

Predictive Modeling of Seasonal Mosquito Population Patterns with Neural Networks [†]

Petros Damos ^{1,2}, José Tuells ² and Pablo Caballero ^{2,*}

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 Ralspeig, 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 (IECE 2021), 1–15 July 2021; Available online: <https://iece.sciforum.net/>.

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.

Keywords: *Culex sp.*; medical entomology, vector borne diseases; public health

Citation: Damos, P.; Tuells, J.; Caballero, P. Predictive Modeling of Seasonal Mosquito Population Patterns with Neural Networks. *Proceedings* **2021**, *68*, x.
<https://doi.org/10.3390/xxxxx>

Published: 1 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vector-Borne Diseases (VBD) are human illnesses caused by parasites, viruses and bacteria that are mostly transmitted by arthropod vectors. The impact to public health due to vector - borne diseases is significant. The major vector-borne diseases account for about

17% of all infectious diseases, while most of them are transmitted through mosquito species. Malaria to date causes globally more than 400,000 deaths every year, most of them are children under the age of 5 [1, 2, 3, 4].

Until recently the highest incidence of mosquito transmitted diseases was observed in the tropical and subtropical regions. However, during the past years several neglected vector-borne diseases that were occasionally sporadic and have been eradicated get to a dynamic reappearance and cause outbreaks in temperate climates. Malaria, for instance, had been degraded while dengue fever was endemic in emerging countries such as sub-Saharan Africa. Nevertheless, several mosquito-borne diseases have emerged in Europe in recent years; these include vivax malaria, West Nile fever, dengue fever, Chikungunya fever and Zika virus etc. [4, 5, 6].

Since there are no vaccines against diseases transmitted by arthropods the principal and most efficient way to prevent them is through knowledge of their population dynamics in order to apply timely vector control measures. Lack of understanding of the importance of proper treatment of arthropods of health importance leads to the ineffectiveness of the measures taken, often leading to risks to public health and waste of public money. Therefore, it is important to develop new tools and methods to be used in decision making and to improve the control of arthropod vectors [7].

Mathematical models provide mean to simulate and predicted the behavior of ecological processes such as arthropod vector dynamics to be used later on for decision making [8, 9, 10]. To date most models used to simulate vector population dynamics are of deterministic nature, and need a priori knowledge of arthropod demographic parameters from Laboratory studies [11, 12]. However, arthropod specific life cycle information is not always available for particular regions nor from Laboratory trials. Moreover, under field conditions real life mosquito population dynamics is actually non-linear, abrupt and noisy.

Artificial Neural Network models, or simply network models, are an alternative candidate of available deterministic to be used in modeling ecological time series and arthropod vector population dynamics particularly [13,14]. ANNs models are using empirical and semi-parametric methods which are inspired by human brain to simulate complex functions [15,16].

One major asset of an ANN model is that is able to perform in an enviable way parallel processing of data and information, which traditional models simply cannot. Additionally, because there is a dynamic and feedback relation between the data used to train the model and its prediction outputs, the model performance can be continuously improved if more data are available to be used for training [16,17]. This is called machine learning [18].

The aim of the current work is to introduce and popularize ANNs in the fields of arthropod vector dynamics and medical entomology to be used to study their dynamics and outline the potentially to be used as decision tools to prevent vector borne diseases. We consider ANN models relevant to simulate the phenology of arthropod vectors and mosquitoes particularly, to understand their dynamics and seasonal population patterns. Moreover, the current mathematical and numerical simulations do not involve laboratory trials, and are based solely on available field abundance data and thus imply economy of time and resources.

2. Materials and Methods

2.1. 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 [18,19]. 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 [19, 20].

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.

