Universitat d'Alacant
Universidad de Alicante

PREVENTING THE WEST NILE VIRUS, FILARIASIS AND ENCEPHALITIS. METHODS FOR PREDICTING THE ABUNDANCE OF CULEX SP IN A MEDITERRANEAN ENVIRONMENT. PREVENCIÓN DEL VIRUS DEL NILO OCCIDENTAL, LA FILARIASIS Y LA ENCEFALITIS. MÉTODOS PARA PREDECIR LA ABUNDANCIA DE CULEX SP EN UN ENTORNO MEDITERRÁNEO

Petros Damos

Tesis **Doctorales**

UNIVERSIDAD de ALICANTE

Universitat d'Alacant
Universidad de Alicante

Departamento de Enfermería Comunitaria,Medicina Preventiva y Salud Pública
e Historia de la Ciencia /Department of Community Nursing, Preventive Medicine
and Public Health and History of Science

Facultad de Ciencias de la Salud/ Faculty of Health Sciences

PREVENTING THE WEST NILE VIRUS, FILARIASIS AND ENCEPHALITIS. METHODS FOR PREDICTING THE ABUNDANCE OF CULEX SP IN A MEDITERRANEAN ENVIRONMENT

PETROS T. DAMOS
SEPTIEMBRE 2021

Tesis presentada para aspirar al grado de
DOCTOR/DOCTORA POR LA UNIVERSIDAD DE ALICANTE
MENCIÓN DE DOCTOR/DOCTORA INTERNACIONAL
Programa de doctorado en Ciencias de la salud

Thesis submitted for the degree of
DOCTORATE FROM THE UNIVERSITY OF ALICANTE
INTERNATIONAL DOCTORAL DEGREE
Doctoral Program in Health Sciences

Dirigida por/ Directed by: Dr Pablo Caballero Pérez

A mis hijos
Theodoros-Grigorios y Alexandros-Marios

To my children
Theodoros-Grigorios and Alexandros-Marios

Contents

List of Acronyms – abbreviations

ANNs: artificial neural networks.

CDC: Center of disease control.

ChiSq: Chi Square.

CMC: conditional Markov chain.

COI: Cone of Influence.

CWT: Cross Wavelett transform.

FTD: focused time deleyed neural network.

GCD: Gratest common divisor.

LM: Levenberg-Marquardt algorithm.

MC: Markov Chain.

MTD: mosquito per trap per day, 26

NARX: non-linear autoregressive network with exogenous inputs.

PM: paraler mode (network training).

SP: series parallel (network training).

WCSS: within cluster sum of square, 20

WNv: West Nile virus, 10

XWT: Cross wavelet transform.

List of Figures

ABSTRACT

Vector born disease account for about one third of all cases of emerging diseases. *Culex sp.*, in particular, is one of the most important mosquito vector transmitting important diseases such as the West Nile virus, filariasis and related encephalitis. Because there are no vaccines available the most effectual means to prevent infections from the above diseases, is to target mosquitos to prevent bites and disease transmission. However, to be effective such a strategy, it is important to predict the temporal change in mosquito abundance as well as to study how it is affected by weather conditions. This diseratiation is devoted on the development of new methods to predict arthropod vector dynamics and with emphasis on the development of stochastic models and computational methods for predicting *Culex sp.* abundance in Northern Greece.

The current dissertation is divided in three parts. The first part explores the non-trivial associations between *Culex sp.* mosquito abundance and weather variables using traditional and straightforward stochastic and novel machine larning techniques. The information from the first part was a prerequisite for developing a series of stochastic prediction models based on the most detrimental factors affecting mosquito abundance. In the second part, a series of conventional and conditional stochastic Markov chain models are applied for the first time to predict the non-linear dynamics of *Culex sp.* adult abundance. In the third part of the dissertation a soft computing approach is introduced to model the population dynamics of *Culex sp.* and a series of autoregressive artificial neural networks are implied. Finaly, the information of the models is extrapolated and a machine learning algorithm is proposed to be used for predicting arthropod vector population dynamics having practical implications for public health decision making.

Based on the current results there was a high and positive correlation between temperature and mosquito abundance during both observation

years (r = 0.6). However, a very poor correlation was observed between rain and weekly mosquito abundances (r = 0.29), as well as between wind speed (r = 0.29), respectively. Additionally, according to the multiple linear regression model the effect of temperature, was significant. The continuous power spectrum of the mosquito abundance counts and mean temperatures depict in most cases similar power for periods which are close to 1 week, indicating the point of the lowest variance of the time series, although appearing on slightly different moments of time. The cross wavelet coherent analysis showed that inter weekly cycles with a period between 2 and 3 weeks between mosquito abundance and temperature were coherent mostly during the first and the last weeks of the season. Hence, the wavelet analysis shows a progressive oscillation in mosquito occurrences with time, which is higher at the start and the end of the season.

According to the correlation results a climate-conditioned Markov Chain (CMC) model was developed and applied for the first time to predict the dynamics of vectors of important medical diseases. Temporal changes in mosquito population profiles were generated to simulate the probabilities of a high population impact. The probabilities achieved from the trained model are very near to the observed data and the CMC model satisfactorily describes the temporal evolution of the mosquito population process. In general, our numerical results indicate that it is more likely for the population system to move into a state of high population level, when the former is a state of a low population level than the opponent. Field data on frequencies of successive mosquito population levels, which were not used for the data inferred Markov Chain modeling, were assembled to obtain an empirical intensity transition matrix and the observed frequencies. The findings match to a certain degree the empirical results in which the probabilities follow analogous patterns while no significant differences were observed between the transition matrices of the CMC model and the validation data (ChiSq=14.58013, df=24, p=0.9324451).

Furter, a soft system computing modeling approach was followed to simulate and predict *Culex sp*. abundances. Three dynamic artificial neural network (ANNs) models were developed and applied to describe and predict the non-linear incidence and time evolution of a medical important mosquito species *Culex sp*. in Northern Greece. The first is a simple nonlinear autoregressive ANN model that used lagged population values as inputs (NAR), the second is an exogenous non-linear autoregressive recurrent neural network (NARX), which is designed to take as inputs the temperature as exogenous variable and mosquito abundance as endogenous. Finally, the third model is a focused time-delay neural network (FTD), which takes in to account only the temperature variable as input to provide forecasts of the mosquito abundance as target variable. All three models behaved well considering the non-linear nature of the adult mosquito abundance data. However, the NARX model, which takes in to account temperature, showed the best overall modelling performances. Nevertheless, although, the NARX model predicted slight better (R=0.623) compared to the FTD model (R=0.534), the advantage of the FTD over the NARX neural network model is that it can be applied in the case where past values of the population system, here mosquito abundance, are not available for their forecasting. This is very important considering that arthropod vector data are not always available as climatic data.

Concluding, the proposed methods for simulating and predicting mosquito dynamics are recommended as viable for modeling vector disease population dynamics in order to make real-time recommendations utile for dynamic health policies decision making. The proposed stochastic models, as well as the current computational and machine learning techniques of this work provide an accurate abstraction of the arthropod vector population progress observed within the dataset used for their generation. The current study may consider also as a new entry point into the extensive literature of ecological modelling, medical entomology, as well as in simulating arthropod vector diseases epidemics. From a public health

standpoint, the current models have the potential to be integrated into a decision support system allowing health policy makers in their planning to initiate specific management actions against the period of high activity of mosquito adults.

RESUMEN

Introducción y planteamiento del problema: Las enfermedades transmitidas por vectores representan aproximadamente un tercio de todos los casos de enfermedades emergentes. El mosquito *Culex sp.*, en particular, es uno de los vectores más importantes que transmite enfermedades como el virus del Nilo Occidental, la filariasis y encefalitis. Debido a que no hay vacunas disponibles, el medio más eficaz para prevenir infecciones de las enfermedades mencionadas apuntando a los mosquitos para prevenir las picaduras y la transmisión de estás enfermedades. Sin embargo, para que una estrategia de este tipo sea eficaz, es importante predecir el cambio temporal en la abundancia de mosquitos, así como estudiar cómo se ve afectado por las condiciones meteorológicas.

Especialmente, durante el 2020, los mosquitos desempeñaron un papel fundamental en todo el mundo actuando como vectores de enfermedades infecciosas, así como en la reaparición y expansión espectacular de enfermedades erradicadas. Principalmente debido a la proliferación de condiciones ambientales favorables y mediante el desarrollo de lugares de agua adecuados para el crecimiento de las larvas, que en su mayoría son producidos por actividades humanas. Las poblaciones de mosquitos se encuentran en gran abundancia y muy cerca de los huéspedes susceptibles. Hasta hace poco, la mayor incidencia de enfermedades transmitidas por mosquitos se observaba en las zonas tropicales y subtropicales. Sin embargo, durante los últimos años, varias enfermedades transmitidas por vectores fueron desatendidas, y en algunos casos fueron ocasionalmente esporádicas, de forma que aparecen dinámicamente y causan brotes también en climas templados. En Europa, especialmente durante la última década, varias especies se han establecido y, por lo tanto, están cambiando el estado epidemiológico de varias áreas con respecto a la aparición de enfermedades transmitidas por vectores.

El virus del Nilo Occidental (WNV por sus siglas en Inglés) es probablemente la enfermedad más importante de transmisión del mosquito *Culex sp.* que actua como vector, sobre todo en el sur de Europa y los países mediterráneos. El WNV, en particular, es un miembro del género Flavivirus y uno de los arbovirus de mayor distribución en el mundo. A pesar de que la Organización Mundial de la Salud resume algunas pruebas de la circulación del WNV a través de los países mediante la Oficina Regional del Mediterráneo Oriental (EMRO), el conocimiento exhaustivo sobre la dinámica poblacional estacional del vector Culex sp. que es el que lo propa y cómo se ve afectado por las condiciones climáticas mediterráneas sigue siendo en gran parte desconocido.

La filaria, por otro lado, son exclusivamente parásitos de los trópicos y subtrópicos, incluida la parte sur del Mediterráneo y Oriente Medio. Estos parásitos provocan la filariasis que es una enfermedad parasitaria en la que se bloquean los vasos linfáticos del abdomen. El mosquito *Culex spp* es el principal vector de los nematodos de la familia Filaroidea, aunque especies pertenecientes a otros géneros como *Anopheles spp*, *Aedes spp* y *Mansonia spp* también pueden transmitir el nematodo.

Durante los últimos años, varios estados europeos han sufrido brotes esporádicos de la enfermedad del WNV, tanto en humanos como en caballos, aunque la epidemia más importante se produjo en 1996 en el sureste de Rumanía. En el sur de Europa, el WNV se ha observado en las especies de mosquitos autóctonos Culex pipiens, incluidas Italia y Portugal. Además, en Grecia, el Culex pipiens se ha identificado como la especie dominante y endófila en las zonas rurales de Macedonia central, incluidas las prefecturas de Imathia, Kilkis, Pella, Pieria y Thessalonika.

Hasta ahora, el medio más eficaz para prevenir la infección por el virus del Nilo Occidental, así como otras enfermedades transmitidas por mosquitos, es apuntar a los mosquitos para prevenir las picaduras y la transmisión de enfermedades. Por tanto, el control de vectores, que se basa en el uso de insecticidas, que es el principal medio de mitigar la propagación de

enfermedades relacionadas. Sin embargo, para que una estrategia de este tipo sea eficaz, es importante predecir el cambio temporal en la presencia de mosquitos, su abundancia y cómo se ve afectada por las condiciones climáticas.

En este contexto, se utilizaron varios modelos matemáticos para conectar los procesos biológicos de la dinámica vectorial y el clima. Hasta la fecha, la mayoría de los modelos epidemiológicos y de población de vectores tienen un carácter determinista y se basan en algunos supuestos básicos para definir los diversos parámetros de la dinámica de vectores y enfermedades en estudio. Sin embargo, la mayoría de las veces estos parámetros se desconocen y se considera necesario estimarlos primero para parametrizar el modelo que se utilizará posteriormente para las proyecciones. Además, debido a los impactos de diversos factores internos y externos, la evolución temporal de los procesos de la población no es lineal y se caracteriza por perturbaciones aleatorias y, en consecuencia, generalmente es difícil analizar y pronosticar la dinámica de la población utilizando solo modelos mecanicistas.

Objetivos y estructura de la tesis: El objetivo principal de la Tesis que presento es prevenir las enfermedades transmitidas por mosquitos, como el WNV, la filariasis y la encefalitis, mediante el control de las poblaciones de los mosquitos que actúan como vector. La manera de hacerlo es mediante la formulación, desarrollo y aplicación de nuevos enfoques de modelado estocástico y métodos de predicción computacional relacionados que permiten predecir la abundacia de las poblaciones de estos vectores de transmisión.

Se hace énfasis en el análisis multivariado para la predicción o modelado de la dinámica poblacional del Culex spp. y variables ecológicas relacionadas y cómo pueden utilizarse como herramientas de decisión para el control de artrópodos vectoriales y la planificación de la salud pública. Con una

predicción de la abundancia de mosquitos, en particular, las autoridades de salud pública podrían predecir la evolución temporal de la abundancia de mosquitos, que es un requisito previo para el manejo exitoso y la prevención de las enfermedades transmitidas por vectores relacionadas.

La tesis actual se divide en tres partes. La primera parte explora las asociaciones no triviales entre *Culex sp.* abundancia de mosquitos y variables climáticas utilizando tradicional y técnicas novedosas y sencillas. La información de la primera parte fue un requisito previo para desarrollar una serie de modelos de predicción estocásticos basados en los factores más perjudiciales que afectan la abundancia de mosquitos. En la segunda parte, se aplica por primera vez una serie de modelos estocásticos de cadena de Markov convencionales y condicionales para predecir la dinámica no lineal de Culex sp. abundancia de adultos. En la tercera parte de la disertación se introduce un enfoque de computación blanda para modelar la dinámica poblacional de *Culex sp.* Están implícitas una serie de redes neuronales artificiales autorregresivas que extrapola la información de los modelos y se proponen métodos computacionales y un algoritmo de aprendizaje automático relacionado para predecir la dinámica de los vectores de artrópodos con implicaciones prácticas para la toma de decisiones de salud pública.

Los Datos: Los datos gratuitos sobre trampas para mosquitos disponibles en el portal abierto de datos de la Unión Europea (EU ODP) (http://data.europa.eu) (Figura 4), que proporciona acceso a datos de las instituciones de la Unión Europea (UE) y otros organismos de la UE que pueden reutilizarse con fines comerciales o no comerciales (Decisión de la Comisión Europea 2011/833 / UE). En particular, usamos datos de trampas para mosquitos adultos de Culex sp. muestreados en 11 lugares del centro de Macedonia-Grecia. Los datos se manejaron como vectores que consistían en intervalos de tiempo cercanos a semanales del número de mosquitos adultos capturados en trampas de CO_2 desde mediados de mayo hasta

septiembre. Para el análisis se utilizaron datos durante tres años de observación sucesivos (2011, 212 y 2013). Los datos climáticos y las temperaturas medias del aire en grados Celsius y los eventos de lluvia en mm fueron obtenidos por el observatorio nacional de Atenas a través de una estación meteorológicahttp://stratus.meteo.noa.gr/front).

Métodos y Resultados: Para establecer si existen posibles asociaciones entre las poblaciones de abundancia de mosquitos y las variables climáticas, se realizó primero un análisis de correlación por pares. Además, se realizó un análisis de regresión lineal múltiple para ajustar la variable dependiente, a las variables independientes más significativas, según el análisis de correlación. Además, para detectar correlaciones no lineales entre la abundancia de *Culex sp* y la temperatura, que fue la variable más influyente, se realizó un análisis de ondículas.

En este análisis, cada ondícula de escala se desplaza en el tiempo a lo largo de toda la serie y se compara con la serie original (abundancia de mosquitos y datos climáticos). El proceso se repite para todas las escalas dando como resultado coeficientes que son una función de la escala de la ondícula y el parámetro de desplazamiento. En este contexto, el análisis de ondículas se utiliza para cuantificar similitudes (es decir, sincronización en el dominio del tiempo con respecto a diferentes escalas) en series de tiempo que se caracterizan por fluctuaciones transitorias.

El análisis temporal reveló fluctuaciones similares para todos los lugares de muestreo y durante ambos años las abundancias estandarizadas de mosquitos indicaron picos de temporada temprana en mayo y junio, seguidos de una disminución en el último trimestre de junio y un resurgimiento durante julio y agosto, especialmente para el segundo año de observación. En particular, la figura 6 muestra los patrones de vuelo estacionales y los errores estándar de la abundancia de Culex adulto. sp. capturados durante 2011 y 2012 en la prefectura de Imathia en el norte de Grecia, así como las autocorrelaciones asociadas para diferentes rezagos

(semanas). Los recuentos representan observaciones semanales de mosquitos atrapados desde mayo hasta mediados de septiembre. La población de mosquitos comenzó a aumentar a fines de mayo y alcanzó su punto máximo a mediados de julio durante 2011. En 2012 se capturó un número considerablemente mayor de mosquitos en comparación con 2011. Además, durante 2012 se observaron dos picos, el primero a fines de junio y el segundo a mediados de agosto. Durante ambos años, las autocorrelaciones no indican fluctuaciones bruscas y / o patrón periódico claro en la serie temporal del mosquito con respecto a los diferentes rezagos. La autocorrelación luego decae a cero, indicando posible la existencia de un proceso de promedio móvil, aunque existe cierta ambigüedad en cuanto a los diferentes patrones observados entre los dos años de observación.

Además y según los resultados actuales, hubo una correlación alta y positiva entre la temperatura y la abundancia de mosquitos durante ambos años de observación ($r = 0,6$). Sin embargo, se observó una correlación suave entre la lluvia y la abundancia semanal de mosquitos ($r = 0,29$), así como entre la velocidad del viento ($r = 0,29$), respectivamente. Además, según el modelo de regresión lineal múltiple, el efecto de la temperatura fue significativo.

Los recuentos de abundancia de mosquitos y las temperaturas medias representan en la mayoría de los casos valores similares para períodos cercanos a 1 semana, lo que indica el punto de menor varianza de la serie de tiempo, aunque aparece en momentos de tiempo ligeramente diferentes. En general, el análisis concordante de ondículas cruzadas mostró que los ciclos inter-semanales con un período entre 2 y 3 semanas entre la abundancia de mosquitos y la temperatura fueron concordantes principalmente durante la primera y última semana de la temporada. En particular, el análisis de ondículas cruzadas para 2011 muestra que los ciclos inter semanales con un período entre 2-3 semanas entre la abundancia de mosquitos (log MTD) y la temperatura (media) fueron concordantes. Esta concordancia es notable principalmente durante la primera y última semana de la temporada (es

decir, mayo , Junio y finales de agosto) pero no durante la mitad de la temporada de verano (julio y principios de agosto). Durante 2012 el análisis de ondas cruzadas mostró patrones análogos, aunque durante el inicio de la temporada (finales de mayo a principios de junio) se observaron dos ciclos inter-semanales con períodos de 1-2 semanas y 6-7.

Por lo tanto, aparece una mayor sincronización entre el mosquito y la temperatura durante el inicio de la temporada (es decir, mayo-junio) y el final de la temporada (es decir, septiembre). Esta sincronización o concordancia es evidente para dos períodos que representan oscilaciones similares entre los eventos máximos de mosquitos y la temperatura por unidad de tiempo. Adicionalmente, valores más altos señalan una sincronización muy alta entre los ciclos de abundancia de mosquitos y las temperaturas medias, en contraste con valores cercanos a 0 que representan una independencia de los ciclos en una escala de tiempo determinada.

Además, hay dos puntos de influencia evidentes; durante el comienzo de la temporada de 2012, en los períodos cercanos al 1 y 7 respectivamente, que representan la región del espectro de ondículas en la que los efectos de borde se vuelven importantes. En estos puntos se definen como el tiempo de plegado electrónico para la autocorrelación de la potencia de la ondícula en cada escala, es decir, para ajustar.

Por tanto, el análisis de ondículas muestra una oscilación progresiva en la aparición de mosquitos con el tiempo, que es mayor al inicio y al final de la temporada. A diferencia de los métodos estándar de análisis, las ondículas pueden proporcionar información útil sobre la estructura de oscilación resuelta en el tiempo de los datos de los mosquitos y la revelación de una asociación no estacionaria con la temperatura.

Otro de los métodos de predición estocásticos que se desarrollaron en esta Tésis, son dos modelos estocásticos de cadenas de Markov para predecir la abundancia del mosquito *Culex sp*. En este contexto se consideran, otra vez,

las capturas de mosquitos así como las variables climáticas relacionadas (temperatura y lluvia). Como en todo proceso estocástico que evoluciona en el tiempo t y se define en un espacio de probabilidad (Ω, F, P). Donde Ω es un espacio muestral, F es un conjunto de resultados en el espacio muestral y P asigna a cada evento de F una probabilidad, P corresponde la la matriz de transición. Si el número de F no es contable, el proceso se indica mediante $\{X(t): t \geq 0\}$ o $\{Xt\}$ t$\geq$0. En el primer caso el proceso se denomina cadena en tiempo discreto y en el segundo, en tiempo continuo. Aquí se consideró el primer caso, ya que los datos se han observado en puntos de tiempo específicos y no de forma continua.

Con base en el análisis de correlación cruzada, se determina que es preferible utilizar la temperatura para condicionar la abundancia de artrópodos vectores, ya que tiene una correlación mucho más fuerte que el impacto de la lluvia como ya hemos visto. Existe una correlación positiva para la temperatura y el retardo de tiempo cero. Por lo tanto, es mejor condicionar la temperatura en el momento mismo y no la temperatura una semana antes o una semana después (es decir, los mosquitos no están rezagados en la temperatura), Además hay una correlación positiva para la temperatura y una relación negativa para la lluvia, ya que probablemente afecta la actividad de vuelo.

Después de utilizar el algoritmo de agrupación de k-medias para determinar en número de estadíos, se han utilizado los k centros de agrupación para posicionar los datos y categorizarlos, considerando primero una situación con dos estados (niveles altos y bajos de población) y luego cinco estados (muy bajo, bajo, intermedio, alto y muy bajo). alto nivel de población).

Hasta la fecha, para una duración de secuencia de menos de 35 semanas, como es el caso que estudiamos, los parámetros del modelo no difieren considerablemente de acuerdo con su contenido de entropía informativa y el sistema se explícita un comportamiento aleatorio en lugar de determinista, lo que sugiere que en un alto grado existe una matriz de transición subyacente que pretedemos determinar.

Para el sistema de cinco estados y la red que se ha construido, las probabilidades de transición muestran que si durante una semana la abundancia de mosquitos es alta (estado 4), existe un 10% de probabilidad de que permanezca en el mismo nivel la próxima semana (estado 4), 30% de probabilidad de que esté en un nivel muy alto (estado 5), 30% de probabilidad de que se mueva a un nivel muy bajo (estado 1) y sin posibilidad de pasar a ninguna otra población niveles (estado 2 y 3). Sin embargo, debe tenerse en cuenta que las probabilidades cero de pasar al estado 2 o al estado 3 podrían ser una indicación de que existe cierta incertidumbre en el proceso de estimación relacionado con el conjunto de datos particular utilizado, ya que si hubiera más datos disponibles, estas probabilidades de transición serían observado con una probabilidad baja en lugar de cero.

Utilizando los datos de las matrices de transición, se ejecutaron simulaciones de Matlab basadas en los modelos de cadena de Markov de dos y cinco estados. En cada momento discreto del proceso de Markov o, utilizando términos ecológicos, después de cada semana de observación, se podría tomar una decisión de control de mosquitos en función de los niveles de población previstos. La base de la decisión será la predicción basada en cómo evoluciona la cadena de Markov en función de los valores de la matriz de transición P. De acuerdo con los resultados de la correlación, se desarrollaron dos modelos, una cadena de MarkoV simple (MC) y una cadena de Markov climatizada (CMC) para predecir aún más la dinámica de Culex sp. adultos, es decir, una cadena de Markov condicionada. Se generaron cambios temporales en los perfiles de población de mosquitos para simular las probabilidades de un alto impacto poblacional. Las probabilidades logradas con el modelo entrenado están muy cerca de los datos observados y el modelo CMC describe satisfactoriamente la evolución temporal del proceso poblacional de mosquitos.

En general, nuestros resultados numéricos indican que es más probable que el sistema de población pase a un estado de alto nivel de población, cuando el primero es un estado de bajo nivel de población.

Datos de campo sobre frecuencias de niveles sucesivos de población de mosquitos, que no se utilizaron para el modelado de MC inferido de datos, se ensamblaron para obtener una matriz de transición de intensidad empírica y las frecuencias observadas. Los hallazgos coinciden hasta cierto punto con los resultados empíricos en los que las probabilidades siguen patrones análogos mientras que no se observaron diferencias significativas entre las matrices de transición del modelo CMC y los datos de validación (ChiSq = 14.58013, gl = 24, p = 0.9324451).

En la tercera parte de la disertación se propuso utilizar tres tipos de redes neuronales artificiales (ANN por sus siglas en inglés) como herramientas matemáticas para simular el funcionamiento complejo de la serie de tiempo ecológica (es decir, la abundancia de mosquitos en relación con la temperatura). Los enfoques están inspirados en el funcionamiento del cerebro humano. Las ANN tienen similitudes con las neuronas biológicas y consisten en un conjunto de neuronas artificiales que interactúan a través de sinapsis. Los modelos de red finales que se presentan tienen la capacidad de procesamiento paralelo de datos y aprendizaje continuo a través de la interacción con el entorno. El grado de interacción entre las sinapsis está determinado por pesos (pesos sinápticos). La red neuronal interactúa con su entorno (es decir, las variables de interés) y los pesos sinápticos cambian constantemente y, por lo tanto, fortalecen o debilitan el poder de cada nodo de interacción. Por lo tanto, la información de las variables externas (es decir, el entorno) se codifica en los pesos sinápticos de la red y le da la capacidad a la ANN para simular el proceso relacionado con esas variables. Para la red, se utilizó un algoritmo de entrenamiento computacional que tiene como objetivo optimizar mediante iteraciones el rendimiento del modelo.

Los tres modelos de redes ANN se utilizaron además para describir y predecir la incidencia no lineal y la evolución temporal de una importante especie de mosquito *Culex sp.* en el norte de Grecia. El primero es un modelo ANN autorregresivo no lineal simple que utilizó valores de población rezagados como entradas, el segundo es una red neuronal recurrente autorregresiva no lineal exógena (NARX), que está diseñada para tomar como entradas la temperatura como variable exógena y la abundancia de mosquitos como endógena. Finalmente, el tercer modelo es una red neuronal de retardo de tiempo (FTD) enfocada, que toma en cuenta solo la variable de temperatura como entrada para proporcionar pronósticos de la abundancia de mosquitos como variable objetivo. En los modelos de ANN aplicados, las predicciones de la dinámica de los mosquitos se realizaron a partir de los valores previos pasados de la serie temporal de abundancia y de los valores presentes y pasados de la entrada de temperatura exógena.

Hasta la fecha, para extraer estas dos variables de entrada clave, analizamos previamente los coeficientes de correlación de diferentes datos meteorológicos con un desfase de tiempo impuesto. Además, se probaron diferentes combinaciones de neuronas ocultas y retrasos para obtener los mejores modelos de ANN. Las estructuras del modelo final han utilizado 10 neuronas ocultas y 2 retrasos (tiempo en semanas), porque están dando resultados satisfactorios después de un entrenamiento preliminar y pruebas de diferentes combinaciones de neuronas ocultas y retrasos. La división de datos se realizó al azar utilizando ambos conjuntos de datos (2011 y 2012) en los que finalmente el 60% de los datos se utilizaron para el entrenamiento NARX (38 pasos de la serie temporal objetivo)

En la mayoría de los casos, los errores de entrenamiento y validación para los modelos ANN disminuyen hasta el periodo resaltado y el mejor estado de rendimiento de validación para el modelo NARX está en 0.388 en la que se minimiza el error cuadrático medio (mse).

Además, el mse de los datos de prueba se minimiza después de 4 iteraciones y permanece estacionario después de ese punto, lo que indica que el modelo

había alcanzado su estado óptimo. Sin embargo, los mejores estados para los datos ocurren después de 3 interacciones, en los que el mse de los datos se minimiza.

En la mayoría de los casos, los tres modelos se comportaron bien considerando la naturaleza no lineal de los datos de abundancia de mosquitos adultos. Sin embargo, el modelo NARX, que tiene en cuenta la temperatura, mostró los mejores rendimientos generales de modelado. El coeficiente de correlación estuvo en niveles aceptables en ambos casos y con respecto al conjunto de datos disponibles ($R = 0,623$ y $R = 0,534$ para los modelos NARX y FTD, respectivamente). Además, teniendo en cuenta la naturaleza abrupta y no lineal de los datos de los mosquitos, las predicciones generales del modelo se encuentran en niveles aceptables cuando se comparan con los datos de abundancia reales. Sin embargo, aunque, el modelo NARX predijo algo mejor($R = 0,623$) en comparación con el modelo FTD ($R = 0,534$),la ventaja del FTD sobre el modelo de red neuronal NARX es que se puede aplicar en el caso de que los valores pasados del sistema de población, en este caso la abundancia de mosquitos, no estén disponibles para su pronóstico. Esto es muy importante teniendo en cuenta que los datos de vectores de artrópodos no siempre están disponibles como datos climáticos.

