for modeling delayed effects and the Distributed Lag Model (DLM) is explained. By extending this concept, the more flexible approach to capture non-linear relations, Distributed Lag Non-Linear Model (DLNM) is introduced. In this work, the DLNM is then integrated in a Generalized Additive Modeling (GAM) framework to achieve a Generalized Additive Distributed (Non-)Linear Model (GADLNM). The Section 5.6 describes the factors that drive the model selection process when selecting the best fitting model for each outcome and climate variable. The implementation in `R` is also shown.

In Chapter 6 the results are presented. The datasets are described in detail (Section 6.1.1 to 6.1.3). The rest of the chapter is subdivided into sections according to the different outcomes. The aim is to get a structured overview over the results. Section 6.2 reports the results for the mosquito outcomes: *WNV positive trap result* and *number of mosquitoes tested*. Section 6.3 reports the health outcomes: *WNND when only including the weather trend perspective* and then also *incorporating the mosquito perspective to explain WNND*. Finally, all results will be discussed, with a consideration of their potential to impact the developing research of the interaction of climate, vector-borne diseases and public health implications (Section 7 and 8).

4 Data

The required data comprises the mosquito surveillance data which is the basis for this project, weather trend data and health data.

4.1 Mosquito Surveillance Data

The mosquito data originates from Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IZSLER) (2025). The aim is the detection of the virus' circulation within the considered region of Emilia-Romagna covering an area of 22,123 km^2. The given data originally included 119 traps (in the region Emilia-Romagna) and 82 traps (in the region of Lombardia). Latter got later excluded due to the unavailability of matching health records and due to the large differences in the collection procedure. Included information are the location of the traps, the amount of mosquitoes that the trap caught and the amount that were tested. If at least one mosquito was tested positive

for having WNV, the result is marked as positive for a specific species in a specific trap for a specific day.

The final mosquito surveillance data included 119 different coordinates (traps) over a 10-year time period between 2014 and 2024. The traps were working everyday, from June until last week of September or October, and overnight, roughly from 5:00 p.m. to 9:00 a.m and collected in the morning. In recent years (since 2020), the Emilia-Romagna region started collecting data from the beginning of May. The geo-referenced stations used attractive traps by carbon dioxide, CO_2-baited suction traps, and were sampled every two weeks (Calzolari et al., 2020; Marini et al., 2020). The species of the sampled mosquitoes were identified and pooled. A maximum of 200 specimens per pool were collected and then screened for WNV detection via PCR (Marini et al., 2020; European Food Safety Authority, 2018).
The dataset was filtered for the species *Culex Pipiens* (Zeller & Schuffenecker, 2004; Garrigós et al., 2023) and reduced to the region of Emilia-Romagna. This allows for combining the mosquito and health data while also guaranteeing a methodological consistency for the collection method throughout the dataset.

Figure 1 depicts the Emilia-Romagna region in Northern Italy and the spatial distribution of the traps, showing that the traps were primarily located in the Northern part of the region, this also applies for each province individually. Except for Ferrera (most North-Eastern province in Emilia-Romagna), where traps were more evenly spread over the entire province. The first collection date was the 2014-05-27 and the last reported date was the 2024-10-17 with a median number of 100 ± 61 mosquitoes (50 and 200 most observed) per trap per measurement.

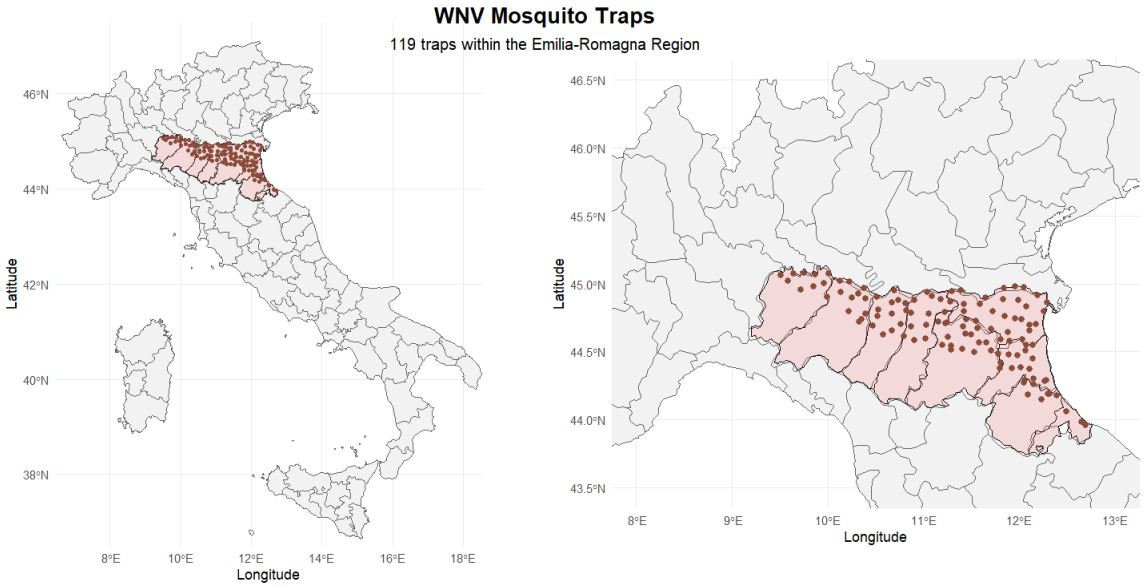


Figure 1: *Map of Italy and traps within the region Emilia-Romagna (highlighted red)*

4.2 Weather Trend Data

The weather trend data incorporates the key climate variables land surface temperature and precipitation over the considered decade (2014 - 2024) to capture changes of the weather over time. The datasets used, originate from different sources (Wan et al., 2021; Funk et al., 2015) all were accessed via Google Earth Engine (Gee) (Gorelick et al., 2017). For the purpose of this study, spatial resolution is more critical than high temporal resolution. Precise geographical accuracy is essential to link environmental conditions with specific trap locations. A spatial resolution between 1 km and 5 km is therefore appropriate. 8-Days or 1-Day data provide an adequate precision for the analysis.

This work includes multiple datasets covering the weather trend aspects. Table 1 gives an overview over the datasets, resolution and used variables.

Table 1: *Overview over the included Weather Trend Datasets*

Dataset	**Availability**	**Resolution**		**Variable**
		Temporal	**Spatial**	
MODIS MOD11A2	2000 to 2025[1]	8-Day	1km	temperature
CHIRPS	1981 to 2025[3]	1-Day	5.5km	precipitation

1: 2000-02-18 to 2025-08-21; **3**: 1981-01-01 to 2025-07-31

The Moderate Resolution Imaging Spectroradiometer (Modis) datasets originate from the Terra (Eos Am-1) and Aqua (Eoc Pm-1) satellites from NASA Goddard Space Flight Center (2025). The element used from the Modis-Mod11a2 (Wan et al., 2021) is the land surface temperature, covering the years 2000 to 2025 with a temporal resolution of 8-Day and a spatial resolution of 1km.
The Climate Hazards group Infrared Precipitation with Stations (Chirps) is a quasi-global rainfall dataset with data covered from 1981 to the present. The data has a spatial resolution of 5.5km and is measured every day (Funk et al., 2015).

4.3 Health Data

The health data comes from the Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IZSLER) (2025) and are available exclusively for the Emilia-Romagna region. The dataset contains individual case records. Included are also detailed information about each patient such as sex, region, age in years, case classification, date of symptom onset, municipality of residence, and death within 30 days. It is also differentiated between Wnv-fever, Wnnd and Wnv-asymptomatic/blood donors. The latter describes patients that were infected by the Wnv but were asymptomatic and were mostly only discovered or tested in the context of a blood donation.

5 Methodology

Environmental stressors often show effects that are delayed in time. To capture the delayed health impacts of those long-term effects, a distributed lag modeling approach was employed. First, the lag structure of exposures was defined to describe how effects accumulate over time. Building on this, a Distributed Lag Model (DLM) was applied to estimate linear lagged associations and then extended to a Distributed Lag Non-linear Model (DLNM) to account for non-linear exposure–response relationships. Finally, this approach was integrated within a Generalized Additive Model (GAM) framework to allow flexible adjustment for confounding factors and outcome distributions. This leads to a Generalized Additive Distributed Lag Non-Linear Model (GADLNM), which is an extenstion of the Generalized Additive Distributed Lag Model (GADLM). The described models build on each other and form the basis for the implementation later.

5.1 Climate Lag Definition

Climate factors such as temperature and precipitation do not only affect the health event, such as diseases or death, on the same day but often occur with a delay (e.g. several days or weeks later). Ignoring lags would underestimate or miss-attribute the true effect of the climate factor on the outcome. The climate lag is a time lag defined as a delay between cause and effect of climate impacts (Wang et al., 2021).

A classic example of lagged climate effects is the relationship between carbon dioxide emissions into the atmosphere and the effect on the climate system. Carbon dioxide is released into the atmosphere while a small amount is absorbed by ocean water and enters the carbon cycle directly. The full effect of the release of the carbon dioxide comes only present over time. Thus, it must be accounted for the delayed effects in order to correctly estimate long-term effects (Price, 2024; Mulhern, 2020; Schwartz, 2000).

We distinguish infinite and finite distributed lags (Berry & Wei, 2013). An infinite DLM assumes that effects decay indefinitely over time and is therefore approximated by transformations. On the other hand, a finite DLM specifies a maximum lag L and the model explicitly includes lags from 0 to L, forming the basis for the analysis. In the following, the maximum lag must be specified prior to estimation based on clinical expertise (Parker, 2013).

According to the literature, the lag is usually determined empirically or chosen based on earlier publications. The two most commonly used information criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Gasparrini, 2014).

$$AIC = 2K - 2\ln(\hat{\mathcal{L}}) \tag{1}$$

$$BIC = K\ln(T) - 2\ln(\hat{\mathcal{L}}) \tag{2}$$

where T is the total number of observations, K the number of model parameters and $\hat{\mathcal{L}}$ is the estimated likelihood for the respective model. By minimizing the criteria AIC or BIC the optimal model can be selected (Gasparrini, 2014).

5.2 Distributed Lag Model

If a linear relationship between predictor and outcome can be assumed, the long-term effects on the response can be described through a DLM. A finite DLM can be written as:

$$y_t = \alpha + \sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} + \sum_{k=1}^{K} \delta_k u_{tk} + \epsilon_t \tag{3}$$

The coefficients β_ℓ are the unknown distributed lag coefficients, which define how the predictor x affects y over time, $x_{t-\ell}$ are the lagged values of the explanatory variable such as temperature or precipitation. The outcome at a given time t may therefore be explained in terms of past exposures $x_{t-\ell}$, where $\ell = [0, \ldots, L]^T$ is the lag vector. L ($< \infty$) defines the maximum lag and ϵ_t the error term. The model assumes that the error terms are independent and identically distributed, with a mean of zero and constant variance, conditional on the predictors included in the model. Other K linear predictors are also included through u_{tk}. Their effects are analogously estimated as the fixed-effect coefficients in mixed models. In the following, however, we ignore the addition of other linear predictors for the sake of easier explanation. Only in Section 5.5 the inclusion of other linear and non-linear predictors into the model are considered and further explained.

$$y_t = \alpha + \sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} + \epsilon_t \tag{4}$$

As described in Gasparrini et al. (2010), the model builds upon a basic time series analysis. Commonly, the outcome y_t is considered to be daily counts from an over-dispersed Poisson distribution. In the past, applications of this model approach focused primarily on short-term effects of climatic factors on health outcomes. Considered factors were air pollution (Schwartz, 2000; Braga et al., 2001) or temperature (Braga et al., 2002). The former studies examined lagged associations between air pollution and daily mortality, with attention to cardiovascular deaths, while the latter investigated the relationship between temperature, and additionally humidity, on mortality. These studies were conducted using data from the United States.

The exposure history at time t can be written as the vector $q_{t.} = [x_t, \ldots, x_{t-L}]^T$. Thus, a transformation of the original n observations leads to the derived lagged matrix $\mathbf{Q}$, which has dimensions $n \times (L + 1)$ and with vector $q_{.1}$ being the first column of the matrix $\mathbf{Q}$ (Gasparrini et al., 2010).

(4) can be further written as:

$$g(\mu_t) = \alpha + \sum_{\ell=0}^{L} s_\ell(x_{t\ell}; \boldsymbol{\beta}) + \epsilon_t \tag{5}$$

where $\mu = E(Y)$ and g is a monotonic link function and the distribution of Y is assumed to belong to the exponential family, like counts from an over-dispersed Poisson distribution or a binary outcome. The smooth function $s()$ is introduced to describe the relationship between response and the main predictor $x_{t\ell}$, defined by the parameters β_ℓ and to further filter out noise and better represent realistic trends in the data.

The function $s()$ can also be represented in matrix notation:

$$s(x_t; \boldsymbol{\beta}) = \boldsymbol{q}_{t.}^T \boldsymbol{C} \boldsymbol{\beta}, \tag{6}$$

where $\mathbf{C}$ is an $(L+1) \times v_\ell$ matrix of basis variables derived from the basis functions v_ℓ to the lag vector (Gasparrini et al., 2010). In order to describe the effects of lagged exposures the matrix $\mathbf{C}$ can be defined. The relationship can be specified through the choice of a basis and the related basis functions. The transformation of the original variable x is therefore known. Typically, splines or polynomials are used here. Non-parametric methods based on generalized additive models can also be introduced here, but for now, we focus on the named parametric and semi-parametric approach. This leads to a better modeling of time dependence as the relation between exposure and effect is allowed to vary over time (Gasparrini et al., 2010).

To summarize, no classic regression can be applied as the outcome is not normally distributed and the predictors are dependent on each other. Therefore the DLNM is needed.

5.3 Basis functions

To specify the lagged-exposure relationship, the choice of basis functions used to model the lag coefficients, is crucial. These functions determine flexibility and smoothness of the lag structure. Polynomial or spline-based approaches are common choices, each differing in how they balance global fit and local adaptability of the lag structure.

5.3.1 Almon Polynomials

The different lag-coefficients β_ℓ need to be estimated for $\ell = 0, \ldots, L$. Estimating all $L+1$ lag coefficients $\beta_0, \beta_1, \ldots, \beta_L$ does not only have a high model complexity but also entails a risk of overfitting. Moreover, the lags are highly correlated, since each variable x_t is determined by $x_{t-\ell}$, resulting in severe multicollinearity. The solution is to fit a smooth polynomial function of degree p over ℓ. The Almon polynomial mitigates this multicollinearity by imposing the smooth polynomial structure on the lag coefficients.

The Weierstrass (1885) approximation theorem says: *Every continuous function defined on a closed interval [a,b] can be uniformly approximated, arbitrary closely by a polynomial function of finite degree p.* Based on this theorem, Almon (1965) suggests that the lag-coefficients can also be approximated through a polynomial of order p. Each lag-coefficient β_ℓ can be represented by a polynomial of degree $p < L$ (Almon, 1965):

$$\beta_\ell = \gamma_0 + \gamma_1 \ell + \gamma_2 \ell^2 + \ldots + \gamma_p \ell^p = \sum_{i=0}^{p} \gamma_i \ell^i \tag{7}$$

This representation (7) can be substituted into (4) (Giles, 2017):

$$\begin{aligned} y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_L x_{t-L} + \epsilon_t \\ &= \alpha + (\gamma_0) x_t + (\gamma_0 + \gamma_1 + \ldots + \gamma_p) x_{t-1} + (\gamma_0 + 2\gamma_1 + \ldots + 2^p \gamma_p) x_{t-2} + \ldots + \\ &\quad (\gamma_0 + L\gamma_1 + \ldots + L^p \gamma_p) x_{t-L} + \epsilon_t \\ &= \alpha + \gamma_0 \boxed{(x_t + x_{t-1} + x_{t-2} + \ldots + x_{t-L})} + \gamma_1 \boxed{(x_{t-1} + 2x_{t-2} + \ldots + L x_{t-L})} + \ldots + \\ &\quad \gamma_p \boxed{(x_{t-1} + 2^p x_{t-2} + \ldots + L^p x_{t-L})} + \epsilon_t \end{aligned} \tag{8}$$

The highlighted (boxed) elements are defined as g_{it} with the according coefficients γ_i:

$$\gamma_i g_{it} = \gamma_i \begin{cases} \sum_{l=0}^{L} x_{t-\ell} & \text{for } i = 0 \\ \sum_{l=1}^{L} \ell^i x_{t-\ell} & \text{for } i \neq 0 \end{cases} \tag{9}$$

with $i = 0, 1, \ldots, p$ and $t = 0, 1, \ldots, T$. Thus, it follows:

$$y_t = \alpha + \gamma_0 \boxed{g_{0t}} + \gamma_1 \boxed{g_{1t}} + \ldots + \gamma_p \boxed{g_{pt}} + \epsilon_t, \tag{10}$$

This approach is called polynomial Dlm (Pdlm) (Zanobetti et al., 2000) as the lag coefficients are imposed to follow a polynomial function. Thus, the lag is not randomly but systematically polynomial distributed.

With $p < L$, the number of coefficients to be estimated is drastically reduced, leading also to a reduced model's complexity. Only $p + 1$ coefficients need to be estimated, compared to $L + 1$ coefficients in the initial distributed lag model (4), with $p + 1 < L + 1$.

The choice of the polynomial degree is not determined by the model itself but needs to be predefined (Giles, 2017). Choosing the degree p leads to a problem of model-selection, potentially introducing bias or inconsistent predictors. Moreover, the polynomial assumption is restrictive. A polynomial degree that is too high can result in overfitting and unstable estimates, while a degree that is too low may lead to underfitting and misspecification bias. Almon polynomials do not perform very well at the extremes. The assumption of the lags being polynomial distributed is the first step to outline non-linear relations. Through the more flexible approach of regression splines and penalized-splines a few disadvantages can be outweighed (Giles, 2017).

5.3.2 Regression Splines and P-Splines Expansion

Introducing splines as the underlying structure for the lag coefficients allows the lag-response curve to fit more closely and flexible at a local level. On the contrary, the Almon polynomials impose a global functional form, which may fail to capture local structures of the lag structure. Splines approximate the function based on piecewise low-degree polynomials, enabling a closer fit to the data.

By adding a truncated power basis, a regression spline function of ℓ looks like:

$$\beta_\ell = \underbrace{\sum_{i=0}^{p} \gamma_i \ell^i}_{\text{global polynomial trend}} + \underbrace{\sum_{m=1}^{M} v_m (\ell - \kappa_m)_+^p}_{\text{local spline deviations}} \tag{11}$$

where $\kappa_1, \ldots, \kappa_M$ is a set of M distinct knots between 0 and L. So, κ_m are joint points or knots, as β_ℓ is a piecewise pth degree polynomial and $(\ell - \kappa_m)_+ = max(0, \ell - \kappa_m)$ (Zanobetti et al., 2000).

This expression enables the construction of a flexible, smooth function that can change slope or curvature at knots. It defines a piecewise polynomial that is zero before the knot and grows polynomial after, weighted by the coefficients v_m.

As regression splines can overfit local fluctuations in the data and lead to a wiggly function that is not good generalizable, a penalty term can be added, resulting in a Penalized Spline. With the addition of the penalty term $\lambda \sum_{k=1}^{K} v_k^2$ and β_ℓ as defined in (11) the estimation of the coefficient is (Zanobetti et al., 2000):

$$\text{minimize} \sum_{t=L+1}^{T} \left(y_t - \alpha - \sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} \right)^2 + \lambda \sum_{m=1}^{M} v_m^2 \tag{12}$$

The coefficients β_ℓ as defined in (11) are now substituted in $\sum_{\ell=0}^{L} \beta_\ell x_{t-\ell}$:

$$\begin{aligned}
\sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} &= \sum_{\ell=0}^{L} \left(\sum_{i=0}^{p} \gamma_i \ell^i + \sum_{m=1}^{M} v_m (\ell - \kappa_m)_+^p \right) x_{t-\ell} \\
&= \sum_{\ell=0}^{L} \left(\sum_{i=0}^{p} \gamma_i \ell^i \right) x_{t-\ell} + \sum_{\ell=0}^{L} \left(\sum_{m=1}^{M} v_m (\ell - \kappa_m)_+^p \right) x_{t-\ell} \\
&= \sum_{\ell=0}^{L} \gamma_i \sum_{i=0}^{p} \ell^i x_{t-\ell} + \sum_{m=1}^{M} v_m \sum_{\ell=0}^{L} (\ell - \kappa_m)_+^p x_{t-\ell}
\end{aligned} \tag{13}$$

Substituting this into the optimization problem (12) follows:

$$\text{minimize} \sum_{t=L+1}^{T} \left[y_t - \alpha - \sum_{\ell=0}^{L} \gamma_i \sum_{i=0}^{p} \ell^i x_{t-\ell} + \sum_{m=1}^{M} v_m \sum_{\ell=0}^{L} (\ell - \kappa_m)_+^p x_{t-\ell} \right]^2 + \lambda \sum_{m=1}^{M} v_m^2 \tag{14}$$

With the help of the penalty term λ overfitting and smoothness can be controlled and coefficients can be shrunk. As $\lambda \to \infty$, the penalty increases and the function approaches

a simple polynomial, ignoring the knots and with $\lambda \to 0$ no penalty is applied, resulting in the initial spline.
The coefficients β_ℓ are expressed as a linear combination of basis functions. While the basis functions provide flexibility in modeling complex shapes, the estimation remains linear in the parameters. This allows the model to capture flexible yet smooth lag structures by combining the basis functions with appropriate coefficients, ensuring smoothness and interpretability simultaneously.

When the degree in (11) is set to $p = 3$, the resulting basis corresponds to a cubic regression spline. The lag coefficients can be written as

$$\beta_\ell = \sum_{i=0}^{3} \gamma_i \ell^i + \sum_{m=1}^{M} v_m (\ell - \kappa_m)_+^3 \tag{15}$$

$$= \gamma_o + \gamma_1 \ell + \gamma_2 \ell^2 + \gamma_3 \ell^3 + \sum_{m=1}^{M} v_m (\ell - \kappa_m)_+^3 \tag{16}$$

To obtain a *natural* cubic spline, additional boundary constraints are imposed. Specifically, the spline is required to fulfill

$$\beta''(\ell) = 0 \text{ for } \ell \leq \kappa_1 \text{ and } \ell \geq \kappa_M \tag{17}$$

which enforces linear behavior beyond the boundary knots and eliminates curvature at the extremes of the lag range.

Natural cubic splines are a popular choice for representing lag structures and exposure-response associations due to their smoothness and stable boundary behavior.

