Impact of climate variability on mosquito occurrence and malaria transmission in Greece

Aristotle University of Thessaloniki

Maria-Chara Karypidou

Thessaloniki 2017

Maria-Chara Karypidou
Geographer

Impact of climate variability on mosquito occurrence and malaria transmission in Greece

Submitted to the School of Geology in the context of the M.Sc. course in
"Meteorology, Climatology and Atmospheric Environment"
Department of Meteorology and Climatology
Date of Oral Defense: 13/12/2017

Supervisor:

Dr. **Eleni Katragkou**, Assistant Professor - Department of Meteorology and Climatology, School of Geology, Aristotle University of Thessaloniki, Greece

Advisory Committee:

Dr. **Antonios Mazaris**, Assistant Professor - Department of Ecology, School of Biology, Aristotle University of Thessaloniki, Greece

Dr. **Adrian Tompkins**, Research Scientist - Earth System Physics, The Abdus Sallam International Center for Theoretical Physics, Trieste, Italy

Impact of climate variability on mosquito occurrence and malaria transmission in Greece

To my beloved parents, Ioannis and Anatoli and to my sweet grandmother, Maria.

Acknowledgements

After my undergraduate journey, the Department of Meteorology and Climatology at the Aristotle University of Thessaloniki followed as a beautiful surprise. I was not aware what a rich continuation that would have been. Of course, that is due to the things I learnt and the people I met. Here I want to express my thanks to some of them.

My time here at the Department as an MSc student would not have been such an enjoyable and stimulating experience, had it not been for my supervisor Dr. Eleni Katragkou. I am immensely thankful for her mentorship, guidance, encouragement and support and very much looking forward to what lies ahead! Also, I would like to thank Dr. Antonios Mazaris for his valuable help and guidance concerning the application of ecological niche models. In addition, I would like to thank Dr. Adrian Tompkins for his bright ideas and his help with VECTRI and also, his kind hospitality during my staying in Trieste.

The current dissertation would not have been possible without the provision of the mosquito data. For this I acknowledge and thank Ecodevelopment S.A. and especially, Dr. Spyridon Mourelatos and Ms. Sandra Gewehr. Furthermore, I would like to thank the Hellenic Center for Disease Control and Prevention for the provision of the epidemiological data.

Additionally, I would like to thank Ms Vasiliki Almpanidou and Dr. Athanasios Tsikerdekis for being always willing to assist and offer an advice. Besides, I would like to thank the staff working in the Grid Infrastructure here in AUTh for their help and more specifically, Ms. Alexandra Charalambidou for always replying with a kind and insightful e-mail to all the technical problems that would arise along the way. Also, I want to thank the "unseen heroes" of the fantastic Stack Exchange community for their massive help with respect to the proper use of R.

Moreover, as this journey of the MSc course comes to its end, I must not forget to express my thanks and appreciation to the people who offered generously their help and support to me, even while being an undergraduate student at the Department of Geography in Mytilini. These people are Prof. Phaedon Kyriakidis, Assist. Prof. Efthymia Kostopoulou, Assoc. Prof. Sotirios Koukoulas and Prof. Georgios Sidiropoulous.

Of course, I would like to extend my greatest gratitude to my dear family for always believing in me and encouraging my endeavors with love.

Above all though, I would like to thank my dearest Lord and Saviour Jesus Christ, *"in Whom are hidden all the treasures of wisdom and knowledge" - Epistle to the Colossians 2:3*

Maria-Chara Karypidou

Abstract

In the current work, five statistical and a dynamical model were employed for the studying of the spatio-temporal attributes of *An. sacharovi*. The data that were employed in the current analysis consisted of climatic data obtained from the WorldClim database that provides gridded climatic data covering the period 1950-2000. Also, eight Regional Climate Model outputs were obtained from the ESGF database, containing evaluation experiments in the context of the EURO-CORDEX initiative at 0.11° spatial resolution. These data were employed for the calculation of certain bioclimatic variables, proposed through the WorldClim project. The information concerning the presence of *An. sacharovi* mosquitoes was extracted from the vector database created and maintained by Ecodevelopment S.A. and the epidemiological data were provided by the HCDCP. The methodological tools applied consist of a certain set of statistical models and a dynamical model. The statistical models provided an estimation of the environmental suitability for *An. sacharovi* based on six explanatory variables and a set of *An. sacharovi* presence locations. Furthermore, a dynamical model was employed, simulating malaria transmission. According to the statistical models applied, the areas characterized as suitable are located over the plain of central Macedonia in northern Greece, over the plain of Thessaly in central Greece, over the coastal areas of northern Greece and over the plain of Serres, also in northern Greece. Additionally, suitable areas are identified over coastal areas in southern mainland Greece, Attica, Peloponnese and over the eastern islands of the Aegean Sea.