Además, debe mencionarse que el rendimiento del modelo es considerablemente mayor al tomar en cuenta solo los datos de entrenamiento (es decir, $r = 0,8$ y $r = 0,62$, para los modelos NARX y FDR, respectivamente) que el rendimiento general final de los modelos. Por tanto, los valores del rendimiento se ven afectados, a la baja, durante la validación y prueba del modelo. Por lo tanto, se puede esperar que el rendimiento del modelo se mejore considerablemente si el tamaño del conjunto de datos de prueba fuera mayor. Sin emabargo, para hacer que el modelo de red sea más eficiente, se ha decidido mantener un conjunto de datos más grande para ser preprocesado para el entrenamiento a pesar de

los retornos más pequeños mostrados para las pruebas y los resultados de validación.

Considerando las actuaciones finales de los modelos de ANN, en la mayoría de los casos, las desviaciones entre pronóstico y valores reales durante ciertos pasos de tiempo están en el rango de -1.4, 1.3 bins que es relativamente bajo y la distribución es alrededor de cero. Además, la frecuencia global del término de error tiene 20 bins. El error de la red neuronal NARX varía de -1.2 (bin más a la izquierda) a 1.03 (bin más a la derecha) mientras que el error de la red neuronal FTD varía de -1.1 (bin más a la izquierda) a 0.9 (bin más a la derecha). Para ambos modelos y especialmente para el modelo NARX, la gran mayoría de las salidas de entrenamiento tienen un error menor y leve entre -0,4 y 0,4. Esto probablemente se deba al hecho de que el conjunto utilizado para el entrenamiento contenía más datos (es decir, el 60% de los datos) que los conjuntos de datos de validación y prueba. Además, las funciones de autocorrelación del error 1 para los modelos ANN se encuentran en límites aceptables. En particular, con un retraso cero, la autocorrelación es igual a la mse, mientras que para las siguientes autocorrelaciones rezagadas, el coeficiente de correlación no excede los intervalos de confianza superior e inferior, salvo en algunos casos. Esto significa que la mayoría de los valores auto-correlacionados rezagados, para ambos modelos, son pequeños y en niveles aceptables considerando que los valores rezagados de cero a 15 (semanas) están entre los intervalos de confianza superior e inferior. En general, la predicción tuvo un buen desempeño en ambos casos, aunque hubo partes en las que los resultados de la producción se desempeñaron peor y especialmente durante el final de la temporada.En gran medida, esto debe abordarse al conjunto de datos particular que estaba disponible y al hecho de que se utilizó un conjunto de datos limitado para la capacitación.

Sin embargo, considerando que la dinámica de la población de mosquitos parecía bastante abrupta, caracterizada por alteraciones no lineales, dado el conjunto de datos limitado, las predicciones generales del modelo se

encuentran en niveles aceptables para ambos modelos. Además, la inclusión de la temperatura como factor exógeno mejoró considerablemente el rendimiento del modelo NARX y los datos predichos siguen en alto grado las observaciones. Tenga en cuenta que tanto los modelos como los datos representan datos reales de la población de mosquitos. Se presenta un algoritmo computacional de ANN final.

El algoritmo describe los pasos, las elecciones iniciales y las rutinas relacionadas (es decir, bucles-decisiones) que se han utilizado para terminar con el modelo ANN de retroalimentación final con línea de retardo en la entrada (es decir, un paso de tiempo - semana). De acuerdo con el algoritmo aplicado, la primera preparación de datos y las pruebas preliminares se realizan para decidir cuál es el mejor conjunto de datos utilizado para el entrenamiento y la validación del modelo. Los conjuntos de datos de validación consisten en la muestra de datos retenidos del entrenamiento, mientras que el conjunto de datos de prueba se usa para ajustar (optimizar) los hiperparámetros del modelo ANN (es decir, tomar pesos de la ANN entrenada y usarla como inicialización para un nuevo modelo siendo entrenado y así sucesivamente). Inicialmente, el proceso comienza seleccionando una pequeña cantidad de neuronas y (es decir, 5-10) con respecto a algunos pesos aleatorios iniciales (p. Ej., aprendizaje supervisado) para las sinapsis y cada vez que se entrena la red da como resultado una solución diferente debido a los diferentes valores iniciales de peso y sesgo, así como a las propiedades de la red (por ejemplo, número de neuronas). El modelo se vuelve a entrenar varias veces para garantizar que tenga una buena precisión hacia una solución óptima basada en una medición de error. El error, como se muestra en la sección de material, se define como la diferencia de la salida de la ANN y la serie de datos deseados externos preespecificados. Finalmente, se estima el error para diferentes estructuras ANN relacionadas con el número de capas ocultas para derivar el modelo final que funciona mejor. El modelo final optimizado se puede alimentar con nuevos datos número de neuronas). El modelo se vuelve a

entrenar varias veces para garantizar que tenga una buena precisión hacia una solución óptima basada en una medición de error. El error, como se muestra en la sección de material, se define como la diferencia de la salida de la ANN y la serie de datos deseados externos preespecificados. Finalmente, se estima el error para diferentes estructuras ANN relacionadas con el número de capas ocultas para derivar el modelo final que funciona mejor. El modelo final optimizado se puede alimentar con nuevos datos número de neuronas). El modelo se vuelve a entrenar varias veces para garantizar que tenga una buena precisión hacia una solución óptima basada en una medición de error. Finalmente, se estima el error para diferentes estructuras ANN relacionadas con el número de capas ocultas para derivar el modelo final que funciona mejor. El modelo final optimizado se puede alimentar con nuevos datos el error se estima para diferentes estructuras ANN relacionadas con el número de capas ocultas para derivar el modelo final que funciona mejor.

Discusión y conclusión. En esta disertación se introducen una serie de métodos novedosos y enfoques de modelado relacionados para predecir la dinámica de la población de mosquitos vectores en un clima mediterráneo. Se utilizaron datos de clima abierto y abundancia de *Culex sp.* para aplicar los métodos de predicción propuestos y demostrar su utilidad en condiciones de campo realistas. Además, considerando que Culex sp. es el principal vector de transmisión del WNV, la filariasis y otras especies de *Culex sp.* encefalitis transmitida, los métodos predictivos de los estudios actuales pueden contribuir a comprender el funcionamiento de la dinámica de los vectores de artrópodos, así como a predecir los períodos de alta actividad de los mosquitos. Esta información es importante para iniciar acciones de control de mosquitos y romper los ciclos de transmisión de enfermedades vectoriales. Desde el punto de vista de la salud pública, el presente estudio contribuye al desarrollo de herramientas de decisión para

ser utilizadas en acciones de control de vectores, así como para iniciar estrategias de manejo de salud pública y prevenir los virus del WNV, la filariasis y la encefalitis. Aunque se probaron varias variables climáticas (es decir, participación, velocidad del viento, dirección del viento) afectan la abundancia de mosquitos, la matriz de correlación de Pearson no indicó correlaciones significativas entre estas variables, pero sí con la temperatura. Además, la regresión lineal múltiple sugirió una asociación positiva entre la abundancia de mosquitos y las temperaturas medias. Esto concuerda con otros estudios que han demostrado que entre los factores climáticos candidatos, la temperatura ejerce, con mucho, la influencia más significativa, probablemente porque el aumento de las temperaturas ambientales aumenta las tasas metabólicas, el rendimiento reproductivo, y el comportamiento de búsqueda de hospedadores de estos vectores y, por lo tanto, afectan la abundancia de mosquitos. Considerando los resultados del análisis de ondículas se reveló que la oscilación de la abundancia de mosquitos está dominada por diversos modos semanales y esto se puede observar en el respectivo espectro de ondículas para ambos años de observación. Sin embargo, los espectros de potencia de ondículas revelaron que las altas frecuencias, tanto para la abundancia de mosquitos como para la temperatura, están en la misma banda de frecuencias altas (≈ 4 semanas) un hecho que no se puede ver al observar la serie de tiempo de los mosquitos per se. Además, el análisis de ondículas cruzadas aplicado a los mismos conjuntos de datos muestra nada más que un modo común de oscilación significativo, principalmente durante el inicio y el final del período de vuelo del mosquito. Por lo tanto, en contraste con las correlaciones simples, el análisis de ondas cruzadas tiene la ventaja de revelar la relación de tiempo probable entre las series de dos tiempos. Por lo tanto, el análisis de ondículas es particularmente adecuado para el análisis de las relaciones entre dos series de tiempo, y especialmente apropiado para seguir el cambio gradual cuando se ve forzado por variables exógenas. Además, está demostrado que el análisis de ondículas funciona mejor que la regresión

lineal, lo que concuerda con otros estudios que muestran que las técnicas clásicas de análisis estadístico pueden ser inadecuadas para modelar series de tiempo no estacionarias. Con base en el análisis de correlación mencionado de la afrenta entre las poblaciones de mosquitos y las variables climáticas, es evidente que la temperatura ejerce un mayor impacto en la fenología del vector de artrópodos adultos en comparación con los eventos de lluvia, a pesar de que esos mosquitos, particularmente prosperan en condiciones de humedad ya que la lluvia afecta indirectamente a la población de mosquitos mediante el aumento de los criaderos.

Por lo tanto, se consideró necesario incluir las variables meteorológicas más influyentes (por ejemplo, solo la temperatura) para mejorar el rendimiento de un modelo MC simple mediante el uso de un modelo CMC. En realidad, se ha encontrado que el modelo de cadena de Markov de la dinámica de la población de vectores de artrópodos que está condicionado por las temperaturas, funciona considerablemente mejor que el modelado estocástico de MC simple de la dinámica de la población de vectores.

Considerando, los diferentes modelos de ANN autorregresivos que se han aplicado para describir la dinámica poblacional adulta de Culex sp. en la mayoría de los casos se han realizado en niveles aceptables considerando la dinámica no lineal de la abundancia de mosquitos adultos. Sin embargo, el modelo que tiene en cuenta la temperatura como variable de entrada exógena, como en el caso del modelo de CM condicionado por la temperatura, se comportó mejor que los demás. En comparación con otros métodos de predicción paramétrica, la principal ventaja de las redes neuronales es, en primer lugar, que almacena el conocimiento y la experiencia del entorno utilizado para su entrenamiento (en este caso, la abundancia de mosquitos y la temperatura), que luego puede recuperar para simular el proceso. En segundo lugar, tiene la capacidad de generalizar, es decir,

En conclusión, los métodos propuestos para simular y predecir la dinámica de los mosquitos se recomiendan como viables para modelar la dinámica de

la población de enfermedades vectoriales con el fin de hacer recomendaciones en tiempo real útiles para la toma de decisiones de políticas de salud dinámicas. Los modelos estocásticos propuestos, así como las técnicas actuales de aprendizaje automático y computacional, de este trabajo proporcionan una abstracción precisa del progreso de la población de vectores de artrópodos observado dentro del conjunto de datos utilizado para su generación. Sin embargo, el estudio actual puede considerarse también como un nuevo punto de entrada a la extensa literatura de modelización ecológica, entomología médica, así como en la simulación de epidemias de enfermedades por vectores artrópodos. Desde el punto de vista de la salud pública, por lo tanto, los resultados del modelo actual podrían sentar las bases para pronosticar los niveles de población de vectores de artrópodos y alertar a las personas pertenecientes a grupos vulnerables y para implementar medidas efectivas de control de vectores para proteger la salud pública durante incidentes de alta presión poblacional. Por último, y desde un punto de vista práctico de la salud pública, vale la pena seguir desarrollando un sistema de apoyo a la toma de decisiones basado en la disponibilidad de los datos que pueda incorporar los modelos epidimiológicos actuales, así como los relacionados, para proporcionar pronósticos de la población de artrópodos en tiempo real. Los sistemas de soporte a la decisión pueden ayudar a automatizar el proceso de toma de decisiones y evaluar las diferentes acciones de gestión sanitaria.

Palabras clave: *Culex sp.*, Apoyo a la toma de decisiones, epidemiología, redes neuronales de cadenas de Markov, salud pública, análisis de ondículas.

Chapter 1: Introduction

1.1   Problem statement

Vector born disease account for about one third of all cases of emerging diseases. During the last year's mosquito vectors, particularly have a worldwide critical role in the outbreak of vector borne diseases as well as in the re-emergence and dramatic expansion of eradicated diseases (Gubler 2010). Mainly because of favorable environmental conditions and through the development of suitable water sites for larval growth, which are mostly produced by human activities, mosquito populations occur at high abundance and in near proximity to susceptible hosts.

Until recently, the highest incidence of mosquito transmitted diseases was observed in the tropical and subtropical areas. Yet, during the past years, several neglected vector-borne diseases, which in a few events were occasionally sporadic, get to a dynamic reappearance and cause outbreaks in temperate climates as well. In Europe, especially during the last decade, several species that have incurred due to importation and are potentially able to biting nuisance, have now become established and therefore changing the epidemiological status of various areas with regard to the occurrence of vector-borne disease (Medlock and Leach 2015). However, among the large number of mosquito species recognized worldwide, only a limited number are causing transmission of diseases and viruses, including species of the genus *Culex*, *Anopheles* and *Aedes* (Harbach and Howard 2007, Harbach et al. 2015).

Among the large number of mosquito species recognized worldwide, only a limited number are causing transmission of diseases and viruses, including

species of the genus *Culex*, *Anopheles* and *Aedes* (Harbach 2007, 2013 and 2015).

Moreover, in contrast to invasive species such as *Aedes spp.* in Europe, the *Culex spp.* complex is probably the most common and has been indicated as the primary vector of flaviroviruses such as the West Nile virus (WNv) (Engler et al. 2013). In particular, species belonging to the genus *Culex* are carriers of important diseases in warm-blooded animals, including humans, equidae, birds and other animals. The diseases they transmit are mainly due to viruses of the arbovirus family such as WNv but also Japanese encephalitis, or St. Louis encephalitis. *Culex spp.* serve also as important vectors nematode which are cause of the filariasis disease as well as protest parasites belonging in the phylum Apicomplexa causing various forms of avian malaria.

Consequently, various European states have suffered from sporadic outbreaks of disease of the WNV, in both human and horses, although the most significant epidemic has been occurring in 1996 in south-eastern Romania (Nikolay et al. 2015). In southern Europe, the WNV has been observed in the indigenous mosquito species *Culex pipiens* including Italy and Portugal (Calzolari et al. 2009, Brugman et al. 2018). Moreover, in Greece, *C.pipiens* has been identified as the dominant and endophilic species in rural areas in central Macedonia including the prefectures of Imathia, Kilkis, Pella, Pieria and Thessaloniki (Papa et al. 2011). Prevention and control of arbovirus infections can be meet mainly through the implementation of vector control strategies during the time of vector activity along with surveillance to determine virus activity and vaccination where possible (LaBeaud et al. 2011, Kading et al. 2019). The global distribution of WNv is shown in Figure 1.

The WNV is maintained in an enzootic cycle between birds (amplifying hosts) and the ornithophilic *Culex spp.* vectors, while humans, horses and other mammals are incidental hosts (Papa et al. 2010). Evidence of WNv has been reported in the majority of the sub-Saharan Africa as well as in North

Africa including Egypt, Maroco and Algeria (Brugman et al. 2018, WHO 2019), while its expansion in Europe has been linked to the presence of infected birds, acting as traveling reservoirs and the major bird migration paths to Europe (Pfeffer and Dobler 2010, Jourdain et al. 2019). The life cycle of WNv is shown in Figure 1a.

Filars are exclusively parasites of the tropics and subtropics including the southern part of the Mediterranean and Middle East. The best known of these diseases is lymph node filariasis which is also referred to simply as filariasis. Filariasis on the other hand is a parasitic disease in which the lymphatics of the abdomen become blocked. It is caused by a group of worms called filariae. *Culex spp* are the main vector of the nematode, although species belonging to other genera such as *Anopheles*, *Aedes spp* and *Mansonia spp* may also transmit the nematode (Paily et al. 2009).

(a)   (b)
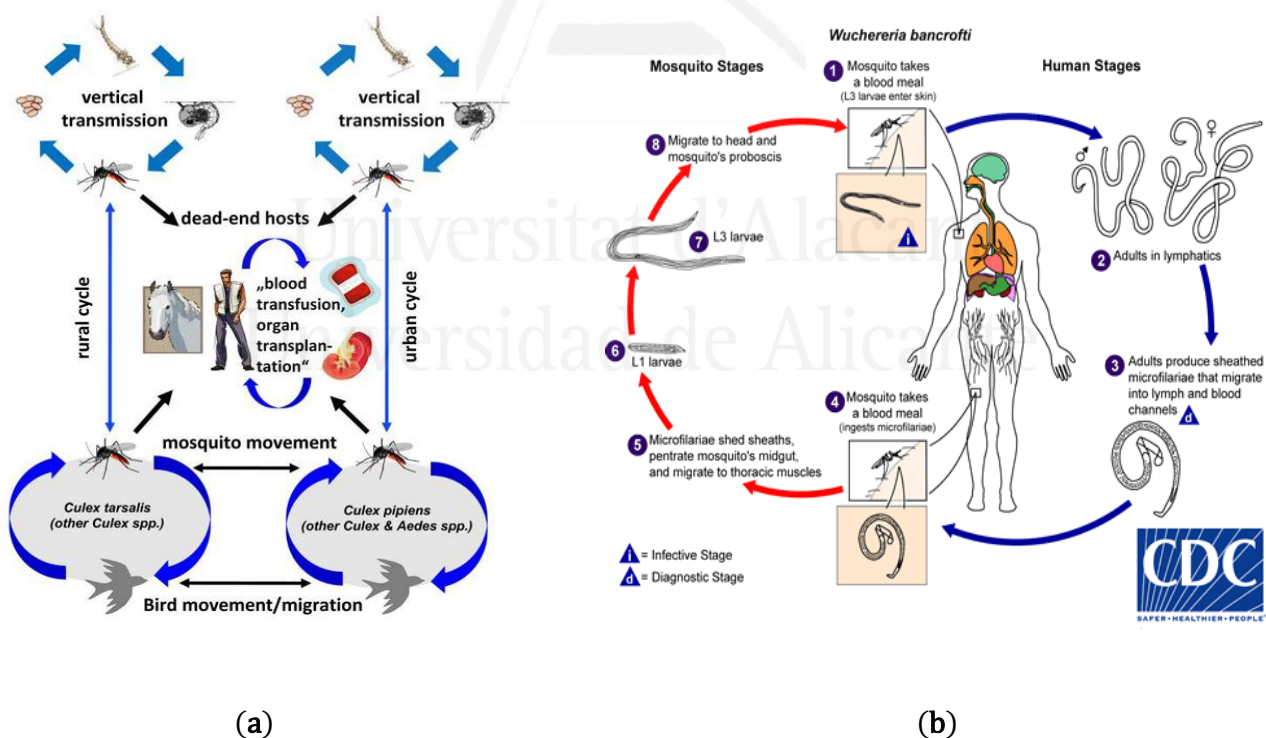
Figure 1.The transmission cycle of the West Nile virus (a) and filariasis (b) (CDC 2021).

The adult nematode worm lives inside the lymph nodes and lymph vessels, which it eventually occludes, causing the striking clinical picture of elephantiasis. The life cycle of filariasis is shown in Figure 2. It is apparent that the transmition cycle breaks if the mosquitoes are eradicated preventing the tranmsition of the desease.

Till now, the most effectual means to prevent infection from West Nile virus, as well as other mosquito borne diseases, is to target mosquitos to prevent bites and disease transmission. Hence, vector control, which relies on the use of insecticides, is the principal means of mitigating the spread of related diseases (Weidhass and Focks 2000). However, to be effective such a strategy, it is important to predict the temporal change in mosquito abundance and how it is affected by weather conditions (Deichmeister JM, Telang 2011, Ferraguti et al. 2013, Lalubin et al. 2013).

Nevertheless, although *Culex spp.* populations have been the object of intense vector control, realized pesticide treatment efficacy remains difficult to prove. This is generally due to the various effects of climatic variability on mosquito population dynamics, the fine scale and regional landscape heterogeneity, as good as due to the high stages of uncertainty in the prediction of population and disease outbreaks. Although climate has a direct effect on mosquito populations there is a complex interrelationship between vectors (e.g. abundance, behaviour, distribution and longevity), their pathogens (e.g. incubation period, replication and lineage) as well as their hosts and their interactions (Lafferty et al. 2001, Lafferty 2009, 2018).

In this context several mathematical models were used to connect the biological processes of vector dynamics and climate (Lord 2004, Lord 2007, Grassly and Fraser 2008, Luz et al. 2010, Eikenberry and Gumel 2018). To date, most epidemiological and vector population models have a deterministic nature and rely on some basic assumptions to define the various parameters of vector and disease dynamics under study (Wonham 2004, Wei at al. 2008, Lewis et al. 2010). Yet, most often these

parameters are unknown thought necessary to be first estimated to parameterize the model to be used later on for projections. Moreover, because of the impacts of various internal and external factors, the temporal evolution of population's processes is non-linear and characterized by random perturbations and, consequently, it is generally difficult to analyze and forecast population dynamics using only mechanistic models (Damos 2015).

On the other hand, stochastic models, recognize that all variables are probabilistic in nature and are handled as such and thus potentially may be employed to model non-linear ecological processes and advance our understanding of vector population dynamics which is utile for public health planning (Damos 2011). Thus, avoiding inherent relationship assumptions and parameter estimation in deterministic models; stochastic models provide a realistic data-based alternative to simulation of complex systems and robust predictions that could make informed and sound decisions.

1.2   Objective of the study

The main objective of the current work is preventing mosquito transmitted diseases, such as WNv, filariasis and encephalitis, through the formulation, development and application of novel stochastic modeling methods. Emphasis is given on multivariate analysis and modeling of *Culex spp*. population dynamics and related ecological variables and how they can be used as decision tools for arthropod vector control and public health planning.

With the mosquito abundance prediction, particularly, the public health authorities could predict the time evolution of mosquito abundance which is a prerequisite for successful management and prevention of related vector transmitted diseases.

To meet the main objective the following specific objectives have been set:

1. To identify the time where there is a high risk of mosquito populations through a typical growth season in a Mediterranean climate, namely Greece and an accurate apprehension of the climate affecting their dynamics to be later used on climate related multivariate modelling. To detect any synchrony and non-linear dependency between mosquito population dynamics and environmental variables using parametric and non-parametric methods and straightforward techniques such as wavelet coherence analysis. This vital in understanding the hazard of the transmission due to the presence of the vector and favourable climate conditions.

2. To present a general stochastic modeling approach enabling the multivariate analysis of vector population dynamics and related ecological variables including temperature and precipitation. Particularly, since fluctuations in mosquito population dynamics appears to a non-linear and to certain degree random, the specific aim was the development, application and testing for the first time a climate sensitive conditional Markov chain model (MCM) for simulating and predicting the population dynamics of arthropod vector dynamics and with the view to be potentially used in arthropod vector control programs and decision making.

3. To develop, implement, train and validate an exogenous non-linear autoregressive neural network (NARX), along with a recurrent neural network (RNN) and to provide a framework utile for the development of expert systems. These models represent a novel class of soft computing modeling methods that can be used as a complement to the MCM model with the view to be integrated in a public health decision support system.

2: Material and Methods

2.1 Detecting the relationship between arthropod vector abundance and climatic variables

2.1.1. The correlation coefficients and linear model

To establish if there are possible associations between the mosquito abundance populations and the climate variables, we first performed the pairwise correlation analysis. Moreover, multiple linear regression analysis was performed to fit the dependent variable, to the most significant, according to the correlation analysis, independent variables (Ganser and Wisely 2013). For the above reason as well as for simplicity, combinations and interaction terms of these variables were not included. The linear models were fitted to each observation year separately, including two climate variables (mean temperature in C^o and rain events in mm/m^2) and one lagged mosquito population variable (lag 1) as indicated by autocorrelations.

2.1.2. Wavelet analysis and the continuous wavelet transform

The wavelet analysis allows the analysis of the relationship between signals, and is particulate appropriate for following gradual change in a time series in forcing by exogenous variables (Cazelles et al. 2007, 2008). By definition, the continuous wavelet transform (CWT) decomposes a signal (time series) over dilated and translated functions known as 'mother wavelets. The mother wavelet is function of zero mean expressed as the function of two parameters one for the time position (Δt) and one for the frequency ($\Delta \omega$) of the wavelet (Cazelles et al. 2008).

The CWT applies a wavelet as a band pass filter to the time series of interest to be analyst. Particularly, the wavelet is stretched in time, t, by varying scale (s), in a way that $n = s \cdot t$, and normalizing (Grinsted et al. 2004). The most common wavelet is the Morlet defined as (Maraun and Kurths 2004):

$$\psi_0(n) = \pi^{-1/4} e^{i\,\omega_0 n} e^{-n^2/2} \tag{1}.$$

Where ω_0 is the frequency along with time is are both without dimensions and here taken to have value $\omega_0 = 6$ to satisfy the admissibility condition which implies that the transform vanishes at zero frequency. The CWT of the candidate time series variable x_n, $n=1,2,...,n$ with time step δ_t, is defined as the convolution of x_n with the scaled and normalized wavelet as follows (Grinsted et al. 2004):

$$W_n^x(s) = \sqrt{\frac{\delta_t}{s}}\; \sum_{n'=1}^{N} x_{n'}\, \psi_0\left[(n'-n)\frac{\delta_t}{s}\right] \tag{2}.$$

The complex argument of $W_n^x(s)$ can be interpreted as the local phase, while the wavelet power is defined as:

$$|W_n^x(s)|^2 \tag{3}.$$

Because the wavelet is not completely localized in time the CWT has edge effects (errors occur at the beginning and the end of the power spectrum of finite-length time series) due to periodicity assumptions of the wavelet transform. These should not be ignored and are defined as a Cone of influence (COI), which also gives a measure of the decorrelation time for a single spike in the time series (Torrence and Compo 1998, Foufoula-Georgiou and Kumar 1996). The decorrelation process reduces autocorrelation within the series and help to define critical change points and detect patterns not explained by seasonal trends. Technically it bounds the region of the wavelet spectrum in which edge effects become important and is defined here as the e-folding time for the autocorrelation of wavelet power at each scale.

Thus, the wavelet analysis includes class of functions which are well localized in time and frequency and exist for a finite duration. A stretched wavelet (large scale factor and low frequency) permits to capture the slowly varying changes in the time series, while a compressed wavelet (low scale factor and high frequency), helps in capturing any abrupt changes.

In the analysis each scale wavelet is shifted in time along the entire length of the series and compared with the original series (mosquito abundance and climate data). The process is repeated for all scales resulting in coefficients that are a function of wavelet scale and shift parameter. In this context wavelet analysis is used to quantify similarities (i.e., synchronisation in the time domain in respect to different scales) in time series which are characterized by transient fluctuations (Tiwari 2016, Sato et al. 2018).

2.1.3. The cross wavelet transforms

Because mosquito and climate data exhibit slowly abrupt changing trends or oscillation punctuated with transients, we apply the cross wavelet transform (XWT) for examining the relationship between mosquito abundances and temperatures in the time- frequency space (Kilpatrick et al. 2004, Grinsted et al. 2004). The XWT can be used to measure the degree of synchronization between two different time series conjointly and how it evolves in time in terms of changes in its behaviour in a given time period (Kilpatrick et al. 2004). In this work the XWT exposes regions in the time (i.e., weeks of observation) with high common power and further reveals information about the phase relationship between mosquitos and temperature to be thought of as the local correlation between the series.

The cross-wavelet transforms (XWT) of two time series Xn and Y_n is defined as (Torrence and Compo 1998):

$$W^{XY}(s) = W_n^X(s)W_n^{Y*}(s) \tag{4}.$$

Where $W_n^{Y*}(s)$ is the complex conjugate of $W_n^Y(s)$. The cross-wavelet power is:

$$|W_n^{XY}(s)| \tag{5}.$$

The confidence intervals for the XWT are derived from the square root product of two chi-square distributions. The probability distribution of the XWT of the two time series for a respective confidence level $Z_v(p)$ is:

$$D\left(\frac{|W_n^X(s)W_n^{Y*}(s)|}{\sigma_X\sigma_Y} < p\right) = \frac{Z_v(p)}{v}\sqrt{P_k^X P_k^Y} \tag{6},$$

where P_k^X and P_k^Y are the theoretical Fourier spectra and σ_X, σ_Y are the standard deviations, respectively. The analysis was performed using R (R Core Team 2017), including the package 'biwavelet' based on the WTC MATLAB package written by Aslak Grinsted and the wavelet program written by Christopher Torrence and Gibert P (Gouhier et al. 2019).

2.2. Stochastic modeling of arthropod vector disease variables through Markov chains

2.2.1. Stochastic process of ecological time series

Let X(t) be the ecological variable (e.g., mosquito population, or climatic variable), which is considered as a stochastic process that evolves in time t and is defined in a probability space (Ω, **F**, **P**). Where Ω is a sample space, **F** is a set of outcomes in the sample space and **P** assigns each event of **F** a probability. If the number of **F** is not countable then the process is denoted by {X (t): t ≥ 0}, or {Xt} t≥0. In the first case the process is called a chain in discrete time and in the second, in continuous time. Here we consider the first case, since data have been observed in specific time points and not continuously.