For computational convenience, the problem can be expressed in matrix notation with

$$\boldsymbol{\theta} = [\alpha,\, \gamma_0, \gamma_1, \ldots, \gamma_p,\, v_1, \ldots, v_K]^\top = [\alpha, \boldsymbol{\gamma}, \boldsymbol{v}]^\top \tag{18}$$

and

$$\boldsymbol{X} = [\mathbf{1}, \boldsymbol{Z}, \boldsymbol{W}] \tag{19}$$

with $\boldsymbol{Z}_t\boldsymbol{\gamma} = \sum_{i=0}^{p} \gamma_i \sum_{\ell=0}^{L} \ell^i x_{t-\ell}$ and $\boldsymbol{W}_t\boldsymbol{v} = \sum_{m=1}^{M} v_m \sum_{\ell=0}^{L} (\ell - \kappa_m)_+^p x_{t-\ell}$. The model is simply $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$. To penalize the spline coefficients, the penalized matrix $\mathbf{P}$, here the identity matrix, is used. The smoothness is controlled by shrinking the spline coefficients. A ridge-type penalty is applied only to the spline coefficients v_m, while the intercept and polynomial terms remain unpenalized.

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{0}_{(p+1)\times(p+1)} & 0 \\ 0 & 0 & \boldsymbol{I}_{M\times M} \end{bmatrix} \in \mathbb{R}^{(p+2+M)\times(p+2+M)} \tag{20}$$

Therefore, it follows the final matrix optimization problem is:

$$\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|^2 + \lambda^2 \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta} \tag{21}$$

$$\text{and with } \hat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda^2 \boldsymbol{P})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \tag{22}$$

The parameters are the estimated coefficients for the intercept, polynomial and penalized spline.

5.4 Distributed Lag Non-Linear Model

The presented DLM assumes a delayed linear relation between exposure and response. However, environmental factors often exhibit a non-linear and delayed effect. This delayed non-linear exposure-response dependency is realized through a Distributed Lag Non-Linear Model (DLNM) (Gasparrini et al., 2010). Unlike DLMs, which assume a linear exposure-response relation, DLNMs allow the exposure-response association now to be non-linear. This is crucial when the considered climatic factors vary across intensity and time since exposure.

DLMs can describe the lag structure of linear effect and the paper from Gasparrini et al. (2010) extended this to models that can describe non-linear effects. DLNMs also consider the fact that the model needs to explain an additional time dimension. The two dimensions that are combined in a cross-basis function are the exposure-response association (e.g. temperature and health event) and the lag-response association (delayed effects of e.g. temperature on the outcome). Conceptually, the cross-basis is a matrix that incorporates both dimensions as a combination of the two according basis functions: one for the exposure dimension and one for the lag dimension. The matrix can be written as:

$$\text{cross-basis} = \left[\begin{array}{c|cccc} & \text{Lag0} & \text{Lag1} & \text{Lag2} & \text{Lag3} \\ \hline 20°\text{C} & a_{10} & a_{11} & a_{12} & a_{13} \\ 21°\text{C} & a_{20} & a_{21} & a_{22} & a_{23} \\ 22°\text{C} & a_{30} & a_{31} & a_{32} & a_{33} \\ 23°\text{C} & a_{40} & a_{41} & a_{42} & a_{43} \end{array}\right] \in \mathbb{R}^{n \times (L+1)},$$

here $n = 4$ is the number of observations, and $L = 3$, the maximum lag length. Each entry a_{ij} represents a unique combination of the ith exposure value and the jth lag, determined by the respective basis functions. It follows that a_{10} represents the basis function combination for 20°C and lag 0, and a_{43} the combination of the basis function for 23°C and lag 3. This matrix is then later included as a predictor, allowing the final model to capture the non-linear shape of the exposure-response relationship and the time course over which the effect of the exposure persists.

Based on the Formula (6) the cross-basis is formally defined as

$$\boldsymbol{W} = \boldsymbol{QC}. \tag{23}$$

$\mathbf{W}$ is the matrix of the transformed variables v_ℓ, allowing for the estimation of the parameter $\boldsymbol{\beta}$. $\mathbf{Q}$ is the matrix of $q_{t.}$. The matrix $\mathbf{C}$ is the matrix of basis variables that were derived from the specific basis functions that were applied to the lag vector.

With that the DLNM can be written as (Gasparrini et al., 2010):

$$s(x_t; \boldsymbol{\eta}) = \sum_{j=1}^{v_x} \sum_{k=1}^{v_\ell} \boldsymbol{r}_{tj.}^T \boldsymbol{c}_{.k} \eta_{jk} = \boldsymbol{w}_{t.}^T \boldsymbol{\eta} \tag{24}$$

where $\boldsymbol{\eta}$ is the corresponding parameter vector, $\mathbf{r}_{tj.}$ equals a vector of lagged exposures for the time t transformed through the jth basis function. $\mathbf{C}$ is the matrix of basis variables derived from the application of the specific basis functions to the lag vector ℓ.

Formally, the DLNM can be written as:

$$y_t = \alpha + \sum_{\ell=0}^{L} f(x_{t-\ell}, \ell) + \sum_{k=1}^{K} \delta_k u_{tk} + \epsilon_t \tag{25}$$

where $f(x, \ell)$ is the smooth non-linear function of both the exposure x and lag ℓ. This can be decomposed into basis functions over exposure and lag dimensions, e.g., spline bases:

$$y_t = \alpha + \sum_{\ell=0}^{L} \sum_{k=1}^{K} \theta_k B_k(x_{t-\ell}, \ell) + \sum_{k=1}^{K} \delta_k u_{tk} + \epsilon_t \tag{26}$$

$B_k(x, \ell)$ are basis functions over exposure and lag θ_k are parameters to be estimated, ϵ_t is the error term (Gasparrini et al., 2010). Other covariates are captured with the expression $\sum_{k=1}^{K} \delta_k u_{tk}$.

Different basis functions can be integrated for the underlying structure of the lag coefficients or exposure variables. These include polynomials (Section 5.3.1), splines (Section 5.3.2), natural cubic splines or linear functions.

The previous chapters describe the basis functions to model the lag structure of exposure effects. As the exposure-response relationships are often non-linear and cannot be adequately captured by linear terms, the basis function approach can be extended for the basis variables for the exposure variables. This allows for non-linear transformations. Polynomials, splines or natural cubic splines are also applicable and well suited for modeling smooth and flexible relationships for climate variables or other continuous exposures. Together with the lag basis, this leads to a bivariate basis representation that jointly captures delayed and non-linear effects.

5.5 DLNM in GAM

Generalized Additive Model (GAM) is an extension of generalized linear models that use smooth functions to flexibly capture non-linear relationships. An embedding from a DLNM within a GAM takes place via including the DLNM cross-basis as a predictor within a GAM framework. This enables the joint modeling of lagged exposure–response associations (via DLNMs) and smooth non-linear temporal or spatial trends (via GAM splines).

GAMs offer a good trade-off between flexibility and interpretability. Complex models and non-linear relationships can be fitted and make good predictions in these cases. With smooths or splines non-linear aspects in data can be captured due to its flexibility. The smooth effects can include the non-linear effects as an univariate, tensor product or tensor interaction smooth. GAMs also incorporate random effects to account for clustered or hierarchical data (Wood, 2010, 2026).

The uncertainty of the model's predictions for specific data points is given by point-wise calculated confidence intervals on the linear predictor scale. These intervals incorporate the variance of all estimated model terms - this includes the intercept, linear effects, and all smoothed and lagged functional components. The confidence interval is formed by multiplying the standard error by a value, according to the desired level of 95%. This interval reflects the point-wise uncertainty for the full model at each observation.

5.5.1 Generalized Additive Distributed Lag Model

Generalized Additive Distributed Lag Model (GADLM) combines the two concepts of DLM and GAM to extend the lag approach with the flexibility of the additive modeling.

The additive structure of GAMs allows to capture complex models in which multiple predictors can be included and modeled as smooth functions or as parametric variables. This flexibility facilitates the inclusion of other relevant control variables, such as additional climatic factors, seasonality, long-term trends or spatial indicators. A key advantage is the ability to model the effect of an additional predictor on the outcome in a non-linear manner while simultaneously controlling for the delayed effect.

Formally, a GADLM can be expressed as (Zanobetti et al., 2000):

$$g(\mu_t) = \alpha + \delta^T u_t + \sum_{j=1}^{d} r_j(h_{tj}) + \sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} \tag{27}$$

where $g(*)$ denotes the link function and $\mu = E(y_t)$ is the expected value at time t, α denotes the intercept and u_t parametric covariates, that influence the model linearly with the corresponding effects δ. The smooth functions of the covariates h_{tj} that influence the model non-linearly are denoted as $r(\cdot)$. The summation $\sum_{j=1}^{d} r_j(h_{tj})$ reflects the additive structure of the GAM. Non-linear relationships such as seasonal or long-term trends can

be included. The distributed lag component $\sum_{\ell=0}^{L} \beta_\ell x_{t-\ell}$, as described in Equation (4) captures, the delayed effect of the exposure variable of interest x_t.

5.5.2 Generalized Additive Distributed Non-Linear Lag Model

To achieve the full flexibility of DLNMs, this work extends the GADLM by introducing non-linear effects into the modeling framework.

This leads to Generalized Additive Non-Linear Lag Models (GADLNMs), which now allow for both, non-linear exposure-response relationships and non-linear lag structures simultaneously.

Extending the Equation (27), the GADLNM can be written as:

$$g(\mu_t) = \alpha + \delta^T u_t + \sum_{j=1}^{d} r_j(h_{tj}) + \sum_{\ell=0}^{L} f(x_{t-\ell}, \ell) \tag{28}$$

The model components remain unchanged, while the final term incorporating the DLNM perspective is altered. The function $f(x_{t-\ell}, \ell)$ denotes a bivariate smooth function of the exposure x and lag ℓ, capturing the joint non-linear dependence across exposure levels and delays. This term explicitly accounts for the non-linearity and the delayed effects of climatic factors (reference to Section 5.4).

The GADLNM thus includes an intercept, linear predictors u_t, smooth functions $r_j(s_{tj})$ for other confounders and a non-linear lag component. This method enables the simultaneous adjustment for multiple confounders while allowing for non-linear exposure and lag effects.

Overall, the introduced GADLNMs provide a highly flexible modeling framework that is well suited for epidemiological applications, where exposure effects are often non-linear and delayed, and controlling for confounding and temporal structure is essential. To realize this method the following section describes the model specifications and selection, followed by the statistical implementation in `R`.

5.6 Model Specification, Selection and Implementation in R

This Section describes the specification, selection and implementation of the statistical models. The conceptual modeling choices are direct linked to their practical implementation in R to ensure traceability.
It needs to be noted, that the model development, lag decision and cross-basis definition go hand in hand as they influence each other. Here, the process of finding the correct lag resulting in developing the final model is described.

1.) Data aggregation: First, the climate, the health and the mosquito data is merged on the basis of the variable `trap ID`, `longitude` and `latitude` or `province`. Health data is aggregated to the level of province within the region as this is the lowest ordinal level to merge with the climate data.

The four outcomes: WNV, Number of Mosquitoes Tested, WNND only with the climate perspective and WNND considering climate and health perspective. They are binary outcomes and the outcome Number of Mosquitoes is a count variable, quasi-Poisson distributed.

2.) Model development: Model development was based on established predictors such as location-specific parameters (province, municipality, and trap coordinates), seasonality and yearly trends (Schwartz, 2000; Zanobetti et al., 2000; Curriero et al., 2002). The climate predictors temperature and precipitation are considered within the cross-basis. Models either include temperature or precipitation as a cross-basis. To extend this, models that included the other climate variable (precipitation or temperature, respectively) as controlling predictors for each lag were tested. This approach allows to control for the other climate predictor across the lag structure without constructing a second cross-basis. For example, in the temperature model with lag of 8, precipitation was included for each lag from 0 to 8, resulting in 9 additional predictors, and vise versa for the precipitation model. This allows to adjust for both climate predictors while avoiding the complexity of including two cross-basis within a single model.

Random effects include location-specific variables, these are trap coordinates (if available) and municipality, to account for spatial variations. The month term is included to control for seasonal patterns, preventing the estimated temperature-outcome association from being confounded by underlying seasonal cycles. Additionally, the predictor *year* is included to capture the effects of climate change, such as increasing temperature, over time. The province is also included as a predictor to account for regional differences and potential variations in climate or environmental factors. Later on the results if the tested mosquitoes were included in the model as well to combine the climate and mosquito perspective when analyzing a health outcome.

These variables were included in the models and different combinations were tested. The inclusion of the precipitation or temperature respectively for each Lag were also part of it.

3.) Lag decision: Based on literature the lag of 3 weeks is commonly recommended, and this was used here as the initial choice (Hart et al., 2024; Moirano et al., 2024). Models with different lags were compared using AIC and BIC, and the best-performing models were selected (Gasparrini, 2014). Lags from 0 weeks to 2 months were considered. This process resulted in a lag of 8 weeks for the outcomes WNV positive, Number of Mosquitoes Tested, and WNND in the mosquito and climate perspective. For the health outcome WNND with only the climate perspective, the model with a lag of 2 weeks performed best.

These models were tested for different lags between 0 and 8. Initially, the testing included lags up until 16, corresponding to $16 \times 7 = 112$ days or roughly 4 months. Only then the maximum lag was reduced to 8 lags meaning 2 months as the performance decreased after 2 months. Further, considering high lags would diminish the interpretability as the time window is only 6 month long.

The measurements AIC, BIC or R^2 were used to assess the performance (Gasparrini, 2014). The best performing models were separated by lags. Based on literature the lag 3 (Hart et al., 2024) and the lag with the best model performance was used.

Resulting in lag 3 and lag 8 when considering the mosquito perspective (Section 6.2 and 6.3.2) and in lag 3 and 2 when considering the health perspective and not including the mosquito perspective (Section 6.3.1).

4.) Knots: The knot placement follows the common practice in the field of climate statistics (Gasparrini, 2014). Here, I choose three knots at 10%, 50% and 90% and two boundary knots at 1% and 99%. Knots are useful because they define where the spline is allowed to bend, providing local flexibility in the fit while preventing the model from becoming overly wiggly and overfitting the noise.

```
knots_temp <- quantile(data$temperature_Celcius,
                       probs = c(0.1, 0.5, 0.90), na.rm = T)
b_knots_temp <- quantile(data$temperature_Celcius,
                         probs = c(0.01, 0.99), na.rm = T)
```

5.) Cross-basis defined: After determining the optimal lag based on AIC / BIC, multiple cross-basis were created. For the four outcomes, cross-basis were generated for the recommended lag of 3 weeks and for the alternative lags (8 or 2 weeks) identified as best-performing. This was done for both precipitation and temperature. A matrix $(n \times (L+1))$ is created for every time t and $\texttt{lag} = 0, \ldots, L$.

The cross-basis were constructed using the 3 knots and 2 Boundary knots of the according exposure variable.

```
cb_temp <- crossbasis(data$temperature_Celcius, lag = 8,
                      argvar = list("ns", knots = knots_temp,
                      Boundary.knots = b_knots_temp),
                      arglag = list("ns", df = 2))
```

6.) DLNMs in GAMs: The cross-basis were then embedded into a GAM, allowing for more flexible predictor-response relationship. GAM support the inclusion of non-linear variables and models using splines.

The cross-basis and linear terms are similarly embedded as a predictor in R with the expression `cb_temp + year`. The smoothing term `s(month, k = 3, bs = "cr")`, is included, defining a penalized cubic spline with three knots evenly spread through the underlying value range. This accounts for the non-linear influence of the variable `month` and therefore here seasonality. Additionally, random effects are included with the expression `s( ..., bs = "re")`. This is used for, if included, location and municipality: `s(longitude, latitude, bs = "re")` or `s(municipality, bs = "re")` (Wood, 2025).

In R, the models follow the overall structure:

```
gam(outcome ~ cb_temp + year + s(month, k = 3, bs = "cr") +
    province + s(lon,lat, bs = "re") +
    s(municipality, by = "re")
```

7.) Model fitting: The models were separately fitted for precipitation and temperature and the lag Perspective of 3 lags or 2/8 lags.

8.) Results visualization: The Cumulative Exposure–Response Curves were produced to visualize the exposure–response relationships over all lags. The Lag-Specific Exposure-Response Curve and Lag-Response Curve at a fixed exposure value show different perspectives of the 3D Exposure-Lag-Response Surface. These visualizations make it possible to separately examine the lag and the exposure perspective. The Contour Exposure-Lag-Response Plot breaks the 3D Surface down to 2D Surface (further, see Section 5.7).

5.7 Conceptual Background for Plot Output

After fitting the DLNM within the GAM framework the interpretation is crucial and mainly done by visualizations. The GAM framework in R already provides ways to visualize the results. This section shall roughly break down the different plots and visualization methods that play a role in Chapter 6. Also, with the help of generalized linear models (GLMs) a simple linear overview is achieved, preposed to the GADLNM results.

The presented visualizations rely on the assumption that the underlying GAM/DLNM adequately captures the true non-linear and delayed exposure-response relationship. Effects are assumed to be additive, smooth across exposure and lag dimensions, temporal stable and consistent with appropriate outcome distributional assumptions. The uncertainty bands reflect the uncertainty of the estimation within the model and do not account for uncertainty that arises from the model choice (Gasparrini et al., 2011).

First, a GAM plot is a partial effect plot, a simple way to display the relation between each variable introduced into the model and the outcome. The non-linear smoothing functions are realized through splines after defining a number of knots. The overall prediction is then achieved through adding this over all variables (Ross, 2025).

Second, an important role plays the Cumulative Exposure-Response Curve. The cumulative effect of the predictor on the outcome over time is shown. A better understanding about the long-term effects is gained, allowing us to make an inference about the relation.

Third, the 3D Exposure-Lag-Response Surface is a three-dimensional plot visualizing the relation between exposure, lag and response and serves as an exploratory analysis rather than a precise quantification. The complex relationship is reduced to a two-dimensional perspective, the Contour Exposure-Lag-Response Plot to understand the complex interaction. These two visualizations can be explored in more detail, when considering a fixed lag or a fixed exposure value. This provides a clearer insight into the layers of the model.

Fourth, breaking down the more complex visualization, the Lag-Specific Exposure-Response Curve shows the relationship between predictor and response variable at a fixed lag. Whereas the Lag-Response Curve considers the relation between Lag and Response variable at a fixed exposure value. In the former, short term effects can be better understood and the latter, the effect of the lag on the response variable is explained while controlling for the effects of other predictor variables.

The increasing/ decreasing effect on the outcome can be observed by the Incidence Rate Ratio (Irr) or Odds Ratio (Or), depending on the outcome's distribution, in comparison to a reference value. That reference value can either be specified manually or, by default, is set to the median exposure value. By using the different plots and perspectives, we are able to gain a better understanding of the relationships between the three-dimensional interactions of the exposure variable, lag and response variable. Allowing to make more informed statements and decisions based on the results (Gasparrini et al., 2011).

6 Results

This project investigates the relationship between climate change, the occurrence of mosquitoes infected with WNV and public health outcome in Northern Italy over a decade from May 2014 to October 2024.

The analysis focuses on the province of Emilia-Romagna as the data is region-wide available for the mosquito dataset as well as the health data. Initially, the considered region was the combined area of the province Emilia-Romagna and Lombardia. However, the collection methods differs strongly between Emilia-Romagna and Lombardia as latter stops collecting the mosquito trap data after one positive result whereas in the Emilia-Romagna region the collection is not stopped. Due to the differences the results are always dominated by the data of the Emilia-Romagna region. In addition, Lombardia does not provide complete health records. Thus, the analysis was solely run on the Emilia-Romagna region.

All datasets are aggregated on a weekly basis across the entire decade and mapped with the available mosquito dataset, which span each year from May/June to September/October. Each climate variable (temperature and precipitation) is an average per week per location. The health data and mosquito data are sums over one week per location, where location is defined by province for the health data and trap coordinates for the mosquito data. To combine the health and weather perspective, the according datasets are aggregated by province as it is the lowest mergeable scale. Otherwise the precise location of the trap is used. The resulting merged datasets contain either mosquito and weather, health and weather, or a combination of mosquito, health and weather variables.

First, the datasets are described in more detail. Providing an overview of the mosquito traps, climate predictors and the health records (Section 6.1.1, 6.1.2 and 6.1.3).

Second, before handling the DLNMs, or GADLNMs respectively, I tested the linear connection of different predictors and the outcome. Therefor I considered GLMs, including the climate predictors independently of each other and combined also with additional predictors.

Third, the GADLNM results were sorted into the Sections 6.2 and 6.3, according to the their outcomes. In the following, the considered outcomes are:
a. positive WNV result (binary per week per measurement within a trap or region),
b. number of tested mosquitoes (number per measurement within a trap or region),
c. WNND; only with weather predictors (binary outcome per week per region) and
d. WNND; with weather and mosquito predictors (binary outcome per week per region).
The construction of the models follow the procedure as described in Section 5.6.

For the construction of the DLNMs a lag of 3 weeks was chosen which is equivalent to 21 days as also found in literature (Moirano et al., 2024; Hart et al., 2024). Additionally, the lag with the best model fit based on AIC and BIC was evaluated. A temperature of 24 $\pm$ 3°C was (additionally to the 50th percentile) used as reference as this is a

favored temperature for transmitting WNV (Heidecke et al., 2025; MacDonald et al., 2024). Precipitation was looked at overall as most provinces have reported very low precipitation.

The results are interpreted based on the GADLNM output and the graphs as described in Section 5.7. For a simple overview GLMs are also considered.

6.1 Descriptive Statistics

6.1.1 Mosquito Surveillance Data

The basis for this analysis is the mosquito surveillance dataset. Over the years, the traps were active starting in the beginning of June until the middle of October (2014 - 2018). Later the collection started already in the beginning of May and ended approximately in the middle of October (2019 - 2024). The months from June till August reported the highest amount of measurements, July and beginning of August reported the highest confirmed positive WNV cases. There were 119 unique traps located in Emilia-Romagna. In Emilia-Romagna 93.82 $\pm$ 3.12 traps were on average per year in place. In total there were 40,030 measurements within Emilia-Romagna. Throughout the decade in Emilia-Romagna 875 (4.2%) positive cases and 20,086 negative test results were reported.

The year 2018 was exceptional, exhibiting more than a twofold increase in positive test results compared to the preceding and subsequent years, which recorded comparatively few positives.