Abbreviations

ANN: Artificial Neural Networks

BRT: Boosted Regression Trees

CCA: Canonical Correspondence Analysis

CIAT: International Center for Tropical Agriculture

CLC: CORINE Land Cover

CORDEX: Coordinated Regional climate Downscaling Experiment

CTA: Classification Tree Analysis

DEM: Digital Elevation Model

DWD: Deutscheur Wetterdienst

ECDC: European Center for Disease Control and Prevention

ECHAM6: European Centre/Hamburg Model

EEA: European Environmental Agency

ENFA: Environmental Niche Factor Analysis

EPSG: European Petroleum Survey Group

ESGF: European System Grid Federation

GAM: Generalized Additive Models

GARP: Genetic Algorithm for Rule-set Prediction

GBIF: Global Biodiversity Information Facility database

GBM: Generalized Boosted Models

GIS: Geographic Information Systems

GLCF: Global Land Cover Facility

GPW: Global Population of the World database

GLM: Generalized Linear Models

GRUMP: Global Rural-Urban Mapping Project

HBI: Human Blood Index

HCDCP: Hellenic Center for Disease Control and Prevention

IPCC: Intergovernmental Panel on Climate Change

MARS: Multivariate Adaptive Regression Splines

ME: Maximum Entropy

MMU: Minimum Mapping Unit

RCM: Regional Climate Model

RCP: Representative Concentration Pathways

RF: Random Forest

WHO: World Health Organization

WMO: World Meteorological Organization

SEDAC: Socioeconomic Data and Application Center

SRE: Surface Range Envelope

SRTM: Shuttle Radar Topography Mission

TAR: Third Assessment Report

TIR: Thermal Infra-Red

VECTRI: VECtor-borne disease community model of the International Centre for Theoretical Physics, TRIeste

WG: Working Group

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Climate and Planetary Health

The climate is known to have a great impact on a plethora of fields and activities, among which the state of human health is also included (IPCC, 2014). The recent research interest in the field of climate science has been justifiably drawn to the ongoing environmental changes and especially to the climatic changes that are in progress (Allan, 2017), due to their great resonance on planetary issues (Parmesan and Yohe, 2002). As a consequence of climate change, the spatio-temporal shifts in environmental and climatic conditions are able to induce further alterations to all the various aspects that climate has an impact on (Thom et al., 2017). This situation is already being observed in the sector of human health (Caminade et al., 2014). However, before proceeding with the discussion concerning the impacts of climate change on the health sector, it is necessary that a definition of climate change be given. According to the Third Assessment Report (TAR) of the IPCC, climate change is defined as following:

"Climate change refers to a change in the state of the climate that can be identified (e.g. using statistical tests) by changes in the mean and/or the variability of its properties, and that persists for an extended period, typically decades or longer. It refers to any change in climate over time, whether due to natural variability or as a result of human activity."

The aforementioned definition alludes to the fact that severe alterations are being observed to a massively dynamic system, such as the climate, whose attributes and mechanisms are not fully understood yet. Such changes inevitably have an impact and can pose further changes and alterations on a wide range of fields, including human health. The fact that even today there remain unanswered questions regarding the climate and its in-lying mechanisms, induces large uncertainties unto the impact modeling of climate change on all the various fields related to climate. The climate-related diseases, as the diseases that are affected by climatic conditions are called, are not an exception to this fact and thus, their modeling results as a perplexing matter.

According to the WHO, it was estimated that in 2012 approximately 23 % of all deaths occurring worldwide, were related to climatic factors, mainly due to the environmental changes taking place, that can function as a force multiplier (Watts et. al., 2016). This type of observed events and their increasing magnitude is the reason why a large amount of effort is currently concentrated on

the process of competent modeling of such dynamic and highly non-linear systems. The purpose of such actions is the construction of efficient early warning systems (Semenza, 2015), especially over vulnerable areas.

A prerequisite to this process, though, is the successful identification of the possible ways through which the climate may exert its impact on the health sector. According to the 5th Assessment Report of the IPCC (2014), conducted by the 2nd Working Group (Impacts, Adaptation, and Vulnerability), there are three fundamental approaches through which climate change displays its impact on the health sector. These delineated pathways are:

1. The Direct impacts, through which extreme weather events are associated directly as the causal effects of health risk (Mora et al., 2017)

2. The Indirect impacts, that are primarily associated with changes in natural systems and ecosystems (Lyons et al., 2012).