Let now *S* be the space created by all possible process values *X(t)* in discrete time. If **S**=(0, 1,...), we refer to a stochastic process with integer

values or a discrete state process, e.g., a population threshold or class that corresponds to the number of mosquitos captured in a day, or a class of mean temperature values for that day etc. Hence, we consider **S** to take real and finite *n* values and this contemplative process is called an n-dimensional stochastic process.

2.2.2. The Markov chain model

The above stochastic process consists of a Markov chain (MC) which is determined by its initial state distribution and a transition probability matrix $\boldsymbol{P}$ of size m is (Damos et al. 2011):

$$\boldsymbol{P}(i,j) = Pr[X(t + \Delta t) = i | (X(t) = j], \textit{t=1,2,...,n} \tag{7}.$$

The simplest kind of discrete variables the transition matrix may have two stages $\boldsymbol{S} = (1,2)$ which we define in their simplest form as a high or low level of the ecological variable (e.g., mosquito vector population, temperature) or occurrence or not occurrence (e.g., rain). A sequence of weekly observations constitutes time series of that discrete variable. For the first order Markov chain, the transition probability to future state depends only on its current state. Thus, knowing that at week *i* the variable X is either in state 1 (low population levels X(i) = a), or state 2 (high population levels X(i) =b) the related transition matrix is:

$$\boldsymbol{P}(i,j) = \left[p_{ij} \right] = \begin{bmatrix} p_{11} \, p_{12} \\ p_{21} \, p_{22} \end{bmatrix} \tag{8}.$$

Where:

$$\begin{cases} p_{11+}p_{12} = 1 \\ p_{21} + p_{22} = 1 \end{cases} \tag{9},$$

considering that $\boldsymbol{P}$ is a row stochastic matrix. By considering m states, $\boldsymbol{S}$=(1,...,m) a higher dimension of the transition matrix is as follows:

$$\boldsymbol{P}(i,j) = [p_{ij}] = \begin{bmatrix} p_{11} & \cdot & \cdot p_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ p_{m1} & \cdot & \cdot p_{mm} \end{bmatrix}, i,j=1,\dots m \qquad (10).$$

Where $0 \leq p_{ij} \leq 1, \sum_{j=1}^{m} p_{ij} = 1, \forall\ i = 1,\dots,m$ and m is the number of states. A state Sj is said accessible from state Si (written Si → Sj) if the defined transition system starting in state Si has a positive probability to reach the state Sj at a certain point, i.e., ∃n> 0: $p_{ij} > 0$. Two states are said to communicate if both Si → Sj and Sj → Si. Moreover, any state Si is considered periodic if any return to state *Si* occur in multiplies of *ki* steps and *ki* is the period and *ki* = GCD {n: Pr(Xn = si |X0 = si) > 0}, where GCD is the greatest common divisor. Thus, for ki = 1 the state Si is aperiodic, else if ki > 1 the state Si is periodic with period ki*.* In other words, a state is periodic if after a fixed number of transitions, ki > 1, the state can only return it itself otherwise it is aperiodic.

2.2.3. Data inferred Markov chain modelling

Because the time evolution of a mosquito population process is too complicated to derive laws and construct parameterizations from first principles, we have used a data-driven method to construct parameterizations by inferring from data.

First the data is classified into different scale states (e.g., population levels) and a matrix is estimated for each scale state. In particular, for *m* states there have to be *m²* matrix entries to be estimated. The transition probability matrix entry $\boldsymbol{P}(i,j)$ is estimate as follows (Dorrestijn et al. 2013):

$$\hat{p}(m,n) = \frac{T(i,j)}{\Sigma_j T(i,j)} \qquad (12).$$

Where $T(i,j)$ counting for the transition from m to n observed states observed in a given data set and $\hat{p}(i,j)$ is accutally the maximum likelihood estimator of $p(i,j)$.

Thus, the Markov chains are 'trained with' data accrue from real observations with the aim of mimicking the observed behavior of the population process afterwards in which a finite state MC is inferred from data by estimating its transition probability matrix:

$$T(i,j) = \sum_t \mathbf{1}[X(t + \Delta t) = i]\mathbf{1}[X(t) = j] \qquad (13).$$

2.2.3. The conditional Markov chain model

For the case we are analyzing, the occurrence and level of mosquito population depend on the physical state of climate conditions. Particularly, since mosquitoes are arthropods and all arthropods are poikilothermic organism, their development and occurrence of states are affected by favorable temperatures and rain conditions. (e.g., favorable versus unfavorable climate). This means that if a MC is used to mimic the process of population occurrence X(t), it can be improved by taking into account the condition of a second process Y(t) which is related to a climate variable (e.g., state transitions of temperature levels or rainy conditions). Under this assumption, probabilities take the following form (Dorrestijn et al. 2016):

$$\boldsymbol{P}_{\gamma}(i,j) = Pr[X(t + \Delta t) = i | (X(t) = j, Y(t) = \gamma], t=1,2,\dots,n \qquad (14).$$

2.2.4. Data inferred conditional Markov chain modelling

If a finite number of states is considered (say five as you will see later), then it is possible to construct a conditional Markov chain (CMC) by estimating a transition probability matrix, $\boldsymbol{P}_{\gamma}$, for each possible state. This can be done by knowing the time evolution of the series for a finite number of states that is used as basis to train the chain model. The transition matrix is estimated as follows:

$$\boldsymbol{P}_\gamma(i,j) = \frac{T_\gamma(i,j)}{\Sigma_\gamma T_\gamma(i,j)} \qquad (15).$$

In which

$$T_\gamma(i,j) = \sum_t \mathbf{1}[X(t+\Delta t) = i]\mathbf{1}[X(t) = j]\mathbf{1}\,[X(t) = \gamma)] \qquad (16).$$

In which $\mathbf{1}$ is the indicator function: $\mathbf{1}(A) = 1$ if A is true and $\mathbf{1}(A) = 0$ if A is false (10), while t runs over time instances in the data set used to train the Markov chain model.

2.2.5. Determination of transitions states and data encoding

In order to work with finite state conditional Markov chains, vector population dynamics and climatic variables must be discrete and coded to a finite number of states. If the data is uniformly distributed, this can be done using tree classification schemes based on pre-defined thresholds.

However, since ecological data are most often not uniformly distributed, choosing thresholds is difficult and could result in classes to which no data is assigned and classes to which almost all data is assigned. Therefore, we have used k-means clustering algorithm, performing preliminary trials, with predefined k:2-6 to overcome the problem of subjective defining the different mosquito population levels [18]. In particular, we considered each set of mosquito population observations x1, x2,...,xn as a d-dimensional real vector and implemented a standard, k-means clustering algorithm aiming to partition the n observations into k sets (k ≤ n) that correspond to discrete population states: $\mathbf{S} = \{S1, S2, ..., Sk\}$ so as to minimize the within-cluster sum of squares (WCSS):

$$arg_S\ min \sum_{i=1}^{k} \sum_{x_j} \left\| x_j - \mu_i \right\|^2 \qquad (17),$$

where μ_i is the mean of points in S_i. Let now n_{ij} denote the number of individuals who were in state i in period t-1 and are in state j in period t. We

can estimate the probability of a mosquito population being in state j in period t given that they were in state i in period t-1, denoted by p_{ij}, using the following formula (Bielecki et al. 2017 , Lafferty et al. 2019):

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \qquad (18).$$

The probability of transition from any given state i is equal to the proportion of mosquitoes that started in state i and ended in state j as a proportion of all individuals in that started in state i.

Thus, using the above scheme the observed behavioural stream of population dynamics is first converted into a symbolic sequence of population states to be used later on the estimation of the transition matrices.

2.2.6. Markov chain modeling and equilibrium distribution

Field data on frequencies of successive mosquito population levels, which were not used for the data inferred MC modeling, were assembled to obtain an empirical intensity transition matrix. Then the empirical transition matrices were generated and the observed frequencies were compared visually, as well as statistically, with that obtained from the MC models.

Two methods have been applied to evaluate the equidistribution between the observed sequences and that of the MC models. First, the homogeneity of the transition matrices was tested using a Chi-square (ChiSq) minimum discrimination statistic test. In particular, we have used the *verifyHomogeneity* function inspired by Kulbach and Kamler [50, 51]. Particularly, considering the time evolution of the mosquito population as a stochastic process that generates: i = 1, 2, . . . , n discrete time Markov chain samples and that the cardinality of the state space is S, the Homogeneity function verifies whether a chain belong to the same unknown one. The

function shows that its test statistics follows a chi-square law and is estimated as follows:

$$2 * \sum_{i=1}^{l} \sum_{j=1}^{S} \sum_{k=1}^{S} f_{ijk} ln \frac{n * f_{ijk}}{f_{i...}f_{.jk}} \sim x^2(r * (r-1)) \tag{19}$$

If there are l realizations of a Markov chain of order 1 with S states, the null hypothesis, H_0, is that the transition probability matrix is the same for all *i* and for every possible pairing of j and k and P(X>Chsq) which is less than or equal to the significant level, $\alpha = 0.05$. In the current work we have considered the case of l=2 chains as two realizations of the S=5-state Markov chains (theoretical and empirical) that are tested for homogeneity. The frequency entries, are: f_{ijk} and i=1,2, j=1,2,... and k=1,2,..,5.

Secondly, the asymptotic distribution properties of the theoretical and the empirical intensity transition matrices were compared using an entropy-based divergence distance measure (Damos 2015). In this case, we considered the mosquito population stochastic process as homogenous and thus starting from an initial distribution of population states a limiting probability, $\boldsymbol{\Pi}_i$, exist:

$$\lim_{n \to \infty} (\boldsymbol{P}_i) = \boldsymbol{\Pi}_i \quad , i=1,2,....,n \tag{20}$$

It follows that:

$$\sum_{i=0}^{\infty} \Pi_i = 1 \tag{21}$$

This is called the normalizing condition. We further associate an entropy:

$$H_i = -\sum_{j=1}^{n} p_{ij} \log p_{ij} \tag{22}$$

H_i represents the average amount of uncertainty of the population system for moving one step ahead being initially in state Si. We are now interested in estimating the average uncertainty of the chain for moving one step ahead of any other initial state, which is (Damos 2015):

42

$$H(X) = H(\boldsymbol{P}) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_i\, p_{ij} \log p_{ij} \qquad (23).$$

Considering the Markov chain process as ergodic (e.g., MC and CMC models as well as the Observed mosquito population process used for model validation), so that:

$$\boldsymbol{H}(X) = \lim_{n\to\infty} (H^t/t) \qquad (24).$$

Finally, I define a distance measure by introducing the following norm:

$$\|\boldsymbol{H}_t - \boldsymbol{H}_{t=0}\|, \text{ t=1,2,....} \qquad (25).$$

Where $\boldsymbol{H}_{t=0}$ the entropy associated with the initial probability distribution $\boldsymbol{\Pi}_{i=1}$ representing the different mosquito population levels and $\boldsymbol{H}_t$, the entropy of each time step ahead t=2,3…

The above scheme quantifies the rate of convergence from a starting non-equilibrium probability distribution towards equilibrium.

2.3 Soft computing of arthropod vector population dynamics

2.3.1. General structure and functioning of ANNs

ANNs were proposed as a mathematical tool to simulate the complex functioning of the human brain. The brain has the ability to parallel processing of data and continuous learning through the interaction with environment. The ANNs has similarities with biological neurons and consist of a set of artificial neurons that interact through synapses (Recknagel 2001, Schmidhuber 2015). The degree of interaction between the synapses is determined by weights (synaptic weights). The neural network interacts with its environment (i.e., variables of interest), and the synaptic weights change constantly and thus strengthening or weakening the power of each interaction node. Thus, the information from the external variables (i.e., environment) is encoded in the synaptic weights of the network and gives the ability to the ANN to simulate the process related to those variables. In order for the network, it used a training algorithm which aims to optimize through iterations the model performance.

The main advantage of neural networks is first that it stores knowledge and experience from the environment used for its training (here mosquito abundance and temperature), which it can then recall to simulate the process. Second, it has the ability to generalize, that is, to extract the basic features of a system characterized by noisy data and complex non-linear processes. The artificial neuron is the structural unit of an ANN at is shown in Fig.1 In this neuron, information always flows in one direction, from left to right, i.e., there is no loop feedback. In the first phase, each input is multiplied by the synaptic weight, w and in the second, the weighted inputs and an externally applied bias threshold factor adds up and gives net input according to an activation potential (Browne 1997).

An ANN can be defined by a different number of neurons that are connected and interact according to their weights. Figure 2a shows a neural

network that consist of three layers: the input layer (variables used to train the model), the hidden layer (which consist of four neurons) and the output layer. Figure 2b shows the most representative activation functions used in ANN models.
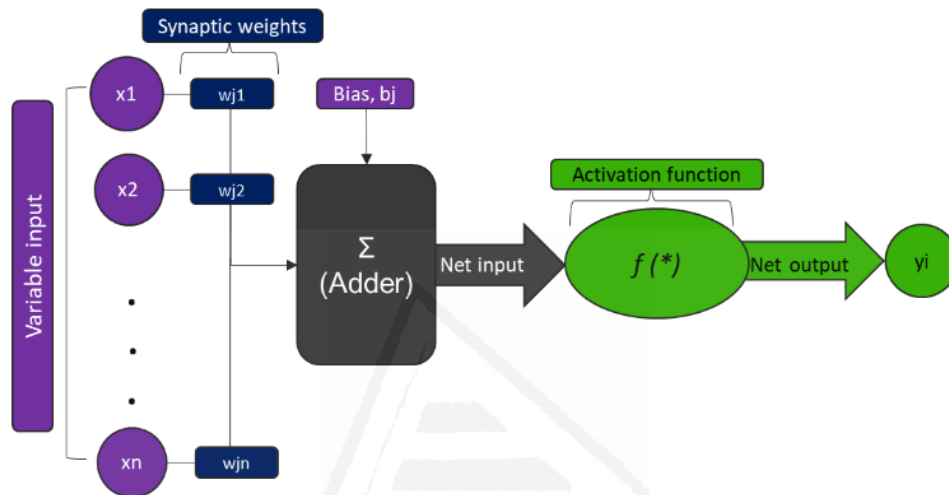
Figure 2. Simplified neuron which consists of the structural unit of an ANN.

(a')   (b')
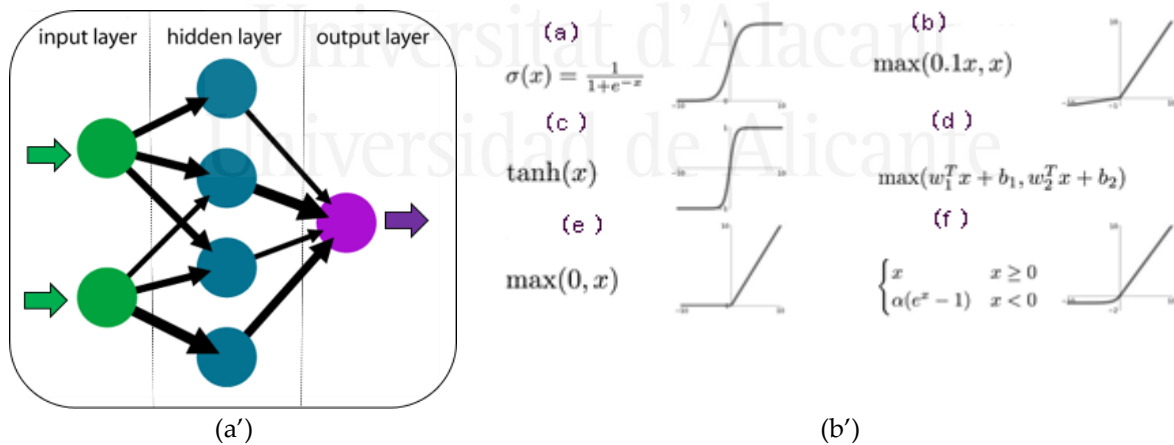
Figure 3. Simplified ANN that consists of three layers: the input layer (variables used to train the model), the hidden layer (which consist of four neurons) and the output layer (a') and a summary of the most common activation functions, f (b'). The functions are (a) (a) sigmoid, (b) Leaky ReLU, (c) tanh(x), (d) Maxout, (e) ReLu, and (f) ELU.

2.3..2 Formulation of the NAR, NARX and FTD Neural networks

Soft computing of the mosquito vector population dynamics was modelled using artificial neural networks (ANNs). Three different autoregresssive models are applied. I have first applied a simle non-linear autoregressive model (NAR). Secondly, a standard non-linear autoregressive network with exogenous inputs (NARX) which is a part of discrete-time non-linear systems which conceptually has feedback connections which enclose the layers of the network and is using the past values for prediction (Bousada et al. 2018). Lastly, a focused time delayed neural network (FTD) was applied and which may considered as a simplified version of the NARX in with the output feedback is eliminated (see below).

The self-regulating non-linear autoregressive neural network (NAR) working with an external mosquito abundance time series is as follows (MatLAB 2021):

$$y(t)=f(y(t-1),y(t-2),...,y(t-p))+e(t) \hspace{2cm} (26)$$

Where y is the mosquito abundance which depends on previous mosquito population values p and et is error term. The model is trained by a sequence of available mosquito abundance data and predicts the population abundance y(t) with data of previous abundances of the same sequence. Further, to find the best model predictions a combination of available mosquito abundance data and different model configuration tests were performed with various delays of mosquito population based on the autocorrelation function (ACF). To date, the ACF reveals how the correlation between any two values of the population sequence changes as the separation changes. Thus, it is a time domain measure and provides a criterion of defining the memory of the population process.

The defining equation for the NARX model with a parallel architecture can be expressed as follows (Beale et al. 2015, Akhtar et al. 2019):

$$y(t) = F([y(t-1), y(t-2), \ldots, y(t-d_y), u(t-1), u(t \qquad (27)$$
$$-2), \ldots, u(t-d_u)]$$

Where $F(\cdot)$, is the mapping (unknown non-linear) function of the neural network, y(t) is the output of the NARX at time step t. $y(t-1), y(t-2), \ldots, y(t-d_y)$ are the past output values of the NARX, $u(t-1), u(t-2), \ldots, u(t-d_u)$ are the exogenous inputs of the NARX, d_u is the number of input delays and d_y is the number of output delays.

Thus, the output of the NARX network $y(t)$, is fed back (close loop) to the input of the network through delays t and thus equation 1 can be described in a compact form as follows (Menezes and Barreto 2006):

$$y(t) = F([y(t), u(t)] \qquad (28).$$

Where $y(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}$ denote the output (mosquito abundance) and the input (temperature) of the model at describe time t, respectively, for different lagged output and input memory orders.

Moreover, a NARX neural network is usually trained in series-parallel (SP) mode first and later on, on a parallel mode (PM). To date in the SP mode only the actual values are taken in to account and form the outputs as follows:

$$\hat{y}(t) = \hat{F}([y(t-1), y(t-2), \ldots, y(t-d_y), u(t-1), u(t \qquad (29).$$
$$-2), \ldots, u(t-d_u)] = \hat{F}([y_{SP}(t), u(t)]$$

In the PM mode, the outputs which are estimated are fed back to the network and are included in the outputs:

$$\hat{y}(t) = \hat{F}([\hat{y}(t-1), \hat{y}(t-2), \ldots, \hat{y}(t-d_y), u(t-1), u(t \qquad (30).$$
$$-2), \ldots, u(t-d_u)] = \hat{F}([\boldsymbol{y}_P(t), \boldsymbol{u}(t)]$$

In the case of the FTD neural network the output memory of a NARX model is set by a zero delay (n_y=0) resulting in a plain neural network architecture which can be described as follows:

$$y(t) = F[u(t-1), u(t-2), \ldots, u(t-d_u)] = F([\boldsymbol{u}(t)] \qquad (31).$$

Where $\boldsymbol{u}(t) \in \mathbb{R}$ is the input regressor (here temperature). Thus, the FTD neural network is a simplified formulation of the NARX model which discards all the dynamic learning information of the output past memories.

2.3.3. Architecture and components of the NARX and FTD neural network

Both, the NARX and the FTD neural networks consist of the input layer and the output layer which are approximating the map function $F(\cdot)$ through an internal architecture known as the Multi-Layer Perceptron (MLP). By definition the classical MLP consist at least of three layers: the input, the hidden and the output layer. If i is the number of neurons in the layer, j is the number of elements in input vector p_j, then each vector of the input layer is connected to each neuron input trough the weight matrix $\boldsymbol{W}$ (Matworks 2020):

$$\boldsymbol{W} = \begin{bmatrix} w_{1,1} \ w_{1,2} \ \dots \ w_{1,j} \\ w_{2,1} \ w_{2,2} \ \dots \ w_{2,j} \\ \\ w_{i,1} \ w_{N,2} \ \dots \ w_{i,j} \end{bmatrix} \qquad (32).$$

Since in most cases the number of inputs to a layer may differ from the number of neurons, the matrix is no necessary *nxn*. For each single layer, each neuron multiplies the input layer p_i given be the previous layer by the weight vector $w_{i,j}$ which yields to the following scalar product: $p_j \cdot w_{i,j}$ (Matworks 2020). The weighted sum of inputs (*netsum*) consists of the transfer potential, θ, which aggregates the inputs and its weights as follows:

The transfer potential passes through a predefined activation function,f, to obtain the output a_i of the following neuron (Boussaada et al. 2018):

$$a_i = f(\sum_{\iota=1}^{n} p_j w_{i,j} + b) \qquad (33)$$

Where i is the index of neuron in the layer, j is the input index of the neural network and b is a bias vector. The output of the NARX neural network has a hyperbolic tangent sigmoid transfer (*tansig*) function in the inner layer

and a pure linear function (*purelin*) in the output layer which are given as follows:

$$f(\theta) = tansig(\theta) = \frac{2}{1 + e^{-2\theta}} - 1 \qquad (34),$$

$$f(\theta) = purelin(\theta) = \theta \qquad (35).$$

The equations which describe the function of the first and the second layers of a NARX and FTD neural network are as follows (Aribowo 2017):

$$a1(t) = \sum_{i=1}^{j} w_{ij}\, p1(t - d1) + b1 \qquad (36),$$

$$a2(t) = \sum_{i=1}^{j} w_{kj}\, a1(t) + b2 \qquad (37).$$

Where *i* and k are the number of neurons, w_{ij} is the weighted input of the network, $p1(t - d1)$ are the lagged inputs of the layer 1, $a1(t)$ is the output of the hidden node, w_{kj} are the weights of the second layer and $a2(t)$ is the output of the kth neuron in the lth layer at the time (t).

2.3.4. Model training, testing and validation

In the applied NARX model, the predictions of mosquito dynamics were performed from the past predicted values of the abundance time series and from the present and past values of the exogenous temperature input. To date, to extract these two key input variables we prior analyzed the correlation coefficients of different meteorological data with an imposed time lag. Moreover, we have used 10 hidden neurons and 2 number of delays-time in weeks, because they are giving satisfactory results after a preliminary training and testing of different combinations of hidden neurons and delays. Data division was performed random using both data sets (2011 and 2012) in which finally 60% of the data were used for NARX training (38 target time series steps), 20% for validation (13 target time

49

series steps) and 20% for testing (13 target time series steps). To date that, the validation datasets consist of the sample of data held back from training, while the test data set is used for fine tuning (optimizing) the ANN model hyper parameters (i.e., taking weights of the trained ANN and use it as initialization for a new model being trained and so on).

Each time step corresponds to the weekly counts of the *Culex sp.* mosquito abundances. The Levenberg-Marquardt (LM) was used as training algorithm in which the network training automatically stops when generalization stops improving, as indicated by an increase in the mean square error (*mse*) of the validation samples which is used as cost function, C defined as:

$$C(w, b) = \frac{1}{n} \sum_{i=1}^{n} (e_i)^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_j)^2 \qquad (38).$$

Where *w* and *b* referred to all the weights and biases in the network, respectively, *n* is the number of training inputs, y_j are the outputs when y_i is the input. The LM algorithm minimize *C* as small as possible by optimizing weights and biases through gradient descent. The partial derivatives of the cost function with respect to any weight w and any bias b were estimated through backpropagation algorithm.

All data analysis was performed using Matlab numerical computing environment and ANNs Simulink toolboxe and related programing language developed by Mathworks (Mathworks 2020).

2.4. Study area, mosquito surveillance data and handling

2.4.1 Study area

The population dynamics of *Culex. sp* was studied along 50 km close to the city of Thessaloniki on the west coast of the Thermaic gulf in Northern Greece. The special geography and climatic conditions favor the

development and evolution of mosquito populations. In especial, the location lies close to the coast and consist of a wetland complex which includes the Lagoon of Kalochori, the estuary of the Gallikos river, the delta of the Axios river, the estuary of the Loudias river, the delta of the Aliakmon River, the wetland of Nea Agathoupoli and the Alyki Citrus wetlands as well as closely related agricultural and semi urban areas. The climate of this area is classified as Csa Mediterranean climatic type (Köppen, 1923) with long, hot, and dry summers (the mean maximum temperature lies often in the range of 29° and 35° degrees of Celsius), relatively mild and rainy winters, and average annual air temperatures of approximately 15°C.

2.4.2 Mosquito Surveillance and climatic data

Free mosquito trap data available from the open European Union Data Portal (EU ODP) (http://data.europa.eu) (Figure 4), which provides access to data from the European Union (EU) institutions and other EU bodies which can be reused for commercial or non-commercial purposes (European Commission Decision 2011/833/EU). In particular, we used adult mosquito trap data of *Culex sp.* sampled from 11 locations in central Macedonia-Greece. Data were handled as vectors which consisted of close to weekly time intervals of the number of adult mosquitoes captured in CO2 traps from mid-May to September. Data during three successive observation years (2011, 212 and 2013) were used for the analysis.

Climate data and mean air temperatures in Celsius and rain events in mm were obtained by the national observatory of Athens through a meteorological station, which is located in Makrohori town which in the same location and latitude and near to the mosquito observation area (http://stratus.meteo.noa.gr/front).
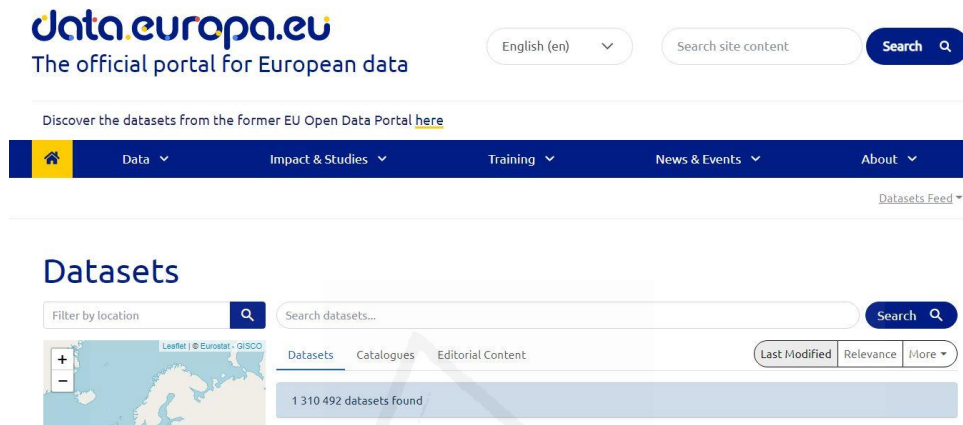
Figure 4. European open access data portal.

2.4.3 Data curation and transformation

Mosquitoes were sampled weekly for a total of nine weeks from May till September. Because of slight differentiations between the times intervals of some of the trap counts, data were transformed to mosquito per Trap per Day (MTD) and thereby averaged over the 11 nearby sampling locations (Figure 5). The index estimates the average number of mosquitoes captured in on the day that the trap is exposed in the field. Its value is the result of dividing the total number of captured flies by the product obtained from multiplying the total number of serviced traps by the average number of days the traps were exposed. The formula is as follows:

$$MTD = \frac{M}{T \cdot D} \tag{39},$$

where: M = Total number of mosquitos trapped, T= Number of serviced traps and D= Average number of days traps were exposed in the field. The

52

function of this population index is to have a relative measure of the size of the insect population in a given space and time (IAEA[1], Trapping-wide fruit fly programs, Vienna 2003). To meet normality assumptions the MTD data were further transformed as follows:

$$MTD\_log = \log(MTD + 1) \qquad (40).$$
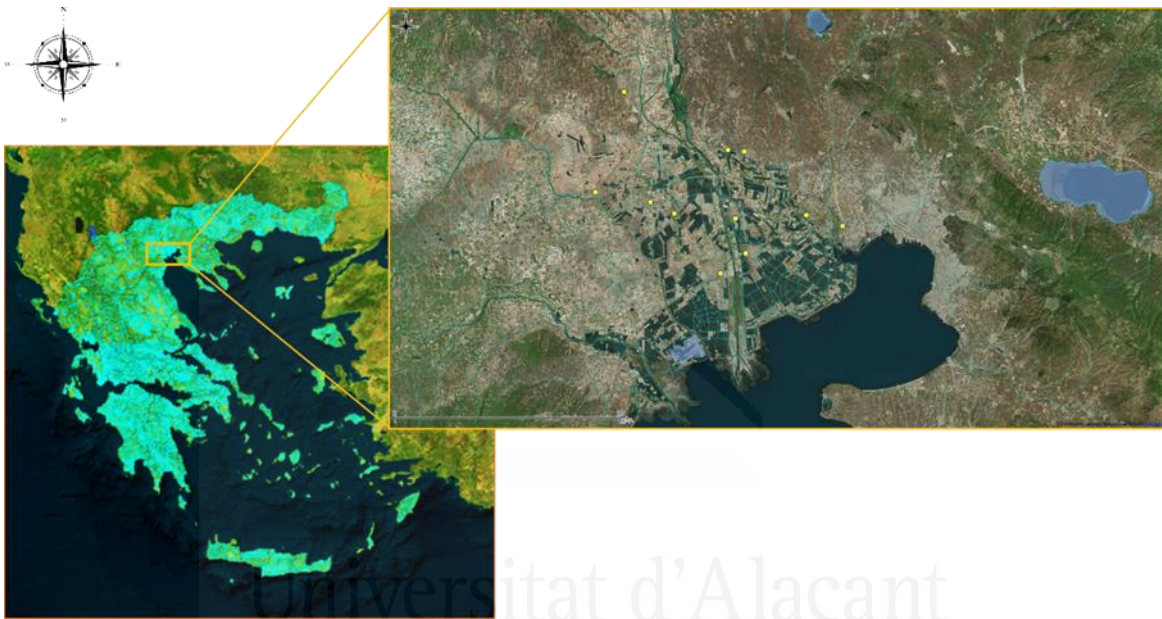
Figure 5. Mosquito surveillance area in Greece (left) and specific *Culex sp.* surveillance locations (in yellow) in the wider area of Thessaloniki municipality in Northern Greece (right).