Provinces Emilia-Romagna	**Average Amount of Measurements per Municipality**	**Number of Municipalities in the data (and total)**	**Number of Traps**
Bologna	228	15 (56)	21
Ferrara (C)	452	13 (21)	30
Forlì-Cesena (C)	151	3 (30)	7
Modena	308	10 (47)	12
Parma	194	9 (44)	11
Piacenza	228	10 (48)	10
Ravenna (C)	175	7 (18)	12
Reggio Emilia	238	12 (45)	13
Rimini (C)	73	3 (27)	3

Table 2: *For each province within Emilia-Romagna, the table reports the average number of measurements per municipality, the number of municipalities included in the dataset (with the total number of municipalities in parentheses), and the number of monitoring traps (Brinkhoff, 2025a). (C) marks the coastal provinces.*

Additionally to Figure 1, Table 2 summarizes the distribution of the traps and covered municipalities across the provinces in the region of Emilia-Romagna. In the Appendix, Figure B.1 shows the distribution of traps in the region of Lombardia. The number of traps cover only a portion of the municipalities and therefore entire region. In Emilia-Romagna,

Rimini shows the lowest rate of traps and lowest average amount of measurements within a municipality. The province Rimini has only 3 out of 27 municipalities present in the dataset. Forlì-Cesena has only 7 traps installed, covering only 3 out of 30 municipalities, but in comparison to Rimini has a higher amount of measurements per municipality. Ferrara and Modena show the highest amount of measurements per municipality. Both have more municipalities in general and municipalities present in the dataset. Especially, Ferrara has with 30 traps more traps installed than any other province.

6.1.2 Weather Trend Data

The climate factors, temperature at 2m above surface level (normally reported in Kelvin but transformed in Celsius) and precipitation mm per day, are considered over the decade 2014 to 2024 for every trap location.

First, a clear mean seasonal trend is visible over the entire Emilia-Romagna region with temperatures until 40°C, in late July, and low temperatures close to 0°C, in January (see Figure 2 and 3, right).

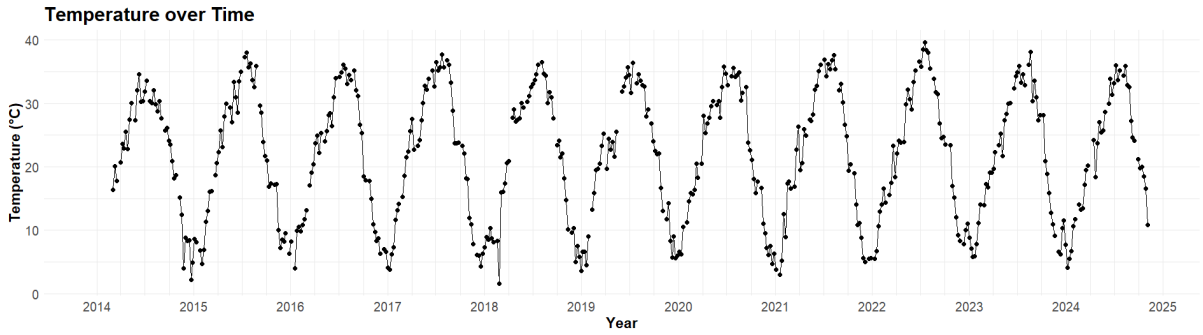


Figure 2: *Temperature over the Decade 2014-2024, visualizing the seasonal pattern*

Second, Figure 3 (left) shows the slight increase of the mean temperature per year from 2014 to 2024. The simple linear regression model reports a positive slope of 0.0578. However, strong shifts in the temperatures are not visible. The annual mean total precipitation is decreasing over the decade. This is mirrored in the reported slope of -0.4905. The year 2021 was the year with the lowest precipitation. In the following year of 2022 higher temperatures with an average of 23°C were reported.

Third, the period from May until October includes the warmest and driest period of the year (see Figure 3, right). The averaged temperatures rise from 26 to 35°C in July and fall to 15°C in October. The total monthly precipitation is lower during the hotter months. June, July and August have a total precipitation below 84 mm. May and September have rising precipitation around 115 mm. October is the month with the highest precipitation, 138 mm.

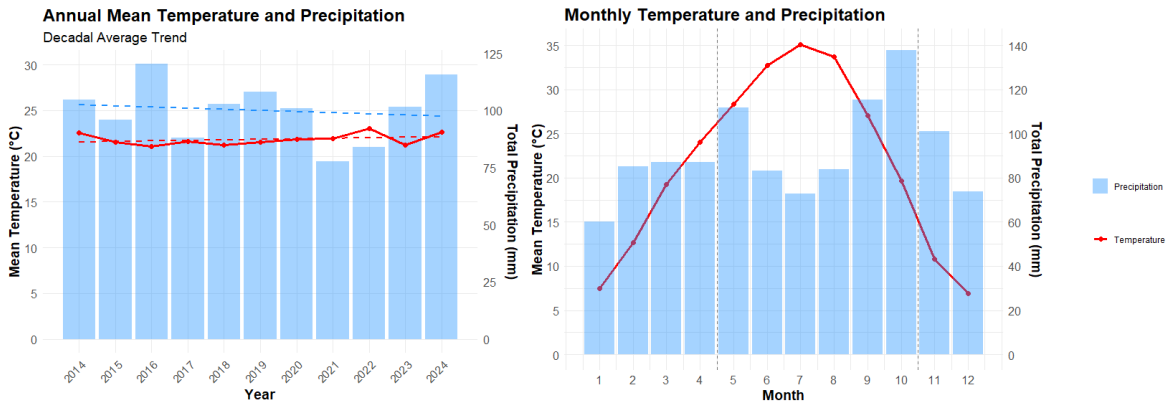


Figure 3: *Averaged Temperature and Precipitation in the region of Emilia-Romagna over the decade 2014-2024 (left) and the seasonal pattern averaged by region (right)*

Fourth, due to the different geographical settings, the climatic patterns can vary per province (see Figure A.1, A.2). From the nine provinces of Emilia-Romagna, Rimini, Forlì-Cesena, Ravenna, and Ferrara all share a coastline along the Adriatic Sea. Further inland towards North-West lie the provinces Bologna, Modena, Reggio Emilia, Parma and Piacenza. Bologna, Modena and Parma are the region's largest provinces by population and economic activity. Coastal provinces along the Adriatic Sea show slightly higher average temperatures over the entire year. The difference is especially visible during the hottest and the coldest season. For the region's largest provinces no clear pattern is observable.

The weather information is extracted for the period where the traps were active and collecting mosquitoes.

6.1.3 Health Data

The health dataset includes cases that occurred within the Emilia-Romagna region from 2014 to 2024 and are in association with a WNV infection. The dataset distinguishes between three types of diseases/infections: WNV-fever, WNND and detected but asymptomatic, mostly via blood donation (for comprehensibility, this group will hereafter refereed to as: asymptomatic). Included is also the information about death within 30 days of symptom onset.

The dataset includes 754 different records, including 172 patients with WNV fever, 469 patients with WNND and 113 asymptomatic patients over the entire decade. In 711/754 (94.3%) cases the disease is classified as confirmed. The cases classified as probable include 22 cases for fever and 21 cases for WNND. A total of 73 (9.68%) patients were reported dead within 30 days of symptom onset. Of these patients, 6 had WNV fever (2 classified as probable) and 67 were diagnosed with WNND (1 classified as probable). WNND was reported in approximately 45 to 76% (median 60%) of the cases each year. The relative amount of deaths per year among patients diagnosed with WNND differs between 0% to

33.3% with a median of 14.97% and the absolute amount of death among all patients diagnosed with WNND is 14.29%.

Figure 4 (left) visualizes that over all years, in August the most cases were recorded, forming 58.89% of the cases. The months July and September report similar percentages with 20.03% and 18.82%. In the years 2018 and 2024 a total of 214 and 220 cases, respectively, were reported (Figure 4, right). This was higher than in any other year. The absolute number of deaths simultaneously increased as the numbers for WNND and for WNV fever increased.

The cases in the dataset range from the start of July to the beginning of October each year. Additionally, in June three entries exist, all WNND cases with one death (one death in 2021 and two cases in 2024). In October, only 14 cases are reported over the years, where 10 are WNND cases, but no death cases.

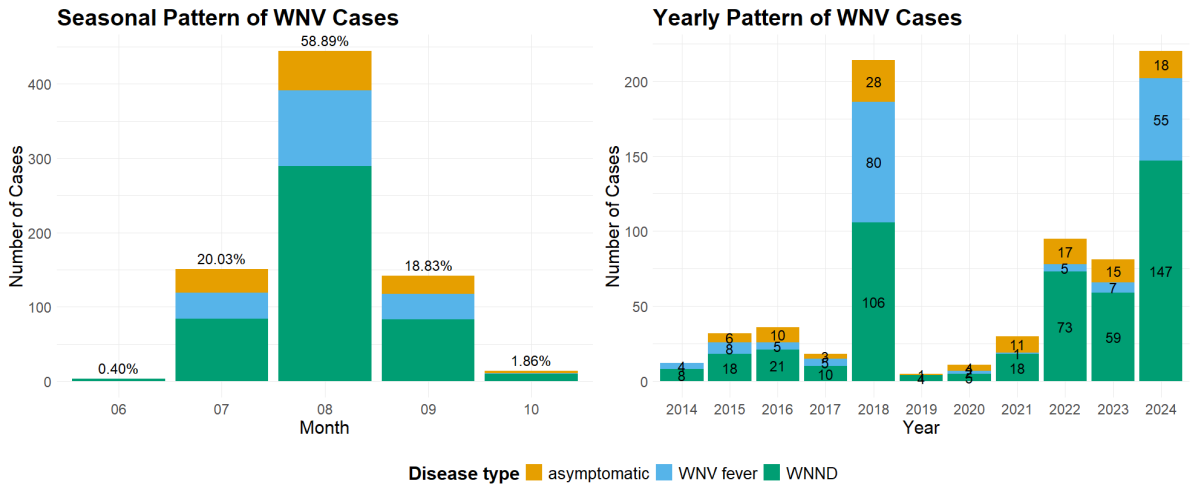


Figure 4: *Annual and Seasonal Pattern for the cases of WNV, distinguishing between the different disease types (WNV fever, WNND and asymptomatic/bloodDonor).*

The mean age over all disease groups is 65.96 years [min: 17, max: 94] and 61% are males. Most reported cases are among males: 61.0% of WNND, 55.2% of fever and 72.6% of asymptomatic cases over the whole decade. Also, as the age increases the more severe the disease diagnosis is. Table 3 shows the age distribution over the different diagnoses and between females and males. For all diagnoses, older males are reported more often. The mean age is higher among individuals who died.

Additionally, it can be noted that the provinces Modena and Bologna reported more cases than other regions, independent of the diagnosis. In the years 2018 and 2024 (with increased reported cases) this effect is especially dominant.

In the following analysis, the health data is aggregated by province and calender week. The aggregated data thus reports only the binary occurence of WNV fever, WNND or asymptomatic cases per week for each province.

Diagnosis	Sex		Age groups 17 - 24	25 - 44	45 - 64	65+	Mean Age
Asymptomatic	M	(72.6%)	1	20	53	8	51.3
	F	(27.4%)	1	7	22	1	48.9
Wnv Fever	M	(55.2%)	5	8	37	45	62.9
	F	(44.8%)	5	11	31	30	58.1
Wnnd	M	(61.0%)	1	7	55	223	72.4
	F	(39.0%)	2	10	37	134	70.2
Death within 30 days	M	(63.0%)	0	0	1	45	79.0; 65.4
	F	(37.0%)	0	0	4	23	78.4; 63.3

Table 3: *Age distribution per sex per Wnv Disease and death within 30 days. The proportion between the sexes per respective diagnosis group is given in brackets. For death are two mean ages stated: first, for patients that died and second, patients that did not die.*

6.2 Model with Mosquito outcome

This section presents results from the spatio-temporal Gadlnm assessing the association between climate variables and mosquito outcomes. The focus are the two mosquito-related outcomes: **a.** positive Wnv result and **b.** number of mosquitoes tested within a trap.

The median temperature is 32.6°C within a range of 11.27 to 43.29. The plots however show a smaller window to ensure smaller confidence intervals. The median daily precipitation is 1.33mm [min: 0, max: 22.3] and the 75th quantile is 3mm. The median temperature or precipitation is used as reference value to interpret the model results.

The presented results are obtained from a Gadlnm. To get a first overview results from a simple generalized linear model are preposed. The tested climate predictors are temperature and precipitation, both considered within the cross-basis functions, respectively. Following the Section 5.6, the lags were chosen based on literature and on the best model performance. Here, the included lags were lag 3 and lag 8 for both outcomes. The results for the Gadlnm are primarily interpretable through different graphical representations (Section 5.7).

6.2.1 WNV positive result

To obtain a first simple impression of the influence of the climate variables, the Glms results are reported first. Temperature has s significant influence in the Wnv test result outcome. With an increase of one degree the odds of receiving a positive test result increase by approximately 10% on average, ceteris paribus. Precipitation does not report a significant influence, when no other variable is included in the Glm. When enlarging the model and including the two climate factors, latitude and longitude, seasonality and year as predictors, the significance changes. Under the assumption that the other variables are held constant, on average following statements can be made: The temperature increases by one degree, the odds of receiving a positive test result increase by the factor 1.22. The

year has also a significant influence, with an additional year, the odds of a positive test result increase by the factor 2.25. With the increase of one latitude degree (a degree more North) the odds of a positive test result increase by 3.75. Only in a simple model including only latitude and longitude, both report a significant effect on the WNV test result with increasing odds for increasing latitude and decreasing odds for increasing longitude.

The following GADLNMs test the influence of the temperature and precipitation on the WNV test results for the lags of 3 and 8, respectively. The models incorporating temperature as a cross-basis also control for seasonality (month), year, province and as random effect: location and municipality. The GADLNMs with precipitation included through the cross-basis function also control for temperature at each lag, seasonality (month), year, province and as random effect: location and municipality.

To start, the GADLNM with temperature as the explanatory cross-basis is reported. At lag 8, the year has a significant influence on the WNV test results. With each additional year, the OR is 1.06. This is also observed for a lag of 3. The provinces Forlì-Cesena, Ferrara (only at lag 8), Ravenna and Rimini report a significant lower OR for reporting a positive test result. The seasonality has a non-linear connection to the positive test results, a quadratic significant influence. The random effect municipality has a high effective degree of freedom of 13.6, a complex spline. Seasonality dominates the variation with a variance of 6.71, explaining an essential part of the model. The random effect municipality is a moderate random effect with a variance of 0.156 (0.16 for lag 3), real differences between the municipalities exist. However, the distribution between the municipalities is not perfectly normal (reporting outliers only on the edges) what aligns with the smaller variance and still statistically significant influence. The spatial difference however does not contribute to the whole model's variance with a variance of ≈ 0.00 for both lags. The interpretation of the temperature cross-basis coefficients is not advisable, as the effects are best understood in a cumulative way, across lags and the exposure range.

Further, similarities are found within the results for the GALDNMs including precipitation as the cross-basis. The named provinces all show a lower OR for lag 8, as seen in the temperature models. The temperature for each lag is additionally included, but had no significant linear influence on the outcome. With one additional year the odds increases by the factor 1.07 on average, ceteris paribus. For both lags, the regions Forlì-Cesena, Ferrara (only at lag 8), Ravenna and Rimini show a significant influence on the outcome of WNV test results with a decreasing odds of experiencing the outcome. Similar to the previous GADLNM results the seasonality modeled over the month, contributes the largest amount to the model's variance. Differences in municipalities are existent and medium with a variance of 0.197 (0.173 for lag 3), not dominating the analysis. Spatial differences through longitude and latitude do not contribute to the model's variance. Again, the interpretation of the cross-basis coefficients is not done here.

A clear quadratic influence of seasonality on the outcome WNV positive test results is visible (see Figure A.3). The months July, August and September show a strong increase

in the OR. August reports an increased OR by 5 times in comparison to the average month with a OR of 1 (middle of June or in October). The difference of precipitation and temperature is very small. The trend of a higher increase of the influence of the temperature is visible during the hotter period and inverted nearer to October. Similar pattern can be observed for the difference between the lags. However, these are trends and not significant differences. For the months May and June a strong decreasing risks, even up until under 1% of the risk in comparison to the average month, can be seen. Both models were adjusted for temperature. Either as cross-basis or as temperature for each lag, when precipitation was considered the cross-basis.

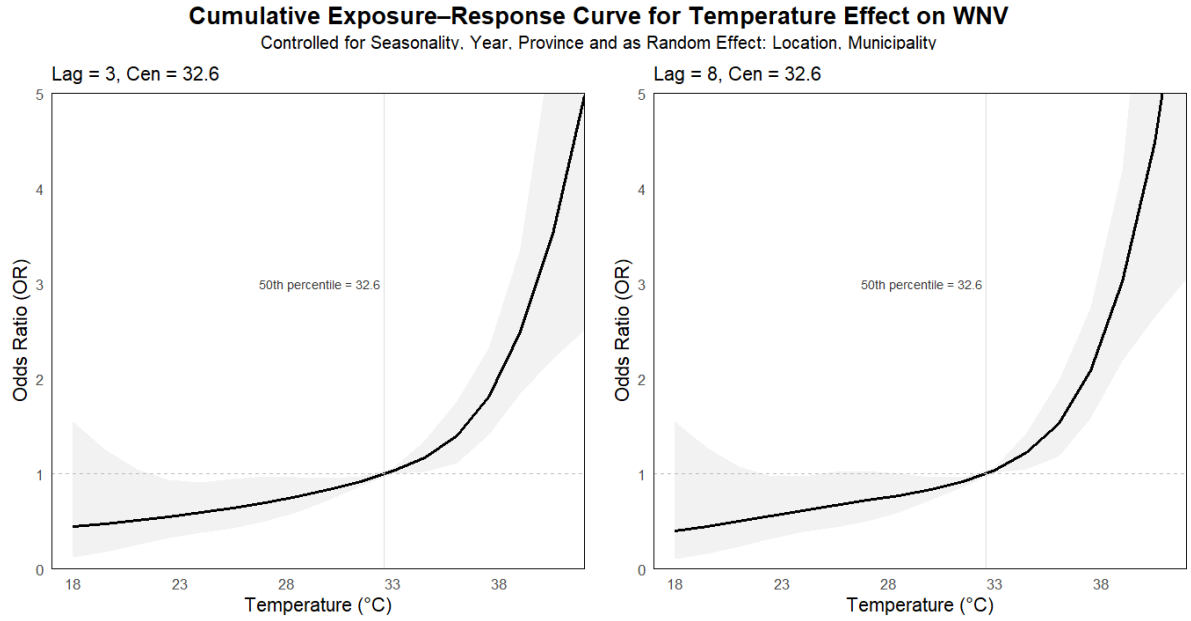


Figure 5: *Cumulative Exposure-Response Curves for temperature effect, OR, on WNV for the lags 3 and 8, with a reference temperature of* 32.6 *°C (the 50th percentile).*

The Cumulative Exposure-Response Curves show a clear trend for both lags: higher temperatures have an higher OR. Temperatures over 32.6°C report a higher risk of experiencing positive WNV test results in comparison to the reference value of 32.6°C. The higher the temperature, the higher the connected risk. At a temperature of 38°C, the OR is increased by approximately 2 times, in comparison to 32.6°C. Lower temperatures between 32.6°C and 21°C show a slightly decreasing OR at lag 3. This can not really be observed for a lag of 8. For temperatures even below 21°C no clear trend is visible.

For precipitation strong uncertainty is present, even increasing with higher total precipitation for both lags (see Figure A.4). No clear pattern is apparent.

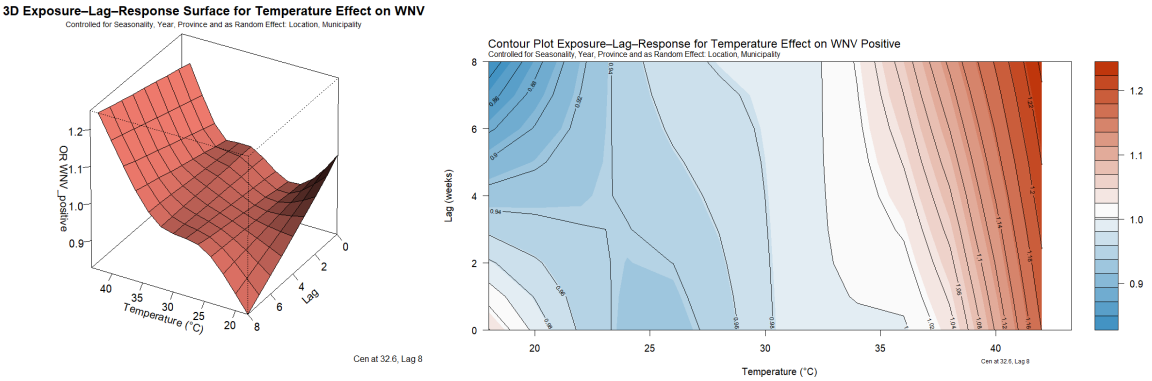


Figure 6: *3D Exposure-Lag-Response Surface and Contour Plot for temperature effect,* Or*, on* Wnv *for the lag 8, with a reference temperature of* 32.6 °C *(the 50th percentile).*

Figure 6 shows, similar to the cumulative plots, that higher temperatures are connected to higher Or especially for higher lags, in comparison to the median temperature. Low temperatures for high lags have a decreasing Or. And low temperatures at lag 0 and 1 show a slightly increased Or.

When fixing the temperature and comparing to the median temperature with a lag of 8, a high level of uncertainty at the edges is reported (see Figure A.5). The uncertainty is only slightly decreasing for higher temperatures.

The 3D Exposure-Lag-Response Surface Plots for the precipitation perspective suggest that higher precipitation at higher lags is connected to higher Ors. And lower precipitation at higher lags associated with lower Ors. However, uncertainty is not included in these plots.

6.2.2 Number of Mosquitoes Tested

The outcome, the number of mosquitoes tested, is a count measure for each trap on each day. Because the variance exceeds the mean, indicating overdispersion, a quasi-poisson distribution was assumed. Accordingly, the fitted Glm and Gadlnm was specified with a quasi-poisson family to appropriately account for overdispersion.

Again, to enable a simplified overview, the results from a Glm are reported. Temperature as well as precipitation have in a univariate Glm a significant influence on the number of mosquitoes tested. Whereas an increase of one degree in temperature is associated with an increased incidence rate of mosquitoes tested by the factor 1.05, an increase of one millimeter of precipitation is connected to a decreased incidence rate ratio of 0.97, ceteris paribus. Extending the model and including the two climate factors, latitude and longitude, seasonality and year as predictors, all predictors have a significant influence. Under the assumption that the other variables are held constant, on average following statements can be made: In the larger model, a one-unit increase for temperature and precipitation are both connected to an increased Irr (1.05 and 1.01, respectively). One degree increase in longitude (towards east), multiplies the incidence rate with 1.03. Latitude (towards north)

has the same tendency, but with an IRR of 1.67. With each additional year, the incidence rate of mosquitoes tested is multiplied by 0.99. Each additional month is associated with an incidence rate of mosquitoes tested multiplied by 0.75.