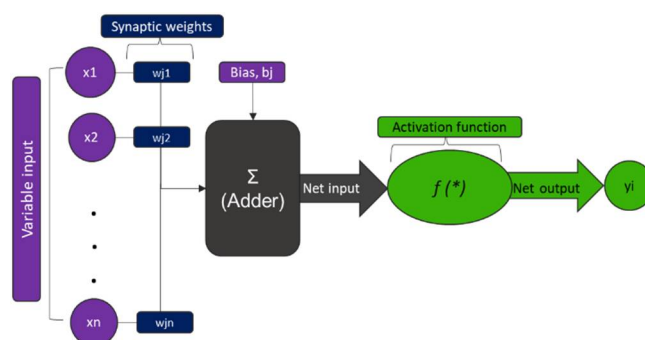


Figure1. Simplified neuron which consists of the structural unit of an ANN.

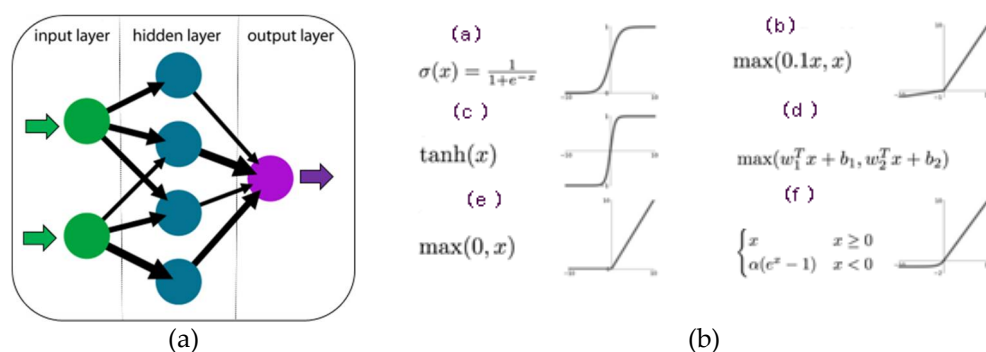


Figure2. Simplified ANN that consists of three layers: the input layer (variables used to train the model), the hidden layer (which consist of four neurons) and the output layer (a) and a summary of the most common activation functions, f (b). The functions are (a) sigmoid, (b) Leaky ReLU, (c) $\tanh(x)$, (d) Maxout, (e) ReLU, and (f) ELU.

2.2. Autoregressive neural network models

For this application we have developed and applied a self-regulating non-linear autoregressive neural network (NAR) working with an external mosquito abundance time series [21]. The model is as follows:

$$y(t)=f(y(t-1),y(t-2),\dots,y(t-p))+e(t) \quad (1)$$

Where y is the mosquito abundance which depends on previous mosquito population values p and $e(t)$ 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 sequency. Furter, 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 sequency changes as the separation changes. Thus, it is a time domain measure and provides a criterion of defining the memory of the population process.

Further we have used a nonlinear autoregressive model with exogenous input (NARX). The model can be defined as:

$$y(t)=F(y(t-1),y(t-2),\dots,u(t), u(t-1), u(t-2), u(t-3),\dots)+e(t) \quad (2)$$

Where y is the studied variable, arthropod vector abundance, and u the exogenous and independent variable. In this study we have considered temperature as the exogenous d independent variable. The above expression says that the information about the exogenous value of u helps to predict y along with the previous values of y . The error term is $e(t)$. We used to two different iteration methods for models training, the Levenberg–Marquart algorithm and the scale conjugate gradient algorithm was used to train the models. This algorithm takes less memory. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples. Moreover, for both neural networks we have used a hyperbolic tangent sigmoid transfer function in the inner layer and a pure linear function in the output layer [21].

The model validation was based on the coefficient determination R^2 of the predicted data in relation to the observed as well the autocorrelations and error distribution. The same method was repeated for different training sets to judge whether the network is appropriate to make accurate predictions. The training follows the reverse transmission algorithm error, where at the end of each training cycle the average square error is evaluated and adjusted the synaptic weights.