---

[1] International atomic energy agency

Chapter 3: Results

3.1. Detecting seasonal transient correlations between populations of the West Nile Virus vector *Culex sp*. and temperature with wavelet coherence analysis.

3.1.1. Adult mosquito seasonal flight patterns

The temporal analysis revealed similar fluctuations for all sampling locations and during both years the standardized mosquito abundances indicated early season peaks in May and June followed by a decline in the last quarter of June and resurgence during July and August especially for the second observation year. In particular, figure 6 shows the seasonal flight patterns and the standard errors of adult *Culex. sp*. captured during 2011 and 2012 in the prefecture of Imathia in Northern Greece as well as the associated autocorrelations for different lags (weeks). Counts represent weekly observations of mosquitoes trapped from May till mid-September. The mosquito population started to increase in late May and peaked in mid-July during 2011. In 2012 a considerable higher number of mosquitos were captured compared to 2011. Additionally, during 2012 there were observed two peaks the first in late June and whilst the second in mid-August. During both years, the autocorrelations do not indicate any abrupt fluctuations and/or clear periodic pattern in the mosquito time series with regard to different lags. Nevertheless, after some perturbations the autocorrelation later decay to zero, indicating possible the existing of a moving average process, although there is some ambiguity regarding the different patterns observed between the two observation years.
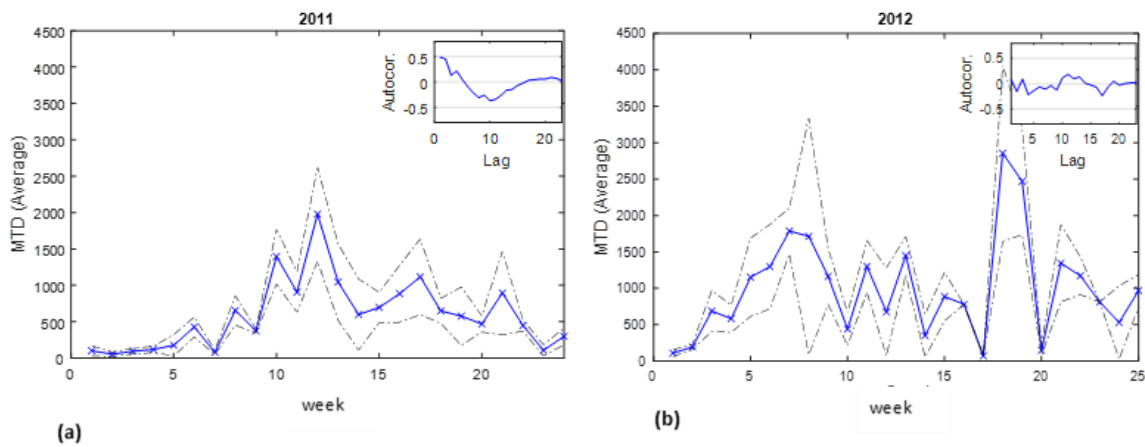
Figure 6. Average (Culex sp.) mosquitos per trap per day (MTD) (straight lines) and standard error (dashed lines) in relation to normalized weekly counts and autocorrelations of in relation to previous count lags.

3.1.2. Pairwise correlations between mosquito abundances and climate variables

In 2011 there was a high and positive correlation between temperature and the average mosquito counts (r=0.60) as well between temperature and laged1 log transformed MTD (r=0.65). Nevertheless, the highest correlation was observed between temperature and the log transformed MTD (r=0.78). Additionally, a very poor correlation was observed between rain events and mosquito abundances (r=-0.32, for the log transformed MTD), as well as between wind speed or wind direction and mosquito captures (r=0.23 and r=-0.14, for the log transformed MTD, respectively) (Figure 7). The same patterns, although with lower values of correlation coefficients were observed during the season of 2012 (Figure 8). In particular, there was a positive correlation between temperature and the average mosquito counts (r=0.46). The highest correlation was observed between temperature and the log transformed MTD (r=0.74). On the other hand, a very poor and negative correlation was observed between rain

events and mosquito abundances (r=-0.11, for the log transformed MTD), as well as between the wind speed or the wind direction and the mosquito captures (r=0.005 and r=0.006, for the log transformed MTD, respectively).
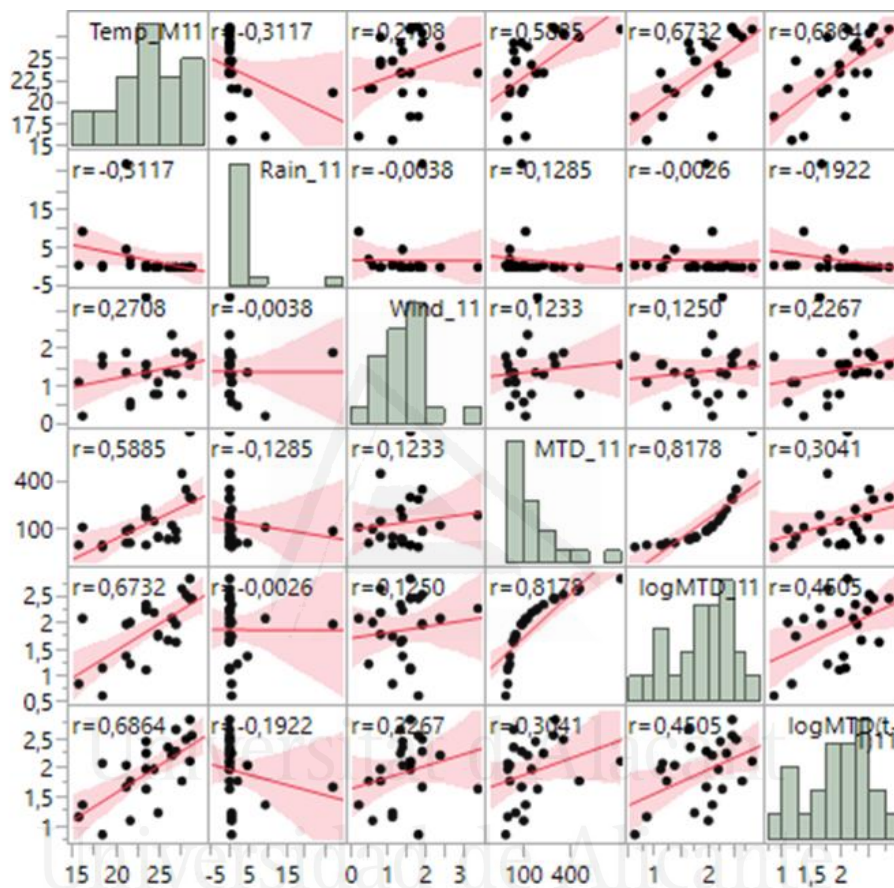
Figure 7. Correlations and linear trend with confidence intervals, between climate variables: mean temperature (Temp_M11, Temp_m_12), Rain (Rain_11, Rain_12), wind speed (Speed_11, SPeed_12), wind direction and Culex sp. adult mosquito counts expressed as mosquito per trap per day (logMTD_11, logMTD-2012), lag1 log transformed mosquito per trap per day (logMTD(t-1)11, log(t-2)12) for 2011, (correlations are estimated by residual maximum likelihood, REML, met
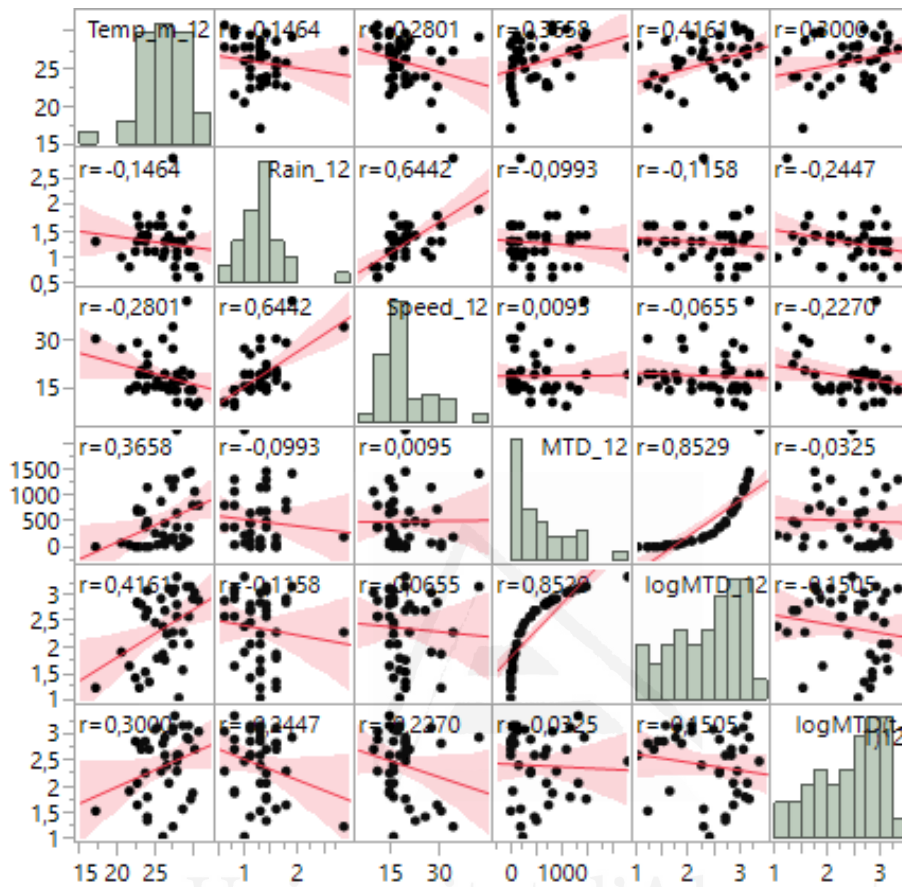
Figure 8. Correlations and linear trend with confidence intervals, between climate variables: mean temperature (Temp_M11, Temp_m_12), Rain (Rain_11, Rain_12), wind speed (Speed_11, SPeed_12), wind direction and *Culex sp*. adult mosquito counts expressed as mosquito per traps per day (MTD_11, MTD_12), log transformed mosquito per trap per day (logMTD_11, logMTD-2012), lag1 log transformed mosquito per trap per day (logMTD(t-1)11, log(t-2)12) for 2012 (correlations are estimated by residual maximum likelihood, REML, method).

Table 1 illustrates the p-values for testing the hypothesis of no correlation between the different expressions of mosquito abundances and the climate variables. Each p value represents the probability of getting a

correlation as large as the observed value by random chance, then the true correlation is zero. During both years of observations, significant p – values (<0.05) are observed between log transformed mosquito abundance (log MTD). Grounded on this relationship the number of variables used to model the mosquito abundance was deduced (see below).

Table 1. p-values of correlations between adult mosquito counts and climatic variables.

		p-value	
Culex spp.	Climate variable	2011	2012
MTD	Mean temperature	**0.003***	**0.015***
	Rain (mm)	0.968	0.309
	Wind speed (km/h)	0.784	0.253
logMTD	Mean temperature	**0.000***	**0.009***
	Rain (mm)	0.406	0.450
	Wind speed (km/h)	0.681	0.484
logMTD(t-1	Mean temperature	**0.022***	**0.041***
	Rain (mm)	0.873	0.881
	Wind speed (km/h)	0.191	0.393

* Significant at the p=0.05 level

3.1.3. Multivariate linear regression model between mosquito vector abundance and climate variables

The overall multivariate linear regression model is represented in terms of a Leverage plot in Figure 9 and 10 are scatter plots of the x predicted residuals against their observed residuals. During both years, 2011 and 2012, the points at the extremes exert lower leverage than points in the middle exert. The parameters of the multiple linear regression models
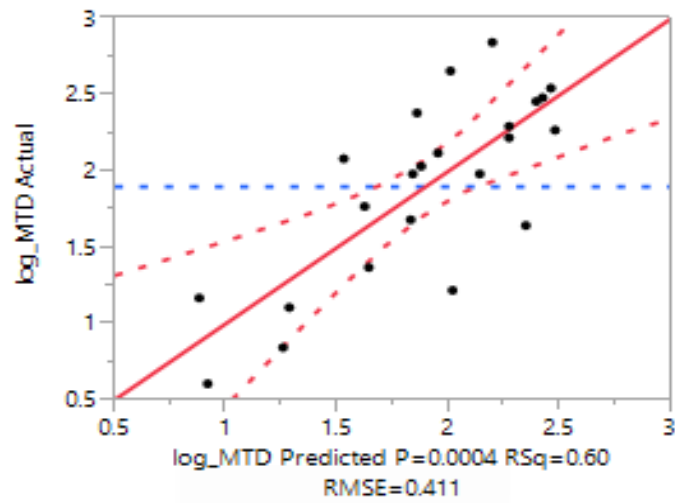
Figure 9. Leverage scatter plot of overall linear regression model performance between actual log(x+1) transformed mosquitos per trap per day and predicted values for 2011 (straight red line: model, dashed red lines: confidence intervals, dashed blue line: midpoint of estimates, RMSE: root mean square error.
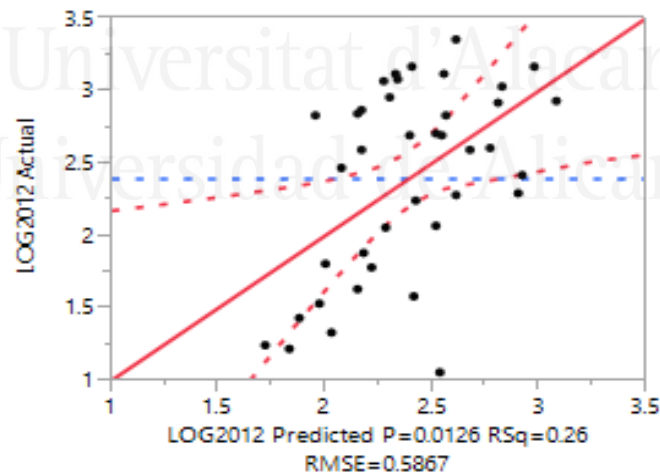
Figure 10. Leverage scatter plot of overall linear regression model performance between actual log(x+1) transformed mosquitos per trap per day and predicted values for 2011 (straight red line: model, dashed red lines: confidence intervals, dashed blue line: midpoint of estimates, RMSE: root mean square error.

testing the associations among environmental variables and mosquito abundance are given in table 2. In general, model performance was higher during 2011 compared to 2012. Nevertheless, during both years (2011 and 2012) the effects of the climate predictor variable and especially temperature, was significant ($p<0.05$). However, model performance was higher for 2011 ($Rsq=0.60$) compared to 2012 ($Rsq=0.26$).

Table 2. Parameter estimates, standard errors (StdError), t Ratio and probability of rejecting the hypothesis the parameter is significant different from zero (Prob>|t|) for the multivariate linear regression models between Culex. sp mosquito abundance (log MTD), and mean temperature, rain and lagged 1 mosquito abundance (logMTD(t-1)) applied for the years 2011 and 2012.

Year	Parameter	Estimate	S.E.	t Ratio	p-value
2011	Intercept	-0.939	0.624	-1.50	0.148
	Temp_M11	0.12	0.033	3.62	**0.001***
	Rain_11	-0.01	0.01	-0.64	0.527
	logMTD(t-1)11	-0.015	0.217	-0.07	0.942
ANOVA					**0.004***
Year	Parameter	Estimate	Std Error	t Ratio	p-value
2012	Intercept	0.698	0.994	0.70	0.487
	Temp_m_12	0.106	0.034	3.12	**0.003***
	Rain_12	-0.2	0.238	-0.84	0.405
	logtMTD(t-1)12	-0.347	0.161	-2.16	**0.037***
ANOVA					**0.012***

* Significant at the p=0.05 level

3.1.4. Continuous wavelet transforms

Figures 11 and 12 shows by terms of a continuous wavelet transform the two variables of consideration for 2011 and 2012, respectively that are expected to be linked according to the analysis so far namely the mosquito abundance and the mean temperatures. The time-scale map shows the correlation between the data variation and the signature of the wavelets used. Thus, the hot colours represent the time point (week) in relation to different periods of the variable (the peaks of the time series) in which they are perfectly synchronized with the mother wavelet for each case (i.e., baseline population peak). It is apparent that a higher synchronization, in respect to different scales (periods), of mosquito abundance with the mother wavelet appears at the end of the time domain (i.e., August to September), whilst the synchronization temperature and the mother wavelet appears at the start of the season (i.e., May to June). The bold black line represents the cone of influence (COI) that delimits the region not influenced by edge effects which may indicate some false periodic effects due to periodicity assumptions of the wavelet transform.

The mean temperatures are a key factor that directly may affect the development and the population dynamics of arthropods such as mosquitos. The continuous power spectrum of the mosquito abundance counts and mean temperatures for 2011 depict the same power, for a period which is close to 1, although appear on a different time scale. During 2012, the wavelet spectra displayed different significant areas (designated by solid dark lines – COI). In particular the mosquito abundance data displayed large significant areas of both the high and low periods (i.e., frequencies), whereas the mean temperature data only on the low period close to 1. These results were to some degree expected considering that the previous correlation analysis and the multivariate regression model showed a poorer influence of temperature on mosquito abundance for 2012 in comparison to

2011. In addition, their location was not consistent and especially for the mosquito abundance.
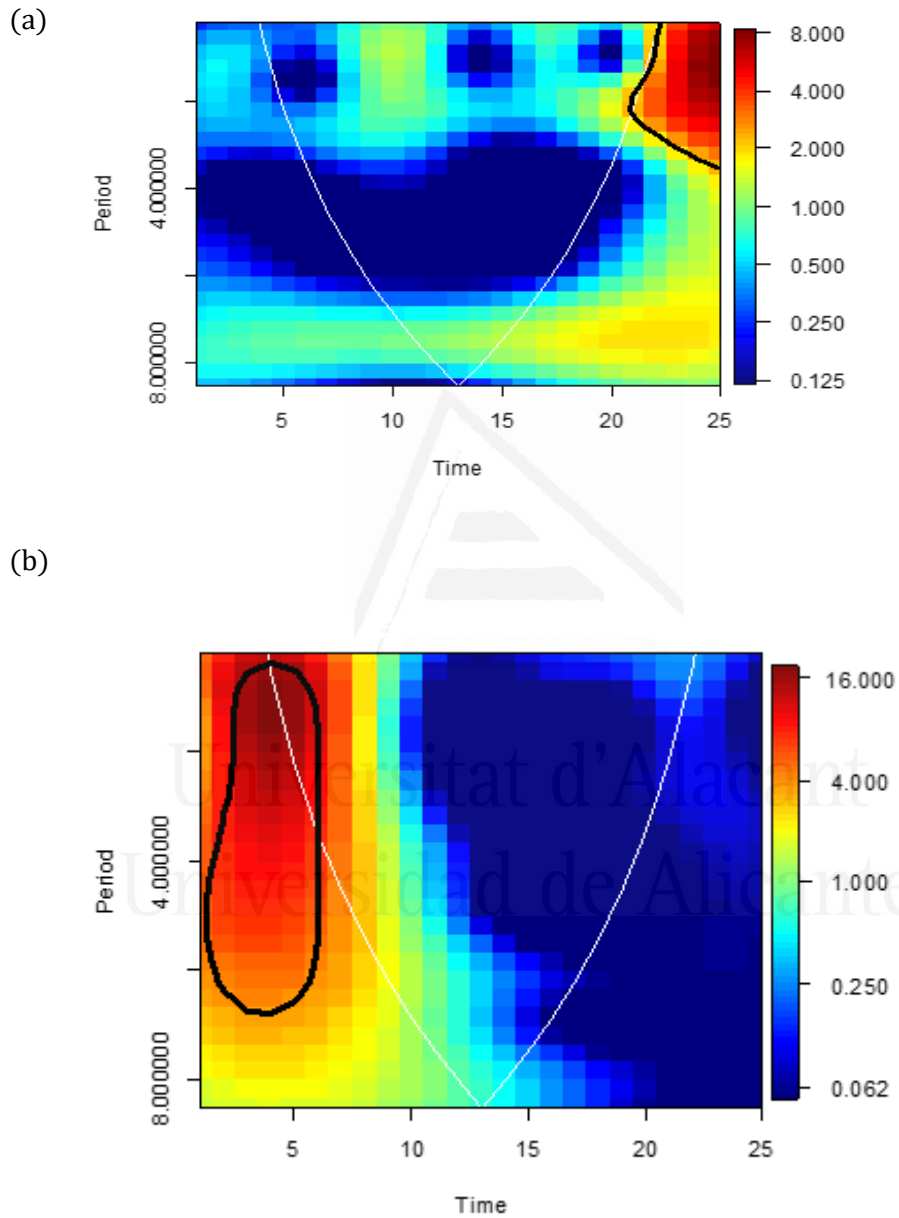
(a)

(b)

Figure 11. Continuous wavelet transforms and power spectrum of log(x+1) MTD (a) and mean temperature (b) for 2011, respective

(a)
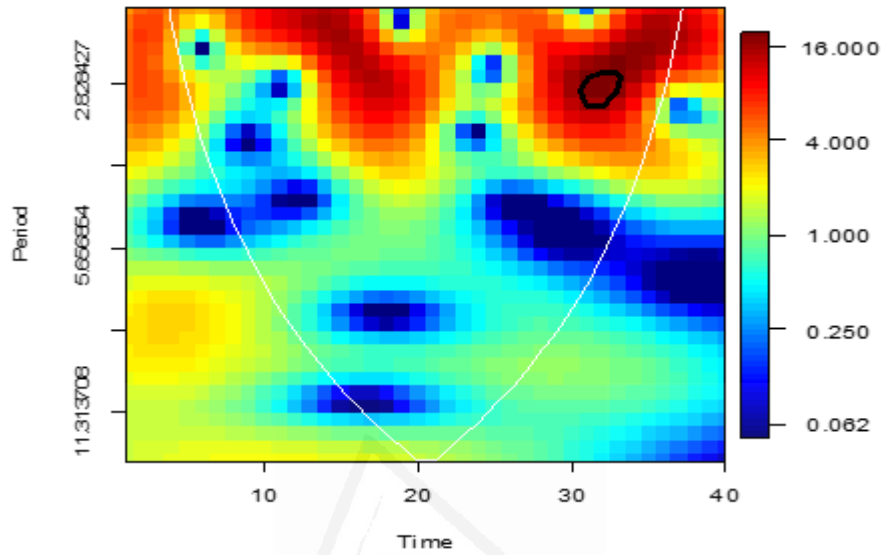
(b)
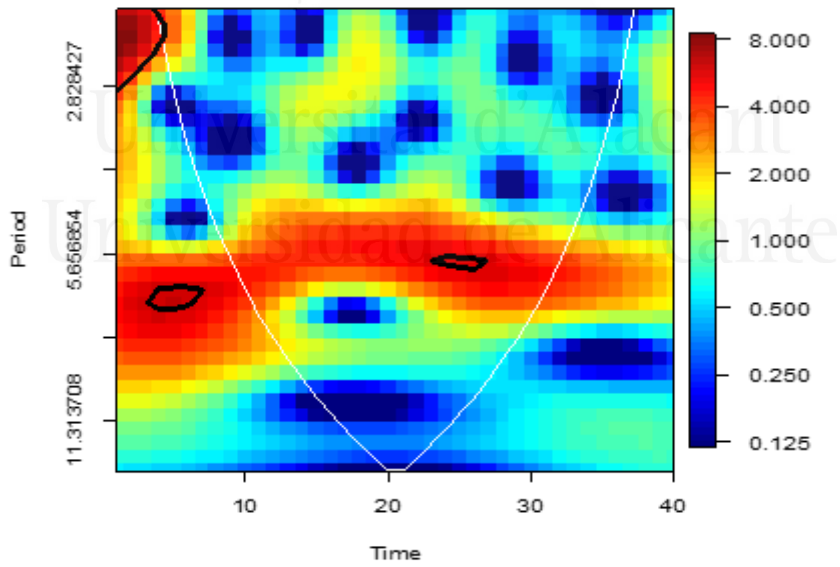
Figure 12. Continuous wavelet transforms and power spectrum of log(x+1) MTD (a) and mean temperature (b) for 2012, respectively.

3.1.5. The cross-wavelet transforms

The cross-wavelet coherence analysis between temperature and adult mosquito population occurrence in respect to weekly sampling intervals (time) are shown in Figure 13. White lines correspond to confidence intervals. In both plots the y-axis represent the time scale, or period measured in weeks, at which mosquito abundance and temperature are coherent (synchronized), while the x-axis represents successive time where counts were performed (note that during 2012 the sampling space was bigger than 2011). Coherence values are given to right of each plot and ranging from zero (blue – cold colours) to one (red – hot colours). The two series (mean temperature and vector mosquito population) share power (i.e., variability - synchronization) in the red-hot regions of the plot. It is apparent that a higher synchronization between mosquito and temperature appears during the start of the season (i.e., May-June) and the end of the season (i.e., September). This synchronisation is apparent for two periods which represent similar oscillations between the peak events of mosquito and temperature per unit of time. Black lines indicate the cone of influence, not affected by edges of the data and are significant coherent (p<0.05).

Moreover, the cross wavelet coherent analysis for 2011 show that inter weekly cycles with a period between 2-3 weeks between mosquito abundance (log MTD) and temperature (mean) were coherent mostly during the first and the last weeks of the season (i.e., May, June and late August) but not during the middle of the summer season (July and early August). During 2012 the cross wavelet coherent analysis showed analogues patterns, although during the early season (late May to early June) there were observed two inter weekly cycles with period 1-2 weeks and 6-7 (Figure 6b). This can be inferred by viewing the y-axis in respect to the x-axis which represents time (in weeks). In particular the y-axis scale indicates the period of the cycles, and the coherence, which is depicted by hot colours, which represent the overlap that are displayed in the mosquito counts and the respective temperatures (Poh et al. 2019). The coherence

can be understood as a special type of correlation estimate, which does not depend upon the normality assumptions of the series. Additionally, higher values point out a very high synchronization between the cycles of the mosquito abundance and the mean temperatures in contrast to values which are close to 0 which represent an independence of the cycles at a given time scale. Moreover, two cones of influence (COI) are apparent during early season of 2012, with periods close to 1 and 7 respectively, representing the region of the wavelet spectrum in which edge effects become important and are defined here as the e-folding time for the autocorrelation of wavelet power at each scale.