3. The human induced impacts, related to undernutrition, population dislocation, mental stress and occupational impacts.

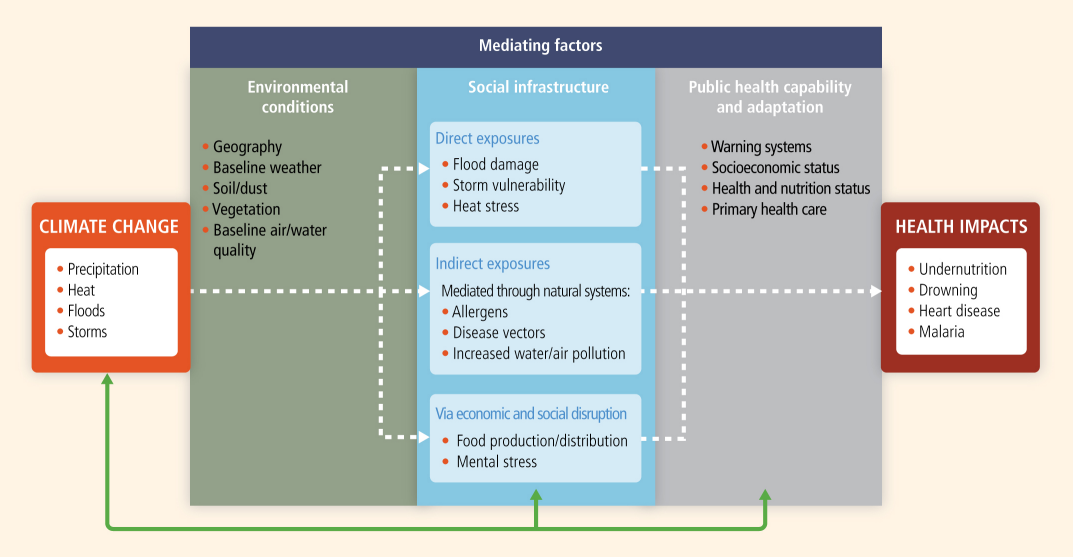


Figure 1.1: Schematic of the possible routes through which climate change may impact the health sector (IPCC, 2014).

As it is obvious from Figure 1.1 and the three pathways discerned by the IPCC, it becomes evident that an efficient solution to the current environmental challenges posing a great threat on human health, could only be approached via an interdisciplinary way (Horton and Lo, 2015). As it is also stated in the 5th IPCC Report (WG2), the geography of a health-related event matters, because:

"Location has an important influence on the potential for health losses caused by climate change (Samson et al., 2011)."

More specifically, the geography of an event is constituted by the ensemble of physical and human conditions prevailing at that particular location. In order for such a research discussion to

be fruitful, scientists from miscellaneous fields need to offer their unique and peculiar perspective in addressing such a challenging issue.

In response to this need, in 2015 the Rockfeller Foundation Lancet Commission initiated a cooperation interested in Planetary Health, which is defined in the following way (Horton and Lo, 2015):

"The achievement of the highest attainable standard of health, wellbeing and equity worldwide through judicious attention to the human systems (political, economic and social) that shape the future of humanity and the Earths natural systems that define the safe environmental limits within which humanity can flourish. Put simply, planetary health is the health of human civilization and the state of the natural systems on which it depends."

This initiative is only indicative of the academic endeavor that is currently in progress with regards to bridging the gap between climate and health. Another initiative of outstanding importance is the establishment of a joint office between the WMO and WHO in 2014 (WMO-WHO, 2016), underlying the cruciality of cooperation between scientists and practitioners in the health sector along with climate scientists. A diagram of the interconnected fabric of climate and health impacts is presented in Figure 1.2 (Watts et. al., 2016). As it is depicted, there are three main disease categories (in purple) that can potentially be affected by climatic changes and additionally, climate change may indirectly cause undernutrition issues or lead to the creation of harmful algal blooms. Also, climatic changes and their impact on numerous natural and societal aspects, may trigger or enhance mental issues.