The models for both lags and for temperature as well as precipitation as included cross-basis, follow the same structure. Within both models it is controlled for seasonality, precipitation or temperature, respectively, at each lag, province, year and as random effect: location and municipality.

The model with temperature as the cross-basis reports significant influence on the outcome. The coefficients however are not further interpreted. The inclusion of precipitation at every lag reports no significant influence. With each year increase, the incidence rate of mosquitoes tested is multiplied by 0.99, when considering lag 8. The region Forlì-Cesena reports for both lags a significant decreased IRR of mosquitoes tested of 0.70. Rimini is also associated with a lower rate of mosquitoes tested, the incidence rate is multiplied by 0.51 (for both lags).

The smoothed term, seasonality, influences the outcome of mosquitoes tested significantly as a quadratic spline. The months June and July have a 50-60% increase in the IRR in comparison to the average month. The month August and the following months show a clear decrease of the IRR. In October, a reduced IRR of 10% in comparison to the average month is reached (see Figure A.6). During the months of June to August, the seasonal effect for temperature and for the lag 3 appears marginally stronger. Nevertheless, these differences are minimal and the described overall seasonal pattern is consistent across the two climate factors and the lags.

The spatial random effects for municipality as well as for latitude and longitude were statistically significant, indicating a systematic heterogeneity by these grouping structures. The spline that models the effect for the seasonality shows over all models for the outcome *number of mosquitoes tested* a clear trend. The random effect for municipality showed a variance of 0.0597 (0.060 for lag 3). In contrast, the random effect for longitude and latitude explain none of the model's variance. A substantial proportion of the model's variance is captured by the seasonal smoothing term, which showed a variance of 0.6104 (0.6417 for lag 3).

The Cumulative Exposure-Response Curves show a strong and significant trend of increasing IRR for temperatures over the centering temperature (32.6°C). The model with lag 3 describes a maximal cumulative IRR of ≈ 2.8 and the model with lag 8 a maximal cumulative IRR of ≈ 3.2. A higher IRR for lag 8 is visible. For temperatures below the 32.6°C, until 21°C, are significantly connected to a lower IRR of maximal 0.8 for both lags.

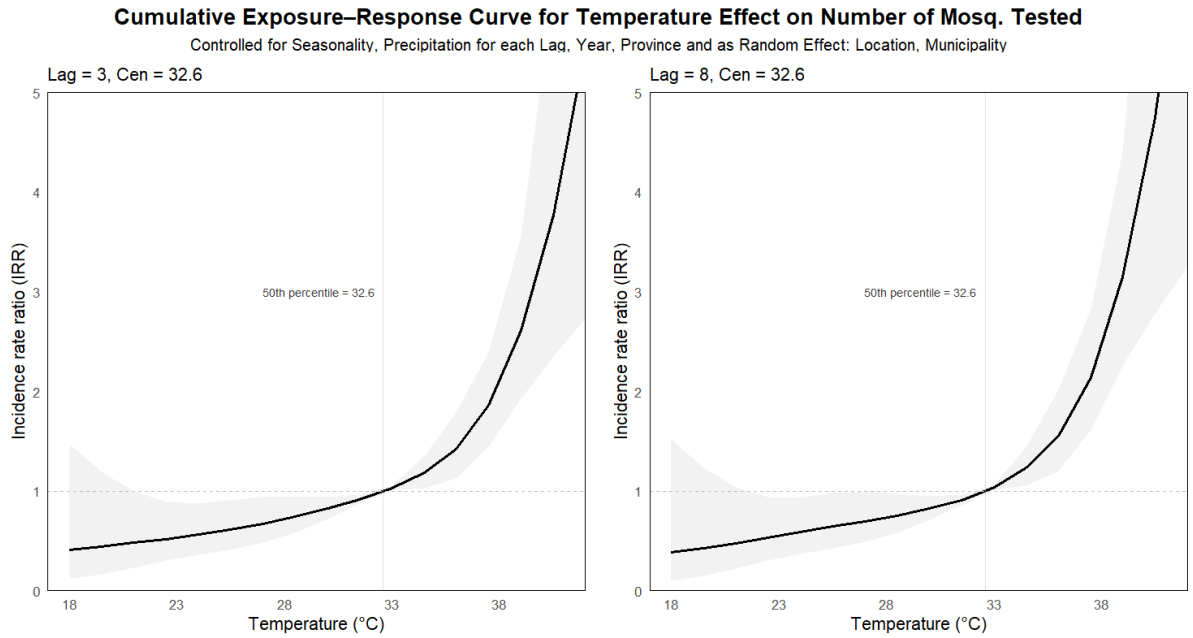


Figure 7: *Cumulative Exposure-Response Curves for temperature effect,* Irr*, on number of mosquitoes tested for the lags 3 and 8, with a reference temperature of* $32.6\,°C$ *(the 50th percentile).*

Again, the Cumulative Exposure-Response Curves for precipitation exhibit an increasing uncertainty with increasing daily precipitation. Wherefore no pattern can be seen or assumed. The 3D Exposure-Lag-Response Surface and Contour representation suggest that with higher precipitation the Irr decreases (Figure A.8). There is a plateau with an Irr of 1 between approximately 1 and 6. This is an overall trend for all lags, but specifically visible in the contour plot for lower lags. However, the uncertainty is here not visible.

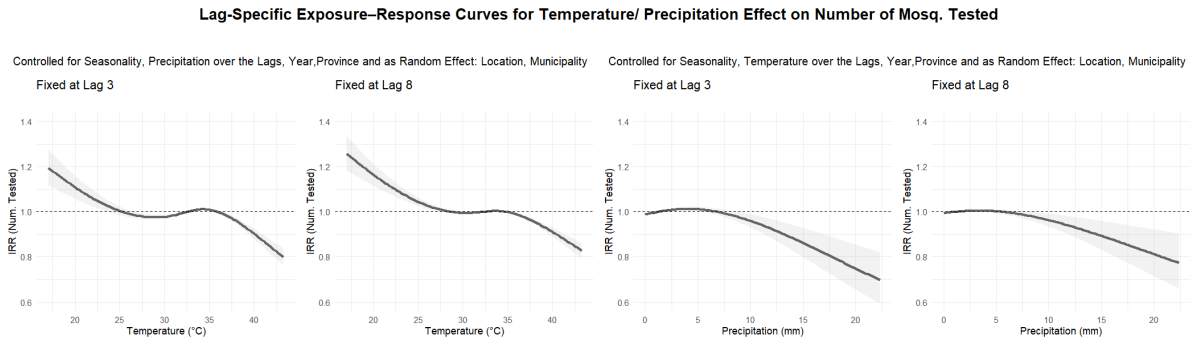


Figure 8: *Lag(3,8)-Specific Exposure-Response Curves for temperature and precipitation,* Irr*, on number of mosquitoes tested, with a reference temperature of* $32.6\,°C$ *and precipitation* $1.33mm$ *(the 50th percentile, respectively).*

Figure 8 shows the Lag-Specific Exposure-Response Curves for the models describing the outcome of number of mosquitoes tested for both lags, 3 and 8, and for bith climatic predictors, temperature and precipitation. A significant pattern of temperature influencing

the IRR can be seen. For temperatures below 27.5°C for lag 8 and 23°C for lag 3 the IRR is increased up until a maximum of 1.28 and 1.19, respectively (with rather small confidence intervals). For lag 3, there is a small decreasing risk between 25°C and 32°C visible. Temperatures higher than 35°C for lag 8 and 36°C for lag 3 show a decreasing IRR until a temperature of 43°C until a IRR of approximately 0.8. The Lag-Specific Exposure-Response Curves for the effects of precipitation on the number of mosquitoes tested look different. For both, lag 3 and lag 8, precipitation higher than 10mm per day is connected to a decreasing IRR as the precipitation amount increases. It needs to be noted that higher precipitation also goes simultaneously with increasing uncertainty.

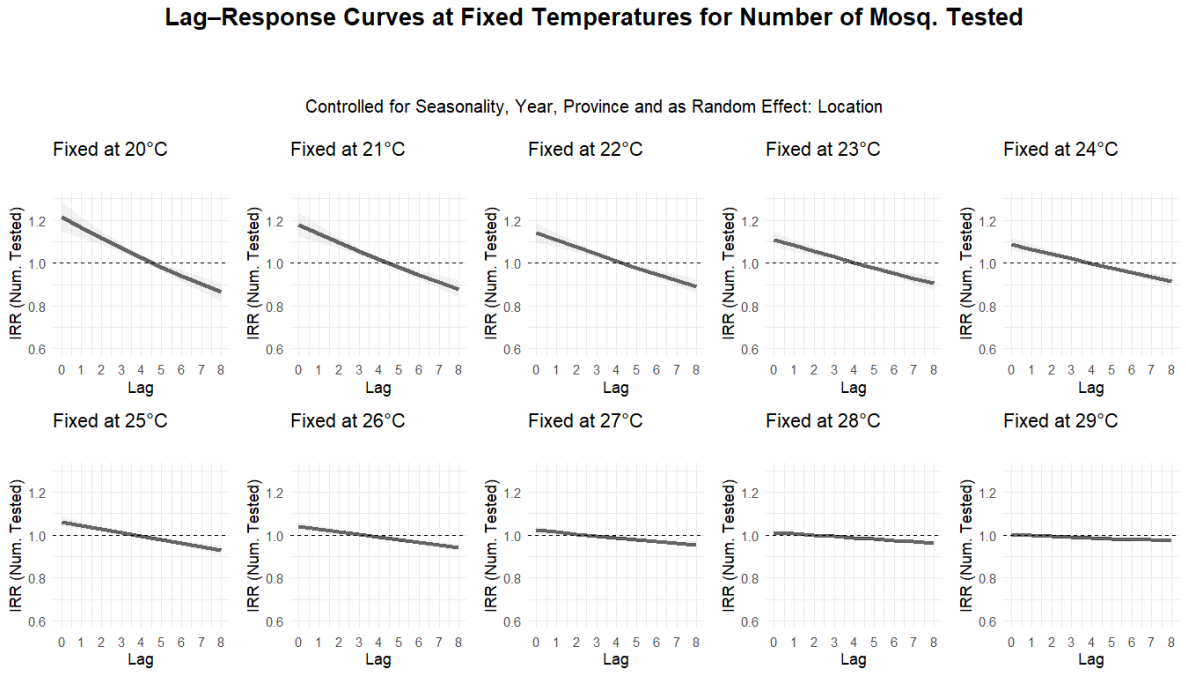


Figure 9: *Lag-Response Curves at fixed temperatures between 20°C and 29°C showing the temperature effect, IRR, on number of mosquitoes tested for the lags until 8, with a reference temperature of* 32.6 *°C (the 50th percentile).*

A clear effect trend can be seen in the Figure 9, where the effect of fixed temperature values on the outcome of number of mosquitoes tested are evaluated for lag 8. In relation to the reference temperature of 32.6°C, lower lags show a higher IRR and higher lags show a decreasing IRR. The direction of the IRR changes at a lag of 4. This pattern becomes less distinct as the fixed temperature increases, meeting the median temperature, the reference value. Similarly, this pattern is observable for lag 3, but with the change shifted to a lag of 1.5.

For fixed precipitation levels, no distinct pattern is visible across the lags or precipitation amounts, with IRRs remaining close to 1. However, there is some indication of a modest decrease in IRRs at lower lags for higher precipitation amounts over 10mm. When evaluating the model at lag 3, a clearer trend emerges in the lag-response curves. Starting

at a lag of 0, IRR decreases with increasing lag. The decline is more pronounced at higher fixed precipitation levels.

6.3 Models with Health Outcome

This section describes the results from the second part of the GADLNMs assessing the association between climate variables and health outcomes. Incorporating the mosquito perspective. The focus is the health-related outcome WNND, considering the two cases: **c.** WNND only with climate predictors (Section 6.3.1) and **d.** WNND with climate and mosquito predictors (Section 6.3.2). Both are binary outcomes per calender week, per region or per trap, respectively.

The median temperature is 32.6°C [min: 9.98, max: 41.6]. Again, the plots show a narrower window to ensure smaller confidence intervals. The median daily precipitation is 1.5mm [min: 0mm, max: 18.84mm] and the 75th quantile is 3.19mm, when only considering the climate perspective. When including the mosquito surveillance data, the median of daily precipitation lies at 1.43mm [min: 0mm, max: 22.34mm]. Both median values are used as reference values if not other specified.

The presented results are obtained from a GADLNM. Prior, to get an overview results from a generalized linear model are stated. The climate factors temperature an precipitation are again considered within the cross-basis functions, respectively. The lags that are considered are lag 3 and lag 2, when testing only with climate predictors (**c.**) and lag 3 and lag 8 when using explanatory variables both from the climate and the mosquito perspective (**d.**). The results of the GADLNMs are mainly interpreted through different graphical representations (Section 5.7).

6.3.1 WNND with Weather Trend Perspective

At first, the effects of different predictors are assessed from a linear perspective using GLMs. The univariate model including temperature as predictor shows a significant increase of the odds by 7% on average when the temperature increases by one degree Precipitation does not significantly influences the WNND outcome. Enlarging the model and including the predictors year, seasonality and province, changes the coefficients. The following statements holds on average and when holding the other variables constant: With one additional degree temperature the odds for experiencing WNND increase by the factor 1.13. Each additional year increase the odds by the factor 1.12. The seasonality is significant and with each additional month the odds strongly increase, by the factor 1.43. The regions Ferrara, Piacenza, Parma, Ravenna and Rimini show significant decreasing odds in experiencing WNND. Rimini stands out with odds multiplied by 0.02.

The GADLNM test the effect of temperature and precipitation on WNND for the lags of 2 and 3. The models having temperature within the cross-basis function include for both lags seasonality (month), precipitation at each lag (only for lag 3), year as predictors

and as a random effect: province. The models with precipitation as the cross-basis also controls for seasonality (month), temperature at each lag, year and includes province as a random effect.

In former, the variable year shows a statistical significant influence on the outcome. An increase of each year for lag 3 and 2, multiplies the odds of experiencing a WNND by 1.12. For lag 3 precipitation at time 0 is significant. With an one-unit increase of precipitation at time 0 the odds decreases by the factor 0.92. Both temperature models have a R^2 of ca. 0.195.

In later, the models with precipitation incorporated within the cross-basis function, the variable year shows a statistical significant influence on the outcome. An increase of each year multiplies the odds of experiencing a WNND with the factor 1.18 for both considered lags 3 and 2. The models have a R^2 of approximately 0.227 and 0.186, respectively. For a lag of 3 the temperature at lag 2 reports also a significant effect, an increase of one unit in precipitation at lag 2 results in an OR of 1.06.

In all the models, the smooth term seasonality quadratically influences the outcome. However, this is clearer for the climatic predictor *precipitation* (see Figure A.9). The figure clearly shows the quadratic pattern of increasing OR for the hotter period with the maximum in August. Models with temperature included within the cross-basis function show a decreased OR in June but increases. Due to the uncertainty no clear trend can be assumed. No distinguishable difference can be assumed for the different lags. The random effect *province* reports a variance of 1.36 for lag 3 and 1.44 for lag 2, respectively. Within the precipitation models even more heterogeneity between provinces exist, a variance of 2.20 and 1.79 is reported, respectively.

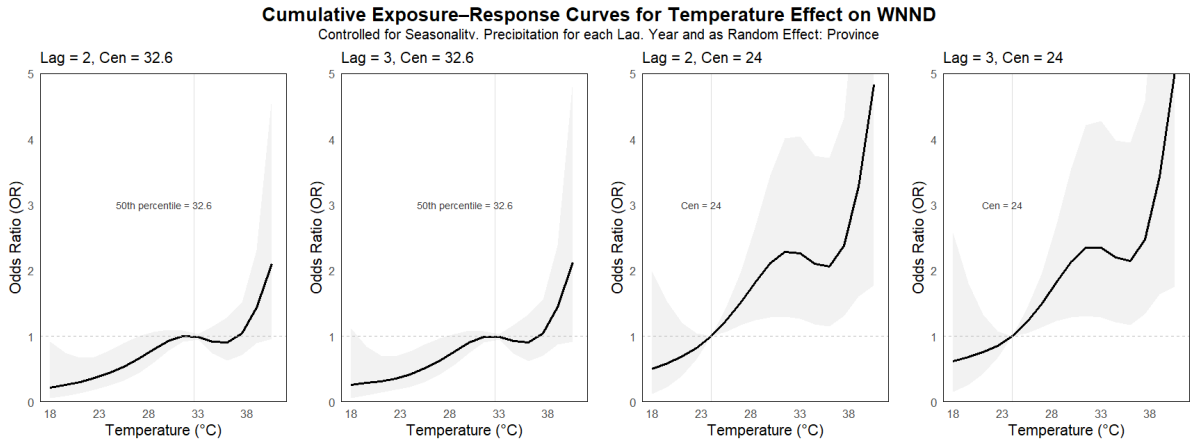


Figure 10: *Cumulative Exposure-Response Curves for temperature effect, OR, on WNND for the lags 2 and 3, with a reference temperature of* 32.6 *°C (the 50th percentile) and* 24 *°C (literature reference).*

Figure 10 shows the changing ORs for the lags 2 and 3 for (**I.**) the 50th percentile and (**II.**) 24°C, the literature reference. In comparison to 32.6°C, lower temperatures under 28 degrees show a decreasing OR and higher temperatures show a tendency of higher

ORs but with high uncertainty. This goes for both cumulative lags. When centering the temperature reference at 24°C, one can see that the OR of experiencing WNND rises with increasing temperature, over 24°C. The uncertainty also increases but the significant trend is over the cumulated lags clearly visible.

For the precipitation models no clear interpretation can be made. Figure A.10 shows big confidence intervals making it not possible to see a trend for the precipitation's effect on the outcome of WNND.

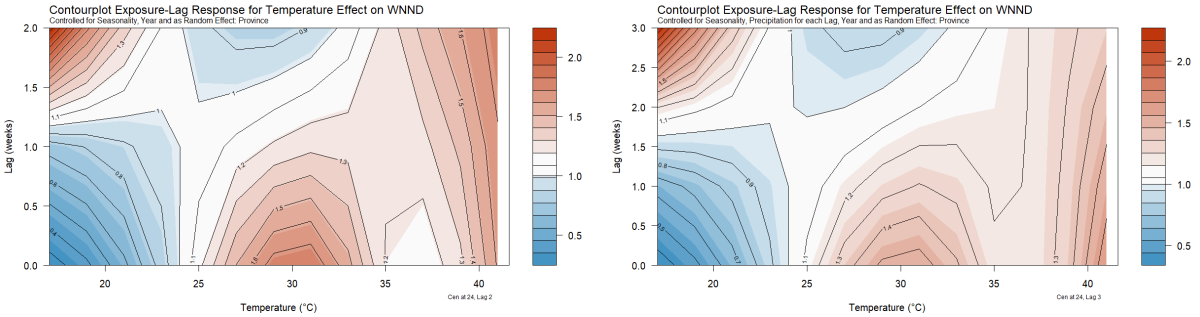


Figure 11: *Contour Exposure-Lag-Response Plot for temperature effect, OR, on WNND for the lag 2 and 3, with a reference temperature of 24 °C.*

The Contour Plots, corresponding to the 3D Exposure-Lag-Response Surfaces in Figure A.11, show a trend for both lags. Temperatures below 20°C show a very strong increase in the OR up until 2, the higher the delay. The opposite is the case for small delays, the OR shrinks to less than 0.5. Temperatures between 25 and 35°C show an increasing OR around 30°C for delays under one week. Temperatures over approximately 38°C show independently of the lag an increasing risk of experiencing a WNND. The same pattern is visible for lag 3.

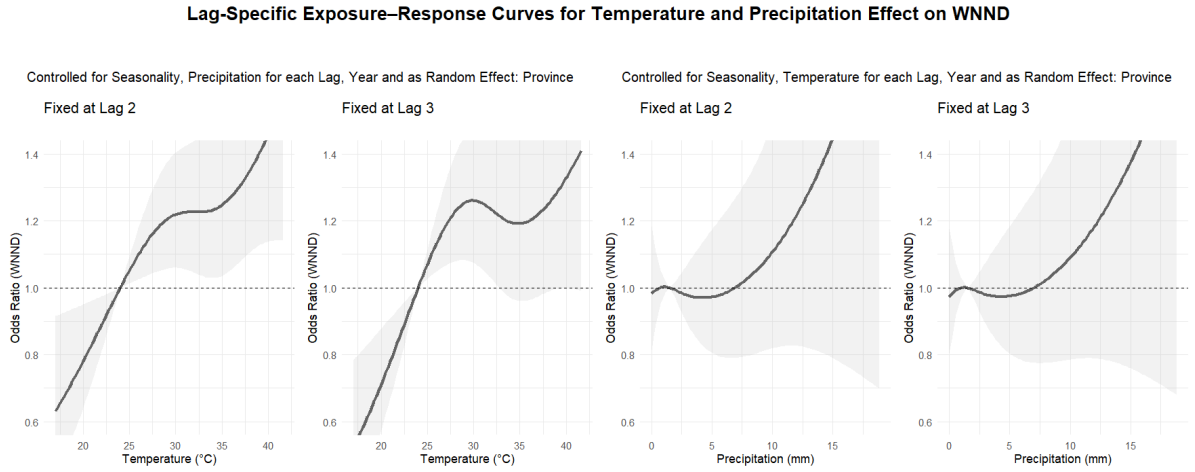


Figure 12: *Lag(2,3)-Specific Exposure-Response Curves for temperature and precipitation effect, OR, on WNND, with a reference temperature of 24 °C and precipitation 1.5mm (the 50th percentile).*

An apparent pattern is visible in the Lag-Specific Exposure-Response Curves for the GADLNMs including temperature as the cross-basis. The effect of increasing temperatures on the odds of experiencing a WNND is increased. For lag 2 and 3 temperatures over 24°C correspond to an increasing OR. For lag 3 temperatures over 32°C have high uncertainty. Temperatures under 24°C have decreasing ORs in comparison to 24°C. Precipitation effects are associated with substantial uncertainty, and the estimated OR are not distinguishable from an OR of 1. Therefore no clear trend is observed.

The Lag-Response Curves at fixed temperatures in Figure A.12 visualize the ORs in comparison to 24°C. Temperatures below 24°C have decreasing risk for small delays. Higher delays have the tendency to increasing risk. For temperatures above 24°C the opposite trend is observed. Higher temperatures correspond at higher lags to lower ORs and at lower lags, the ORs is increased compared to 24°C. Precipitation has strong uncertainty, wherefore no clear trend is visible.

6.3.2 WNND with Weather Trend and Mosquito Perspective

Finally, in addition to the previous section, the mosquito perspective is added to the model to explain the outcome of the occurrence of WNND per calender week per region.