(a)

(b)

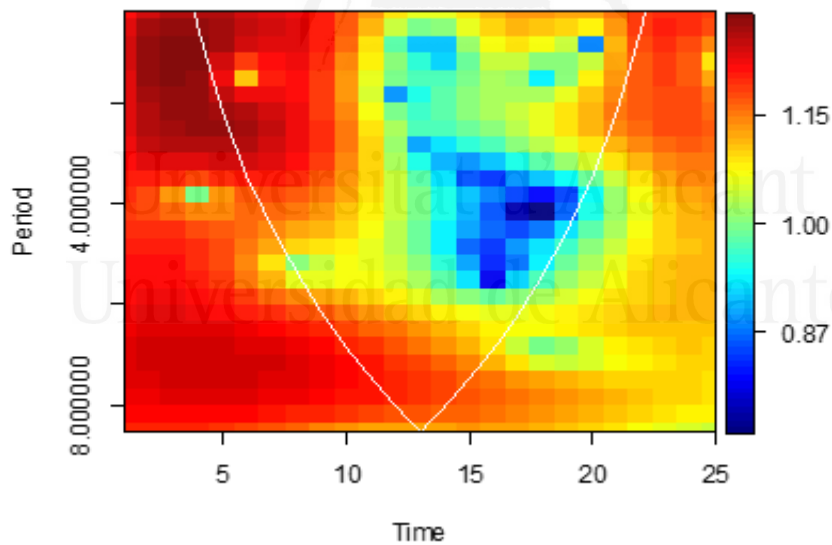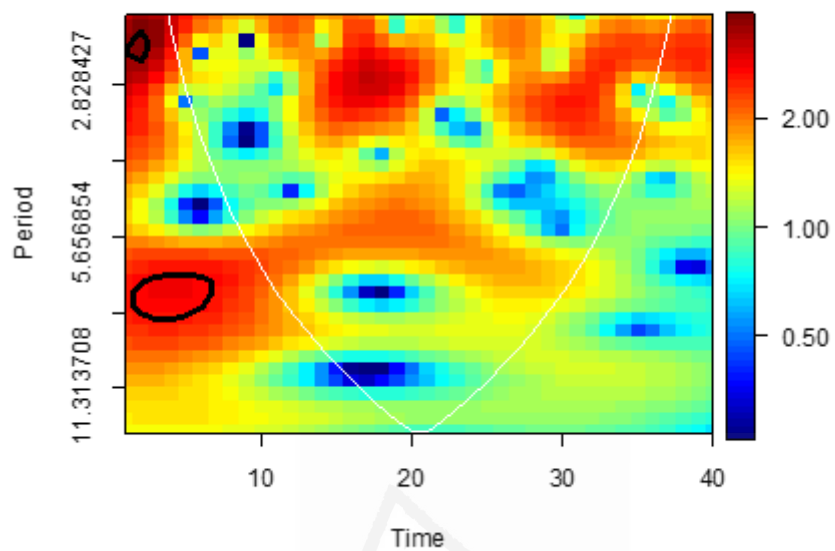Figure 13. Cross-wavelet coherence analysis between temperature and adult mosquito population occurrence in respect to weekly sampling intervals (time).

3.2. A climate conditioned Markov Chain model

3.2.1 Mosquito population dynamics and cross correlation with whether variables

Successive mosquito catches through the observation period and meteorological data are presented in Figure 14. These data correspond to the standardized weekly counts of *Culex* mosquito species (MTD), as well as the mean temperatures (^{o}C) and rainfall incidence (mm) recorded in two successive seasons (2011 and 2012) of typical mosquito activity (May through September) under a typical temperate climate in Northern Greece. As presented in Figure 1(a) there are similarities in the time evolution of the three variables, but it is not easy to evaluate the degree to which this occurs. Nevertheless, based on the cross-correlation analysis, we determine that it is preferable to use temperature to condition the abundance of arthropod vectors, as it has a much stronger correlation than the impact of rain (Figure 15). For the cross-correlation analysis all the available data were used in a combined way to detect inherent correlations and increase the validity of the results.

There is a positive correlation for temperature and that the cross-correlation at time lag zero is highest for temperature. Hence, it is best to condition on temperature at the moment itself and not temperature a week earlier or a week later. The cross-correlation at time lag zero is highest for temperature (i.e., Mosquitos are not lagging temperature), and there is a positive correlation for temperature and a negative relationship for rain.

This is in consistence with previous study analysis (section 3.1), suggesting that higher temperatures result in more adult mosquito populations, while the opposite occurs for rain which probably affects flight activity.
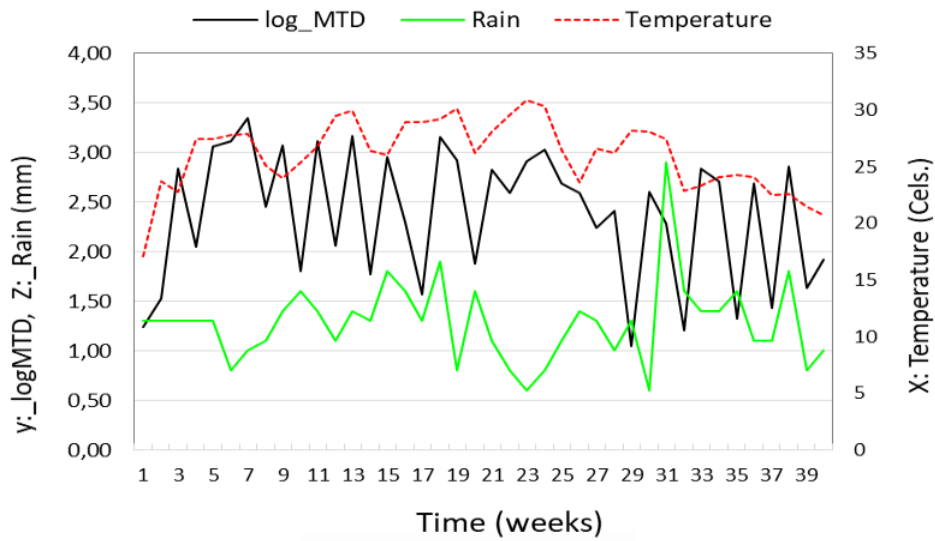
Figure 14. Observed time series of mosquito abundance (Y: log_MTD), rain levels (Z: Rain) and mean temperature (X: Temperature (^{o}C). The data correspond to two consecutive periods of mosquito activity (combined periods of 2011 and 2012).
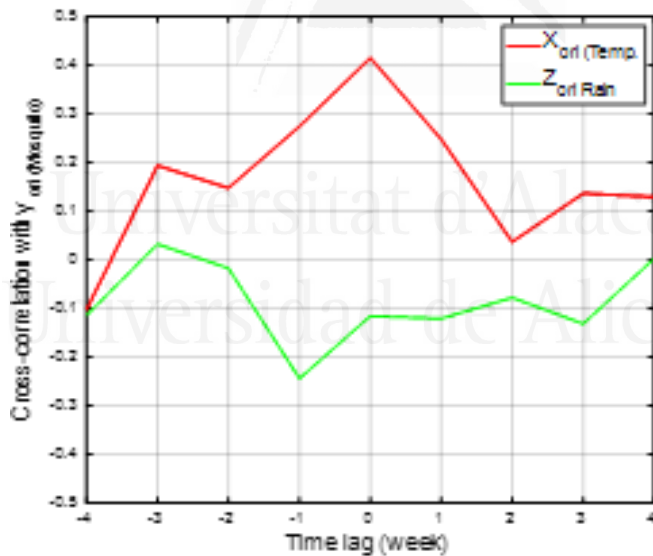
Figure 15. Cross correlation between mosquito abundance and Temperature (Red line) and between mosquito abundance and Rain (Green line) as function of time lag (weeks).

Based on the time lagged correlation analysis, we derive that it is best to condition on temperature at the moment itself and not temperature a week earlier or a week later. However, we cannot exclude the possibility that in reality mosquitos could lag the temperature for shorter time intervals (i.e., by maybe one single day), but this cannot be detected since the data consist of averages over a week.

3.2.2. Data partitioning, population transition networks and model training

After using k-means clustering algorithm, the k cluster centers have been used for positioning the data and considering first a case of two states (high and low levels of population) and then five states (very low, low, intermediate, high and very high population level). The same process has been applied to temperature data. First, I consider a case of two states (high and low temperature levels) and then five states (very low, low, intermediate, high and very high temperature level).

Moreover, because we expect that transition probabilities may affected by the length of the input data, we have performed a preliminary analysis, using different lengths of input variables, to find the point from which the size of the sequence does not differentiate the transition probabilities of the mosquito population. To date, for a sequence length>35 weeks the model parameters do not differ considerable according to their informational entropy content and the system explicit a random behavior rather than deterministic suggesting that to a high degree an exact underlying transition matrix exists.

Figure 16 depicts a directed graph (or trained network) which represents the actual transition matrix of the mosquito abundance systems. The criterion of determining how many states to use in the Markov chain depends on both the characteristics of the time series data (range) and the criterion used to limit the classes (e.g., k-means) as previous described. In

this work I have decided to use first the simplest case, having two clusters as well as the example of having five states based on the data range distribution and k-means clustering statistics. Thus, the transition matrix and the relative graph shown were built for a vector population system with two states (high and low) and five stages (very low, low, intermediate, high and very high) (Figure 16a and b, respectively).

(**a**)   (**b**)
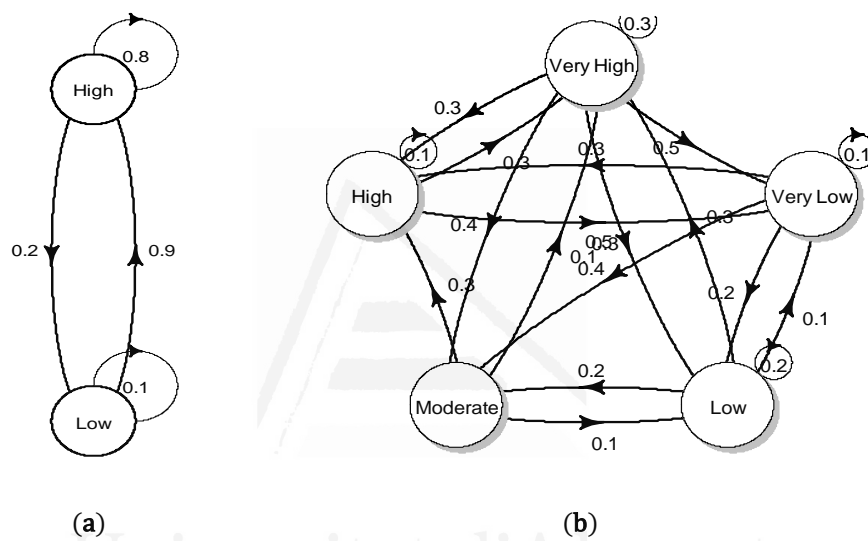
Figure 16. Transition matrix of the mosquito abundance system illustrated in terms of a directed graph (or network). (a) The transition matrix has been constructed for two states (high and low) and (b) for five stages (very low, low, intermediate, high and very high).

The values that are shown represent the probability of transition from one state to another in the form of an arrow. States are represented as vertices (or nodes) whereas transitions are represented as directed edges (or links). This representation scheme allows the population system of mosquitoes to change from state 1 (e.g.: node i – low population) to state 2 (e.g., node j high population) along the k edges of the graph through a path of length k from i to j. For the five states system, for instance, the transition probabilities show that if for a week the mosquito abundance is high (state 4), there is a 10% chance that it

will remain at the same level the next week (state 4), 30% chance that it will be at a very high level (state 5), a 30% chance that it will be move at a very low level (state 1) and no chance to move to any of the other population levels (state 2 and 3) (Figure 2b). Yet, it should be noted that the zero probabilities to move to state 2 or state 3 could be an indication that there is some uncertainty of the estimation process related to the particular data set used, since if more data were available these transition probabilities would be observed with a low probability instead of zero.

3.2.3 Markov chain model realizations of mosquito population levels

Utilizing the data from the transition matrices, Matlab simulations were executed based on the two and five states Markov chain models, respectively (Figure 17 and 18). The charts show the weekly sequence of the mosquito population levels with two and five states, where the stochastic process X (t) remains in the same state, or, moves from one state to another, depending on the probabilities of the transition matrix. At each discrete moment of the Markov process or, using ecological terms - after each week of observation, a mosquito control decision will be made depending on the predicted population levels. The basis of decision will be the prediction based on how the Markov chain evolves based on the values in the transition matrix P. The division into five population states, compared with only two, results in sections of the series where the X(t) process presents large deviations from the probability range of values as it evolves over time. By making predictions over time, in order to make decision-making actions, it is our intention to capture the time point where the probability of a high population increases and to avoid any action if the prediction shows that the probability value of the observed variable will decrease. Otherwise, if the

71

forecast indicates that the observed arthropod vector population variable will remain in its present state over a longer period (i.e., low or moderate population level), it is of little practical interest as it is difficult in that situation to decide to undertake control action.
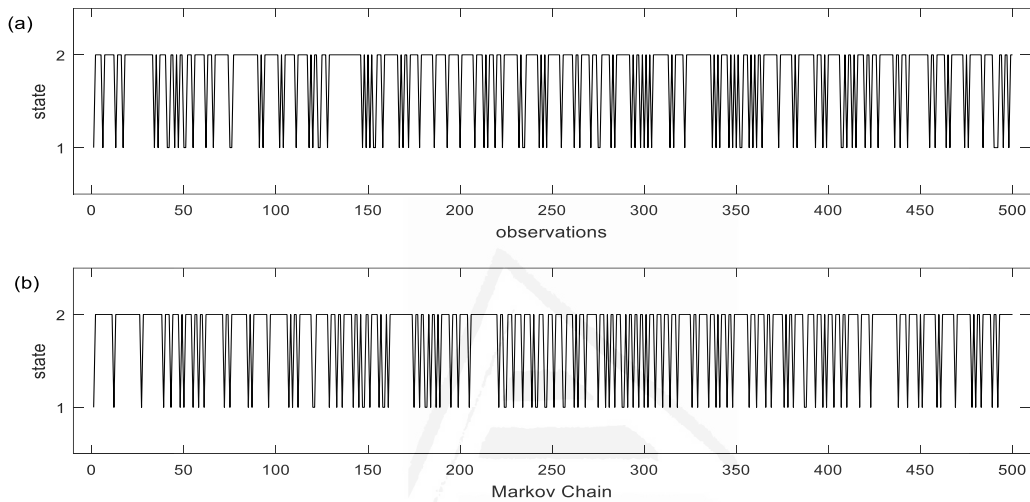
Figure 17. (a) Long term sequence of observation of the mosquito population process having two states, (b): realization of a Markov Chain trained model on these observations.
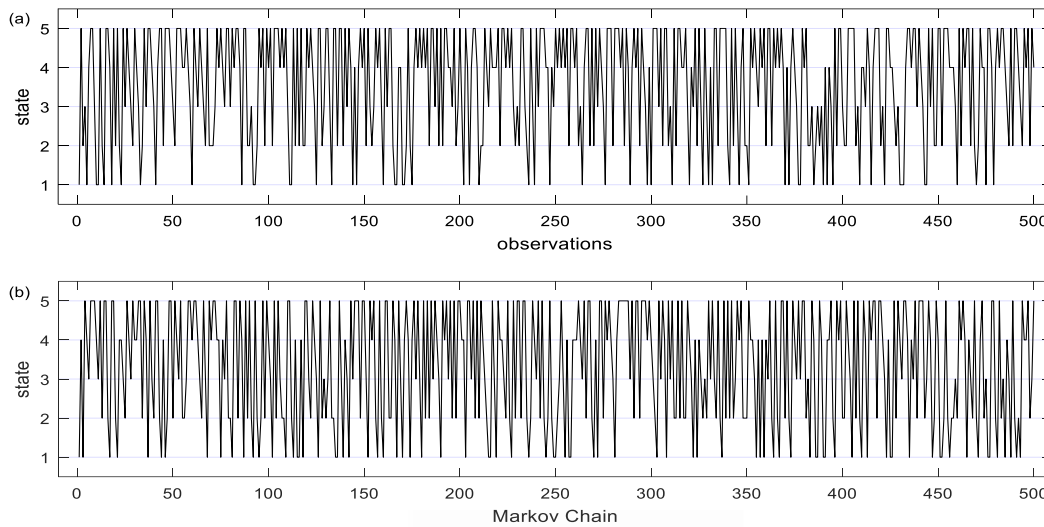
Figure 18. (a) Long term sequence of observation of the mosquito population process having five states, (b): realization of a Markov Chain trained model on these observations.

3.2.4. Conditional Markov chain model realizations of mosquito population levels

Taking into account the temperature-dependent Markov chain, the situation is different from the previous example because an additional variable is known from the model which can improve simulation performance and population prediction efficiency (Figure 19). The different states are; State 1: very low mosquito population or temperature levels, state 2: low mosquito population or temperature levels, state 3: moderate mosquito population or temperature levels, state 4: high mosquito population or temperature levels, state 5: very high mosquito population or temperature levels). Each time period corresponds to an observation for a finite number of successive seasons. The initial state at time 0 corresponds to the first observation made. Due to the relatively limited data set and to avoid less accurate transition probabilities the CMC model in this study was trained and tested on the same two season's data. The actual observations that have

73

been used to run the model (Figure 19a) is very close to the realizations generated by the conditional Markov Chain model which was trained on these observations (Figure 19c) in contrast to the realization of the non-conditioned of the Markov Chain model (Figure 19b). Again, the prediction is built solely on empirical data from the past and never from the future which is very important to decision-making. Say for example, that at a given time point, but before making any control action decision, we know the states of the process in the two preceding moments, then we can judge to implement an action against the vector if the population level prediction is high. In summary the simulation results with the conditional Markov chain model we came up with more promising results than those with the simple Markov chain.
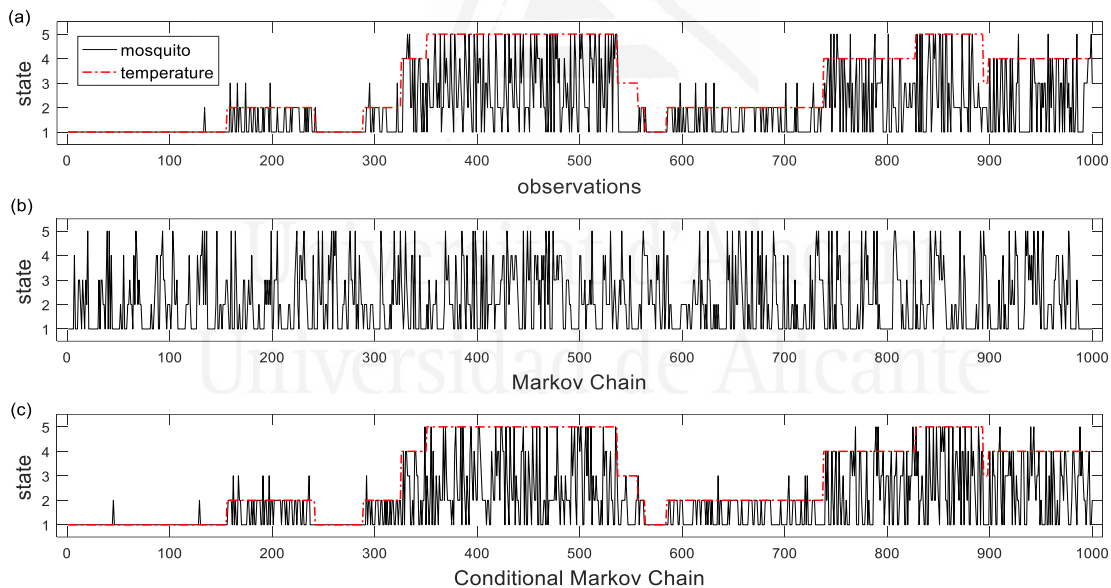
Figure 19. (a) Long term sequence of observation of the mosquito population process having five states (straight black line) and the process of a large-scale temperature having five states (red dash – dotted line), respectively and acting as background variable (b): realization of a Markov Chain trained model on these observations and (c) realization of a conditional Markov Chain trained model on these observations.

3.2.5 Validation of Markov chain models and homogeneity

Figure 20a and 20b depicts the sequence of realizations generated from the trained MC and MCM models and that of the empirical intensity transition matrix, respectively and which were created using data that were not used for MC model training. The MC model and the empirical realization follow a similar pattern although there are also slight deviations in some time points.

(**a**)
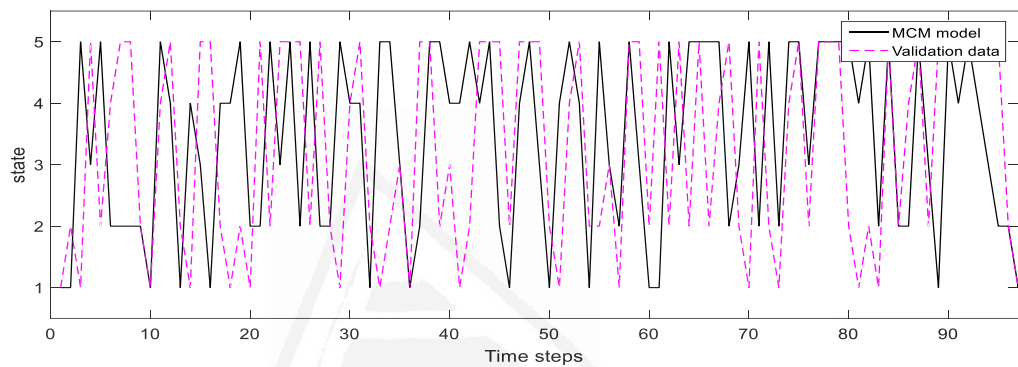
(**b**)
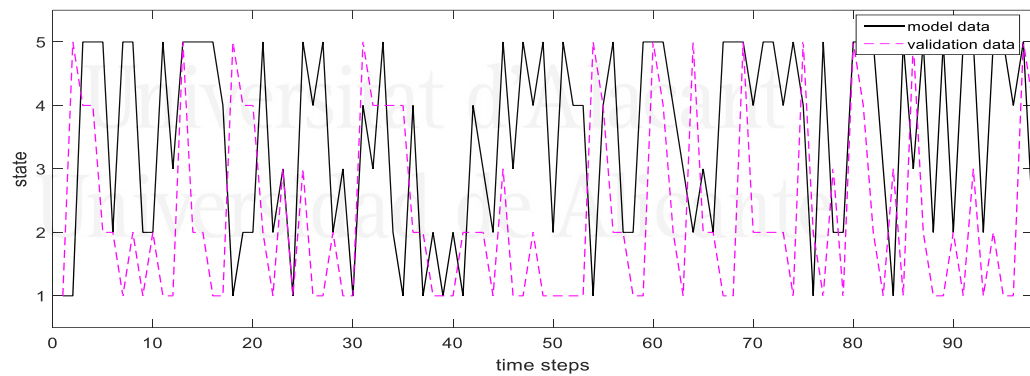
Figure 20. Medium term sequence of observations of the mosquito population process having five states generated from the trained model (black strait line) and that of the empirical intensity transition matrix. (a) Markov chain model and (b) conditional Markov chain model (pink dashed line) which were resembled using data that were not used for the model training.

These deviances could be justified by the fact that the amount of data available for model evaluation was relatively small. Nevertheless, over all

the general model patterns fit well to that of the observed probabilities and this is in accordance with the results of the Chi-square test which test the null hypothesis that both realizations are homogenous, that is, they come from the same matrix of transition probabilities. Particularly, there were no significant differences between the transition matrix of the MC model and the transition matrix used for validation (ChiSq=18.736, df=24, p=0.765). Additionally, no significant differences were observed between the CMC model and the validation matrix (ChiSq=14.58, df=24, p=0.932).

3.2.6. Limiting probabilities of Markov chain models and stationary distribution.

The convergence of the three Markov chain (MC, CMC models and observed transitions) are illustrated in Figure 21a, b and c, respectively. Our findings match to certain degree the empirical results in which the probabilities follow analogous patterns. For instance, by looking at the plot we observe that all probabilities convergence fast and the final probabilities are analogous to the stationary distribution. Yet, the chain which was developed using the empirical data shows slight deviations and especially for low population levels, although they finally converge with a consistent rate.

Figure 21d illustrates the similarities of the time evolution of the entropy, which is related to the average probability of the MC, CMC and the validation data, starting from the initial probabilities of mosquito population levels towards equilibrium (e.g., stationary distribution for each case). The Markov chain models have shown very similar convergence patterns towards equilibrium.

(a)   (b)
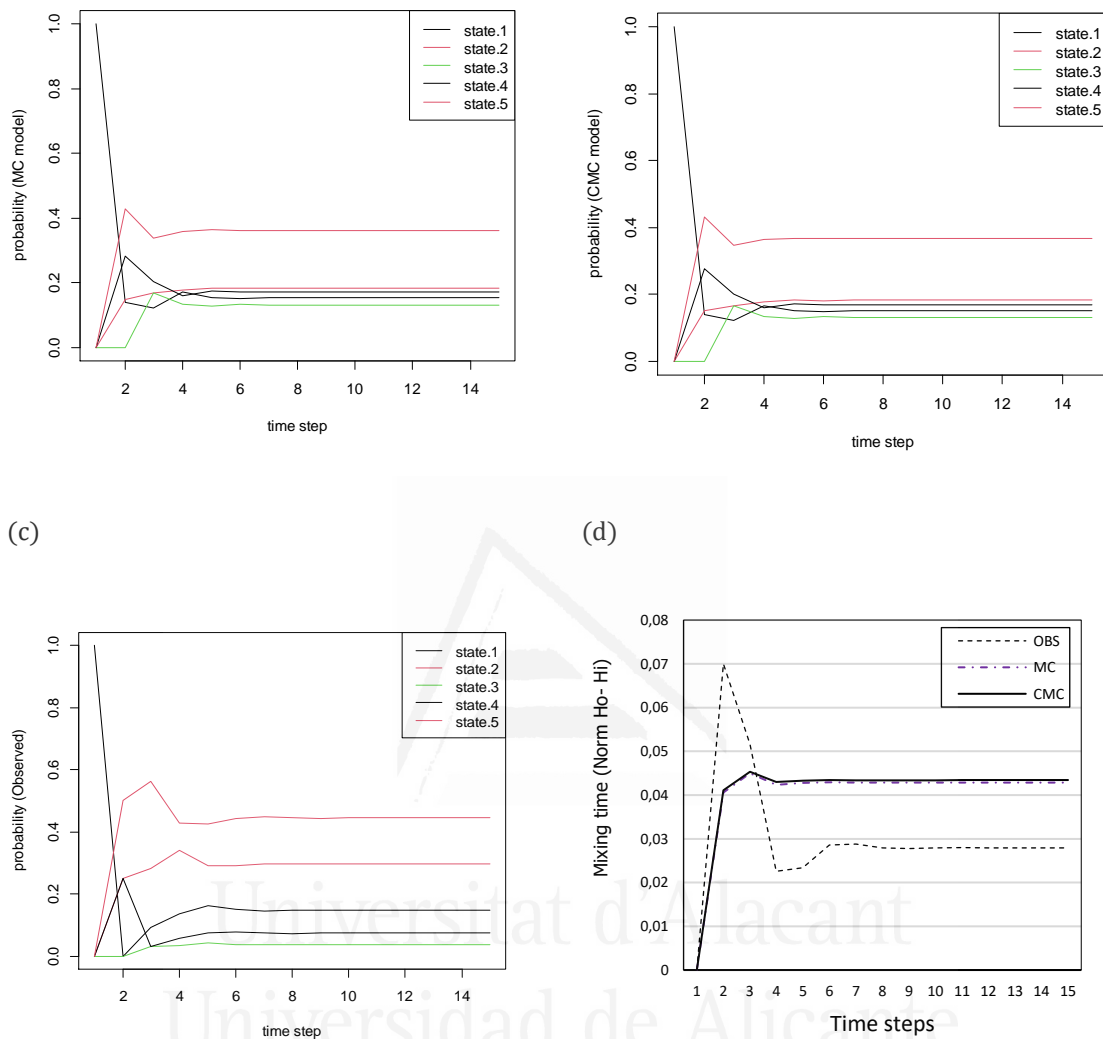
Figure 21. Limiting probabilities of a 5 states Markov chain according to the MC model (a), the CMC model (b) and the empirical MC of the observed validation data (c) as well as the mixing times towards the steady-state probability (d). The steady state represents the equilibrium distribution when the mosquito population dynamics is considered as ergodic process.

A different representation of the conditional data driven Markov chain transition matrix (e.g., direct graphic) used to project the temporal development of the mosquito population system is shown in Figure 22. This illustration indicates of how and to what extent the system evolves after

77

many steps (probabilities are not indicated). Based on this representation, we conclude that it is more likely for the system to go into a state of high population level, when the former is a state of a low population level. On the other hand, there is a lower probability to remain in the same state of low or even high population pressure, especially in the case of the five-state transition matrix. So, the transition matrix can be the fundament for the decision on a control method in a given moment – either a high or moderate arthropod vector population level, depending on the probability of occurrence of a state lying below or above the current level.
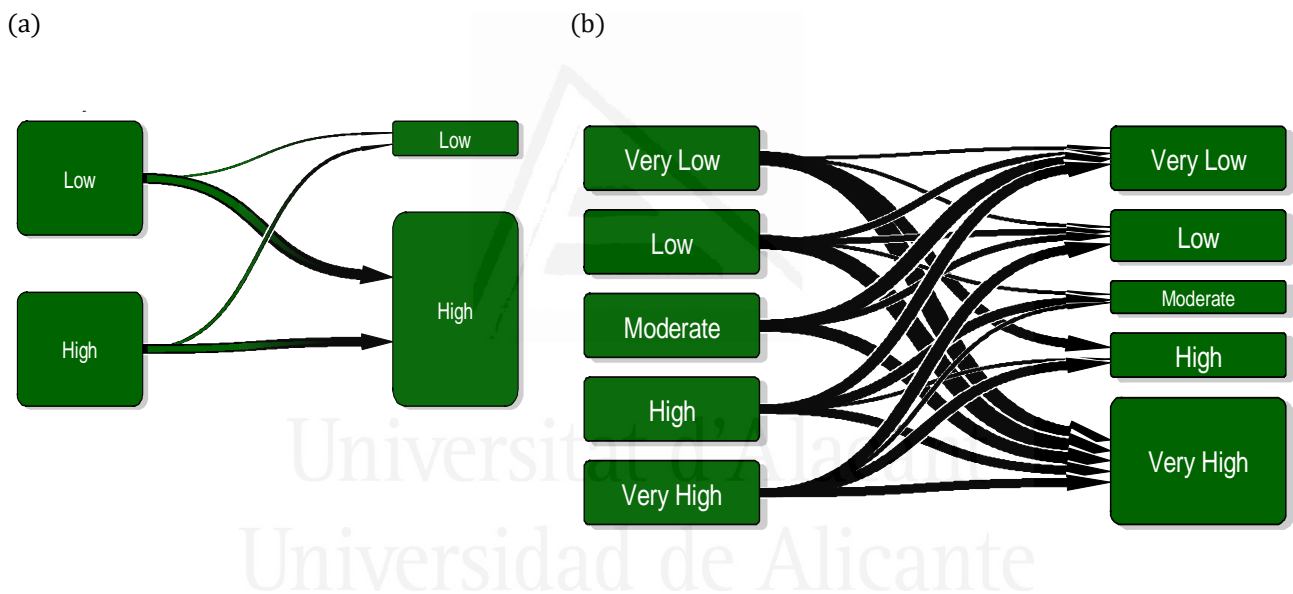
Figure 22. Mosquito Population Transition Network with two states (a) and five states (b) depicting state transition probabilities indicated by different arrows (thicker arrows indicate greater probability of transitions).