2.2. Mosquito surveillance data

Public mosquito trap data available from the open European Union Data Portal (EU ODP) (<http://data.europa.eu>, accessed on 3 May 2019) were used for the study analysis. Mosquito surveillance data included adult *Culex sp.* which were captured CO2 traps from mid-May until September and during two successive observation years (2011 and 2012). Data were sampled from 11 closely related locations in central Macedonia and Greece. The observation area includes semi-urban areas and agricultural landscapes with similar habitat characteristics. Data were pooled in order to have an ecological time series of *Culex sp.* populations. Climate data, and in particular, mean air temperatures, were obtained by the national observatory of Athens through a meteorological station, which was located in Makrochori town, which was in the same level and nearby the mosquito observation area (<http://stratus.meteo.noa.gr/front>, accessed on 2 April 2020). Data were handled as vectors, which consisted of close-to-weekly time intervals of the number of adult mosquitoes captured and were normalized before the analysis.

3. Results

Figure 1a, shows the mosquito abundance autocorrelation function and Figure 1b the autocorrelation function for the temperature in respect to different time lags. In both cases the autocorrelations do not indicate any abrupt fluctuations and/or clear periodic pattern in the mosquito time series with regard to different time lags. For temperature, particularly, the autocorrelation after some perturbations later decay to zero, indicating possible

the existing of a moving average process, although there is some ambiguity regarding the different patterns observed between mosquito abundance and temperature. Based on the autocorrelations we conclude that one- or two-weeks lag values should be taken into account in modeling mosquito population dynamics.

Figure 2a and 2b show the architecture of the NAR and the NARX model respectively. Both models predict the mosquito population value based on 1- and 2-week previous population values (i.e., delay1 and delay2). However, the NARX model takes into account the temperature variable as exogenous factor, additionally to the previous mosquito population values, to simulate the mosquito population dynamics.

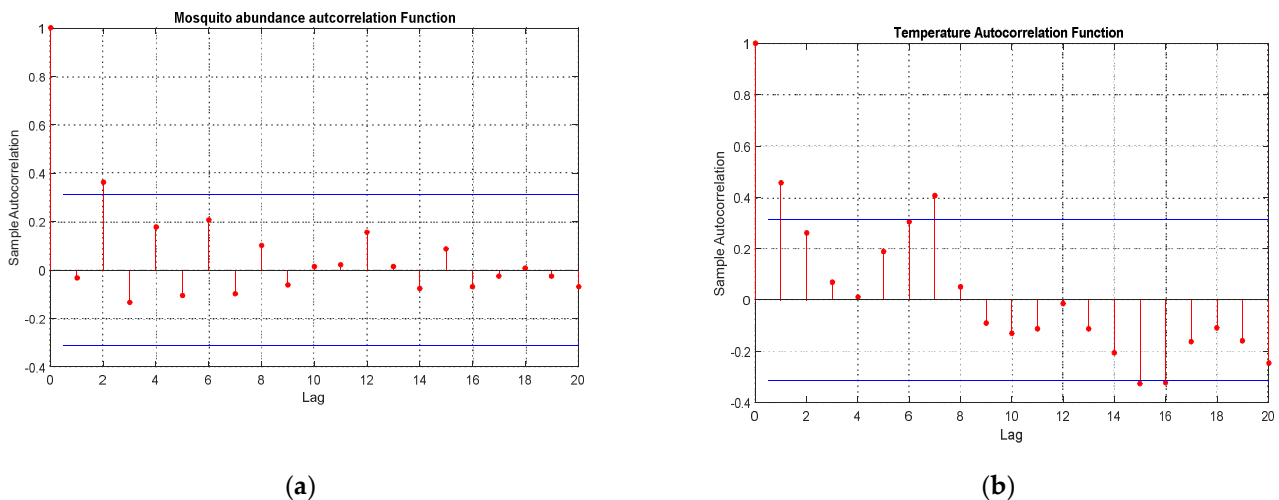


Figure 1. Autocorrelations of mosquito abundance data (a) and autocorrelations of temperature (b) in respect to different time lags (weeks).