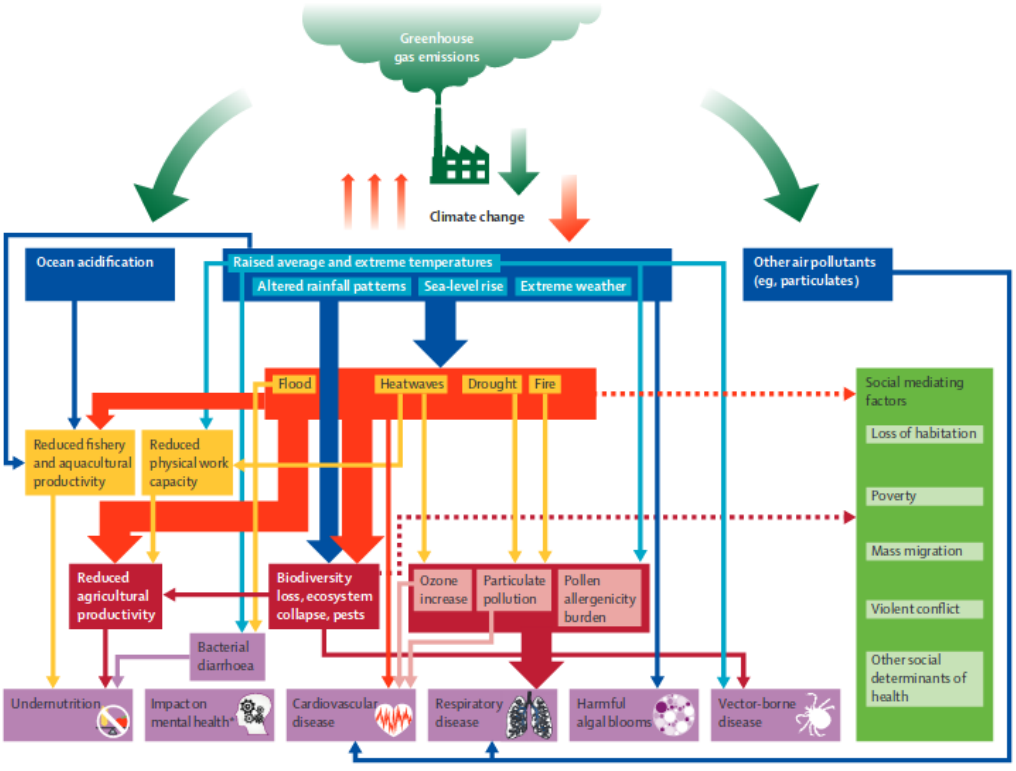


Figure 1.2: The impact of climate change on health (Watts et al., 2015).

The point of interest in the current project is the impact of climate on a certain category of diseases that are called "vector-borne diseases". According to ECDC, vector-borne diseases are:

"Infections transmitted by the bite of infected arthropod species, such as mosquitoes, ticks, triatomine bugs, sandflies, and blackflies."

More specifically, the vector-borne disease in concern is malaria, which is a mosquito-borne disease (transmitted via mosquitoes).

As it is shown in Figure 1.2, vector-borne diseases are primarily affected by the change in the temperature regime over a region and the shift in ecosystem balances due to biodiversity loss, landscape alterations and the overall ecological disturbances caused by extreme weather events and raised average and extreme temperatures.

The question concerning the impacts of climate change on health, is a very timely research question, with a very rapidly growing amount of information flowing in and out of the impact modeling process (Bush et al., 2017). Nevertheless, the assessment of climate change impacts is highly dependent on the existence of information regarding the relation of the disease in concern and the current climate (Parham and Michael, 2010). If such information is not available, then climate change impact modeling is not possible. In many cases, this relation is highly non-linear, posing severe difficulties and introducing considerable uncertainties in modeling even the current state of the disease. A summarized table concerning the relation of climate and certain vector-borne diseases, along with the associated confidence levels is presented in Figure 1.3 (IPCC, 2014).

Disease	Area	Cases per year	Climate sensitivity and confidence in climate effect	Key references
Mosquito-borne diseases				
Malaria	Mainly Africa, SE Asia	About 220 million		WHO (2008); Kelly-Hope et al. (2009); Alonso et al. (2011); Omumbo et al. (2011)
Dengue	100 countries, esp. Asia Pacific	About 50 million		Beebe (2009); Pham et al. (2011); Astrom et al. (2012); Earnest et al. (2012); Descloux (2012)
Tick-borne diseases				
Tick-borne encephalitis	Europe, Russian Fed., Mongolia, China	About 10,000		Tokarevich et al. (2011)
Lyme	Temperate areas of Europe, Asia, North America	About 20,000 in USA		Bennet (2006); Ogden et al. (2008)
Other vector-borne diseases				
Hemorrhagic fever with renal syndrome (HFRS)	Global	0.15–0.2 million		Fang et al. (2010)
Plague	Endemic in many locations worldwide	About 40,000		Stenseth et al. (2006); Ari et al. (2010); Xu et al. (2011)

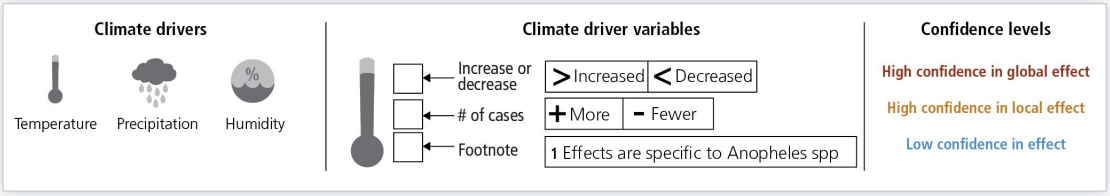


Figure 1.3: Relation of climate and certain vector-borne diseases (IPCC, 2014).