3.3. Soft computing of the dynamics of a medically important mosquito vector using recurrent and dynamic artificial neural networks

3.3.1. Network architecture

The NARX neural network is a nonlinear auto-regressive model with exogenous Inputs. Figure 23 is a graphical illustration of a NARX network, in a parallel identification mode, with *du* input and dy output delays. The NARX neural network structure has an input layer which consists of the mosquito abundance counts and the temperature recording counts, which are connected through the weight matrix to each of the 10 neurons which consist of the hidden layer. The model has been generated for two input delays of 1and 2 weeks respectively, for each of the two variables (mosquito counts and mean temperatures). The results of the hidden layer are linked through the summation function in the output layer.

An abbreviated dynamic model structure, in a parallel mode, of the overall NARX neural network for the input layer (a) and the output layer (b) according to the Mat Lab Simulink ANN system model construction process is shown in Figure 24. In this structure the network simulation data (the input of the model) consist of 2 concurrent vectors: $p1 = [12]$ and $p2=[21]$, where p1, is the mosquito abundance vector which corresponds to weekly counts of *Culex sp.* adult stages and p2 is the respective mean temperature vector. The FTD neural network architecture has the same topology with NARX model but without the lagged mosquito input variable and therefore it consists of a feedforward network with a tapped delay line at the input. is set dy=0, then the NARX network is reduced to a plain FTD neural network architecture.

The model is applied to predict the population (a{2}) of a medical important mosquito species (*Culex. sp*), from previous temperature recording values (Delays 1) of exogenous inputs (p{1}) and previous (Delays 2) mosquito

population values (p{2}). Each element of the input and output network is
connected to each neuron through a weighted matrix (W).
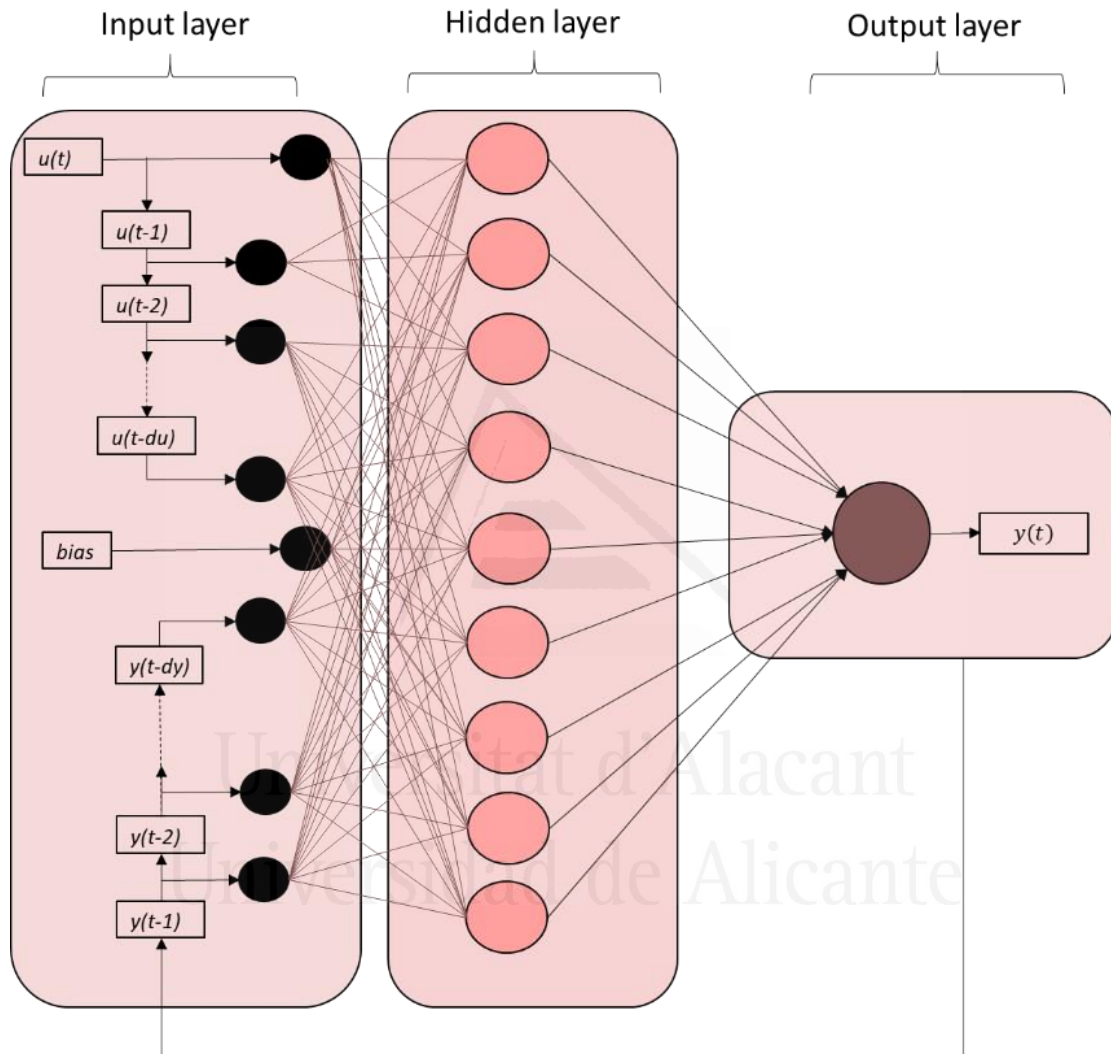
Figure 23. Graphical illustration of the NARX network with du input and dy output
memory and a number of neurons in the hidden layer. Note that if the output memory

Elements of layer 1, such as its bias (b{1}), net input, and output have a
superscript 1 to indicate that they are associated with the first layer, while
those of layer 2 have superscript 2. The FTD neural network has the same

topology without the p{2} mosquito input variable as well the related delays and weights (netsum: transfer potential θ, tansig: hyperbolic tangent sigmoid transfer function, purelin: pure linear function transfer function).

(a)

(b)
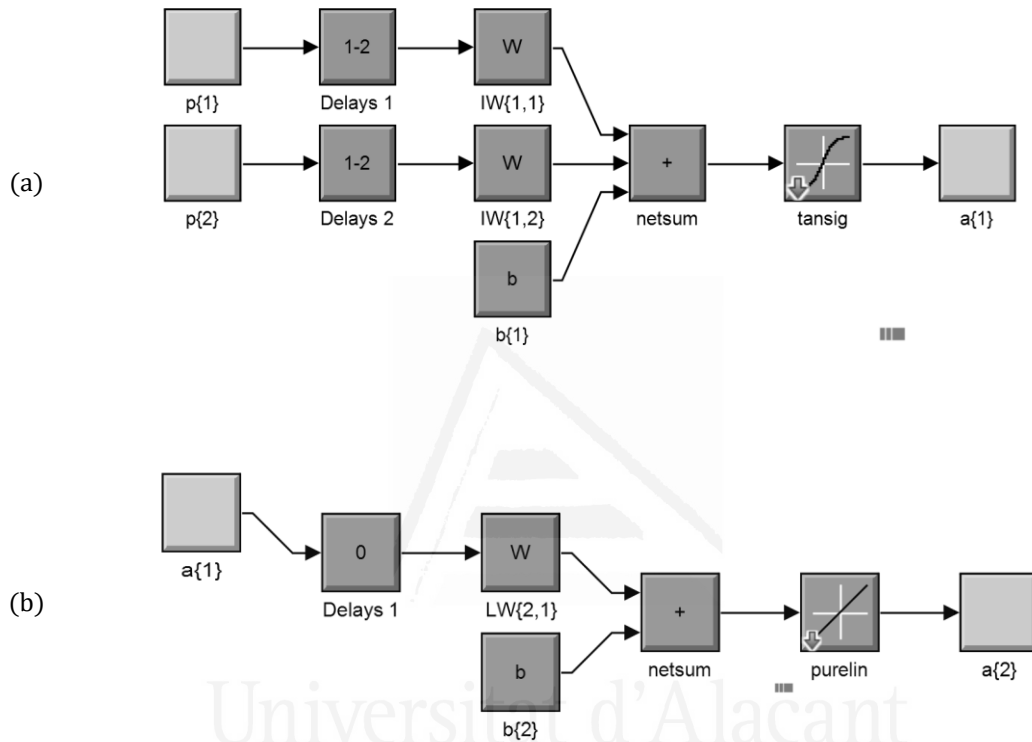
Figure 24. Abbreviated dynamic model structure, in a parallel mode, of the overall NARX network for the input layer (a) and the output layer (b) according to the Mat Lab Simulink ANN system model construction process (details in text).

3.3.2. Model training and validation

Figure 25a and Figure 25b show the variation of the *mse* of the training, validation and test data in respect to a successive number of iterations (epochs), for the NARX and the FDR neural network models, respectively. The three curves have a similar overall trend but the train data. Moreover, it can be seen that training and validation errors for the NARX model decrease until the highlighted epoch and the best validation performance

81

state is in 0.388 at epoch 3 in which the *mse* is minimized. Additionally, considering that validation error does not increase before this epoch informs that overfitting has not occurred. The *mse* of the test data has a similar pattern and is minimized after 4 iterations and remained stationary after that point, which indicates that the model had reached its optimal state. However, the best states for the train data occurring after 3-time steps (epochs), at which the *mse* of the test data is gradual minimized.

Figures 25 and 26 shows model performance in terms of regressions between the output and the target data sets (i.e., training, validation, testing and overall) for the NARX (Figure 25) and the FTD model (Figure 26). In most cases the model performs well considering that the data are in the in the vicinity of the diagonal. The correlation coefficient was at acceptable levels in both cases and in respect to the available data set (R=0.623 and R=0.534 for the NARX and FTD models, respectively). Moreover, considering the non-linear and abrupt nature of the mosquito data the overall model predictions are in acceptable levels when comparted to the actual abundance data.

(a)   (b)
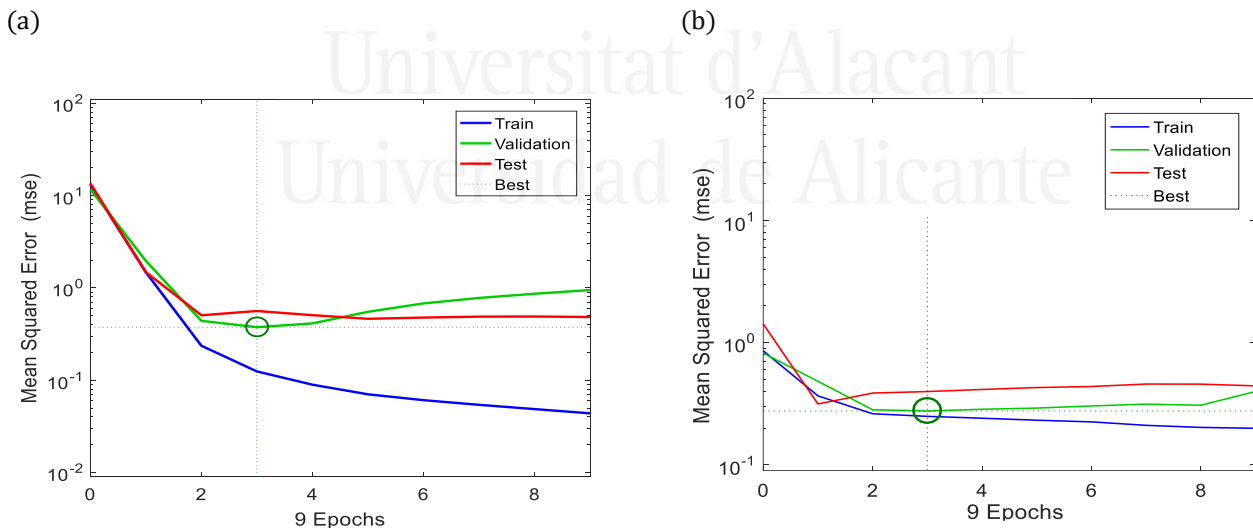
Figure 25. NARX (a) and FTD (b) neural network training, validation and testing performance. Note that the best validation performance for the NARX model is 0.388 at epoch 3 and for the FDR is 0.276 at epoch 3.

82

In addition, it should be mentioned that the model performance is considerable higher by taking inn to account only the training data (i.e., r=0.8 and r=0.62, for the NARX and FDR models, respectively) and that the final overall model performance values is affected by the lower validation and model testing performances. Thus one can expect that the model performance could be considerably improved if the test data set size was higher. However, to make the network model more efficient I have decided to keep a larger data set to be preprocessed for training despite the smaller returns showed for the testing and validation performances.
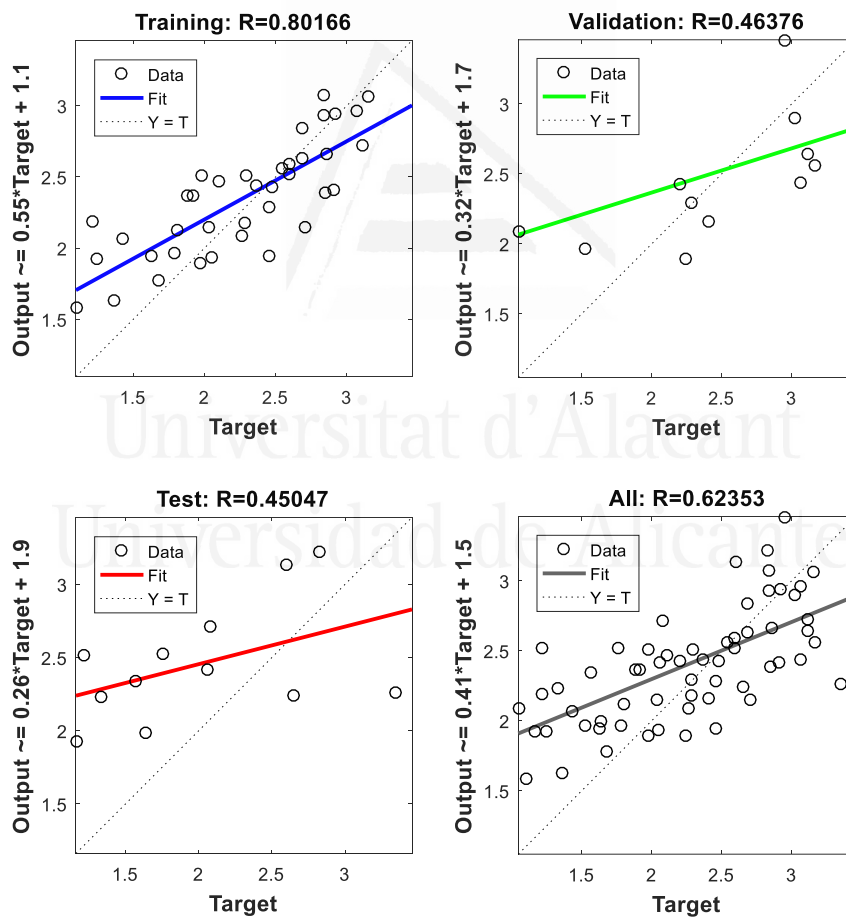
Figure 26. NARX neural network training, validation and testing performance.
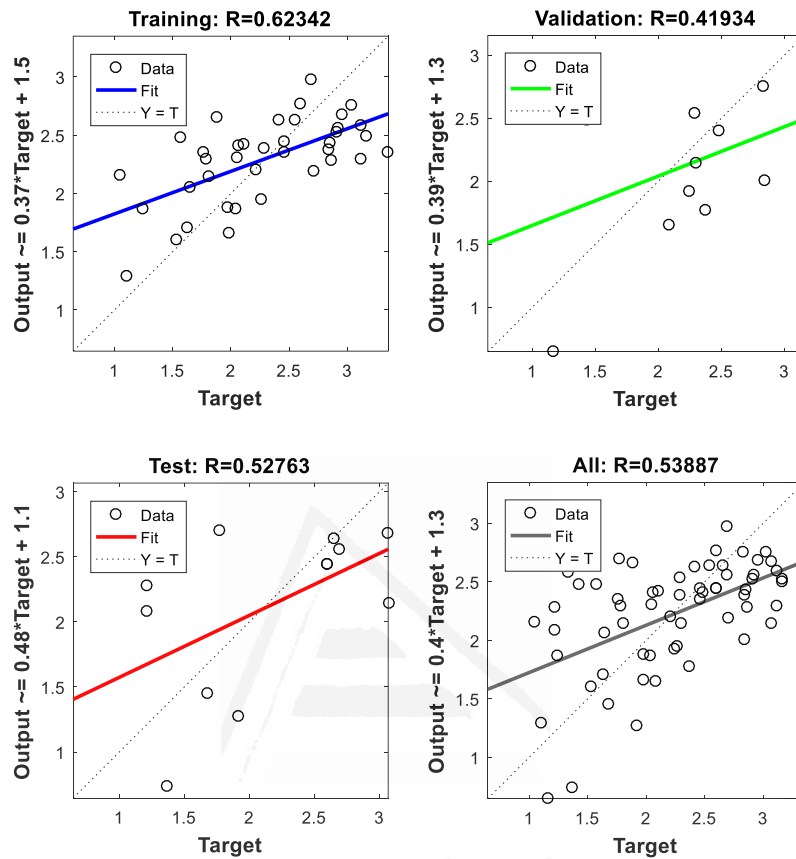
Figure 27. NARX neural network training, validation and testing performance.

3.3.3 Overall model performances

Figure 28a and Figure 28b show the variation of the *mse* of the training, validation and test data. The time scale corresponds to weekly time intervals (from mid-May till September). In general, the prediction-output trend performs well although there are time steps where the prediction results are not ideal and the reason for that is that the amount of available data is relatively small. However, for the first model (NARX model) there are some cases which show high values with low target values and thus positive bias,

while in the second (FTD model) there are few low output values with high target values suggesting a negative bias. Nevertheless, in most cases the deviations during certain time steps are in the range of -1.4, 1.3 which is relatively low and the distribution is around zero. Moreover, the overall frequency of the error term is shown in Figure 29 which is an error histogram chart having 20bins.

(a)

(b)
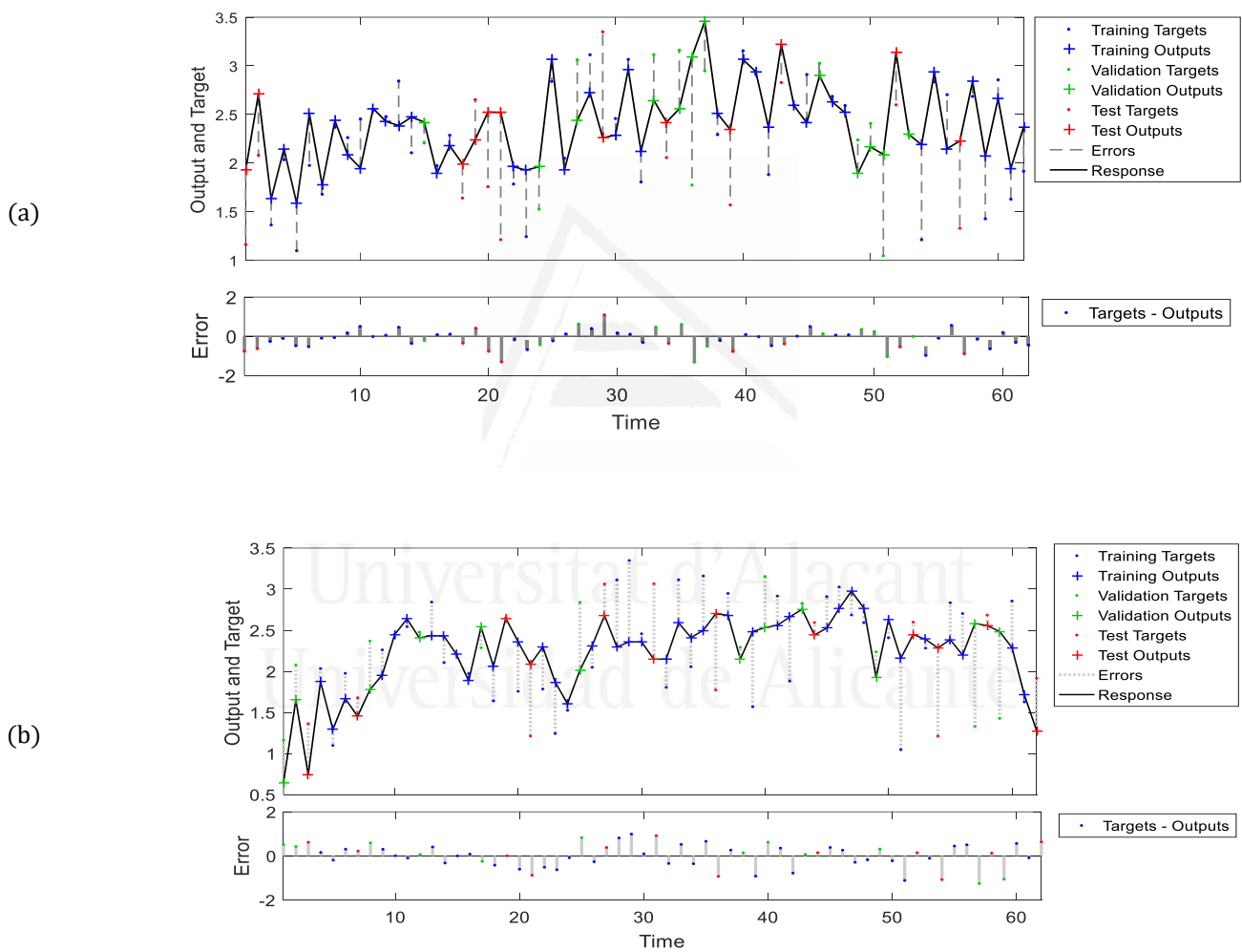
Figure 28. Response of the NARX (a) and FTD (b) neural network model output to the mosquito population time series and error. The model training is performed in open loop (i.e., parallel series architecture), including the validation and testing step and later on, after training it is transformed to a closed loop for the multistep-ahead prediction.

85

The number of samples from each data set is represented by a vertical bar. The error of the NARX neural network ranges from -1.2 (leftmost bin) to 1.03 (rightmost bin) while the error of the FTD neural network ranges from -1.1 (leftmost bin) to 0.9 (rightmost bin). For both models and especially for the NARX model the vast majority of the training outputs have smaller error and slight between -0.4 and 0.4. This is due to the fact that the set used for training contained more data (i.e., 60% of data) than the validation and test data sets.

Figure 30 shows the autocorrelation function of error 1 for the NARX (Figure 30a) and the FTD (Figure 30b) model, respectively, in relation to different time lags and related confidence limits. At zero lag the autocorrelation equals the *mse*, while for the succeeding lagged autocorrelations, the correlation coefficient not exceed the upper and lower confidence intervals but for some cases. This means that the most of the lagged self – correlated values, for both models, are small and in acceptable levels considering that values lagged from zero till 15 (weeks) are between the upper and the lower confidence intervals.

Figure 31 presents the response outputs of the best fitted model, namely the NARX model, to the *Culex sp.* population time series as well as the observed data, respectively. In general, the prediction-output performed well in both cases, although there were parts were the output results performed less well and especially during the end of the season.

To a high degree this should be addressed to the particular dataset that was available and the fact that a limited data set was used for training. Nevertheless, considering that mosquito population dynamics appeared quite abrupt, characterized by non-linear alterations, given the limited data set, the overall model predictions are in acceptable levels for both models. Moreover, the inclusion of temperature as exogenous factor improved considerable the NARX model performance and the predicted data follow to

a high degree the observations. Note that both, models and data, represent actual mosquito population data.

(a)

(b)
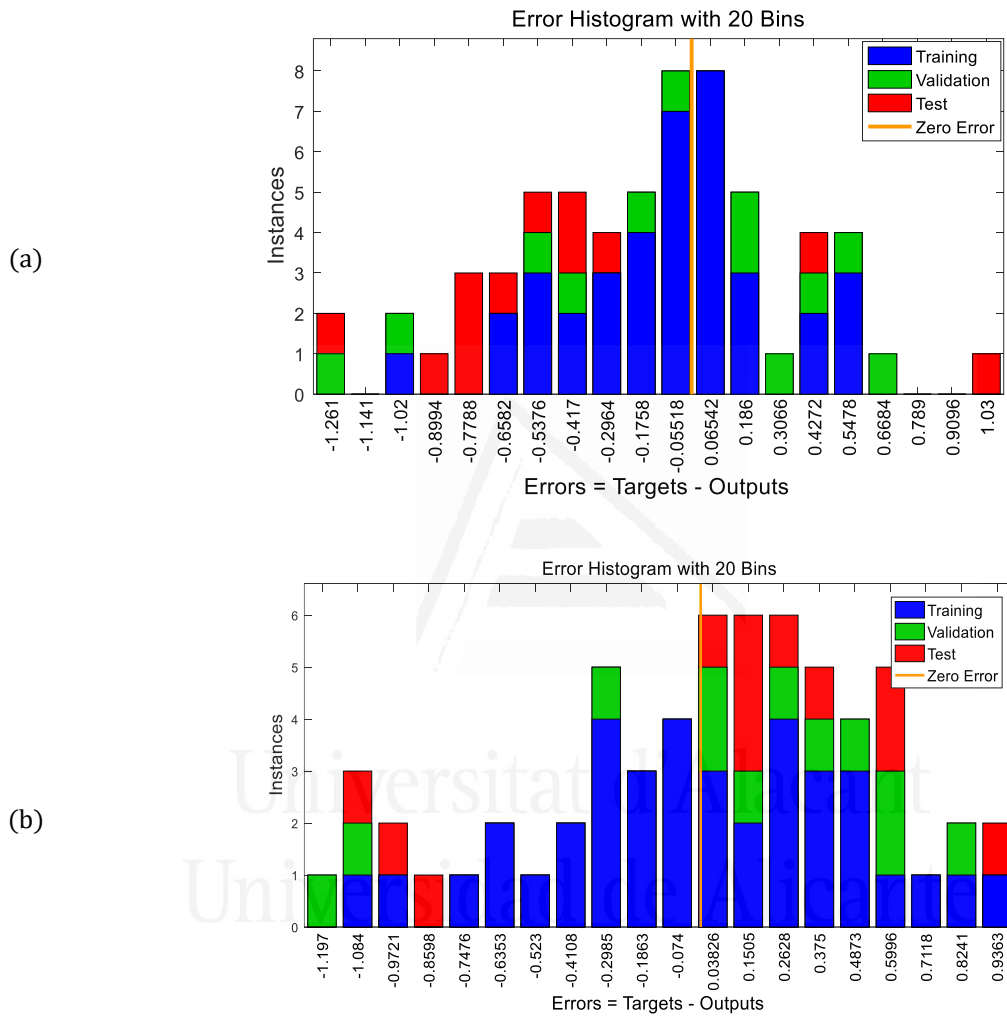
Figure 29. Error histogram chart having 20bins for the NARX (a) and the FTD (b) neural network.
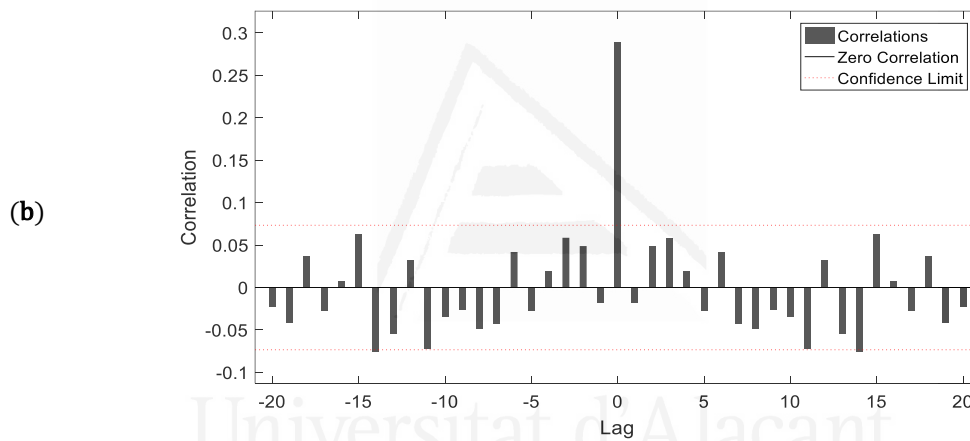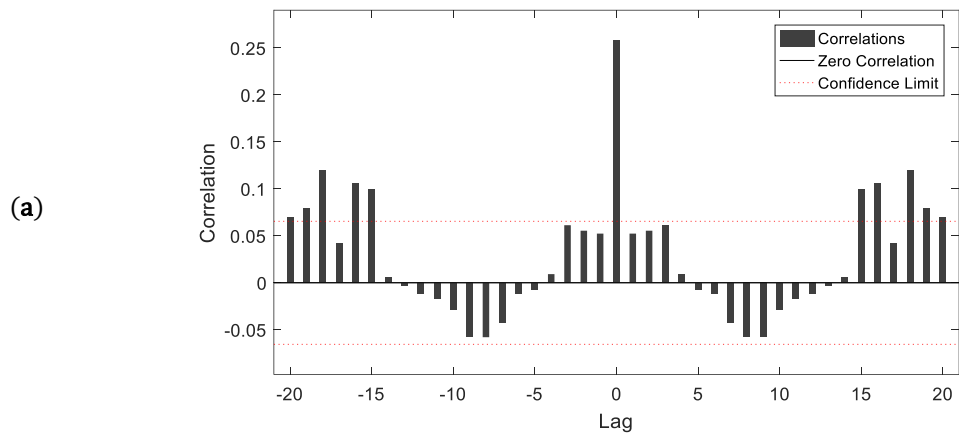
Figure 30. Autocorrelation values of the NARX (a) and the FTD (b) model in respect to different time lags and related confidence limits.
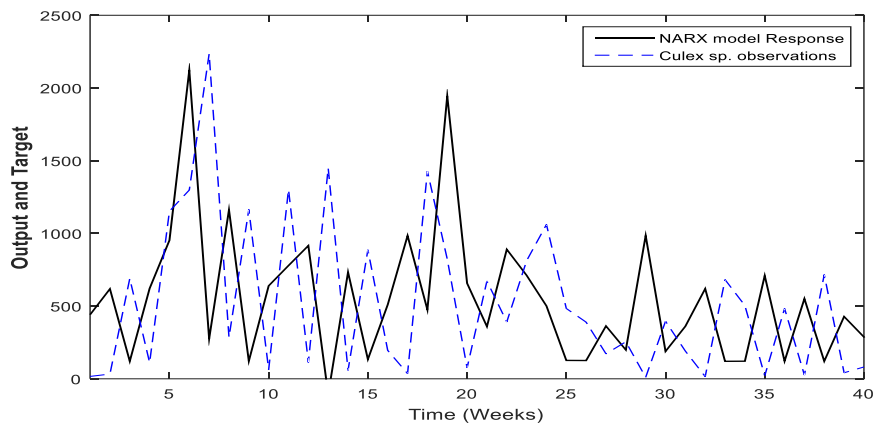
Figure 31. Clear response of the NARX neural network model output to the mosquito population time series.