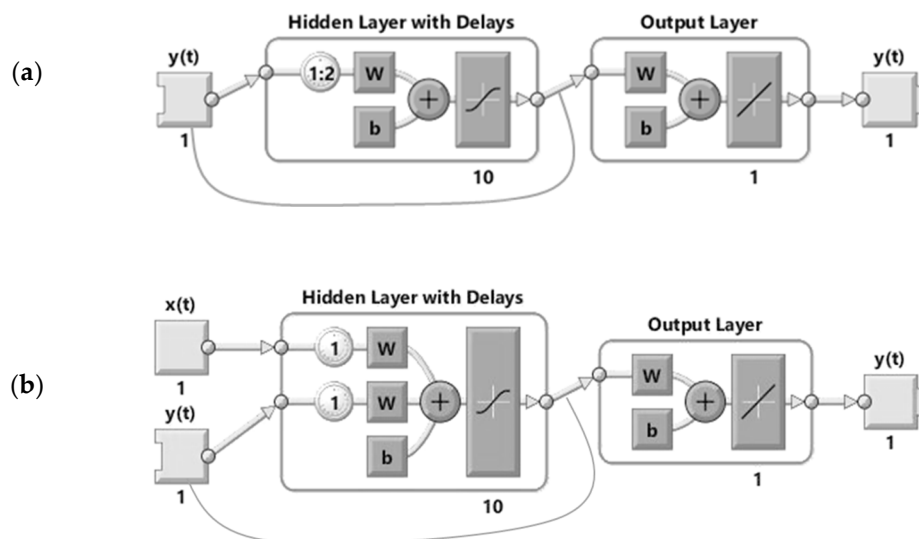


Figure 2. Architecture of the non-linear autoregressive network model (NAR) (a) and the non-linear autoregressive model with exogenous outputs (b). $y(t)$: mosquito abundance variable, $x(t)$: temperature variable, W : weight, b : bias, 1:2, delays 1 and 2, 10: number of hidden neurons. Both models are using as activation functions a hyperbolic tangent sigmoid transfer function in the inner layer and a pure linear function in the output layer. The network structure was generated with the MatLab neural network toolbox [21].

Figures 3a and 3b, presents the response outputs of the NAR and NARX neural network 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 where 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.