The interest of the current work emphasizes on assessing the current relation of climate and malaria, primarily by assessing the impact of climate on the occurrence and activity of an effective malaria vector (*An. sacharovi*) that is present over the study region (Greece). The ideal goal of the current work would be to extract some quantitative results, describing the relation between the current climate and the disease, so that future studies concerning the impact of climate change on the activity of *An. sacharovi* over Greece, would be possible and meaningful.

1.2 The issue in concern: The Global burden of Malaria

Malaria is a vector-borne disease that is caused by the *Plasmodium* parasite which is transmitted through the bite of the female *Anopheles* mosquito. Malaria is listed among the most well-studied vector-borne diseases (Leedale et al., 2016), due to the great burden that it causes, resulting to more than 200 million infections and causing more than 500.000 deaths every year (WHO, 2016). Approximately 70 % of those deaths occur in children under the age of 5. Malaria is mostly prevalent in the sub-Saharan Africa and SE Asia, but historically malaria was present in the extra-tropics (Figure 1.4), until it was eradicated through the application of malaria and vector control programs. Nevertheless, recent re-emergences of malaria in places where it was previously eliminated, draws attention from a variety of research fields, aiming to investigate the driving forces that caused such events. Briefly, the scientific consensus agrees on the fact that malaria resurgence is caused by a perplexing set of biotic and abiotic factors (Semenza, 2015).

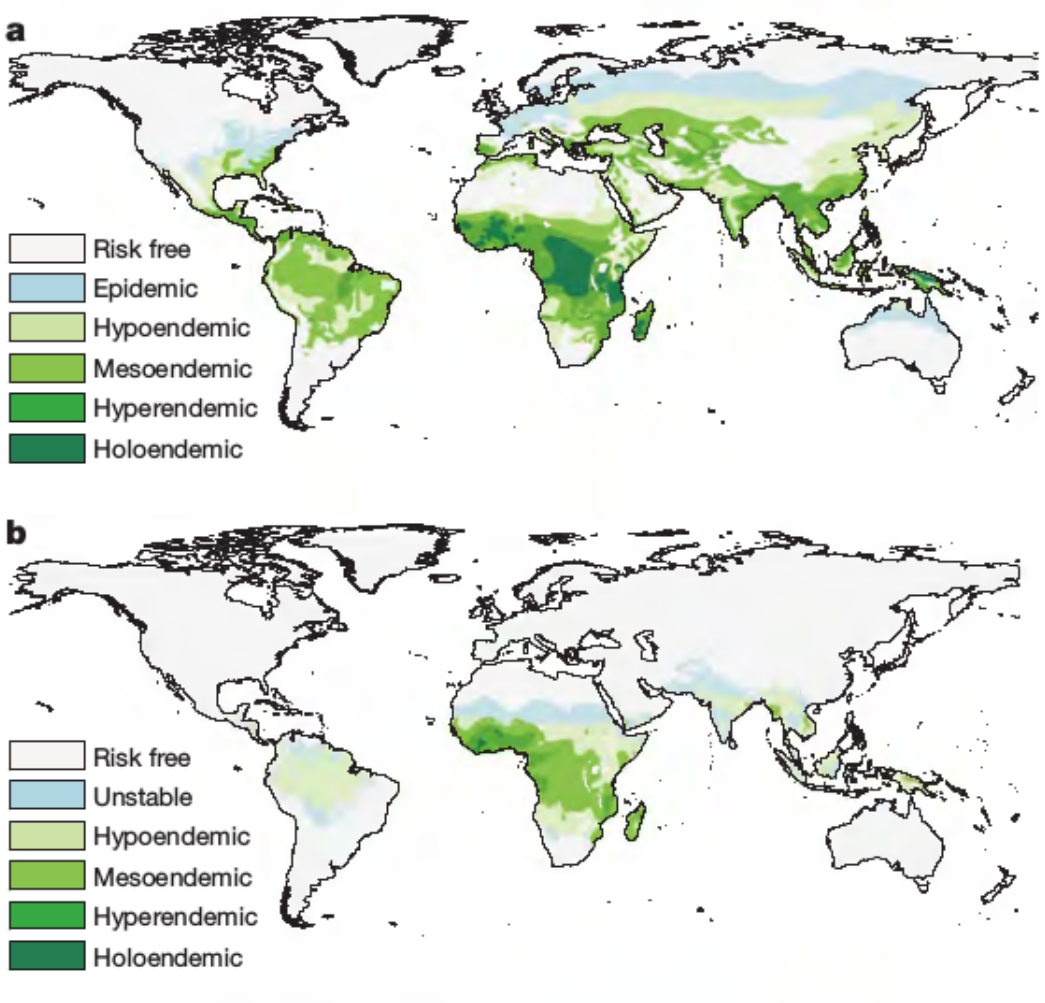


Figure 1.4: Global malaria endemicity. a: Pre-intervention endemicity (1900), b: Contemporary endemicity (2007) (Gething et al., 2010). The state at which the disease is constantly present within a population is characterized as "endemic"

Particular emphasis has been given on the impact of climate on malaria (Pascual et al., 2008).