An initial overview can be obtained through GLMs. The models presented in Section 6.3.1 is now extended by the predictor *WNV test results*. The following is stated on average, when holding the other predictors constant. The health outcome WNND is significantly influenced by the predictors *temperature* and the *WNV test result*. With an increase of one degree the odds of experiencing WNND is increased by the factor 1.10. A positive WNV test result results in an OR of 1.63. Including precipitation in the model, does not alter the parameters. With each additional millimeter precipitation significantly alters the odds by the factor 0.95. To expand the model, the predictors *year* and *seasonality* are added. With each additional month, the odds of experiencing WNND are significantly changed by the factor 1.36.

The influence of temperature and precipitation, for the lags 3 and 8, is tested on the outcome WNND. The GADLNMs including the temperature within the cross-basis function also control for seasonality, precipitation at each lag, year, WNV test results, province and as random effects location and municipality. This holds for lag 3 and 8. The model with precipitation included as cross-basis for lag 3 and 8, also controls for seasonality, temperature at each lag, year, WNV test results, province and as random effects location and municipality.

First, the GADLNM incorporating temperature as the cross-basis is considered. A positive WNV test result multiplies the odds by the factor 1.50 for a lag of 3 and by 1.32 for a lag of 8. For the predictor *year* an OR of 1.07, at lag 3, and an OR of 1.04, at lag 8, is observed. Suggesting an increase of the odds of experiencing WNND of 7 or 4%, respectively, annually. Precipitation has neither for a lag of 8 nor for a lag of 3, at any given delay an additional significant effects on the output. The regions Ferrara, Piacenza,

Parma, Ravenna, Reggio Emilia, Rimini have a significant decreasing effect on the odds of experiencing WNND, independent of the lag. Modena stands out as it significantly increases the odds of experiencing WNND, with an OR of 2.44 for a lag of 3 and 2.92 for a lag of 8. Next, precipitation is included as the cross-basis in the model. A positive WNV test result translates to an OR of 1.63, when considering lag 3. This is effect is not as strong for a lag of 8 with a OR of 1.35. An additional year leads to an increase of the odds by the factor 1.08, for lag 3, and 1.05 for lag 8. The included temperature predictors for each lag does not report an additional significant effect on the outcome, independent of the lag. The regions Ferrara, Piacenza, Parma, Ravenna, Rimini have a significant decreasing influence on the odds of experiencing WNND. Again, the region Modena stands out, as the odds are increased by the factor 2.00, for a lag of 8 and 1.78, for a lag of 3.

The seasonality exhibits a global significant quadratic effect on WNND, for both lags in both the temperature and precipitation model. However, the effect is associated with substantial pointwise uncertainty, and monthly estimates are not locally distinguishable from zero (see Figure A.13). This relates probably to the model's definition, including multiple explanatory predictors and a low effective degree of freedom for the seasonality term. At lag 3, the temperature model has a R^2 of 0.254 and the precipitation model has a R^2 of 0.237. At lag 8, this only slightly changes for both: the temperature model has a R^2 of 0.276 and the precipitation model a R^2 of 0.244. The variance of the spatial random effect municipality level ranges between 0.106 (lag 3) and 0.151 (lag 8), whereas the variance associated with longitude and latitude is 0.0. In contrast, substantially larger differences are observed between months, with estimated variances of 3.028 (lag 3) and 3.289 (lag 8).

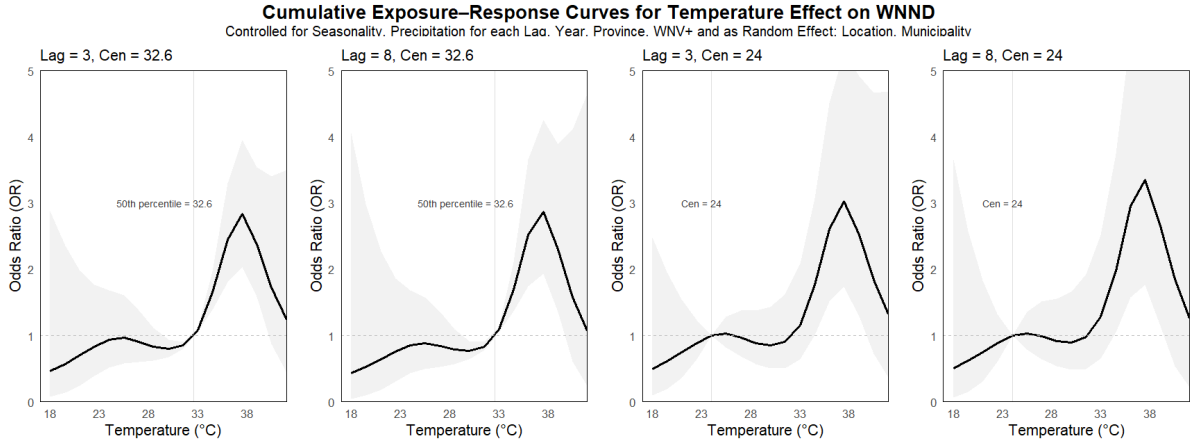


Figure 13: *Cumulative Exposure-Response Curves for temperature effect, OR, on WNND for the lags 3 and 8, with a reference temperature of 32.6 °C (the 50th percentile) and 24 °C (literature reference).*

The Cumulative Exposure-Response Curves show higher ORs for higher temperatures. For temperatures between 32.6°C and 40°C the odds are significantly increased. The odds of experiencing a WNND are increased by a factor up until approximately 2.00. When

considering the median as the reference value, the temperatures ranging from 30°C to 32.6°C show a decreasing OR. This trend is equally for both lags.

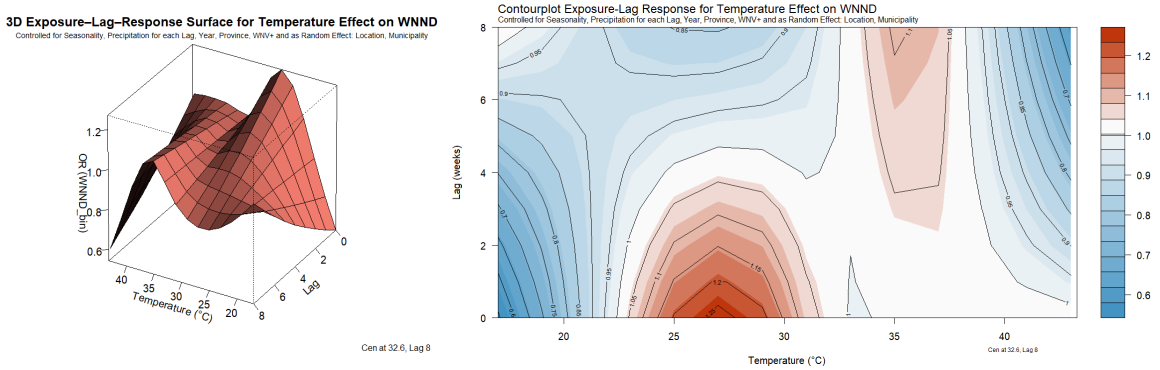


Figure 14: *3D Exposure-Lag-Response Surface and Contour Plot for temperature effect, OR, on WNND for the lag 8, with a reference temperature of 32.6 °C (the 50th percentile).*

The 3D Exposure-Lag-Response Surface and Contour Plot for the temperature effect on the outcome WNND takes the WNV positive test results into account and show the effects in comparison to the centering median temperature of 32.6°C. Temperatures between 23°C and 32°C are associated with an increasing OR for lags under 4 weeks up until an OR of 1.25 for temperatures of 27°C to 28°C at a lag of 1. Low temperatures and low lags translate into a decreasing OR as well as high temperatures and high lags until an OR of 0.6. Temperatures of approximately 35°C to 38°C show a slight increase in the odds with an OR up until 1.1. For temperatures in the same interval, and lags below 3, show a similar OR to the OR at the reference value of 32.6°C.

Figure A.15 and A.16 show the 3D Exposure-Lag-Response Surface and Contour Plot of the precipitation effect on WNND when taking the WNV test results into account. The pattern of the increasing OR is contrarily for lag 3 and lag 8. For former, high precipitation over 15mm at lags below 0.5 are associated with higher odds of experiencing WNND. For later, high precipitation over 15mm relates to a strong increase in the OR until an OR of 3.5 at highest precipitation and highest lag. A decreasing trend is visible for high precipitation at lower lags. An OR of 0.2 is reported. The association of higher precipitation with higher OR depends on the specific lag for the two models.

The temperature and precipitation effects on the outcome WNND are shown in the Lag-Specific Exposure-Response Curves in Figure 15 (left). The course of the curve is more pronounced for a lag of 8 than a lag of 3.0 At lag 3, lower temperatures, below 22.5°C, are related to decreasing odds in comparison to 32.6°C, however the confidence interval is large. For other temperatures is no clear pattern visible. At lag 8, temperatures between 25°C and 32.6°C result in higher ORs up until 1.1 for 28.7°C in comparison the the median temperature of 32.6°C. Temperatures between 32.6°C and 36.25°C are associated with slightly lower odds. Higher temperatures over 43°C show a trend towards an increase in the OR.

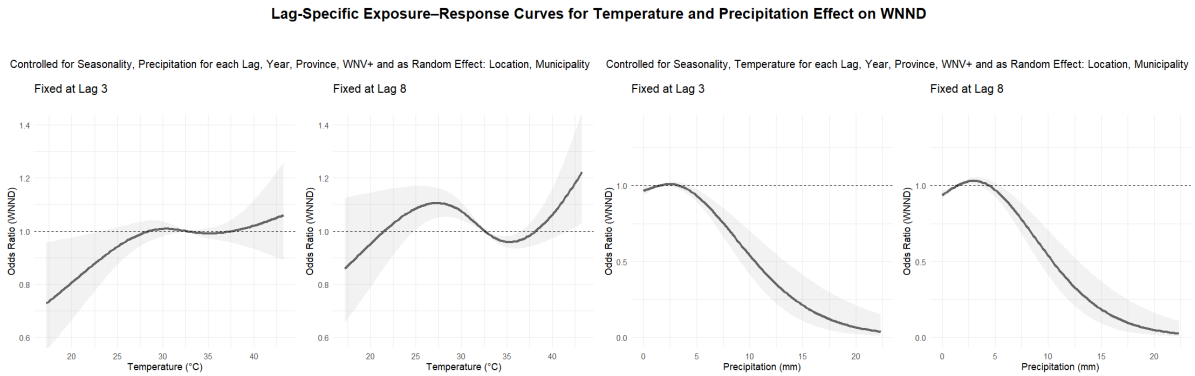


Figure 15: *Lag(3,8)-Specific Exposure-Response Curves for temperature and precipitation effect,* Or*, on* Wnnd*, with a reference temperature of* 32.6 °C *and precipitation* 1.43*mm (the 50th percentile, respectively).*

The Figure 15 (right) suggests that the higher the precipitation (over 5 mm) the lower the Or, in comparison to a median precipitation level of 1.43 mm for the lags 3 and 8. A slight increase over an Or of 1 is visible for 2.5 mm precipitation at a lag of 8. At a precipitation of roughly 10 ± 2.5 the Or is approximately halved. The Or is quartered at a precipitation of circa 15 ± 2.5 for lag 8 and 15.5 ± 3.1 for lag 3. Even higher precipitation results in Ors close to zero.

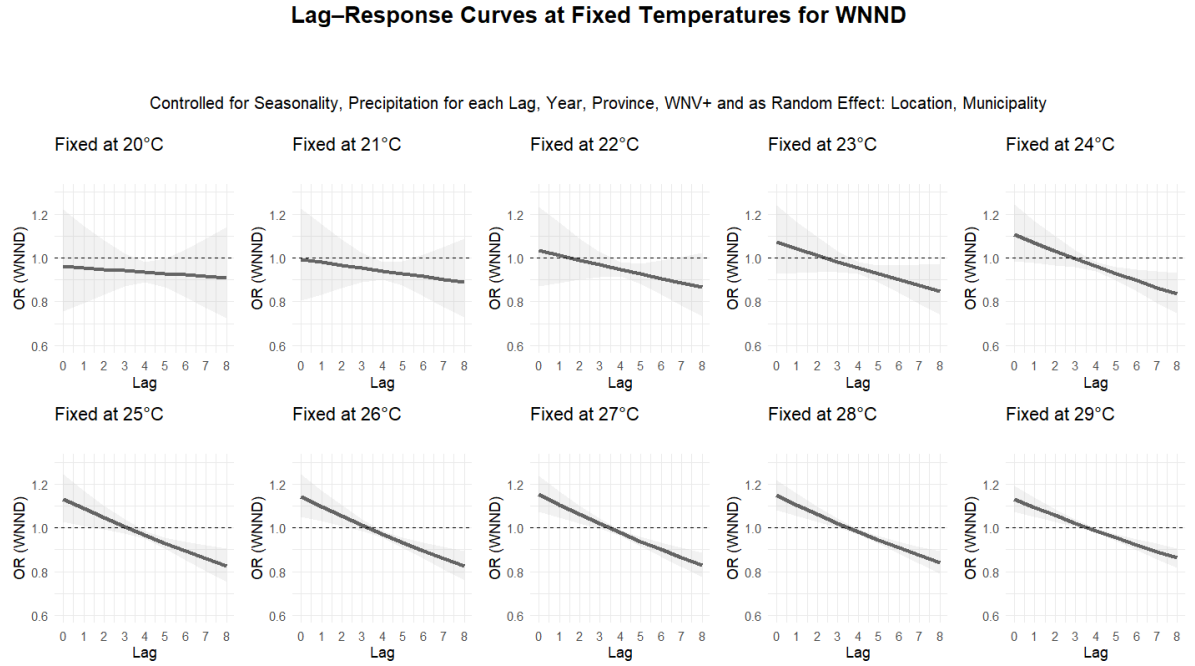


Figure 16: *Lag-Response Curves at fixed temperatures between* 20 °C *and* 29 °C *showing the temperature effect,* Or*, on* Wnnd *for the lags until 8, with a reference temperature of* 32.6 °C *(the 50th percentile).*

Further, the Figure 16 shows the Gadlnm fixed at temperature values from 20°C to 29°C, showing a trend of lower lags (lags 0 to 3) with higher Ors and higher lags (lags 4 to 8)

with lower ORs. This effect is even more pronounced for 26°C to 27°C. An OR of 0.85 at lag 8 is visible, when comparing to a reference temperature of 32.6°C. The odds of experiencing WNND are increased at a lag of 0 by the factor of maximal 1.1 for the temperatures over 24°C. Considering the GALDNM estimated with lag 3 the same clear lag-response pattern is not visible (see Figure A.17). For temperatures till 26°C the odds are decreased for the lags 1 and 2 to a minimum OR of 0.9. Temperatures over 26°C do not depict a clear trend.

For the precipitation models the lag-response association can be seen in Figure A.18 and A.19 for lag 8 and 3, respectively. In the model estimated at lag 8 one can see that the higher the precipitation (over 5mm) the more pronounced is the effect: lower lags and precipitation over 5mm relate to a lower OR. The same effect can be observes at a precipitation rate of 0mm and a trend of slightly increasing incidence rate for higher lags (lags 6 to 8). The precipitation model estimated at lag 3 shows a different tendency. For precipitation values from 0mm to 4mm show no significant difference to the reference value of 1.43mm. Higher precipitation amounts are associated to decreasing odds of experiencing WNND with increasing lags (1 to 3).

7 Discussion

This thesis investigated the relationships between the decadal weather trend (2014-2024), mosquito occurrence and human public health outcomes related to WNV infections in Northern Italy, using a DLNM. WNV infections represent a current growing health concern, with increasing case numbers across Europe. It is projected to rise further in the next decades. Climate change, particularly rising temperatures, contributes to the expanding of affected areas and seasonal activity windows of mosquito vectors.

The main findings of this analysis indicate that the weather-related factors, temperature and precipitation, affect the mosquito occurrence and human WNND outcome in delayed and non-linear ways. Temperature and precipitation play important roles in explaining the dynamics of mosquito occurrence and WNV infections in humans. Clear seasonal patterns were observed for both temperature and precipitation. Seasonality was also present in the resulting GADLNMs.

Higher temperatures were cumulatively associated with higher effects on all studied outcomes. In particular, temperatures between 33°C and 40°C were associated with a significantly increased risk of WNND, with stronger effects observed at lower lags. Conversely, lower temperatures reported lower cumulative risks for WNV positive mosquito infections (approximately below 30°C), the number of mosquitoes tested (approximately below 32°C), and WNND (approximately below 28°C). This is more evident for lower lags.

The effects of precipitation were more uncertain. However, a tendency towards decreasing risks with increasing precipitation amounts above 5mm were observed across outcomes. This overall trend was largely independent of the lag value when looking at fixed lags. Lower precipitation levels, 2mm to 5mm, were associated with slightly increased number of mosquitoes tested, more pronounced at lower lags. For WNND, the medium precipitation was linked to small increases in risk, with effects being more distinct for higher lags.

Mainly, WNND cases were considered within this project. Literature reports that 80% of WNV infections are asymptomatic and therefore represent the majority of the cases (ECDC, 2024; Robert-Koch-Institut, 2023). However, this distribution is not reflected in the analyzed data. This discrepancy can be attributed to the fact that asymptomatic infections are difficult to diagnose and are therefore largely underreported. Patients with more severe symptoms, particularly WNND, are more likely to get correctly diagnosed and recorded. Therefore, the proportion of cases observed differ from the percentage reported in literature. Consequently, this analysis mainly focused on WNND cases.

The previous finding, that elderly are more affected, can be supported (ECDC, 2024). The severer the diagnosis, the more distinct this is. Within this analysis especially older men were more affected. The mortality rate of approximately 17% among patients with WNND patients, as reported in literature (ECDC, 2024; Brown et al., 2007; Zou et al., 2010), is broadly consistent with the findings from the data analyzed for Northern Italy.

In the presented data, an overall mortality rate of 14.29% and a close annual median mortality of 14.97% is reported.

Mosquito surveillance data formed the basis for this analysis. Mostly, the traps are located in the Northern half or upper third of each province. Only Ferrera is more evenly covered with traps. This can affect the measuring results. Further, the geographical variation across provinces can be observed: coastal provinces, bordering the Adriatic Sea, exhibit slightly higher temperatures on average compared to areas inland. Reasons for that can likely be the moderating maritime influences. Local climate dynamics are drivers of thermal variation across the region. Differences in population size and the associated concentration of economic activity in larger urban areas, do not appear to have a perceptible effect on temperature patterns.

The climatic conditions in Northern Italy are characterized by higher temperatures, occasionally reaching up to 42°C, and accordingly low and infrequent precipitation. Emilia-Romagna extended its mosquito collection window starting from 2020 onward. This shift may already reflect an adaptive decisions to climate change, as earlier seasonal warming likely encouraged an earlier start of mosquito collection. This trend is consistent with findings that transmission windows are enlarging and mosquitoes were indeed collected in the extended period (ECDC, 2025).

A clear seasonal quadratic pattern is evident in the data. The odds of a positive Wnv test result increase substantially from July through September, indicating a pronounced seasonal peak in Wnv activity during these months - in particular, the end of July and the beginning of August. In contrast, May and June are associated with significantly reduced Ors, suggesting lower transmission risk during these early summer months. The seasonal pattern observed in the model assessing the number of mosquitoes tested reveals a different temporal trend. While the Or of positive Wnv mosquitoes peaks in August, the Irr for the number of mosquitoes tested reaches its maximum in June, followed by a decline in the beginning of August.
The peak of the number of mosquitoes therefore precedes the peak in the positive Wnv mosquitoes, indicating that mosquitoes are present in large numbers before the viral circulation intensifies. With increasing temperature, a higher proportion of mosquitoes test positive for Wnv. The most human infections are reported in the months July till September, especially August, suggesting a small delay between the positive Wnv mosquitoes and the human incidences. Although the temporal overlap between mosquito infections and human cases is evident, the increase in positive Wnv results consistently slightly precedes the rise in human infection and especially Wnnd cases. This seasonal pattern in number of mosquitoes tested and positive Wnv remains consistent across the different models and exposures within the Gadlnm. Thus, this pattern is assumed to capture the mosquito occurrence and Wnv infected mosquitoes properly. However, only the Gadlnm incorporating the weather trend data, health records and mosquito data exhibits a strong uncertainty, limiting the seasonal interpretation.

In general, higher temperatures are associated with higher risks across all outcomes. This pattern is especially pronounced for the health-related outcomes. Specifically, the OR of experiencing WNND increased for temperatures between 25°C and 35°C. When additionally incorporating the mosquito perspective, this elevated OR shifts to a lower temperature range, spanning 23°C to 32°C. Independent of the reference value and specific delay, higher temperatures are cumulatively connected to an increased OR of experiencing WNND. Over all outcomes, this relation is more pronounced for lower lags but remains mostly robust over different lags, indicating that the effect is not mainly driven by temporal delay but the biological response to warming conditions.

This finding is supported by established biological evidence: key processes in the WNV transmission cycle are accelerated by elevated temperatures. Higher temperatures enhance viral replication rates within mosquitoes, the growth rates of vector populations and shorter incubation times (Paz, 2015). This underscores the findings that increased temperatures are associated with higher mosquito occurrence and thus infection risks. While lower temperatures lead to a prolonged life cycle and longer living periods, mosquitoes' life cycle are completed in a faster manner when exposed to higher temperatures, but show shorter lifespans (Kiarie-Makara et al., 2015). A threshold is suggested beyond which high temperatures may suppress mosquito activity, possibly due to heat stress or reduced survival. Exceeding temperature levels of approximately 43°C were observed to lead to the death of larvae of *Aedes aegypti* (Kiarie-Makara et al., 2015). This trend is also coherent in the mosquitoes behavior in this analysis. On the other hand, when considering a delay of 3 or 8 weeks in the data, it is very likely that the higher temperature period is followed by a period with decreased temperatures.

Evidence from studies on *Chikungunya* virus suggests that mosquitoes reared at higher temperatures may adapt better to the increased temperature, which also enhances their susceptibility to arbovirus pathogens. This is also plausible for the *Culex pipiens* mosquitoes, the vector of WNV. In addition, temperature may act as a climatic selection factor, favoring mosquito populations that are both better adapted to the temperature and simultaneously more susceptible towards the arbovirus (Kiarie-Makara et al., 2015).

Higher precipitation amounts are generally associated with a lower risk of WNV and lower precipitation are associated with slightly increased risks, a suggested pattern by the lag-specific results. One possible explanation is that higher amounts of precipitation may wash away relevant nutrients which are essential for the mosquito larvae's development. However, the literature reports contradictory findings regarding the effect of precipitation on the development of *Culex pipiens* (Paz, 2015). In contrast, precipitation amounts over the average, have also been shown to increase mosquito abundance, as more temporary aquatic habitats for larval development become available. This can not be concluded from the presented model.