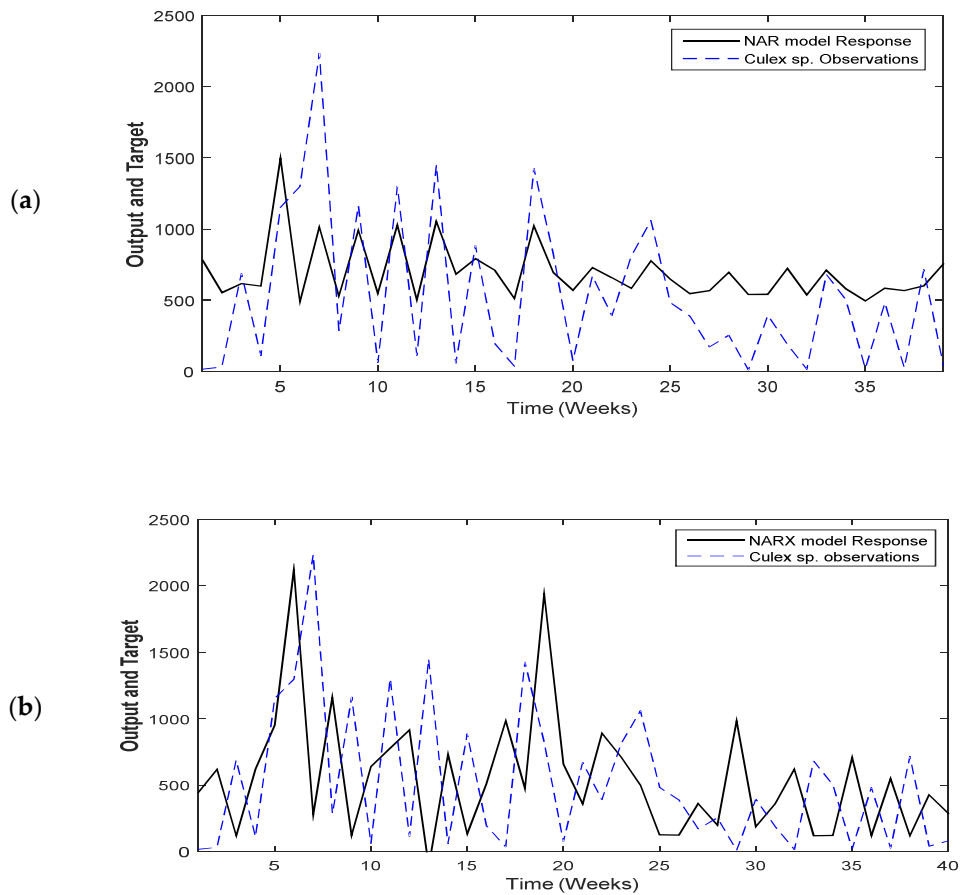


Figure 3 Response of the NAR and the NARX neural networks respectively model output to the mosquito population time series.

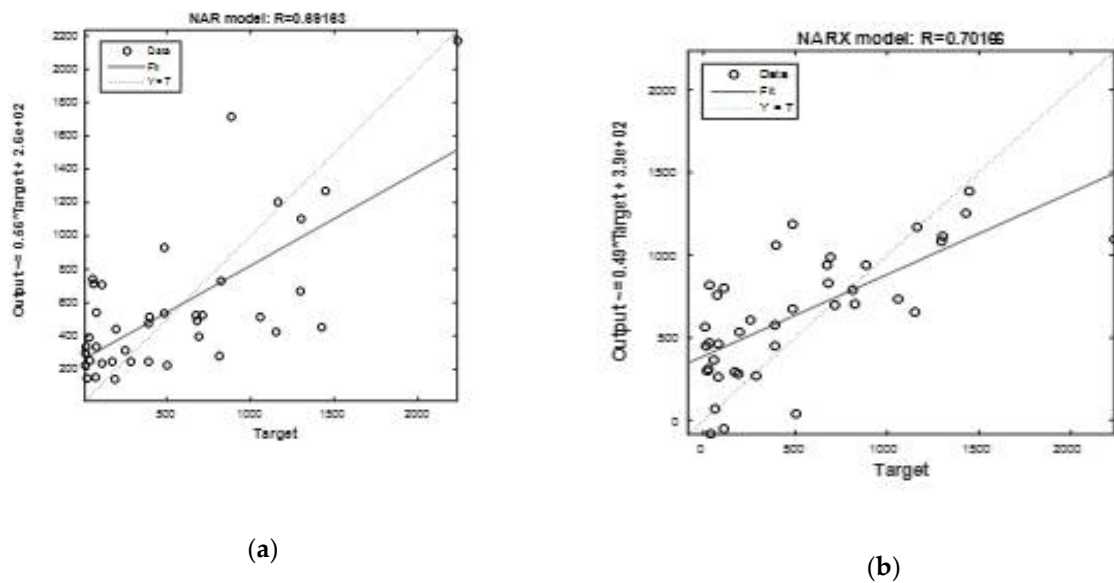


Figure 4 Response of the NAR and the NARX neural networks respectively model output to the mosquito population time series.

Figure 4a and 4b, shows the overall model performances for the NAR and the NARX models, respectively. For both model predictions are at acceptable levels and the coefficient determination was $r=0.69$ and $r=0.7$, for the NAR and the NARX model, respectively.