Such a focus is justifiable and expected, as the qualitative relation between the disease and climate was known for centuries. For instance, the greek word for malaria is "elonosia", meaning the disease of the marshes. In addition, the origin of the word "malaria" translates to bad air in latin (mala = bad and aria = air), suggesting that even the etymology of the words indicate that people in early centuries had an intuitive information that, somehow, the disease is related to certain environmental conditions. This observation still constitutes the basis for the study, modeling and potential forecasting of the disease and huge improvements have been marked so that the pre-existing qualitative knowledge on the relation between climate and malaria may become robustly quantitative.

According to Gething et. al. 2011 a temperature suitability index constructed for *P. vivax* globally, maps well beyond the area of the tropics, delineating regions under the possible impact of malaria occurrence. As it is observed in Figure 1.5, extensive areas in the tropics are categorized as suitable for malaria transmission, while in the extra-tropics the temperature suitability index appears to be more localized and limited in geographic extent. Nevertheless, successful transmission of the disease is also dependent on the occurrence of an effective malaria vector. As it is depicted in Figure 1.6, the areas of the extra-tropics where the temperature suitability index is at medium magnitudes are simultaneously areas where malaria vectors are present. Hence, at any given case at which an infected person is introduced into those areas, transmission could potentially occur.

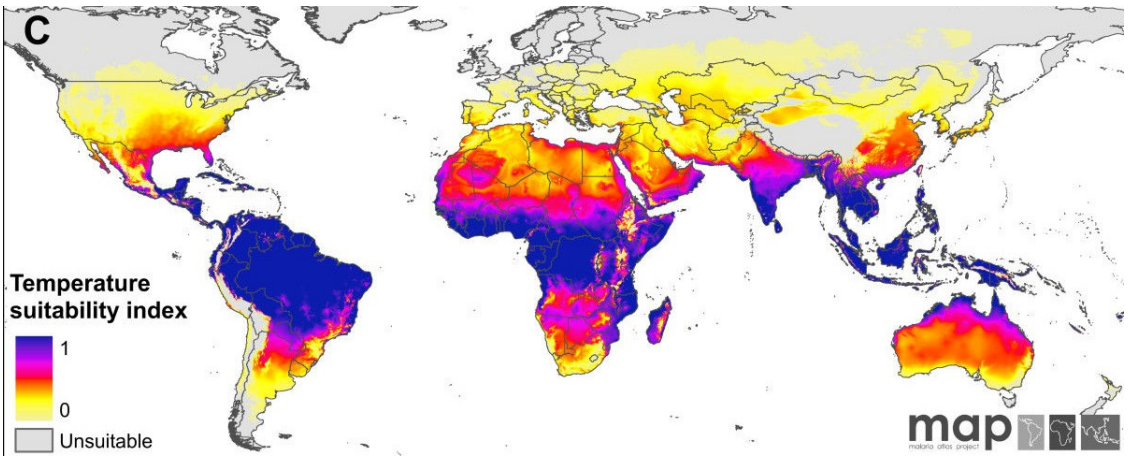


Figure 1.5: The temperature suitability index for P. vivax (Gething et al., 2011).

In this context, the Mediterranean was historically a region where malaria was observed. As it is shown in Figure 1.5, the calculated index displays areas of increased suitability over the Mediterranean and also, according to Figure 1.6, there are *Anopheles* species present, thus historical malaria is explainable over the region of the Mediterranean. More specifically, Greece -which is the focus of the current study- is an area suitable both for vector species capable for transmitting malaria and for *P. vivax* and in fact, malaria was known to be endemic in Greece since 400 BC (Mandyla et. al. 2011). This matter emphasizes the fact that the necessary deterministic circumstances for malaria transmission are present over Greece and at any given instance that