3.3.4. Soft computing algorithm and extension for decision support

Figure 32 shows the procedure that have been followed to develop the ANNs model as well as an extension which can be potentially be generated to be used for vector eradication programs and related health management actions decision making. The ANNs model development provide a robust method for analyzing the past data and to be later used to forecast the arthropod vector population dynamics in respect to real time data (i.e., temperatures).

The algorithm describes the steps, initial choices and relate routines (i.e., loops-decisions) that have been used to end up with the final feedforward ANN model with tapped delay line at the input (i.e., one time step - week).

First data preparation and preliminary testing is performed to decide upon the best data set used for model training and validation. The validation datasets consist of the sample of data held back from training, while the test data set is used for fine tuning (optimizing) the ANN model hyper

89

parameters (i.e., taking weights of the trained ANN and use it as initialization for a new model being trained and so on).

Initially, the process starts by selecting a small number of neurons and (i.e., 5-10) in respect to some initial random weights (e.g., supervised learning) for the synapses and each time the network is trained it results to a different solution due to the different initial weight and bias values as well as network properties (e.g., number of neurons). Note that different divisions of data in to training, validation and testing may result also to different model performance. The model is retrained several times to ensure it has good accuracy towards an optimal solution based on an error measurement. The error, as shown in the material section, is defined as the difference of the output of the ANN and the pre-specified external desired data series. The error is estimated for different ANN structures related to the number of hidden layers to derive the final model which performs best. The optimized Final model can be fed with new data *per se*, or to be retrained the model and to predict the values of future time steps.
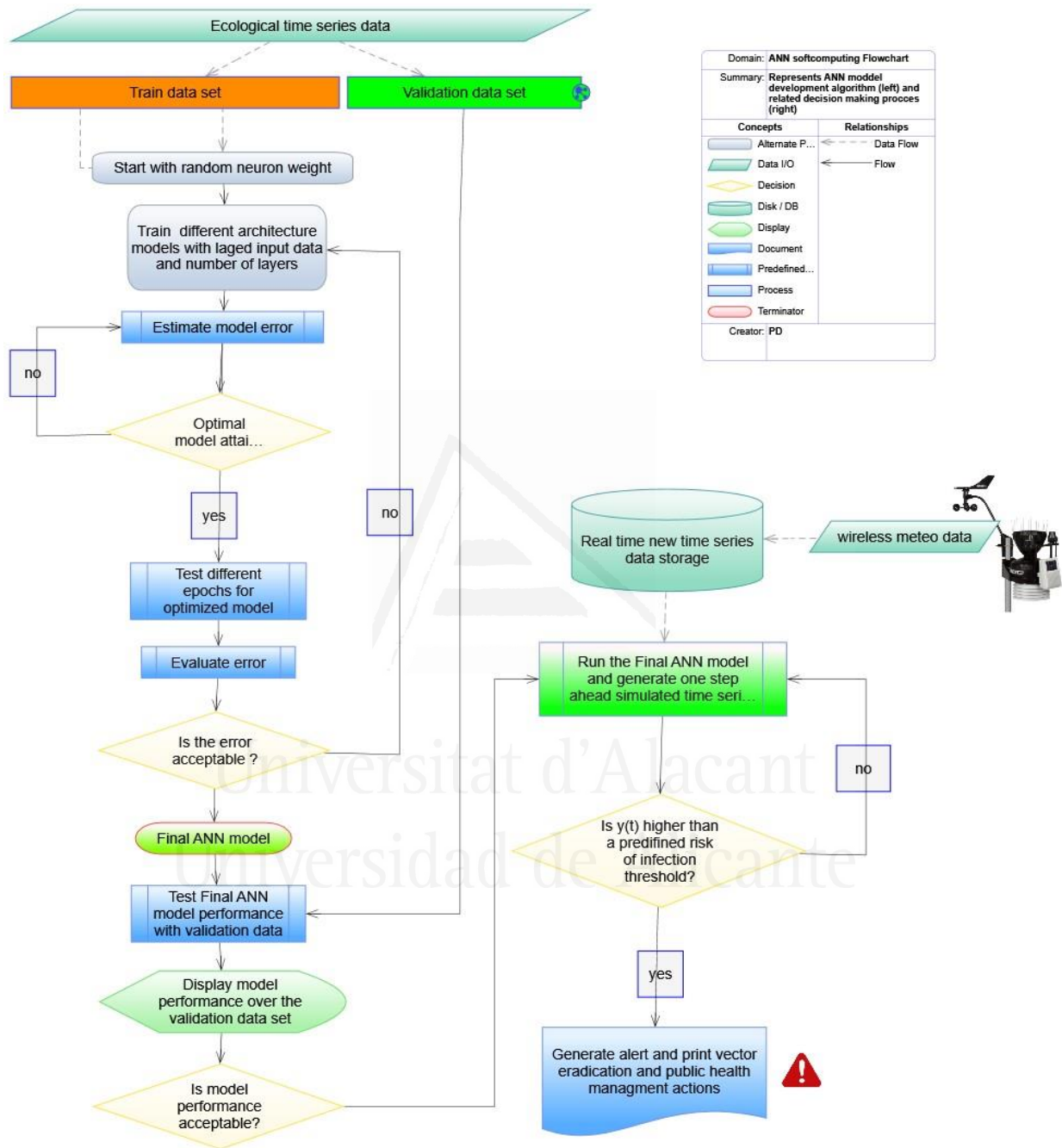
Figure 32. Graphical illustration of the logical operations followed to develop the dynamics autoregressive ANNs models for predicting adult mosquito population dynamics (left). Real time data can be used later to forecast the arthropod vector population dynamics and or to retrain the model under new circumstances (right).

Chapter 4: Discussion

In this work a series of novel methods and related modeling approaches are introduced in predicting mosquito vector population dynamics in a Mediterranean climate. Open climatic and *Culex sp.*, abundance data were used to apply the proposed prediction methods and demonstrate their utility under realistic field conditions. Furthermore, considering that *Culex sp.* is the major transmission vector of the WNV, Filariasis as well as other *Culex sp.* transmitted encephalitis, the predictive methods of the current studies may contribute to the understanding of the functioning of arthropod vector dynamics as well as to predict the periods of high mosquito activity. This information is important to initiate mosquito control actions and break the vector diseases transmission cycles. From a public health standpoint, the current study contributes to the development of decision tools to be used vector control actions as well as for initiating public health management strategies and preventing WNV, f(ph)ilariasis and encephalitis viruses.

Bearing of the issue of the effect of the climate variables, temperature, as expected, had a strong influence on mosquito temporal population dynamics. Particularly, although several climatic variables were tested (i.e. participation, wind speed, wind direction) whether they affect mosquito abundances the Pearson's correlation matrix did not indicate significant correlations among these variables but temperature. Moreover, the multiple linear regression suggested a positive association between mosquito abundance and mean temperatures. This is in accordance with other studies which have shown that among candidate climatic factors temperature exerts by far the most significant influence (Ganser and Wisely 2013) probably because increased environmental temperatures increasing

metabolic rates, reproductive output, and host-seeking behavior of these vectors and thus affect mosquito abundance (Shone et al. 2006).

Nevertheless, although environmental temperatures are the most important abiotic variables affecting insect physiology, development and bio ecology (Reinhold et al. 2018) the wavelet analysis revealed transient relations between adult mosquito captures and temperatures. On the opposite, despite the innate perception that precipitation levels may have a strong effect on indigenous population dynamics, the analysis do not explicit such direct effect for both years since we in all cases we observed a weak negative association with participation. This may be related to the fact that the rainy condition may cause an early indirect effect on population dynamics providing suitable wet sites for breeding. In addition, the particular region of research is reach in water resources since it is characterized by irrigated cultivations of high agricultural intensity providing suitable breeding sides throughout the summer season. Temperature is noted to have a stronger influence rather than participation on abundances of other mosquito species such as *Anopheles arabiensi*s and was the main driver of a population model (Abiodum 2016).

Considering the results of the wavelet analysis revealed that the oscillation of the mosquito abundance are dominated by diverse weekly modes and this can be observed in the respective wavelet spectrum for both observation years. The wavelet power spectrums however revealed that the high frequencies, for both mosquito abundance and temperature are in the same band of high frequencies ($\approx$4 weeks) a fact which cannot be seen by observing the mosquito time series *per se*. Moreover, the cross-wavelet analysis applied to the same data sets shows nothing but a significant common mode of oscillation mostly during the start and the end of the mosquito flight period. Thus, in contrast to simple correlations, the cross-wavelet analysis has the advantage to revel the time were likely relationship between the two-time series. Thus, wavelet analysis is particular suitable

for the analysis of the relationships between two time series, and especially appropriate for following gradual change in forcing by exogenous variables (Cazelles et al. 2008). Moreover, it is demonstrated that wavelet analysis works better than linear regression, which is in agreement with other studies which show that classical statistical analysis techniques may be inadequate in modeling non-stationary time series (Simões et al. 2013).

The major advantage of wavelet transforms over statistical correlation is as it can be applied to analyse time series that incorporate non-stationary power at many different frequencies. This is feasible by decomposing a time series into a time-frequency space to determine both, the dominant modes of variability and how these vary in time. Moreover, to analyze relationships in time-frequency space between two time series the cross-wavelet transform (XWT) and wavelet coherence (WTC) can be applied, while the phase angle statistics can be applied to acquire confidence in causal relationships and to test physical relationships between the time series (Ehelepola et al. 2015). The XWT is also preferred over the furrier transforms since it does not represent abrupt changes efficiently because it represents data as a sum of sine waves which are not localized in time or space.

Based on the affront mentioned correlation analysis between mosquito populations and climatic variables, it is apparent that temperature exert a higher impact on the adult arthropod vector phenology compared to rain events, despite those mosquitoes, particularly thrive in wet conditions since rain indirectly affects the mosquito population through increase breeding grounds. Therefore, it was judged as necessary to include the most influential meteorological variables (e.g., only temperature) to improve the performance of a simple MC model through the use of a CMC model instead. Actually, it is found that the Markov chain model of arthropod vector population dynamics which is conditioned over temperatures, performs considerable better than single MC stochastic modeling of vector population dynamics.

Thus, after the importance of the meteorological conditions was found, it is apparent that once the population reach a high state during a week, there is a very high probability to remain in this state by the following week and so on. Moreover, considering that time evolution of temperature states is quite analogous of that of the arthropod vector states, once can conclude that if we register increased probabilities of higher temperature states there is also an increased probability to observe very high mosquito levels. Thus, a part of the CMC mode results, modeling only temperatures through MC model may be proved very utile in judging whether during the same period the mosquito population is also high. This is very important for public health management and vector eradication programs considering saving costs and time for the establishment of mosquito surveillance programs over different areas.

Summarizing the modelling approach and related simulations, it is worthwhile to look at some of the features of the conditional Markov chain modelling method under consideration. Classical conditional Markov chain models (also known as Linear-Chain Conditional Random Fields in the literature e.g., Lafferty et al. 1999) are defined very recently by Bielecki and Rutkowski (2004), for applications in finance and insurances. This is done in response to the need for modeling dependence between dynamic systems in cases when some conditional properties of a system are important and should be accounted for (Bielecki et al. 2017). Hence, conditional Markov Chains are defined as a versatile class of discriminative models for the distribution of a sequence of hidden or latent states conditional on a sequence of observable variables and are used in this study for the first time in modeling *Culex sp.* dynamics.

Considering, the different autoregressive ANN's models that have been applied to describe the adult population dynamics of *Culex sp.* in most cases they have been performed in acceptable levels considering the non-linear dynamics of the adult mosquito abundances. However, the model that takes in to account temperature as exogenous input variable, as in the case

of the temperature conditioned MC model performed better than the others. Structurally, the NARX recurrent network receives the sequence of two external inputs as well as the recurrent output layer state, while the NAR and FTD network not. To the best of my knowledge, the development and application of the current ANNs is one of the first of its kind in modeling arthropod vector dynamics. One advantage of the ANNs models, over the MC model, is that they can be used in cases where the data series are short. Additionally, since they are non-parametric methods, they do not depend strongly on normality assumptions. To date, a practical limitation of MC models is that the initial transition matrix may be affected by the length of the sequence.

Moreover, considering the soft computing approach, the networks that have been applied belong to the general class of recurrent dynamic neural networks (RNNs), which in contrast to ANNs, are designed to take a series of inputs with no predetermined limit on size and to memorize prior inputs while generating an output. Although, the NARX model predicted slight better in compared to the FTD model, the differences in model performance are in general low. A practical implication of this fact in model development is that temperature can be used as the main input contributor of the ANN to predict population abundance and that the inclusion of previous mosquito population abundance does not improve dramatically the model performance. As a result, the advantage of the FTD over the NARX neural network model is that it can be applied in the case where past values of mosquito abundance are not available.

One other advantage of using ANNs models, over time series models (linear and non-linear), is the fact that the cases were predictions are performed form a random sample from the same population as the time periods about which one makes the prediction. Additionally, the performance of autoregressive time series models is also affected in situations of limited data availability, while the true shape of data distribution is unknown (Jain and Kumar 2006, Damos 2015, Damos 2016).

On the other hand, one limitation of ANN models is that there is no set method for the construction of the network architecture (Eftekhar 2005). For instance, the development of the final model structure, which is presented in the current study, is the result of numerous prior combinations of candidate ANNs model structures (i.e., number of neurons and hidden layers) and input variables (i.e., lagged climate and mosquito abundance data). One other explanatory limitation of ANNs is that the analysis generates weights, instead of standardized coefficient parameters, which are difficult to interpret and often not present as they are in regression analysis (Baxt 1995). Therefore, one of the most criticized features in ANN's is the lack of interpretability at the level of individual variables (Ohno-Machado and Rowland 1999). Moreover, since the ANN learning performance has been checked against the disjoint set of data that was available (i.e., test set), it is of fundamental importance to choose an appropriate training set size, and to provide representative coverage of all possible conditions for modeling.

Nevertheless, from a practical standpoint the purpose of the proposed methods for modeling mosquito dynamics was to predict the next value of mosquito abundance taking into account past values of input variables as well all past predictions of the model to improve its forecasting efficacy. Based on the modelling results, the mosquito populations have a certain period of high activity in a temperate climate (i.e., population peaks of high abundance), which can be further used to initiate specific management actions against periods of high activity of mosquito adults.

Compared to other recent studies applying univariate ANNs to model underlying population abundance trajectories and don't take in to account the effect of other dynamics variables to model population process with structure or interactions (Goodacre et al. 1996, Lee et al. 2016). For cold blood species, however, such as mosquito and other arthropod vectors, temperature is considered as predominant factor affecting their life history traits (Shapiro et al. 2017, Reinhold et al. 2018). As a result, the addition of

temperature has considerably improved predicting performance of both, the CMC as well as the NARX model.

This is consisted with other studies as well, which show temperature as major factor affecting mosquito development rates and related dynamics (Samy et al. 2016). Meta-analytic results, for instance, performed on another mosquito species *Aedes aegypti*, indicated that the environmental factor of temperature is sufficient to explain development rate variability and factors such as diet should never be considered to the exclusion of temperature in modeling development (Couret and Benedict 2014). In addition, related dengue virus transmission is influenced by the amplitude and pattern of daily temperature variation (Lambrechts et al. 2011), while development and survival rate of both *Anopheles* mosquitoes and the Plasmodium parasites that cause malaria depend on temperature, making this a potential driver of mosquito population dynamics and malaria transmission (Beck-Johnson et al. 2013). In addition, for West Nile Virus in *Culex pipiens* increasing temperatures may accelerate transmission of WNV, as demonstrated by Kilpatrick et al. (2008).

Because mosquitoes that feed on both birds and mammals, they are referred to as bridge vectors for WNV between an infected reservoir (birds) and mammalian incidental host (Kilpatrick et al. 2005). Additionally, since the epidemiology of WNV maintained in nature depends in an enzootic transmission cycle between birds and mosquitoes, their infection rates are suppressed by the absence mosquito (Chancey et al. 2015). Moreover, there is a delay in human and horse cases because due to incubation and because they are considered incidental hosts which do not participate in the WNV lifecycle since they do not develop sufficient viremia to infect mosquito vectors (van der Meulen et al. 2005). Thus, control management actions against the mosquito bridge vector is critical in breaking the epidemiological cycle of vector borne diseases.

Thus, the understanding of mosquito phenology and the description of its population dynamics is essential for the prevention of vector borne

diseases and to initiate proper management actions (Diuk-Wasser et al. 2006, , Wijaya et al. 2019). Thus, to facilitate this understanding it seems reasonable to build mathematical models of increasing complexity that reflect some true state of the time evolution and dynamics of natural populations of mosquitoes (Hacker et al. 1973).

In specific, the current modelling approaches may tend to contribute very significantly to the improvement of accuracy of population data description and particularly that of mosquito abundances which due to their specific life cycle are most often are characterized by abrupt outbreaks.

Chapter 5: Conclusion

5.1 Overall conclusion

This work introduces novel prediction methods and based on stochastic and soft computing techniques for modeling the population of arthropod vectors, to date the world's deadliest animal, which accounts for 80% of human vector-borne diseases. Surely, the most important asset of the current research is the introduction of novel analysis and prediction methods for *Culex sp.* and finally, the evolution of a series of models which potentially can be applied as a method for detecting and forecasting the periods in which there is a high probability of vector population persistence.

Moreover, the current results provide a strong evidence of the presence of transient relationships between meteorological variables and mosquito abundance and may support an ecological justification based on the fact that mosquito populations are able to find refuge or even decrease their flying behavior when environmental conditions are extreme. The results of the wavelet analysis, for example, suggest also that mosquito dynamics may not be stationary over the full range of sampling points which is an important information for selecting appropriate modeling methods and is generally not captured by the regression analysis.

The current study shows also that modeling of the Markov chain is useful for simulating future population dynamics of arthropod vectors and that modeling performances are improved with the addition of temperature. This is probably one of its first kind application in modeling arthropod vector population dynamics. Furthermore, the estimation of the transition matrix through the use of empirical data first to define the system states and later on for training the Markov chain model, is a principal step for the simulation of realistic vector population projections and without the need

for defining differential equations and related state variables. Therefore, the CMC model that is proposed in this study might prove very suitable in public health decision making and particularly helpful for vector eradication programs.

Considering the ANNs models, in contrast to traditional statistical models (i.e. autoregressive models, multivariate regression models), that have been used to model arthropod vector abundance and related disease dynamics (Wang et al. 2011, Earnest et al. 2012, Simões et al. 2013, Archana 2017, Karen et al. 2019), the ANNs model that have been applied in the current study have as main asset their ability of their neurons (i.e. sub model of different weights) working simultaneously, but independent from each other (Saproska et al. 2014). Especially, in the cases were the outcome variable (i.e., here the mosquito abundance) is affected by more factors (i.e., temperature, previous mosquito abundance and more), these can be independently introduced and taken in to account by the network in terms of its weight during the learning process (or training).

In the context of prevention to mitigate the effects of arthropod vector dynamics on population health there is an urgent demand for warning systems that will aid Public Health structures in vector control decision making. The current model results thus could form the basis to forecast arthropod vector population levels and to alarm people belonging to vulnerable groups and to implement effective vector control measures to protect public health during incidents of high population pressure.

5.2 Recommendations and prospects of future work

The presented results are promising, although it should be remembered that they were obtained under certain assumptions, such as the stability of the particular study environment and the conditioning over only one climate variable, namely temperature. However, it might be possible that additional ecological factors may affect mosquito population

dynamics in a more complex manner and due to the limitations of the current dissertation have not been taken into account. For instance, among them is the possibility of a parallel influence of two or more climate variables on the mosquito population dynamics or even a more substantial influence of lagged population values.

Another future direction worth verifying would be the calculation of a multivariate semi Marcov conditional model with different orders. Although this is a bigger challenge and it is not yet known if it is too promising, the result should not be pre-judged before testing.

Therefore, the proposed predictive methods and related modeling approaches, although valuable inn the frame of the current study, have the potential to be further improved. This can be achieved not only by including more more variables, but also by extending and testing the prediction performance of the current models in to new areas.

In essence, all the same, the current subject field can be seen as valuable since it's not just the first of its sort, but also encouraging further modelling and exploration of the behavior of more complex vector population systems.

Finnally and from a practical public health standpoint, it is worth to develop a web based decision support system which can incrorporate the current, as well as related apidimiological models, to provide real time artropode population forecasts. DSS systems can help to automize the decision-making process and evaluate the different health management actions.

6. References

Abiodun G.J. , Maharaj R. , Witbooi P. , Okosun K.O. Modelling the influence of temperature and rainfall on the population dynamics of Anopheles arabiensis. Malaria J., 15 (2016).Article Number: 364

Akhtar, M.; Kraemer, M.U.G.; Gardner, L.M. 2019. A dynamic neural network model for predicting risk of Zika in real time. *BMC Med.*, *17*, 171.

Aribowo, W. 2018. An Adaptive Power System Stabilizer Based on Focused Time Delay Neural Network. *J. Teknosains* , *7*, 67–73.

Baxt, W. 1995. Application of artificial neural networks to clinical medicine. *Lancet*, *346*: 1135–1138.

Beale, M.H.; Hagan, M.T.; Demuth, H.B. 2015. *Neural Network ToolboxTM, Users Guide*; The MathWorks, Inc.: Natick, MA, USA, 2015.

Beck-Johnson, L.M., Nelson, W.A., Paaijmans, K.P., Read, A.F., Thomas, M.B., Bjørnstad, O.N. 2013. The Effect of Temperature on Anopheles Mosquito Population Dynamics and the Potential for Malaria Transmission. *PLoS ONE* 2013, *8*, e79276.

Bielecki TR, Jakubowski J, Niewęgłowski M. 2017. Conditional Markov chains: Properties, construction and structured dependence. Stoch Process their Appl. 127:1125–70.

Bielecki, T.R.; Jakubowski, J.; Niewęgłowski, M.2017. Conditional Markov chains: Properties, construction and structured dependence. *Stoch. Process. their Appl. 127*: 1125–1170.

Bielecki, T.R.; Rutkowski, M. 2004. Modeling of the Defaultable Term Structure: Conditionally Markov Approach. *IEEE Trans. Automat. Contr. 49*: 361–373.

Bielecki, T.R.; Rutkowski, M. 2004. Modeling of the Defaultable Term Structure: Conditionally Markov Approach. *IEEE Trans. Automat. Contr. 49*: 361–373.

Boussaada, Z.; Curea, O.; Remaci, A.; Camblong, H.; Bellaaj, N.M. 2018. A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. *Energies*, *11*, 620.

Browne, A., 1997. Neural Network Analysis, Architectures and Applications. Institute of Physics Publishing, Bristol and Philadelphia (1997).

Brugman VA, Hernández-Triana LM, Medlock JM, Fooks AR, Carpenter S, Johnson N. 2018. The role of *Culex pipiens* L. (Diptera: Culicidae) in virus transmission in Europe. *Int. J. Environ. Res Publ. Health*, 15: 389.

Calzolari M, Bonilauri P,Bellini R, Albieri A, Defilippo F, Maioli G, *et al.* 2010. Evidence of Simultaneous Circulation of West Nile and Usutu Viruses in Mosquitoes Sampled in Emilia-Romagna Region (Italy) in 2009. 2010; PLOS ONE 5(12): e14324. https://doi.org/10.1371/journal.pone.0014324

Cazelles B. , M. Chavez, D. Berteaux, F. Menard, J.O. Vik, S. Jenouvrier, *et al. 2008.* Wavelet analysis of ecological time series Oecologia, 156, 287-304.

Cazelles B. , M. Chavez, G.C. Magny, J. Guégan, S. Hales. 2007. Time-dependent spectral analysis of epidemiological time-series with wavelets. J. R. Soc. Interface, 4, pp. 625-636

Chancey, C., Grinev, A., Volkova, E., Rios, M. 2015. The global ecology and epidemiology of West Nile virus. *BioMed research international, 2015*, 376230. https://doi.org/10.1155/2015/376230

Couret, J.; Benedict, M.Q. 2014. A meta-analysis of the factors influencing development rate variation in *Aedes aegypti* (Diptera: Culicidae). *BMC Ecol.* 2014, *14*, 3.

Damos P, Rigas A, Savopoulou-Soultani M. 2011. Application of Markov Chains and Brownian motion models on Insect Ecology. In: Earnshaw RC, Riley EM. Brownian Motion: Theory, Modelling and Applications. pp. 71–104. Nova Science Publications.

Damos, P. 2015. Mixing times towards demographic equilibrium in insect populations with variable age structures. Theoretical Population Biology, 103, 93-102.

Damos, P. 2015. Mixing times towards demographic equilibrium in insect populations with variable age structures. Theoretical Population Biology, 103, 93-102.

Damos, P. 2016. A stepwise algorithm to detect significant time lags in ecological time series in terms of autocorrelation functions and ARMA model optimisation of pest population seasonal outbreaks. *Stoch. Environ. Res. Risk Assess., 30*, 1961–1980.

Deichmeister J..M, Telang A. 2011. Abundance of West Nile virus mosquito vectors in relation to climate and landscape variables. *J Vector Ecol.* 6:75–85.

Diuk-Wasser, M. Vourch G., Cislo P., et al. 2010. Field and climate based model

for predicting the density of host-seekinng nymphal Icodes scapularis, an important vector of tock-borne disease agents in the eastern United States. Global Ecol. Biogeogr.:19, 504–514.

Dorrestijn, J.; Crommelin, D.T.; Siebesma, A.P.; Jonker, H.J.J.; Selten, F. Stochastic convection parameterization with Markov chains in an intermediate complexity GCM. *J. Atmos. Sci, 73:* 1367–1382, doi:10.1175/JAS-D-15-0244.1.

Eftekhar, B.; Mohammad, K.; Ardebili, H.E.; Ghodsi, M.; Ketabchi, E. 2005. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med. Inform. Decis. Mak. 5*: 3.

Ehelepola N.D.B. , Ariyaratne K. , Buddhadasa W.M.N.P. , Ratnayake S. , Wickramasinge M. 2015. A study of the correlation between dengue and weather in Kandy City, Sri Lanka (2003—2012) and lessons learned. Infect Dis Poverty. 4: p. 42

Eikenberry, S.E., Gumel, A.B. 2018. Mathematical modeling of climate change and malaria transmission dynamics: a historical review. *J. Math. Biol.* 77, 857–933.