4. Discussion

In this work we have demonstrated how ANNs can be applied to model the population dynamics of arthropod disease vectors and mosquitoes particularly. We have used self-regulating prediction neural networks the first without and the second with temperature as an external auxiliary time series. Based on the results obtained from each training it is further judged whether the network is appropriate to make a prediction or not. To date that whenever the network is re-trained it generate different results, as the training algorithm follows an iterative process which in each test converges to different results. Therefore, a combination of available data and tests were performed with various parameters in order to train the network and to derive to the best fitting models.

Based on the results, both models performed at acceptable levels and described the temporal evolution of mosquito population dynamics. However, the NARX model, which takes in to account temperature as exogenous factors, performed better compared to the simple NAR model. This is in accordance with previous studies that have demonstrated that the emergence and outbreak of vector-borne diseases is to high degree related of alterations in current environmental conditions and temperatures particularly [22], but also to other factors; including climate change and favorable environment conditions for vector breeding, economic downturns affecting health policies, travel and human migrations [23, 24].

Furthermore, climate change particularly is one of the most important causes of vector born disease emergence since it affects directly the population dynamics and geographic expansion of arthropod vectors and should be therefore taken in to account when modeling arthropod vector dynamics [22,23]. Arthropod vectors, such as mosquitoes, are poikilotherms species and any increase of mean temperatures may cause direct effects on their development, number of generations and geographical dispersion. Additionally, the temporal and spatial changes in temperature, precipitation and humidity that occur under

different climates may affect the biology and ecology of vectors and intermediate hosts and consequently the likelihood of disease transmission [25].

It is possible that all these factors to affect the final mosquito population dynamics and its abrupt dynamics as appeared under natural field conditions. The time horizon of the prediction, in the context of this research, is set to one week which corresponds to a regular time line used in most entomological studies, although can also not strictly defined. In the literature, predictions that have been made by a well-trained network can reach the number of already known values and this has been also observed in this study. However, in practice the ANN model performance depends on the nature of the data time series, the proper network configuration as well as the data set used for validating and model testing. Thus, current evidence suggests that inter-annual *Culex* sp. population dynamics and climate variability have a direct effect of their temporal evolution and abundance and can be predicted through ANNs.

In conclusion, an overall evaluation of the proposed model results and the factors used for their development and training, suggest that they have a strong potential to be used to predict the non-linear population dynamics of arthropod vectors. From a public health point of view, they are utile to decide whether to implement integrated vector management, allowing the development of better pharmaceutical and preventive methods and for designing effective public health management policies in local and regional scales.

Author Contributions: Conceptualization and investigation, P.D.; methodology, P.D. and P.C.; software, data curation, and formal analysis, P.D., writing—original draft preparation, P.D.; writing—review and editing, J.T. and P.C. J.T., supervision, P.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Non-applicable