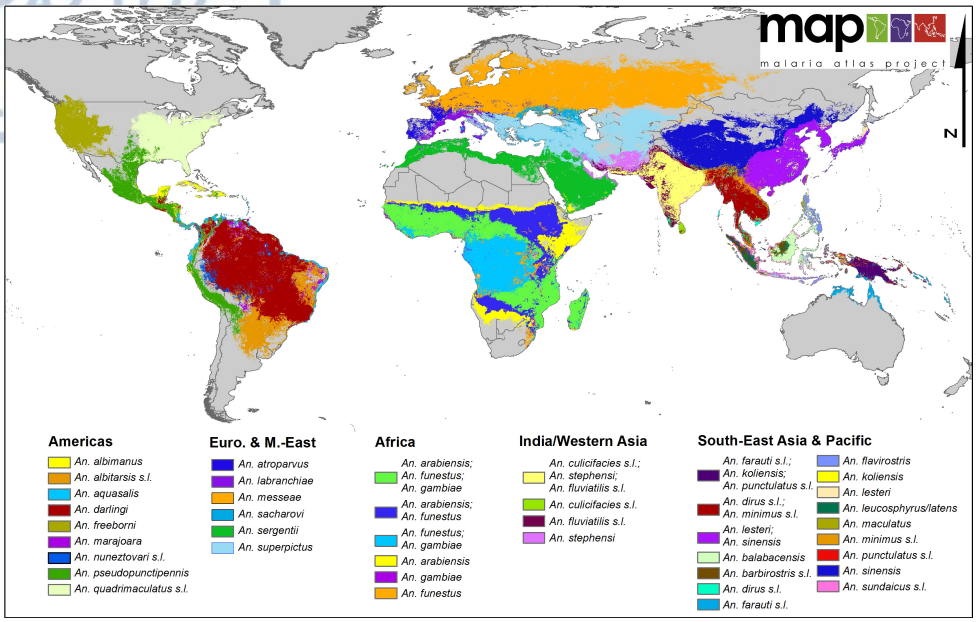


Figure 1.6: Efficient malaria vectors (Sinka et al., 2012).

certain stochastic requirements are met, an outbreak may happen, as it did in years 2009, 2010, 2011 and 2012 (Danis et al.,2013). The stochastic requirements referred to above are primarily a function of population movement and of the control measures applied, which in turn are a function of further socio-economic factors, which will not be discussed in the current analysis.

1.3 Malaria in Greece

Malaria constituted a source of significant health threat in Greece during the first decades of the 20th century, causing approximately 1-2 million infections annually, throughout the period 1905-1940 (Vakali et al., 2012). During this period malaria cases were observed through the whole extent of mainland Greece (Figure 1.7). Urban areas and areas located in high elevations were excluded from this malaria intensive status (Vakali et al., 2012). Along with the high burden of malaria, the dawn of the 20th century was accompanied by the formation of the Greek Anti-Malaria League in 1905, by two eminent personalities, Constantinos Savvas, Professor of Hygiene and Microbiology and president of the League and Dr. Ioannis Kardamatis, a pediatrician.

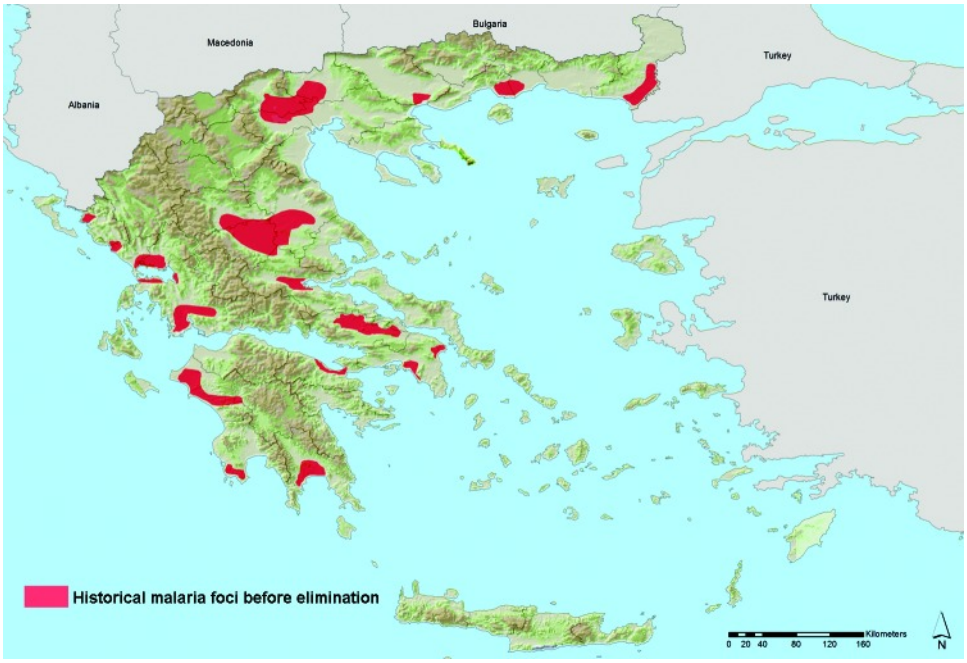


Figure 1.7: Areas of historical malaria transmission (Sudre et al., 2013)

The work of the Greek Anti-Malaria League was monumental. Through the well-instrumented and massive distribution of quinine in 1907, malaria prevalence dropped impressively to 47 % (from 100 %) and further on to 13 % the following year until it reached 9 % in 1909. During the same year, the first attempt to assess the geographic aspects of malaria was undertaken by the League in 1909, by publishing the manual named "Malaria in Greece and Crete", which also included the first map of malaria occurrence (Figure 1.8). In 1924 the League, under the supervision of Dr. Kardamatis, published the work "Statistical maps of swamps and frequency of malaria in Greece" (Tsiamis et al., 2013) which was a further attempt to identify the geographic status quo of malaria in northern Greece (Figure 1.9).