Lower precipitation, approximately between 2mm and 5mm, was associated with increased risks at specific lags. This may be attributable to the limited access to water, facilitating

closer contact between hosts and mosquitoes, thereby accelerating the epizootic cycle and increasing WNV incidence. This mechanism has already been associated with population outbreaks of mosquito species (Paz, 2015), leading to higher numbers of infected mosquitoes and, consequently, higher numbers of infections in humans.

There are large confidence intervals when evaluating the effects of precipitation, particularly for cumulative effects. Northern Italy, and the Emilia-Romagna region, generally experiences low total precipitation. This is especially relevant in this analysis given that the mosquito surveillance period includes the driest phases of the year, during which the daily precipitation averages are often near zero. The maximum averaged observed precipitation is 20mm. As this study considers weekly averages at the regional or trap level, high precipitation values are further mitigated. The limited number of higher precipitation during the mosquito surveillance period, underscores the challenge in estimating the role of precipitation in the presented WNV dynamics as this contributed to higher uncertainty in the models' estimates for precipitation related effects.
A particularly strong uncertainty could be observed in the cumulative exposure-response curves. The effect and the uncertainty increase with longer delays, as impacts are aggregated over multiple lags. Models estimating the effects of precipitation exposure need to be interpreted with caution, especially when considering extreme values. As a result, it is difficult to report an overall clear trend or change in WNV infections connected with precipitation.

The results suggest that shorter delays are associated with more pronounced effects. Mosquito-related outcomes exhibit stronger effects at shorter delays. A delay of 3 weeks is connected to a clearly significant (cumulative) trend, with higher temperatures corresponding to increased risks. Also, lower temperatures are associated with a significant lower number of mosquitoes tested and fewer positive WNV test results. These findings align with results reported in literature and with the discussed biological evidence. Higher temperatures are connected to faster mosquito life cycles, due to the enhanced viral replication rates. This retroactively supports the initial decision to consider a delay of 3 weeks.

In contrast, health-related outcomes appear largely independent of delay, as the reported results for the delays of 2, 3 and 8 weeks show the similar overall trend. An exception is the increased risk at medium temperature for WNND, which is associated with longer delays. The high uncertainty is present independent of the delays. Across all delays, higher precipitation is connected to lower risks for WNND. At higher lags, a lower risk for very low precipitation levels was found for WNND, when considering both, the weather and mosquito perspective. Despite the uncertainty resulting from limited higher precipitation amounts, it is noteworthy, that the effects of medium and very low precipitation are only observed at a delay of 8 weeks. The clear seasonal trend remains consistent across all delays.

While the findings contribute to a better understanding on relationships between weather trend, mosquitoes and human health, several limitations of this analysis should be acknowledged.

First, the health records are subject to underreporting. Severer diagnoses such as WNND are more likely to be detected and are as a consequence, overrepresented. In addition, the analysis is restricted to the reported positive cases, implying that no explicit control group is available. This limitation applies to both, the mosquito surveillance data and the health records.

Second, more climatic factors influence the mosquito occurrence and WNV infections. While temperature represents a major driver of mosquitoes development and transmission cycles, other variables may also play an important role. In particular, the relation between temperature and precipitation or relative humidity is interesting to further consider. Indicators such as vegetation index, wetness index, evaporation and heat index can also offer a better insight in the disease transmission. Due to increasing complexity, this analysis focused on temperature and precipitation as main climatic factors, even though additional predictors are potentially informative and available.

Third, the uncertainty associated with the model selection process is not explicitly reflected in the presented results, which may lead to an underestimation of the overall uncertainty. The AIC was used in order to select the best-fitting models. However, the method faces critic in the context of complex modeling frameworks, such as DLNMs, as it tends to favor more complex models. Although the AIC selects the best-fitting model by explaining the greatest amount of variation with as little predictors as possible, it tends to favor more complex models. In the context of this project, models incorporating a cross-basis with higher lags may still be preferentially selected. This largely influences the stability and interpretability of the estimated effects. Other possible decision criteria are: the quasi-AIC, which is only advisable for overdispersed outcomes, BIC, which prefered on average the same models within this project, the Watanabe–Akaike information criterion (Watanabe, 2010), which is an extension of the AIC or cross-validation.

Fourth, the modeling framework involves a large number of parameters. Embedding a DLNM into a GAM introduces multiple modeling choices, each of which can substantially influence the results. In this project, several parameters were pre-defined in order to limit model flexibility and to maintain the focus on the lag specification and model selection. However, this approach introduces uncertainty that is not explicitly quantified. A simulation study could provide a valuable insight into the sensitivity of the results to these modeling choices. On the other hand, such a simulation study would substantially increase the model's complexity due to the large number of parameters and their interaction.

Despite these limitations this project provides a valuable insight in the transmission dynamics and the effects of the weather trend on the WNV infections.

8 Conclusion

In order to effectively fight the growing threat posed by WNV it is essential that the effects of climate on mosquito occurrence, spread and public health must be better understood.

The thesis contributes to research by implementing all three pillars into a DLNM and allowing to adjust for the delayed and non-linear effects. The embedding into a GADLNM made it possible to jointly model the delayed and non-linear impacts of the climatic variables, but also including other potentially non-linear predictors. The thesis extended the GADLM framework with the inclusion of non-linear predictor terms (GADLNM).

The findings are mostly consistent with biological evidence. Higher temperatures are associated with increased risks of experiencing WNND and higher proportions of WNV-positive mosquitoes as replication rates are enhanced. However, a threshold of approximately 40°C appears to suppress mosquito activity. In comparison, lower temperatures exhibit slightly decreased risks for both, mosquito- and health-related outcomes. These effects are more evident for shorter delays, aligning with research on the relationship between weather trend and mosquito dynamics.

The effects of precipitation are rather difficult to interpret, as low and infrequent precipitation result in only few extreme observations. In general, higher precipitation is connected to a decreasing risk of WNND as essential nutrients for mosquito development can be washed away. Medium precipitation amounts (2mm to 5mm) are associated with slightly increased risks across all considered outcomes. This is more pronounced at a delay of 8 weeks. For the number of mosquitoes tested, the risk is also slightly increased at the shorter delay of 3 weeks. This may be linked to the reduced availability of standing water, which can enhance the mosquito-bird interactions.

The relation of weather trend and mosquito dynamics on WNND reports stronger effects for longer delays, suggesting a delayed pathway from weather trend through WNV-infected mosquitoes on human health.

Several limitations should be acknowledged. The WNV-infection is underreported, wherefore this project focused on WNND cases. Uncertainty related to the model selection can not be fully captured in the reported results. Finally, additional climatic factors influence the outcomes and need to be considered in future research.

Despite the limitations, this project provides a better insight in WNV occurrence and WNND in a Mediterranean climate described through temperature and precipitation. It therefore offers a basis for improving climate-informed public health interventions in comparable regions across Europe to build resilience against future outbreaks and supporting the development of adaptive public health strategies. This analysis underscores the importance of mosquito surveillance and climate-related health risks. This not only holds for WNV but all other vector-borne diseases who are expected to expand or emerge in Europe over the next decades.

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, *33*(1), 178–196.

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, *17*(6), 624–631.

Berry, K., & Wei, S. (2013). Distributed lag models. Tech. rep., Department of Economics and Finance, University of Wyoming. Advised by Professor David Aadland.

Braga, A. L., Zanobetti, A., & Schwartz, J. (2001). The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 us cities. *Journal of Occupational and Environmental Medicine*, *43*, 927–933.

Braga, A. L. F., Zanobetti, A., & Schwartz, J. (2002). The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. *Environmental Health Perspectives*, *110*(9), 859–863.

Brinkhoff, T. (2025a). City Population: Emilia-Romagna (Italy): Population Statistics, Charts and Map.
URL `https://www.citypopulation.de/en/italy/emiliaromagna/` 2025-01-07

Brinkhoff, T. (2025b). City Population: Lombardia (Italy): Population Statistics, Charts and Map. `https://www.citypopulation.de/en/italy/lombardia/`. 2025-01-07

Brown, J. A., Factor, D. L., & Tkachenko, N. (2007). West nile viremic blood donors and risk factors for subsequent west nile fever. *Vector-Borne and Zoonotic Diseases*, *7*, 479–488.

Calzolari, M., Angelini, P., & Bolzoni, L. (2020). Enhanced west nile virus circulation in the emilia-romagna and lombardy regions (northern italy) in 2018 detected by entomological surveillance. *Frontiers in Veterinary Science*, *7*, 243.

Centers for Disease Control and Prevention (CDC) (2022). Mosquito life cycle: Culex pipiens, cx. quinquefasciatus, and cx. tarsalis. `https://www.cdc.gov/mosquitoes/pdfs/culexlifecycle-p.pdf`. 2026-01-17

Cotar, A. I., et al. (2016). Transmission dynamics of the west nile virus in mosquito vector populations under the influence of weather factors in the danube delta, romania. *EcoHealth*, *13*, 796–807.

Curriero, F. C., Heiner, K. S., Samet, J., et al. (2002). Temperature and mortality in eleven cities of the eastern united states. *American Journal of Epidemiology*, *155*, 80–87.

ECDC, E. (2024). West nile virus infection. `https://www.ecdc.europa.eu/en/west-nile-virus-infection`. 2025-07-15

ECDC, E. (2025). World mosquito day 2025: Europe sets new records for mosquito-borne diseases. `https://www.ecdc.europa.eu/en/news-events/world-mosquito-day-2025-europe-sets-new-records-mosqu` 2025-12-02

European Centre for Disease Prevention and Control (ECDC) (2021). Surveillance, prevention and control of west nile virus and usutu virus infections in the eu/eea. `https://www..europa.eu/sites/default/files/documents/Surveillance_prevention_and_control_of_WNV_and_Usutu_virus_infections_in_the_EU-EEA.pdf`. 2025-07-15

European Food Safety Authority (2018). Field sampling methods for mosquitoes, sandflies, biting midges and ticks – vectornet project 2014-2017. Tech. Rep. EN-1435, EFSA Supporting Publications.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*, 150066.

Garrigós, M., Garrido, M., Panisse, G., Veiga, J., & Martínez-de la Puente, J. (2023). Interactions between west nile virus and the microbiota of culex pipiens vectors: A literature review. *Pathogens*, *12*(11), 1287.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine*, *33*, 881–99.

Gasparrini, A., Armstrong, B., & Kenward, M. (2010). Distributed lag non-linear models. *Statistics in Medicine*, *29*(21), 2224–2234.

Gasparrini, A., Armstrong, B., & Kenward, M. G. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, *43*(8), 1–20.

Giles, D. E. (2017). Explaining the almon distributed lag model. Blog post on "Econometrics Beat: Dave Giles' Blog", `https://davegiles.blogspot.com/2017/01/explaining-almon-distributed-lag-model.html`.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. `https://doi.org/10.1016/j.rse.2017.06.031`.

Hart, J. E., Hu, C. R., Yanosky, J. D., Holland, I., Iyer, H. S., Borchert, W., Laden, F., & Albert, C. M. (2024). Short-term exposures to temperature and risk of sudden cardiac death in women: A case-crossover analysis in the nurses' health study. *Environmental Epidemiology*, *8*(4), e322.

Heidecke, J., Wallin, J., Fransson, P., Singh, P., Sjödin, H., Stiles, P. C., Treskova, M., & Rocklöv, J. (2025). Uncovering temperature sensitivity of west nile virus transmission: Novel computational approaches to mosquito-pathogen trait responses. *PLOS Computational Biology*, *21*(3), 1–35.

Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IZSLER) (2025). Bruno ubertini. `https://www.izsler.it/`. 2025-08-26

Kiarie-Makara, M. W., Ngumbi, P. M., & Lee, D.-K. (2015). Effects of temperature on the growth and development of *culex pipiens* complex mosquitoes (diptera: Culicidae). *IOSR Journal of Pharmacy and Biological Sciences*, *10*(6 Ver. II), 01–10.

MacDonald, A. J., Hyon, D., Sambado, S., Ring, K., & Boser, A. (2024). Remote sensing of temperature-dependent mosquito and viral traits predicts field surveillance-based disease risk. *Ecology*, *105*(11), e4420.

Marini, G., et al. (2020). A quantitative comparison of west nile virus incidence from 2013 to 2018 in emilia-romagna, italy. *PLoS Neglected Tropical Diseases*, *14*(1), e0007953.

Moirano, G., et al. (2024). Short-term effect of temperature and precipitation on the incidence of west nile neuroinvasive disease in europe: a multi-country case-crossover analysis. *The Lancet Regional Health – Europe*, *48*, 101149.

Mulhern, O. (2020). The time lag of climate change. `https://earth.org/data_visualization/the-time-lag-of-climate-change/`. 2025-12-26

Nanni Costa, A., et al. (2011). West nile virus: the italian national transplant network reaction to an alert in the north-eastern region, italy 2011. *Eurosurveillance*, *16*(41), pii=19991.

NASA Goddard Space Flight Center (2025). About MODIS. `https://modis.gsfc.nasa.gov/about/`. 2025-08-26

Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., Huang, A., Rosenberg, A., Greenberg, A., Sherman, M., et al. (2001). The outbreak of west nile virus infection in the new york city area in 1999. *New England Journal of Medicine*, *344*, 1807–1814.

Orton, S. L., Stramer, S. L., & Dodd, R. Y. (2006). Self-reported symptoms associated with west nile virus infection in rna-positive blood donors. *Transfusion*, *46*, 272–277.

Parker, J. (2013). Chapter 3: Distributed-lag models. Online PDF. Economics 312: Theory and Practice of Econometrics, Spring 2013. 2025-12-26

Paz, S. (2015). Climate change impacts on west nile virus transmission in a global context. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1665), 20130561.

Price, V. (2024). Climate lag. `https://www.ebsco.com/research-starters/environmental-sciences/climate-lag`. 2025-12-26

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 4.0.2.

Robert-Koch-Institut (2023). *Falldefinition: West-Nil-Fieber (West-Nil-Virus)*. Robert-Koch-Institut. RKI-Falldefinition gemäß § 11 IfSG zur epidemiologischen Überwachung von Infektionskrankheiten. 2025-12-23

Ross, N. (2025). Generalized additive models in r: Chapter 2 – interpreting and visualizing gams. `https://noamross.github.io/gams-in-r-course/chapter2`. 2025-01-04

Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology*, *11*(3), 320–326.

Smithburn, K. C., Hughes, T. P., Burke, A. W., & Paul, J. H. (1940). A neurotropic virus isolated from the blood of a native of uganda. *The American Journal of Tropical Medicine*, *s1-20*(4), 471 – 492.

SurveillanceMoustiques (2026). Life cycle. `https://mosquitosurveillance.be/life-cycle`. 2026-01-20

Van Rossum, G., et al. (2024). *Python Language Reference, Version 3.12*. Python Software Foundation.

Wan, Z., Hook, S., & Hulley, G. (2021). MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V061. NASA LP DAAC, `https://doi.org/10.5067/MODIS/MOD11A2.061`. 2025-09-10

Wang, P., Zhang, X., Hashizume, M., Goggins, W. B., & Luo, C. (2021). A systematic review on lagged associations in climate–health studies. *International Journal of Epidemiology*, *50*(4), 1199–1212.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594. 2026-01-23.

Weierstrass, K. (1885). Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, (II).

Wood, S. (2025). *Random effects in GAMs*. 2025-10-21.

Wood, S. N. (2010). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36.

Wood, S. N. (2026). *gam.models: Specifying Generalized Additive Models*. R documentation for the `mgcv` package, R-manual, version 1.9-4. 2026-01-03

Zanobetti, A., Wand, M. P., Schwartz, J., & Ryan, L. M. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, *1*(3), 279–292.

Zeller, H. G., & Schuffenecker, I. (2004). West nile virus: An overview of its spread in europe and the mediterranean basin in contrast to its spread in the americas. *European Journal of Clinical Microbiology and Infectious Diseases*, *23*(3), 147–156.

Zou, S., Foster, G. A., Dodd, R. Y., Petersen, L. R., & Stramer, S. L. (2010). West nile fever characteristics among viremic persons identified through blood donor screening. *Journal of Infectious Diseases*, *202*, 1354–1361.

Appendix A: Figures

A.1 Seasonal Weather Trend Patterns

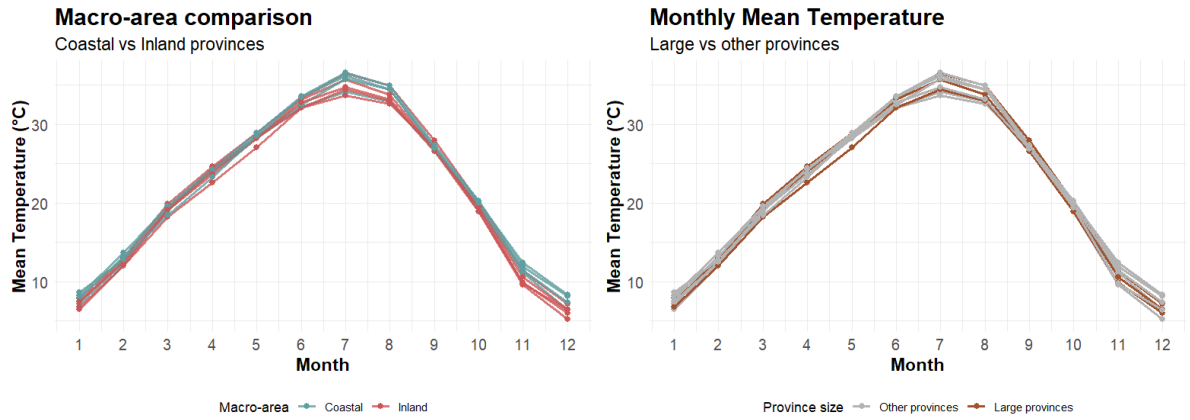


Figure A.1: *Seasonal pattern of the averaged temperature by province of Emilia-Romagna. Then, categorized into a coastal and inland macro-area (left) and differentiated by the province's inhabitant size (right). The provinces Rimini, Ferrara, Ravenna and Forlì-Cesena have a coastline. The provinces Bologna, Modena and Parma are labeled as large provinces.*

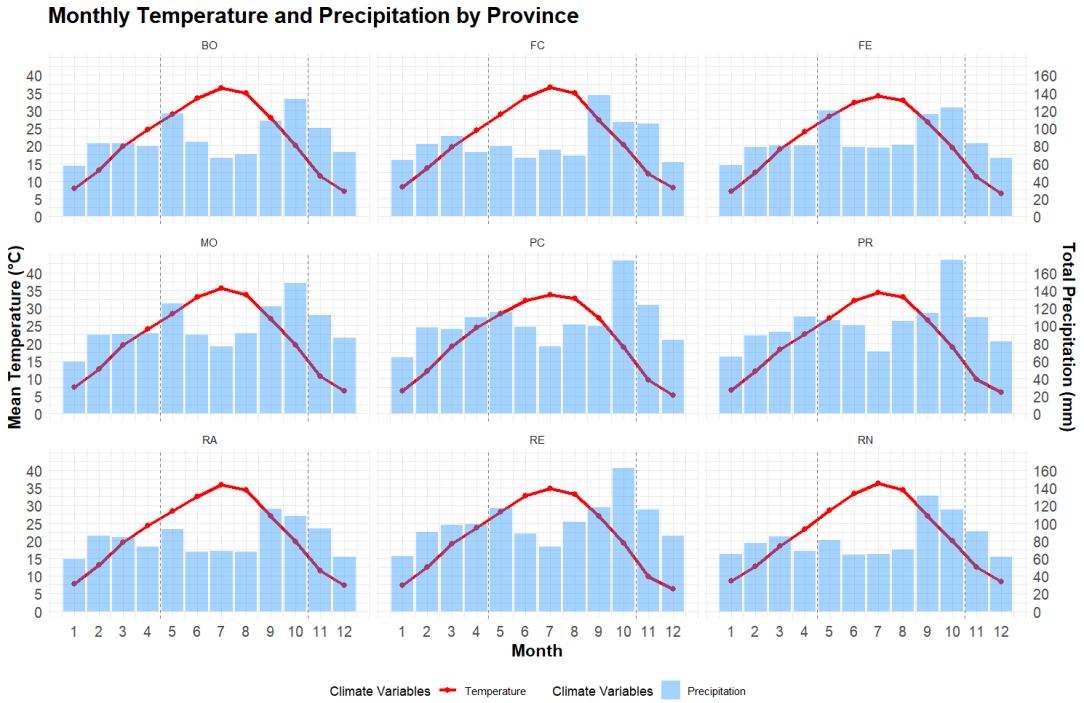


Figure A.2: *Seasonal pattern of the averaged temperature and total precipitation by provinces of Emilia-Romagna (BO = Bologna, FC = Forlì-Cesena, FE = Ferrara, MO = Modena, PC = Piacenza, PR = Parma, RA = Ravenna, RE = Reggio Emilia, RN = Rimini).*

A.2 Additional Figures for the WNV Response

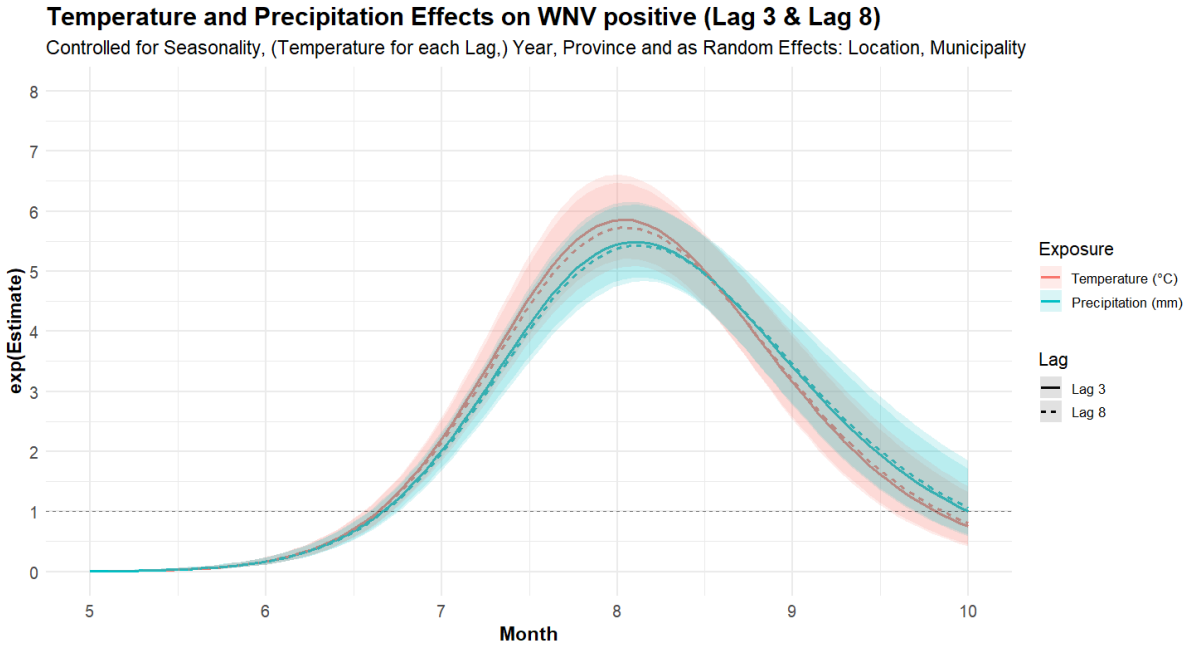


Figure A.3: *The effect of the smoothing term seasonality,* Or*, on the outcome* Wnv*, divided for cross-basis predictor (temperature and precipitation) and lag (3 and 8).*