Informed Consent Statement: Non- applicable

Data Availability Statement: Publically available datasets were analyzed in this study. This data can be found here: [<http://data.europa.eu> (Accessed on 3 May 2019)] and here: [<http://stratus.meteo.noa.gr/front> (Accessed on 2 April 2020)].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Mosquito Borne Diseases. 2019. Available online: https://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/ (Accessed on 20 April 2021).
2. Gratz, N.G. *The Vector- and Rodent-Borne Diseases of Europe and North America: Their Distribution and Public Health Burden*; Cambridge University Press: Cambridge, UK, 2006.
3. Brugman, V.A.; Hernández-Triana, L.M.; Medlock, J.M.; Fooks, A.R.; Carpenter, S.; Johnson, N. The Role of *Culex pipiens* L. (Diptera: Culicidae) in Virus Transmission in Europe. *Int. J. Environ. Res. Public Health* **2018**, *15*, 389.
4. World Health Organization and UNICEF, "Global vector control response 2017-2030," 2017
5. Gubler, D.J. The Continuing Spread of West Nile Virus in the Western Hemisphere. *Clin. Infect. Dis.* **2007**, *45*, 1039–1046.
6. Rocklöv J., Dubrow R., "Climate change: an enduring challenge for vector-borne disease prevention and control," *Nature Immunology*, vol. 21, no. 5, pp. 479–483, 2020.
7. Yang G-J, Brook BW, Bradshaw CJA (2009) Predicting the Timing and Magnitude of Tropical Mosquito Population Peaks for Maximizing Control Efficiency. *PLoS Negl Trop Dis* 3(2): e385. doi:10.1371/journal.pntd.0000385
8. Reed, Hanna, "Mathematical Models of Mosquito Populations" (2018). Honors Undergraduate Theses. 299.
9. Mageni, M.L., Penny, M.A., Smith, Th., Chitnis N., Mathematical modelling of mosquito dispersal in a heterogeneous environment. *Math. Biosc.*, **2013**, *241*, 198-216.
10. Diuk-Wasser, M.A.; Brown, H.E.; Andreadis, T.G.; Fish, D. Modeling the Spatial Distribution of Mosquito Vectors for West Nile Virus in Connecticut, USA. *Vector-Borne Zoonotic Dis.* **2006**, *6*, 283–295.
11. Westbrook CJ, Reiskind MH, Pesko KN, Greene KE, Lounibos LP. Larval environmental temperature and the susceptibility of *Aedes albopictus* Skuse (Diptera: Culicidae) to Chikungunya virus. *Vector Borne Zoonotic Dis.* 2010;10(3):241-247. doi:10.1089/vbz.2009.0035
12. Rueda, L.M.; Patel, K.J.; Axtell, R.C.; Stinner, R.E. Temperature-Dependent Development and Survival Rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). *J. Med. Entomol.* **1990**, *27*, 892–898.

13. Chon, T.S.; Park, Y.S.; Kim, J.M.; Lee, B.Y.; Chung, Y.J.; Kim, Y. Use of an artificial neural network to predict population dynamics of the Forest–Pest pine needle gall midge (*Diptera: Cecidomyiida*). *Environ. Entomol.* **2000**, *29*, 1208–1215.
14. Lee, K.Y.; Chung, N.; Hwang, S. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecol. Inform.* **2016**, *36*, 172–180.
15. Akhtar, M.; Kraemer, M.U.G.; Gardner, L.M. A dynamic neural network model for predicting risk of Zika in real time. *BMC Med.* **2019**, *17*, 171.
16. 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.
17. Bianconi, A.; Von Zuben, C.J.; Serapião, A.B.S.; Govone, J.S. Artificial neural networks: A novel approach to analysing the nutritional ecology of a blowfly species, *Chrysomya megacephala*. *J. Insect Sci.* **2010**, *10*, 58.
18. Recknagel, F. Applications of machine learning to ecological modelling. *Ecological Modelling* **146** (2001) 303–310
19. Schmidhuber, J. “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
20. Browne, A., *Neural Network Analysis, Architectures and Applications*. Institute of Physics Publishing, Bristol and Philadelphia (1997).
21. Matlab, R2015b; The Mathworks Inc.: Natick, MA, USA, 2015.
22. Damos, P.; Caballero, P. Detecting seasonal transient correlations between populations of the West Nile Virus vector *Culex* sp. and temperatures with wavelet coherence analysis. *Ecol. Inform.* **2021**, *61*, 101216.
23. Rocklöv J., Dubrow R., “Climate change: an enduring challenge for vector-borne disease prevention and control,” *Nature Immunology*, vol. 21, no. 5, pp. 479–483, 2020.
24. Cosner C, Beier JC, Cantrell RS, et al. The effects of human movement on the persistence of vector-borne diseases. *J Theor Biol.* **2009**;258(4):550-560. doi:10.1016/j.jtbi.2009.02.016
25. Morin C. W., Comrie A. C., “Modeled response of the West Nile virus vector *Culex quinquefasciatus* to changing climate using the dynamic mosquito simulation model,” *International Journal of Biometeorology*, vol. 54, no. 5, pp. 517–529, 2010.