Engler, O., Savini, G., Papa, A., Figuerola, J., Groschup, M. H., Kampen, H., Medlock, J., Vaux, A., Wilson, A. J., Werner, D., Jöst, H., Goffredo, M., Capelli, G., Federici, V., Tonolla, M., Patocchi, N., Flacio, E., Portmann, J., Rossi-Pedruzzi, A., Mourelatos, S., … Johnson, N. 2013. European surveillance for West Nile virus in mosquito populations. *International journal of environmental research and public health*, *10*(10), 4869–4895. https://doi.org/10.3390/ijerph10104869

Ferraguti M, Martínez-de la, Puente J, Muñoz J, Roiz D, Ruiz S, Soriguer R, Figuerola J. 2013. Avian *Plasmodium* in *Culex* and *Ochlerotatus* mosquitoes from southern Spain: effects of season and host-feeding source on parasite dynamics. *PLoS One.* 2013;6(6):e66237. doi: 10.1371/journal.pone.0066237.

Foufoula-Georgiou E., Kumar. P. 1995. Wavelets in geophysics, 373, Academic Press (1995)

Ganser C. , S.M. Wisely 2013. Patterns of spatio-temporal distribution, abundance, and diversity in a mosquito community from the easten Smoky Hills of KansasJ Vect. Ecol., 38, pp. 229-236

Ganser C., Wisely S.M. 2013. Patterns of spatio-temporal distribution, abundance, and diversity in a mosquito community from the eastern Smoky Hills of Kansas. Journal of Vector Ecology, 38: 229-236.

Goodacre, R., Neal, M.J., Kell, D.B. 1996. Quantitative Analysis of Multivariate Data Using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra. *Zent. Bakteriol.*, *284*: 516–539.

Gouhier T.C. , A. Grinsted, V. SimkoPackage 'Biwavelet' R-Core Team

Grassly NC, Fraser C. 2008. Mathematical models of infectious disease transmission. *Nat Rev Microbiol.* 6: 477–487.

Grinsted A. , J.C. Moore, S. Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinerar processes in geophysics. 11: 561-566J. R. Soc. Interface, 4: 625-636.

Gubler D.J. 2010. The Global Threat of Emergent/Re-emergent Vector-Borne Diseases. In: Atkinson P.W. (eds) Vector Biology, Ecology and Control. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-2458-9_4

Hacker, C.S, Scott D. W., Thompson J.R. 1973. Time series analysis of mosquito population data. J. Med. Ent., 10: 533-543.

Harbach R. E.2015. *Mosquito Taxonomic Inventory. Valid Species List.* http://mosquito-taxonomic-inventory.info/valid-species-list [accessed 12 February 2015]

Harbach RE. 2013. Genus *Anopheles* Meigen, 1818. http://mosquito-taxonomic-inventory.info/genus-emanophelesem-meigen-1818-0 [accessed 23 February 2015]

Harback R.E. 2007. The Culicidae (Diptera): a review of taxonomy, classification and phylogeny. Zootaxa 1668:591–638.

Host F.G. 2018. Range, Host–Virus Interactions, and Virus Transmission. *Viruses.* 2018;101-134. doi:10.1016/B978-0-12-811257-1.00005-X

Reinhold, Joanna M.; Lazzari, Claudio R.; Lahondère, Chloé. 2018. "Effects of the Environmental Temperature on *Aedes aegypti* and *Aedes albopictus* Mosquitoes: A Review" *Insects* 9, no. 4: 158. https://doi.org/10.3390/insects9040158

Jain, A.; Kumar, A.M. 2006. An evaluation of artificial neural network technique for the determination of infiltration model pa-rameters. *Appl. Soft Comput. 6*: 272–282.

Jourdain, E., Gauthier-Clerc, M., Bicout, D., Sabatier, Ph. 2007. Bird Migration Routes and Risk for Pathogen Dispersion into Western Mediterranean Wetlands. Emerging Infectious Diseases 13:365-72

Kading RC, Abworo EO, Hamer GL. 2019. Rift Valley Fever Virus, Japanese Encephalitis Virus, and African Swine Fever Virus: Three Transboundary, Vector-Borne, Veterinary Biothreats with Diverse Surveillance, and Response Capacity Needs. *Front Vet Sci.*, 6:458.

Kilpatrick, A.M., Meola, M.A., Moudy, R.M., Kramer, L.D. 2008. Temperature, Viral Genetics, and the Transmission of West Nile Virus by *Culex pipiens* Mosquitoes. *PLOS Pathog.* 2008, *4*, e1000092.

KilpatrickA.M. , L.D. Kramer, S.R. Campbell, E.O. Alleyne, A.P. Dobson, P. Dasza, D. Maraun, J. Kurths. 2004. Cross wavelet analysis: significance testing and pitfalls Nonlinear Process. Geophys., 11: 505-514.

Köppen, W.1923.Die Klimate der Erde: Grundriss der Klimakunde. Walter de Gruyter Germany, Berlin (1923).

Kullback S., Kupperman M., Ku H. 1962. Tests for Contingency Tables and Marlkov Chains. Technometrics, 4: 573–608.

LaBeaud AD, Bashir F, King CH. 2011. Measuring the burden of arboviral diseases: the spectrum of morbidity and mortality from four prevalent infections. *Population Health Metrics* 9:2014.

Lafferty J.,McCallum A., Pereira F. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proc. of the 18th IEEE International Conference on Machine Learning (ICML).

Lafferty KD. 2009. The ecology of climate change and infectious diseases*. Ecology*, 90: 888–900.

Lafferty, J.; Mccallum, A.; Pereira, F. 1999. Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data Abstract. 1999, *2001*, 282–289.

Lalubin F, Delédevant A, Glaizot O, Christe P. 2013. Temporal changes in mosquito abundance (Culex pipiens), avian malaria prevalence and lineage composition. *Parasit Vectors*. 6:307.

Lambrechts, L., Paaijmans, K.P., Fansiri, T., Carrington, L.B., Kramer, L.D., Thomas, M.B., Scott, T.W. 2011. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA*, *108*: 7460–7465.

Lee, K.Y.; Chung, N.; Hwang, S. 2016. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecol. Inform. 36*, 172–180.

Lewis M.A, Krkosek M., Wonham M. 2010. Dynamics of Emerging Wildlife Disease. Fields Institute CommunicationsVolume00,1-22.

Lord CC. 2004. Seasonal population dynamics and behaviour of insects in models of vector-borne pathogens. *Physiol Entomol.* 29:214–222.

Lord CC. 2007. Modeling and biological control of mosquitoes. *J Am Mosq Control Assoc.* 23:252–264.

Luz P.M., Struchiner C.J., Galvani A.P. 2010. Modeling transmission dynamics and control of vector-borne neglected tropical diseases. *PLoS Negl Trop Dis*. 4:761.

Maraun D , J. Kurths 2004. Cross wavelet analysis: significance testing and pitfalls. Nonlinear Process. Geophys., 11: 505-514.

Matlab, R2015; The Mathworks Inc.: Natick, MA, USA, 2015.

Matworks 2020. Neural Network Architectures. Available online: https://fr.mathworks.com/help/deeplearning/ug/neural-network architectures.html (accessed on 20 April 2021)

Medlock J.M., Leach SA. 2015. Effect of climate change on vector-borne disease risk in the UK. Lancet Infect Dis., 15:721-30. doi: 10.1016/S1473-3099(15)70091-5. Epub 2015 Mar 23. PMID: 25808458.

Menezes, J.M.P.; Barreto, G.A. A New Look at Nonlinear Time Series Prediction with NARX Recurrent Neural Network. In Proceedings of the Ninth Brazilian Symposium on Neural Networks, Ribeirao Preto, Brazil, 23–27 October 2006.

Nikolay B. 2015. A review of West Nile and Usutu virus co-circulation in Europe: How much do transmission cycles overlap? *Trans. R. Soc. Trop. Med. Hyg,* 109:609–618. doi: 10.1093/trstmh/trv066.

Ohno-Machado, L.; Rowland, T. 1999. Network Applications in Physical Medicine and Rehabilitation1. *Am. J. Phys. Med. Rehabil., 78*: 392–398.

Paily KP, Hoti SL, Das PK. A review of the complexity of biology of lymphatic filarial parasites. *J Parasit Dis*. 2009;33(1-2):3-12. doi:10.1007/s12639-009-0005-4

Papa A, Bakonyi T, Xanthopoulou K, Vázquez A, Tenorio A, Nowotny N. Genetic characterization of West Nile virus lineage 2, Greece, 2010. *Emerg Infect Dis*. 2011;17:920-922.

Papa A, Xanthopoulou K, Gewehr S, Mourelatos S. 2011. Detection of West Nile virus lineage 2 in mosquitoes during a human outbreak in Greece. *Clin Microbiol Infect*. 2011; 17:1176–80 10.1111/j.1469-0691.2010.03438.x

Pfeffer M, Dobler G. 2010. Emergence of zoonotic arboviruses by animal trade and migration. *Parasit Vectors* 3:35.

Recknagel, F. 2001. Applications of machine learning to ecological modelling. Ecological Modelling 146 (2001) 303 – 310

Reinhold, J.M.; Lazzari, C.R.; Lahondère, C. 2018. Effects of the Environmental Temperature on *Aedes aegypti* and *Aedes albopictus* Mosquitoes: A Review. *Insects* 2018, *9*, 158.

Samy, A.; Elaagip, A.H.; Kenawy, M.A.; Ayres, C.F.J.; Peterson, A.T.; Soliman, D. 2016. Climate Change Influences on the Global Potential Distribution of the Mosquito *Culex quinquefasciatus*, Vector of West Nile Virus and Lymphatic Filariasis. *PLoS ONE* 2016, *11*, e0163863.

Sato T. , R. Kajiwara, I. Takashima, T. Iijima 2018. Wavelet correlation analysis for quantifying similarities and real time estimates of information encoded or decoded in signaltial oscillatory brain waves. Wavelet theory and its applications (2018), 10.5772/intechopen.74810. Edited by Radhakrishnan S

Schmidhuber, J. 2014. Deep learning in neural networks: An overview," Neural Networks, 61, 85–117.

Shapiro, L.L.M.; Whitehead, S.A.; Thomas, M.B. 2017. Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. *PLoS Biol.* 2017, *15*, e2003489

Shone S.M. , Curriero F.C. , Lesser C.R. , Glass G.E. 2006. Characterizing Population Dynamics of Aedes sollicitans (Diptera: Culicidae) Using Meteorological Data. J. Med. Entomol. 43:393-402.

Simões T.C. , Codeço C.T. , Nobre A.A. , Eiras A.E. Modeling the non-stationary climate dependent temporal dynamics of *Aedes aegypti*. PLoS One, 8 (8) (2013), Article e64773,

Spedicato G. A., Kang T.S., Yalamanchi S.B., Yadav D. 2021. The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R.

Tiwari A.K. , M.I. Mutascu, C.T. Albulescu. 2016. Continuous wavelet transform and rolling correlation of European stock markets. Int. Rev. Econ. Finance, 42: 237-256.

Torrence C., G.P. Compo. 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc., 79: 61-78.

van der Meulen, K.M., M. B. Pensaert, and H.J. Nauwynck, 2005: West Nile virus in the vertebrate world. Arch. Virol. 150, 637–657.

Wei H.M., Li X.Z., Martcheva M. 2008. An epidemic model of a vector-borne disease with direct transmission and time delay. ournal of Mathematical Analysis and Applications 342:895-908.

Weidhaas D., Focks D. 2000). "Management of arthropod borne diseases by vector control," in *Medical Entomology*, eds Eldrigde B., Edman J. (Boston: Kluwer Academic Publishers, 539–564

WHO 2019. Mosquito borne diseases. https://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/

Wijaya K. P., Götz T., Soewono E. 2015. Advances in mosquito dynamics modeling. Mathematical Methods in the Applied Sciences 39: 4750-4763.

Wonham, M.J., de Camino-Beck, T., Lewis, M.A. 2004. An epidemiological model for West-Nile virus: Invasion analysis and control application, Proceedings of the Royal Society of London B 271: 501–507

7. Annex

7.1. Congresses presentations and published abstracts

7.1.1 Damos, P. Caballero, P., Dorrestijn, J. 2021. Markov chain modelling of ecological time series data with emphasis on arthropod disease dynamics. 19th Conferennce of the Applied Stochastic Models and Data Analysis International Society ASMDA, 1-4 June 2021, Athens, Greece.

BOOK of ABSTRACTS

ASMDA 2021

and

Demographics 2021 Workshop

http://www.asmda.es/asmda2021.html

Markov chain modeling of ecological time series data with emphasis on arthropod disease vector dynamics

Petros Damos[1], Pablo Caballero[1] and Jesse Dorrestijn[2]

[1]Department of Preventive medicine, Public Health and History of Science, Faculty of Health Science, University of Alicante, Carretera San Vicente s/n03690 San Vicente del Raispeig, Alicante, Spain
[2]Delf University of Technology, Delf, Netherlands

Several mathematical and standard epidemiological models have been proposed in studying vector transmitted infectious disease dynamics. However, most models are of deterministic nature and are not able to estimate other relevant metrics such as the probability of vector population emergence as well as the probability and expected time to reach certain population and/or infection state. Here we are focusing in stochastic modeling of ecological and epidemiological time series data using Markov chains (MCs) and are particularly interested in estimating transition probabilities (TPs) which are conditioned with other non-trivial stochastic

111

variables. The aim is to generalize a formulation of conditional Markov chain models (CMSs) for predicting probability transition estimates of vector populations and disease outcomes at discrete future time steps. In this context, first we present the basic principles and assumptions behind Markov chain modeling approach with an intuitive interpretation of the integration of conditional Markov chains (CMCs). We also include an approach for estimating the probability of state transitions between consecutive time steps from typical ecological and medical survey data. Finally, we demonstrate the usefulness of the current approach by training a CMCs model from existing vector disease data to obtain a simulated probability sequence of the variable of interest. Understanding and predicting vector population and related disease dynamics, is crucial for gaining insight into the abundance and dynamics of arthropod disease vectors, and for the design of effective vector control strategies. We conclude that discrete-time Markov chain technique is recommended as viable for modeling vector disease population dynamics in order to make real-time recommendations utile for dynamic health policies decision making.

Although, the Markov models generated in this work provide an accurate abstraction of the vector disease progress observed within the dataset used for their generation we envision the current approach as an entry point into the extensive literature and potential applications of Markov chains in ecological modeling and vector diseases dynamics simulation

Universitat d'Alacant
Universidad de Alicante

7.1.2 Damos, P., Tuels, J., Caballero, P. 2021. Predictive modelling of seasonal mosquito population patterns with neural networks. 1St International Electronic Conference on Entomology, 1-15 July, https://iece.sciforum.net/

Mosquito species are considered important vectors of many diseases in humans, companion animals, and livestock. There is a great need to understand their dynamics and to develop methods for predicting their abundances. However, the population dynamics of mosquitoes are often complex displaying non-linear dynamics. Thus, making it difficult to be modeled using linear statistical approaches. In this project, we explored the seasonal population patterns of mosquito populations in a Mediterranean environment in Northern Greece using straightforward machine learning techniques such as Artificial Neural Networks (ANNs). To train, validate and test the network model we have used 2 years weekly counts of adult mosquito data including *Culex sp.*, a major vector of the West Nile virus and related encephalitis diseases. The model training was performed in an open-loop (i.e., parallel series network architecture), including the validation and testing step and later on, after training, it was transformed to a closed-loop for the needs of a multistep-ahead mosquito abundance prediction. Determined by the autocorrelation function, one of the final models is using as inputs one week lagged values of mosquito abundances and was able to capture the adult seasonal mosquito patterns in most cases at acceptable levels. We conclude that ANNs suggest an important candidate for modeling and predicting the seasonal abundance of mosquito data since it is suitable for modeling noisy and incomplete ecological data, with no specific assumptions to be made about the underlying relationships and which are solely determined through data mining. However, we are also looking forward to improving the particular model performance using new data sets since it is of fundamental importance to choose an appropriate training set size and to provide representative coverage of all possible conditions to capture accurately the patterns of ecological time series. Nevertheless, despite the limitations of the current study, this work contributes to knowledge of the seasonal functioning of arthropod vector dynamics and contributes towards the development of decision tools to be used in the preventive management of the transmission cycle of vector-borne diseases.

Invitation for oral presentation:

12 July 2021		
Live Session 4: Session "Biodiversity, Ecology and Evolution" and Session "Medical and Veterinary Entomology"		
Time: 9:00am (CEST) ǀ 03:00am (EDT) ǀ 15:00pm (CST Asia)		
Speaker	**Time (CEST)**	**Presentation Topic**
Session Chair António Onofre Soares	9:00-9:25	The introduction of *Harmonia axyridis* in the Azores islands: why was it not successful?
Invited Speaker Paulo A. V. Borges	9:25-9:50	Biodiversity erosion in islands: quantifying impacts using long-term monitoring of forest arthropods in Azorean Islands
Short Break 9:50-10:00		
Session Chair Andre B. B. Wilke	10:00-10:30	Urbanization and the rise of vector mosquitoes and arbovirus transmission
Invited Speaker Guido Favia	10:30-11:00	Microbial symbiosis in mosquito vectors
Invited Speaker Petros Damos	11:00-11:30	Predictive modelling of seasonal mosquito population patterns with neural networks

7.2. Published work in scientific Journals with impact factor

7.2.1 Damos, P., Caballero, P. 2021. Detecting seasonal transient correlaations between populations of the West Nile Virus vector Culex sp. and temperatures with wavelet coherence analysis. Ecological Informatics, 61, 101216, https://www.sciencedirect.com/science/article/pii/S1574954121000078

Detecting seasonal transient correlations between populations of the West Nile Virus vector *Culex* sp. and temperatures with wavelet coherence analysis

Petros Damos [a,b,*], Pablo Caballero [a]

[a] Department of Preventive Medicine and Public Health and History of Science, Faculty of Health Science, University of Alicante, Alicante, Spain
[b] Pharmacy Department, University General Infectious Disease Hospital of Thessaloniki AHEPA, Aristotle University of Thessaloniki, Greece

ARTICLE INFO

Keywords:
Vector borne diseases
Mosquito
Climate
Correlations
Wavelets
Public health

ABSTRACT

Culex sp. is one of the most important mosquito disease vector and climate is considered to be a key factor affecting its population dynamics. In this study we use straightforward techniques based on correlations and wavelet analysis to determine the non-trivial associations between *Culex* sp. mosquito abundance and weather variables (lagged population abundance, mean temperatures, rain and wind speed) in Northern Greece during two successive years. In particular, mosquito abundances were examined for collinearity using Pearson's correlation matrices and redundant or low correlated variables were excluded from further analyses. Multiple linear regression analysis was performed to fit normalized mosquito abundance weekly counts as the dependent variable to the independent variables including: mean temperatures, rain as well as lagged mosquito populations. There was a high and positive correlation between temperature and mosquito abundance during both observation years ($r = 0.6$). However, a very poor correlation was observed between rain and weekly mosquito abundances ($r = 0.29$), as well as between wind speed ($r = 0.29$), respectively. Additionally, according to the multiple linear regression model the effect of temperature, was significant. The continuous power spectrum of the mosquito abundance counts and mean temperatures depict in most cases similar power for periods which are close to 1 week, indicating the point of the lowest variance of the time series, although appearing on slight different moments of time. The cross wavelet coherent analysis showed that inter weekly cycles with a period between 2 and 3 weeks between mosquito abundance and temperature were coherent mostly during the first and the last weeks of the season. Hence, the wavelet analysis show a progressive oscillation in mosquito occurrences with time, which is higher at the start and the end of the season. Moreover, in contrast with standard methods of analysis, wavelets can provide useful insights into the time-resolved oscillation structure of mosquito data and accompanying revealing a non-stationary association with temperature.

7.2.3 Damos, P., Tuells, J. Caballero, P. 2021. Soft computing of a medically important arthropode vector with autoregressive recurrent and focused time delay artificial neural networks. Insects, 2021. https://www.mdpi.com/2075-4450/12/6/503

Article

Soft Computing of a Medically Important Arthropod Vector with Autoregressive Recurrent and Focused Time Delay Artificial Neural Networks

Petros Damos [1,2,*], José Tuells [2] and Pablo Caballero [2]

1 Department of Community Nursing, Preventive Medicine and Public Health and History of Science, Faculty of Health Science, University of Alicante, 03080 Alicante, Spain
2 University General Hospital of Thessaloniki, AHEPA, Kiriakidi 1, 54621 Thessaloniki, Greece; tuells@ua.es (J.T.); pablo.caballero@ua.es (P.C.)
* Correspondence: damos@agro.auth.gr or petrosdamos@gmail.com or pd21@alu.ua.es

Simple Summary: Arthropod vectors are responsible for transmitting a large number of diseases, and for most, there are still not available effective vaccines. Vector disease control is mostly achieved by a sustained prediction of vector populations to maintain support for surveillance and control activities. Mathematical models may assist in predicting arthropod population dynamics. However, arthropod dynamics, and mosquitoes particularly, due their complex life cycle, often exhibit an abrupt and non-linear occurrence. Therefore, there is a growing interest in describing mosquito population dynamics using new methodologies. In this work, we made an effort to gain insights into the non-linear population dynamics of *Culex* sp. adults, aiming to introduce straightforward soft-computing techniques based on artificial neural networks (ANNs). We propose two kind of models, one autoregressive, handling temperature as an exogenous driver and population as an endogenous one, and a second based only on the exogenous factor. To the best of our knowledge, this is the first study using recurrent neural networks and the most influential environmental variable for prediction of the WNv vector *Culex* sp. population dynamics, providing a new framework to be used in arthropod decision-support systems.

Abstract: A central issue of public health strategies is the availability of decision tools to be used in the preventive management of the transmission cycle of vector-borne diseases. In this work, we present, for the first time, a soft system computing modeling approach using two dynamic artificial neural network (ANNs) models to describe and predict the non-linear incidence and time evolution of a medically important mosquito species, *Culex* sp., in Northern Greece. The first model is an exogenous non-linear autoregressive recurrent neural network (NARX), which is designed to take as inputs the temperature as an exogenous variable and mosquito abundance as endogenous variable. The second model is a focused time-delay neural network (FTD), which takes into account only the

7.2.4 Damos, P., Tuels, J., Caballero, P. 2021. Predictive modelling of seasonal mosquito population patterns with neural networks. Proccedings of the 1St International Electronic Conference on Entomology, 1-15 July, https://iece.sciforum.net/ (accepted for publication)

Predictive modeling of seasonal mosquito population patterns with neural networks [†]

Petros Damos [1,2], José Tuells [2] and Pablo Caballero [1,*]

[1] Department of Community Nursing, Preventive Medicine, Public Health and History of Science, Faculty of Health Science, University of Alicante, Alicante, Carretera San Vicente s/n 03690 San Vicente del Raispeig, Alicante; (P.D) pd21@alu.ua.es, (J.T) tuells@ua.es, (P.C) pablo.caballero@ua.es
[2] University General Infectious Disease Hospital of Thessaloniki AHEPA, Aristotle University of Thessaloniki, Greece, petrosdamos@gmail.com
e-mail@e-mail.com
[*] Correspondence: petrosdamso@gmail.com
[†] Presented at the 1st International electronic conference on Entomology, 01/07/2021-15/07/2021.

Abstract: Mosquito species are considered important vectors of many diseases in humans, companion animals, and livestock. There is a great need to understand their dynamics and to develop methods for predicting their abundances. However, the population dynamics of mosquitoes are often complex displaying non-linear dynamics and thus, making it difficult to be modeled using linear statistical approaches. In this project, we explored the seasonal population patterns of mosquito populations in a Mediterranean environment in Northern Greece using straightforward machine learning techniques such as Artificial Neural Networks (ANNs). To train, validate and test the network model we have used 2 years weekly counts of adult mosquito data including *Culex sp.*, a major vector of the West Nile virus and related encephalitis diseases. The model training was performed in an open-loop (i.e., parallel series network architecture), including the validation and testing step and later on, after training, it was transformed to a closed-loop for the needs of a multistep-ahead mosquito abundance prediction. Determined by the autocorrelation function, one of the final models is using as inputs one week lagged values of mosquito abundances and was able to capture the adult seasonal mosquito patterns in most cases at acceptable levels. We conclude that ANNs suggest an important candidate for modeling and predicting the seasonal abundance of mosquito data since it is suitable for modeling noisy and incomplete ecological data, with no specific assumptions to be made about the underlying relationships and which are solely determined through data mining. However, we are also looking forward to improving the particular model performance using new data sets since it is of fundamental importance to choose an appropriate training set size and to provide representative coverage of all possible conditions to capture accurately the patterns of ecological time series. Nevertheless, despite the limitations of the current study, this work contributes to knowledge of the seasonal functioning of arthropod vector dynamics and contributes towards the development of decision tools to be used in the preventive management of the transmission cycle of vector-borne diseases.

7.2. Unpublished work-unnder consideration

7.2.1 Damos, P., Dorrestijn, J, Thomidis, T., Tuells., T, Caballero, P. 2021. A climate conditioned Markov chain model for predicting the dynamics of medical important diseases vectors

Abstract: Understanding and predicting mosquito population dynamics is crucial for gaining insight into the abundance of arthropod disease vectors, and for the design of effective vector control strategies. In this work, a climate-conditioned Markov Chain (CMC) model was developed and applied for the first time to predict the dynamics of vectors of important medical diseases. Temporal changes in mosquito population profiles were generated to simulate the probabilities of a high population impact. The probabilities achieved from the trained model are very near to the observed data and the CMC model satisfactorily describes the temporal evolution of the mosquito population process. In general, our numerical results indicate that it is more likely for the population system to move into a state of high population level, when the former is a state of a low population level than the opponent. Field data on frequencies of successive mosquito population levels, which were not used for the data inferred MC modeling, were assembled to obtain an empirical intensity transition matrix and the observed frequencies. Our findings match to a certain degree the empirical results in which the probabilities follow analogous patterns while no significant differences were observed between the transition matrices of the CMC model and the validation data (ChiSq=14.58013, df=24, p=0.9324451). The proposed modeling approach is a valuable eco-epidemiological study. Moreover, compared to traditional Markov chains, the benefit of the current CMC model is that it takes into account the stochastic conditional properties of ecological related climate variables. The current modeling approach could save costs and time in establishing vector eradication programs and mosquito surveillance programs.

Keywords: Stochastic process; Forecasting techniques, Decision Making, Disease Vectors, West Nile virus, Public Health.

7. 3. Activities and courses taken during the doctorate

7.3.1 2018-2019

3.1 DOCUMENTO COMPROMISO DOCTORAL

Tipo:DOCUMENTO COMPROMISO DOCTORAL Fecha de creación:28/02/2019

Estado actualizado por:Ángeles Sirvent Ramos (28/02/2019)Mostrar detalle actividad

3.2 ACTIVITY 1: Tools for the administration and recovery of information

Tipo:ACTIVIDADES FORMATIVAS COMUNESFecha de creación:05/03/2019Estado actualizado por:Pablo Caballero Pérez (15/05/2019)

3.3 ACTIVIDAD 1: HERRAMIENTAS PARA LA GESTIÓN Y RECUPERACIÓN DE LA INFORMACIÓN

Tipo:ACTIVIDADES FORMATIVAS COMUNESFecha de creación:14/05/2019Estado actualizado por:Ángeles Sirvent Ramos (14/05/2019)

3.4 actividad formativa específica SEMINARIOS J - PROYECTOS DE TESIS - 10 May

Tipo:AFTOE: SEMINARIO J (PROYECTOS DE TESIS).Fecha de creación:24/05/2019Estado actualizado por:Pablo Caballero Pérez (20/06/2019)

3.5 ACTIVIDAD 2: FINES Y OBJETIVOS DE LA INVESTIGACIÓN

Tipo:ACTIVIDADES FORMATIVAS COMUNESFecha de creación:27/05/2019Estado actualizado por:Ángeles Sirvent Ramos (25/05/2019).

3.6 ACTIVIDAD 3: MODELOS DE COMUNICACIÓN CIENTÍFICA

Tipo:ACTIVIDADES FORMATIVAS COMUNESFecha de creación:28/05/2019Estado actualizado por:Ángeles Sirvent Ramos (28/05/2019)

3.7 ACTIVIDAD 4: MODELOS DE TRANSFERENCIA DEL CONOCIMIENTO
Tipo:ACTIVIDADES FORMATIVAS COMUNESFecha de creación:26/06/2019Estado actualizado por:Ángeles Sirvent Ramos (25/06/2019)

7.3.2. 2019-2020

3.8 Detailed PhD Research Plant Report 2019-2020
Tipo:DOCUMENTO COMPROMISO DOCTORALFecha de creación:25/03/2020Estado actualizado por:Pablo Caballero Pérez (07/07/2020)

3.9 CURSO FORMACION ONLINE SCOPUS –APIS
Tipo:ACTIVIDADES FORMATIVAS OPTATIVAS DEL PROGRAMA DE DOCTORADOFecha de creación:25/03/2020Estado actualizado por:Pablo Caballero Pérez (07/07/2020)

7.3.3. 2020-2021

3.10. Recognition of specific activity1 through participation in research project
Tipo:AFTOE:¿CÓMO FINANCIAR UN PROYECTO DE INVESTIGACIÓN?Fecha de creación:01/05/2021Estado actualizado por:Pablo Caballero Pérez (01/05/2021)

8. Research internships

1/1/2020-12/12/2020

International Hellennic University. Department of Nutritional Sciences and Dietetics, Greece.