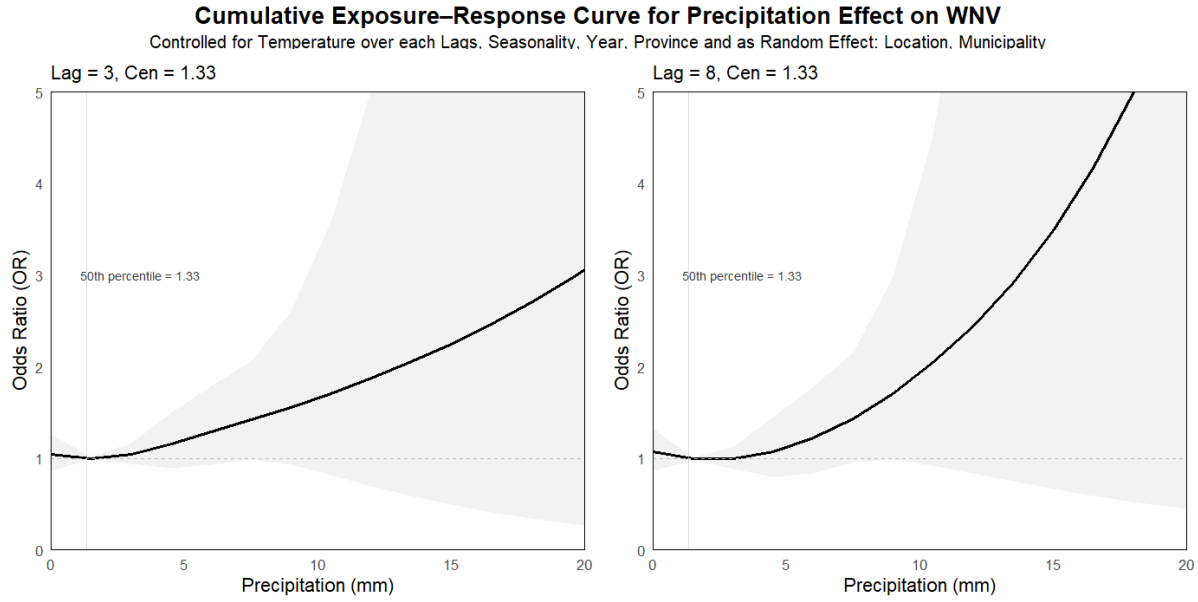


Figure A.4: *Cumulative Exposure-Response Curves for precipitation effect,* Or*, on* Wnv *for the lags 3 and 8, with a reference daily precipitation of* 1.33*mm (the 50th percentile).*

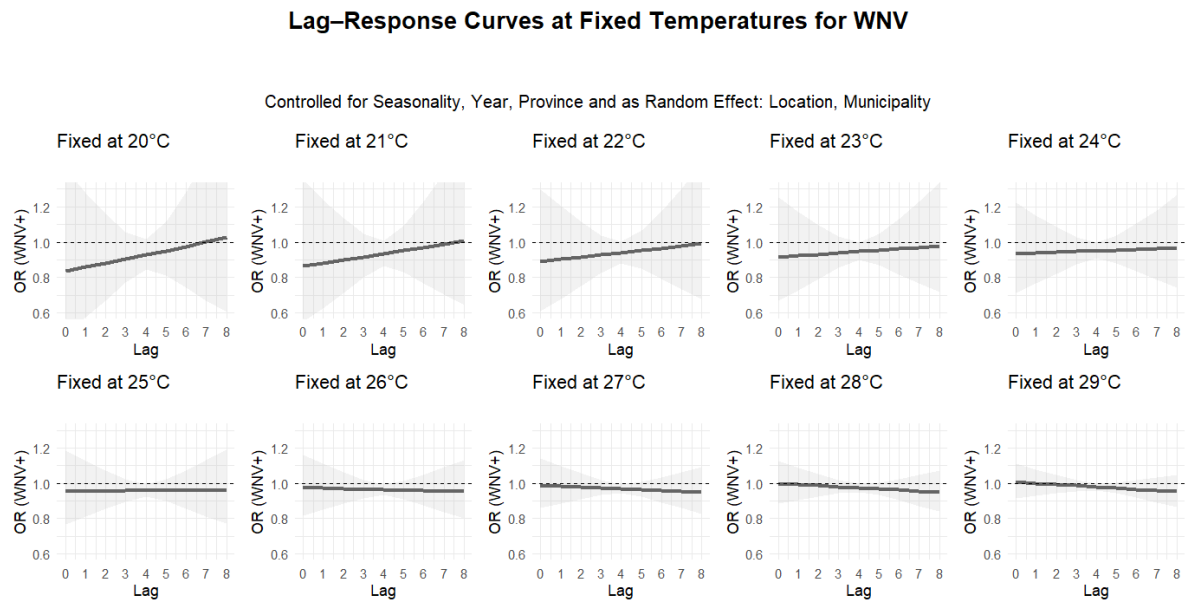


Figure A.5: *Lag-Response Curves at fixed temperatures between 20°C and 29°C showing the temperature effect,* Or*, on* Wnv *for the lags until 8, with a reference temperature of 32.6 °C (the 50th percentile).*

A.3 Additional Figures for the Number of Mosquitoes Tested Response

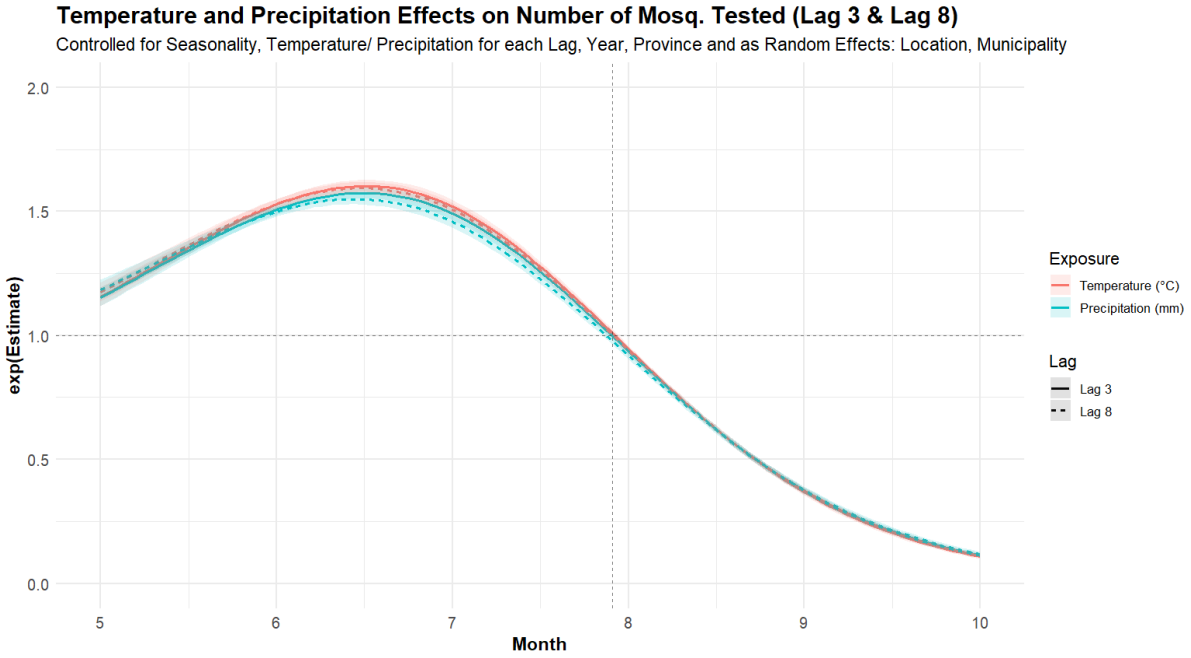


Figure A.6: *The effect of the smoothing term seasonality,* Irr*, on the outcome number of mosquitoes tested, divided for cross-basis predictor (temperature and precipitation) and lag (3 and 8).*

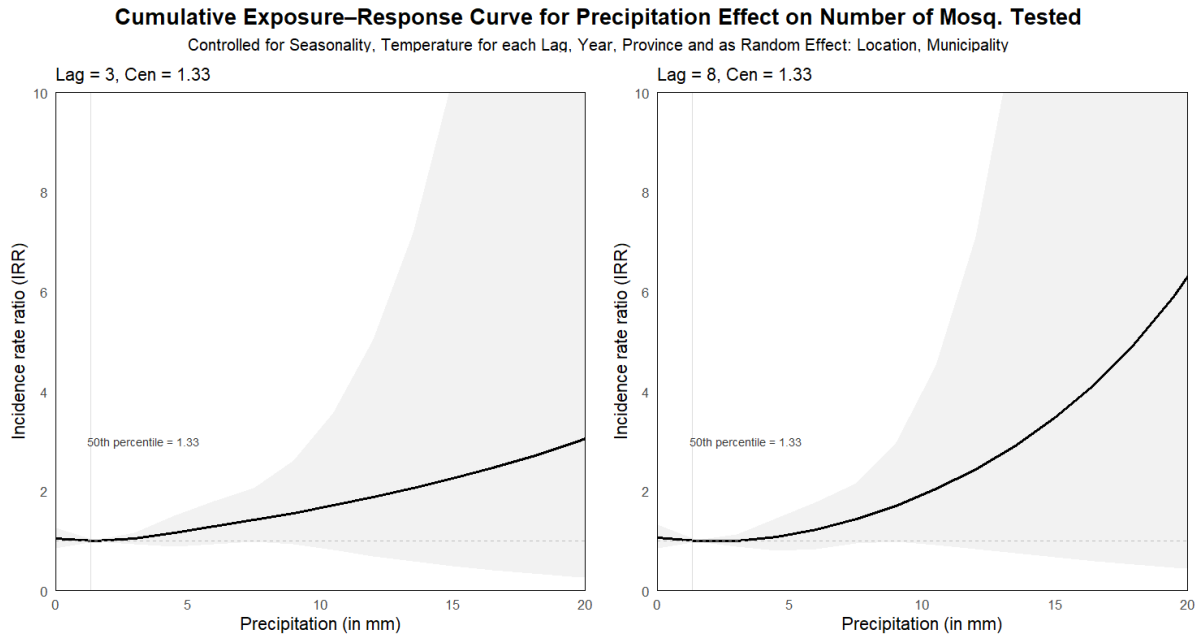


Figure A.7: *Cumulative Exposure-Response Curves for precipitation effect,* IRR*, on number of mosquitoes tested for the lags 3 and 8, with a reference daily precipitation of* 1.33*mm (the 50th percentile).*

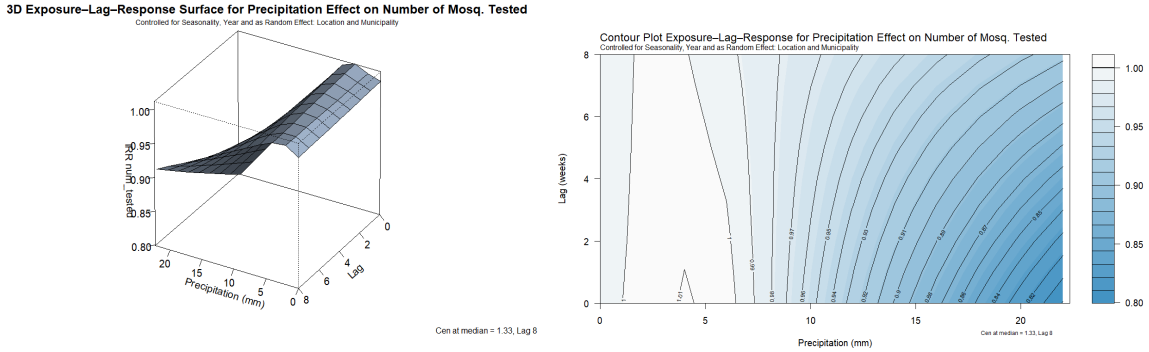


Figure A.8: *3D Exposure-Lag-Response Surface and Contour Plot for precipitation effect,* IRR*, on number of mosquitoes tested for the lag 8, with a reference precipitation of* 1.33*mm (the 50th percentile).*

A.4 Additional Figures for WNND Response incorporating Weather Trend Perspective

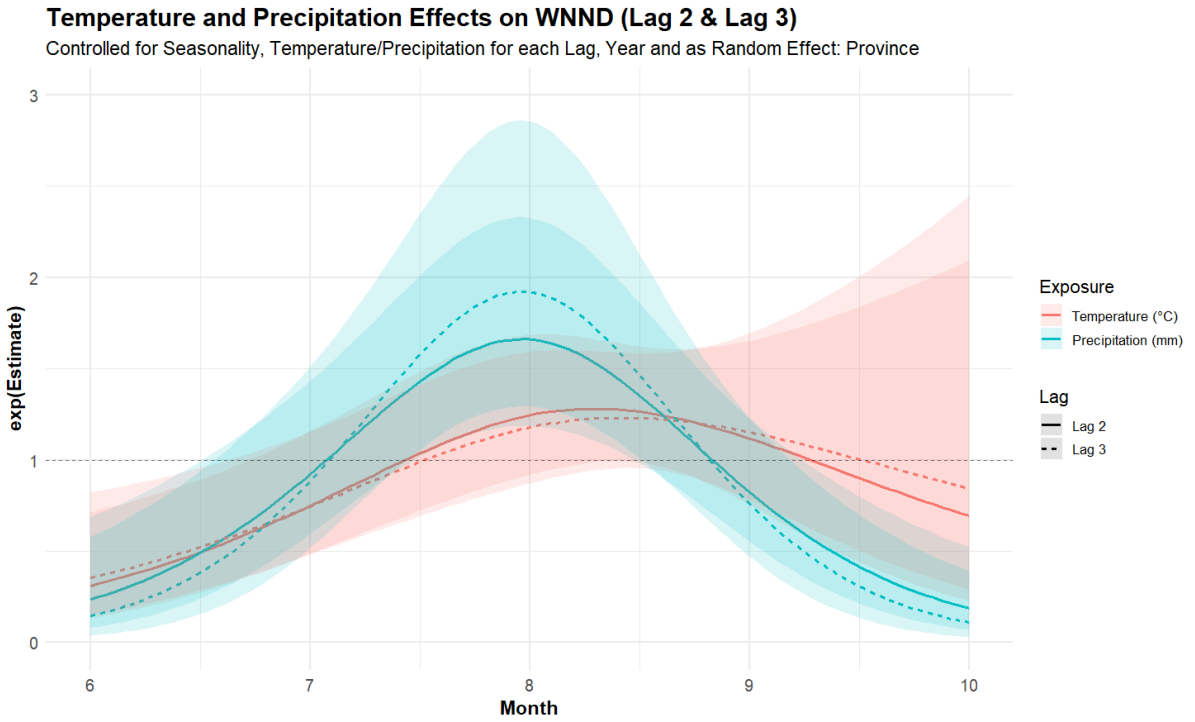


Figure A.9: *The effect of the smoothing term seasonality,* OR*, on the outcome* WNND*, divided for cross-basis predictor (temperature and precipitation) and lag (2 and 3).*

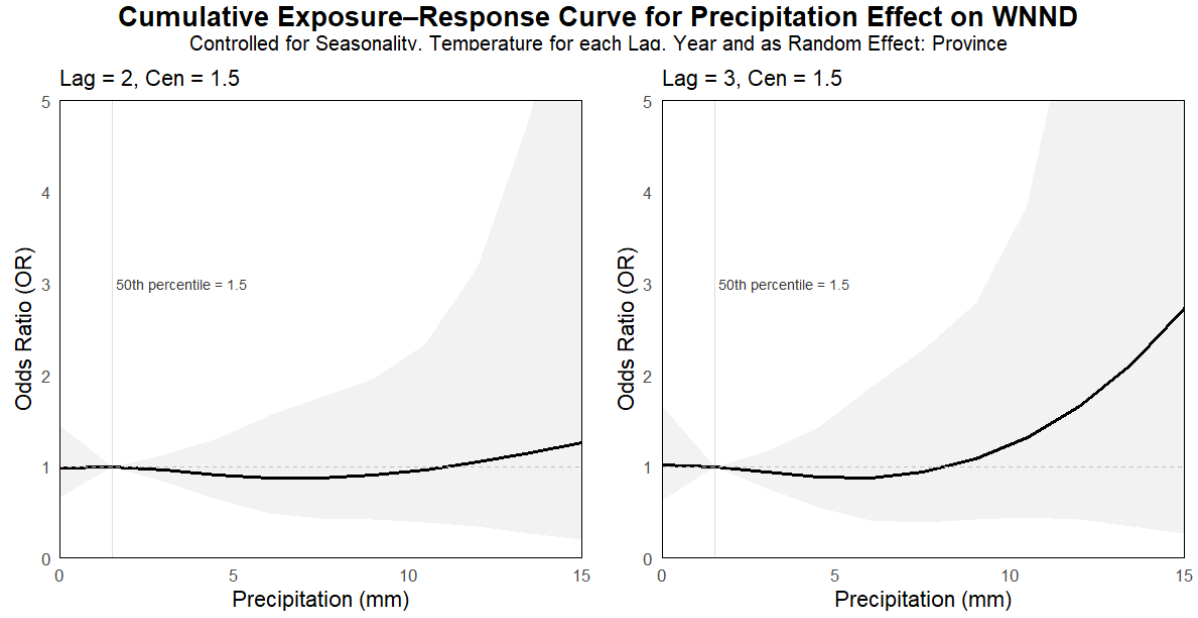


Figure A.10: *Cumulative Exposure-Response Curves for precipitation effect,* OR*, on* WNND *for the lags 2 and 3, with a reference daily precipitation of* 1.5*mm (the 50th percentile).*

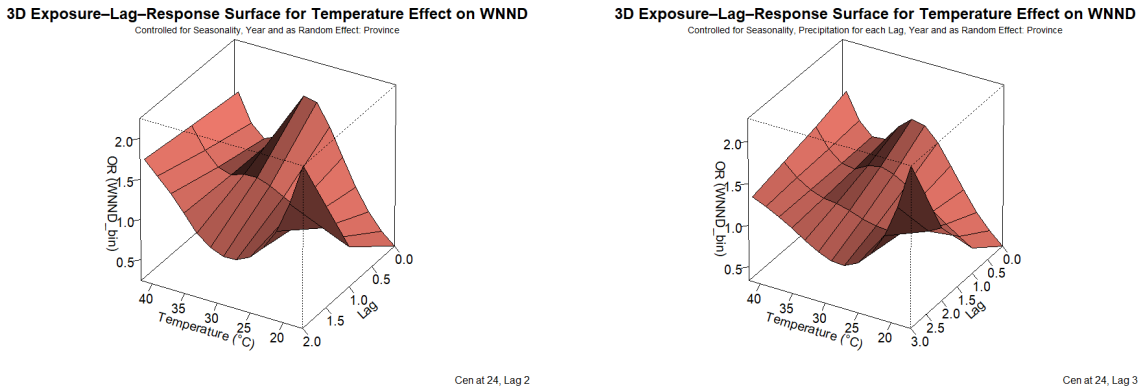


Figure A.11: *3D Exposure-Lag-Response Surfaces for temperature effect,* OR*, on* WNND *for the lag 2 and 3, with a reference temperature of 24 °C.*

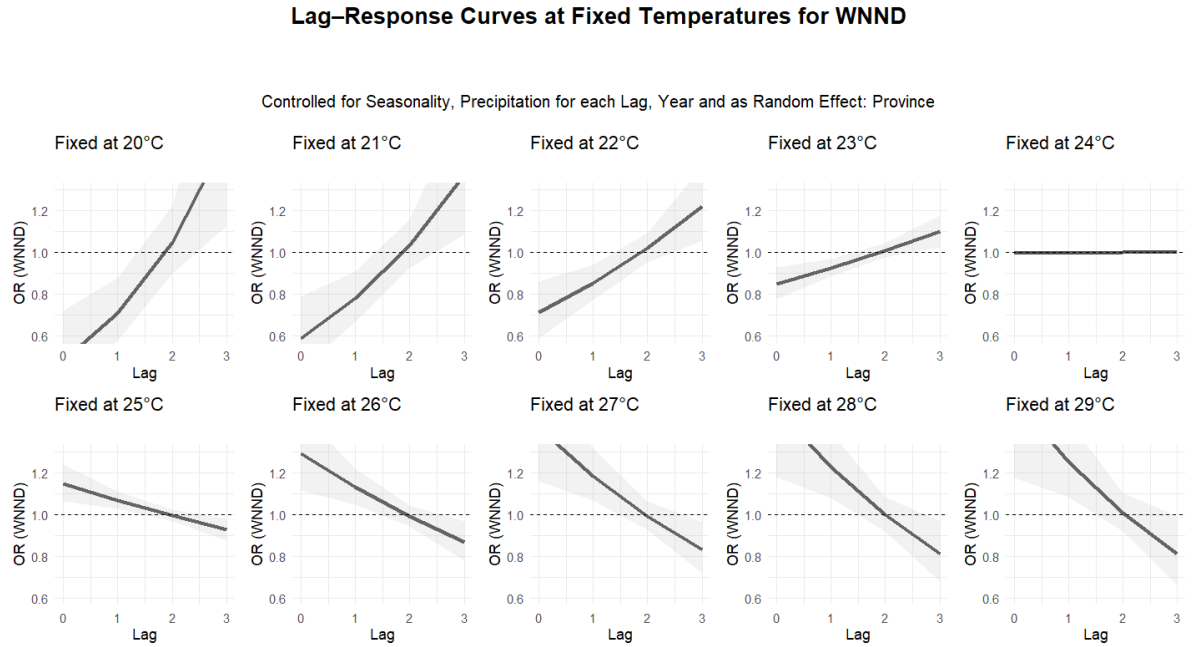


Figure A.12: *Lag-Response Curves at fixed temperatures between 20°C and 29°C showing the temperature effect,* OR*, on* WNND *for the lags until 3, with a reference temperature of 24 °C.*

A.5 Additional Figures for WNND Response incorporating Weather Trend and Mosquito Perspective

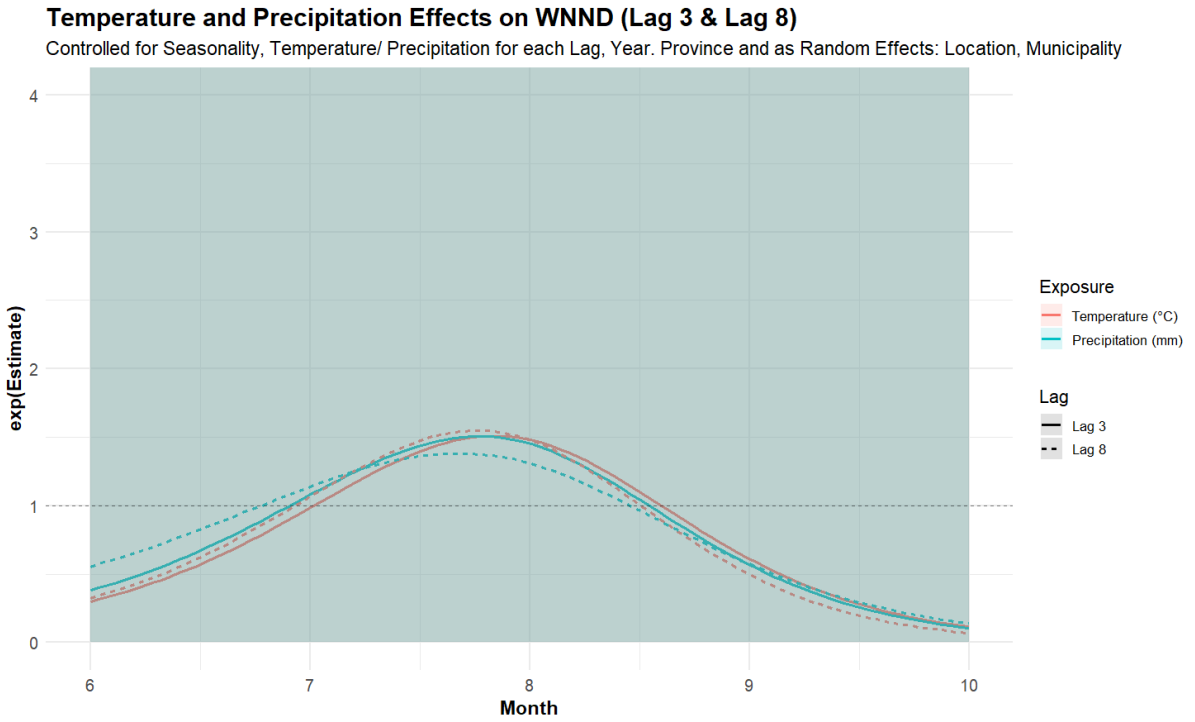


Figure A.13: *The effect of the smoothing term seasonality,* OR*, on the outcome* WNND*, divided for cross-basis predictor (temperature and precipitation) and lag (3 and 8).*

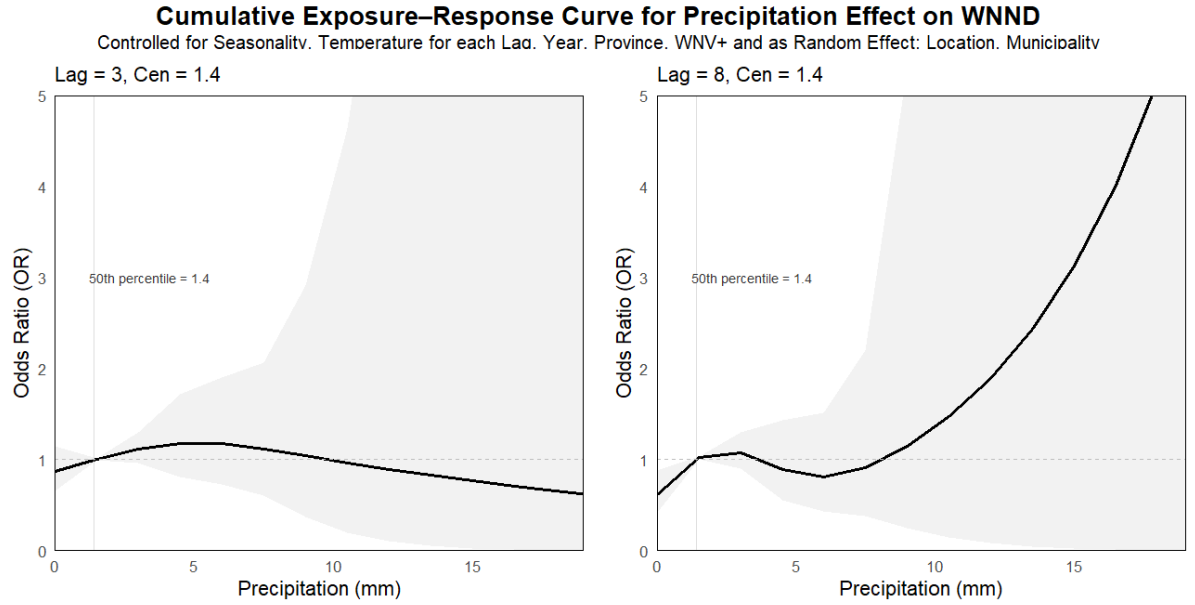


Figure A.14: *Cumulative Exposure-Response Curves for precipitation effect,* OR*, on* WNND *for the lags 3 and 8, with a reference daily precipitation of* 1.4*mm (the 50th percentile).*

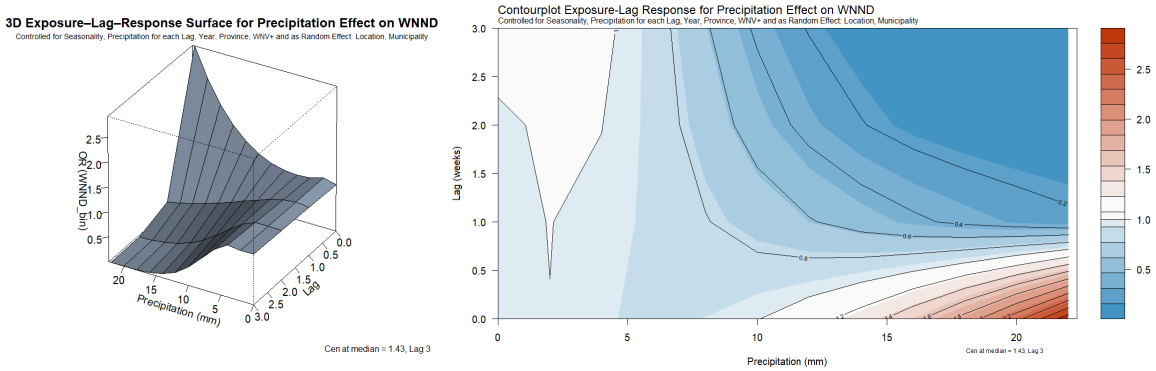


Figure A.15: *3D Exposure-Lag-Response Surface and Contour Plot for precipitation effect,* OR*, on* WNND *for the lag 3, with a reference precipitation of* 1.43*mm (the 50th percentile).*

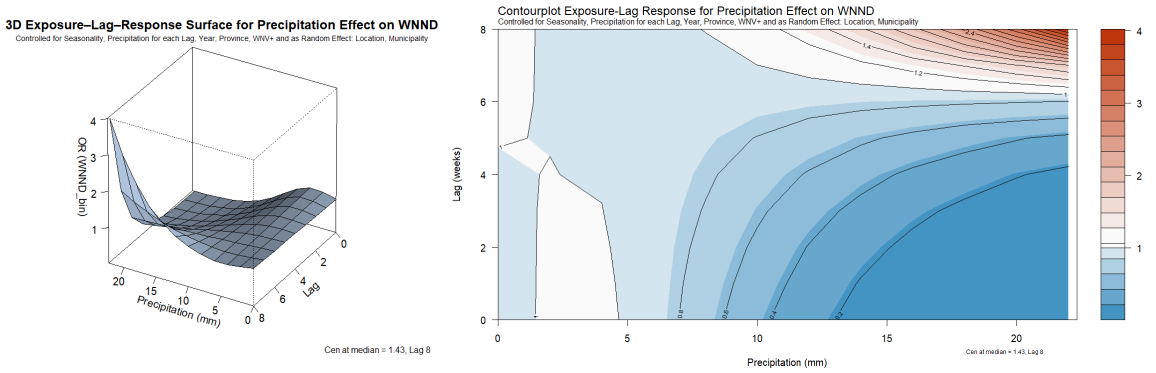


Figure A.16: *3D Exposure-Lag-Response Surface and Contour Plot for precipitation effect,* OR*, on* WNND *for the lag 2, with a reference precipitation of* 1.43*mm (the 50th percentile).*

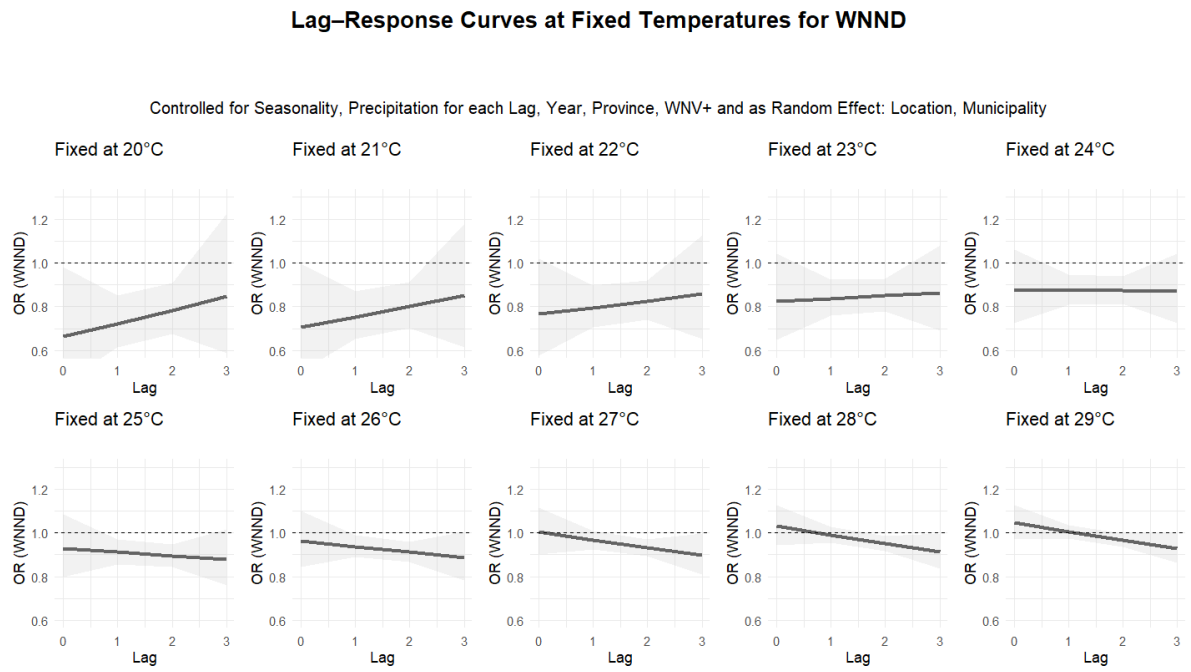


Figure A.17: *Lag-Response Curves at fixed temperatures between 20°C and 29°C showing the temperature effect,* OR*, on* WNND *for the lags until 3, with a reference temperature of* 32.6 *°C (the 50th percentile).*

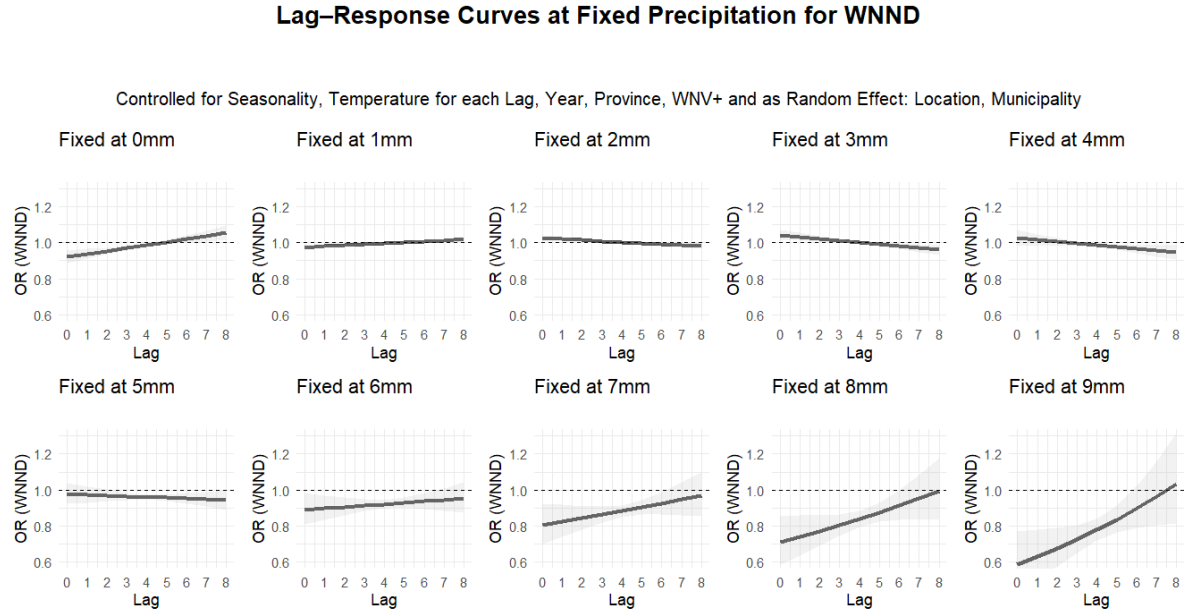


Figure A.18: *Lag-Response Curves at fixed precipitation between 0mm and 9mm showing the precipitation effect,* OR*, on* WNND *for the lags until 8, with a reference precipitation of* 1.43*mm (the 50th percentile).*

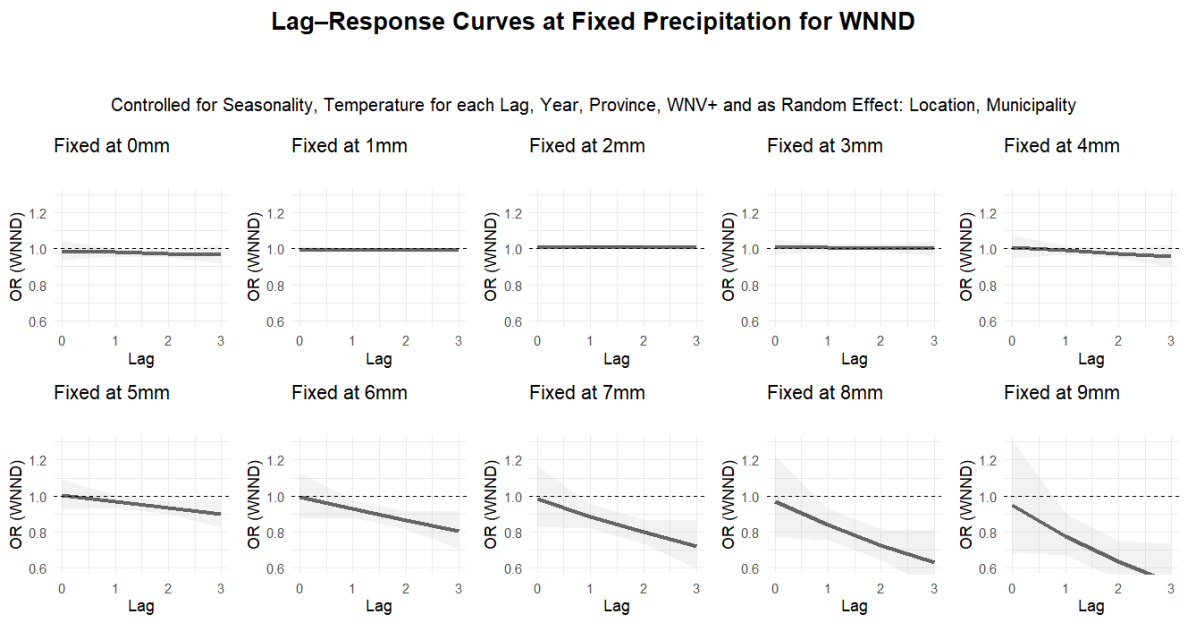


Figure A.19: *Lag-Response Curves at fixed precipitation between 0mm and 9mm showing the precipitation effect,* OR*, on* WNND *for the lags until 3, with a reference precipitation of* 1.43*mm (the 50th percentile).*

Appendix B: Additional Remarks on the Lombardia Region

In Lombardia 82 traps were located, resulting in 201 unique traps in Emilia-Romagna and Lombardia together. On average 41.91 ± 4.2 traps were in place in Lombardia per year.In total there were 40,030 measurements.

Throughout the decade in Lombardia 189 positive test results (3.0%) and 6,093 negative test results were reported. Lombardia had a different measurement than Emilia Romagna, which led to the exclusion of the province. In comparison to Emilia-Romagna, Lombardia stopped collecting once reaching a pre-defined threshold of mosquitoes within a trap.

The year 2018 was exceptional in both regions, exhibiting more than a twofold increase in positive test results compared to the preceding and subsequent years, which recorded comparatively few positives. In Emilia-Romagna 187 measurements were made (68.9 measurements ±24.03 for the other years) and in Lombardia 42 measurements (17.36 ± 11.01) were made within the year 2018.

Provinces Lombardia	**Average Amount of Measurements per Municipality**	**Number of Municipalities in the data (and total)**	**Number of Traps**
Bergamo	37.8	6 (243)	6
Brescia	56.5	12 (205)	12
Como	17.8	3 (147)	3
Cremona	51.8	10 (113)	10
Lecco	17.3	3 (84)	3
Lodi	48.4	5 (60)	5
Mantova	65.0	14 (64)	14
Milano	41.9	11 (133)	11
Monza e Brianza	32.5	1 (55)	1
Pavia	60.6	11 (185)	11
Sondrio	14.0	4 (77)	4
Varese	19.8	2 (136)	2

Table B.1: *For each province within Lombardia, the table reports the average number of measurements per municipality, the number of municipalities included in the dataset (with the total number of municipalities in parentheses), and the number of monitoring traps (Brinkhoff, 2025b).*

Table B.1 shows that each province in Lombardia comprises more municipalities than those in Emilia-Romagna. However, less traps are installed. As a result, only a fraction of all municipalities in the region is covered. Consequently, the validity of results regarding the Lombardia region is therefore lower.

Appendix C: Abbreviations

AIC	Akaike Information Criteria
AUC	Area Under the Curve
BIC	Bayesian Information Criteria
CDC	Centers for Disease Control and Prevention
CHIRPS	Climate Hazards Center Infrared Precipitation with Stations; dataset for precipitation measurements
CHIRTS	Climate Hazards Center InfraRed Temperature with Stations; dataset for temperature measurements
DLM	Distributed Lag Model
DLNM	Distributed Lag Non-Linear Model
ECDC	European Centre for Disease Prevention and Control
GADLNM	Generalized Additive Distributed Lag Non-Linear Model
GADLM	Generalized Additive Distributed Lag Model
GAM	Generalized Additive Model
GEE	Google Earth Engine
GLM	Generalized Linear Model
IRR	Incidence rate ratio
IZSLER	Istituto Zooprofilattico Sperimentale Della Lombardia e Dell'Emilia Romagna
MODIS	Moderate Resolution Imaging Spectroradiometer
OR	Odds Ratio
PCR	Polymer Chain Reaction
PDLM	Polynomial Distributed Lag Model
PET	Potential Evapotranspiration
P-Spline	Penalized Spline; a mathematical function used in curve fitting and smoothing
WNND	West Nile Neuroinvasive Disease
WNV	West Nile Virus

Appendix D: R Project Access

This entire project is based on data handling and calculation with `R` (R Core Team, 2020). The data extraction from GEE was done via a python script (Van Rossum et al., 2024).

The repository contains the code and documentation necessary for running and understanding the project. One can access the GitLab repository for the project at the following link: `https://gitlab.gwdg.de/laura.boehme/master_mosquitos`

The repository is organized into several key directories and files:

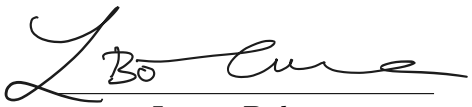
- `/A_Preprocessing/`: Contains three folders - `data`, `describing` and `gee_climate`:
  - The `data` and `describing` folders include, for each individual dataset as well as the merged datasets, scripts for data cleaning (`data_*.R`) and data description (`describing_*.R`).
  - The `gee_climate` folder includes the extracted coordinates and the corresponding Python scripts used for downloading the data from GEE.
  - This directory also includes `R` scripts for function definitions (`helper.R` and `helper_dlnm.R`).
- `/B_DLNM/`: Contains scripts related to the DLNM, organized into two subfolders:
  - `Lag_testing/`: Scripts (`dlnm_Lag_Testing_*.R`), adapted to each outcome, used for the model selection based on lag and exposure.
  - `DLNM_Scripts/`: Full DLNM scripts (`DLNM_*.R`) for each outcome, including predictor selection after the preselected lag structures.
- `/B_DLNM_Outputs/`: Contains the four subdirectories, corresponding to the studied outcomes `WNV`, `nT`, `WNND` and `WNND_Mosquito`, each including the resulting output plots.
- `/C_PlotsMaps/`: Includes a script to visualize the traps in Italy and a script to describe the trap measurements over the decade (`traps_mosquitos_map.R` and `year_mosquitos_plot.R`).
- `/D_extend/`: Includes an auxiliary script, including the definition of quantile-based knots (`quantileknots.R`).
- `/E_images/`: Stores images related to `climate`, `health`, `mosquitoes` and `maps`, organized into correspondingly named subfolders.

The `README.Rmd` file provides a more detailed overview of the project structure, data sources, analysis workflow, is located in the root.

Affidativ

I hereby declare that I have produced this thesis independently without using other aids than those listed. I have clearly identified the sources of any sections from other works that I have quoted or given in essence. Further, the code for the `python` and `R` scripts is written on my own.
Artificial Intelligence tools were used solely to support efficient coding and language refinement, without generating scientific content, analyses, results, or complete text passages. Full responsibility for the content of this thesis remains with the author.
I have complied with the guidelines on good academic practice at the University of Göttingen. I am aware that failure to comply with these principles will result in the examination being graded "nicht bestanden", i.e. failed.

Göttingen, January 31, 2026



Laura Böhme