After almost two decades, during 1946-1960, remarkable effort was laid on anti-malaria programs and practices, which eventually resulted to Greece being declared free of malaria in 1974 (Tseroni et al., 2015). The successful implementation of malaria control actions was achieved by the detection of both passive and active cases, efficient treatment of all incidents caused by *P. vivax*

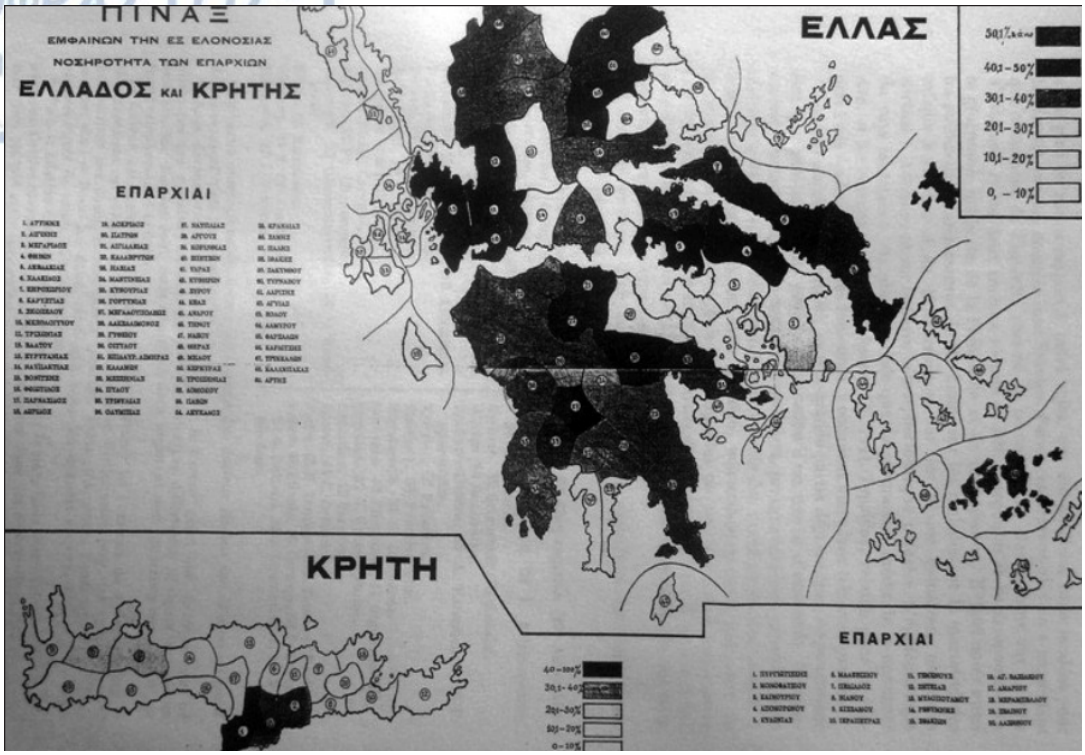


Figure 1.8: The first epidemiological map of the splenic index in Greece and Crete (Tsiamis et al., 2013). The splenic index is used as a metric of the state of malaria transmission inside a population and refers to the enlargement of the spleen (Chavez et al., 2011)

and systematic control of the vectors by means of applying practices such as the indoor residual spraying (IRS), larviciding and the exploitation of larvivorous fishes in water bodies (Danis et al., 2013).

Malaria was estimated to be present in the greek region from approximately the 5th century BC, as it is addressed in the writings of Hippocrates who referred to a disease that was causing fever and was related to the existence of marshes (Kousoulis et al., 2013). An interesting aspect in Hippocrates's narrative though, is that he states that the occurrence of the disease was observed in autumn and winter (Kousoulis et al., 2013), while it is known that climatic conditions are suitable for malaria transmission during summer and autumn. Also, regardless of the fact that there are not records or entomological archives concerning the *Anopheles* species present in ancient Greece, it is safe to assume that there has not been a significant change on this aspect from the period of the 5th century BC until the 20th century (Kousoulis et al., 2013) when entomological data were eventually available. The kind of *Plasmodium species* responsible for malaria occurrence in ancient Greece remains unknown, but from Hippocrates's description of the clinical attributes of the patients, it is assumed that the main pathogen was either *P. vivax* or *P. malariae* (Kousoulis et al., 2013).

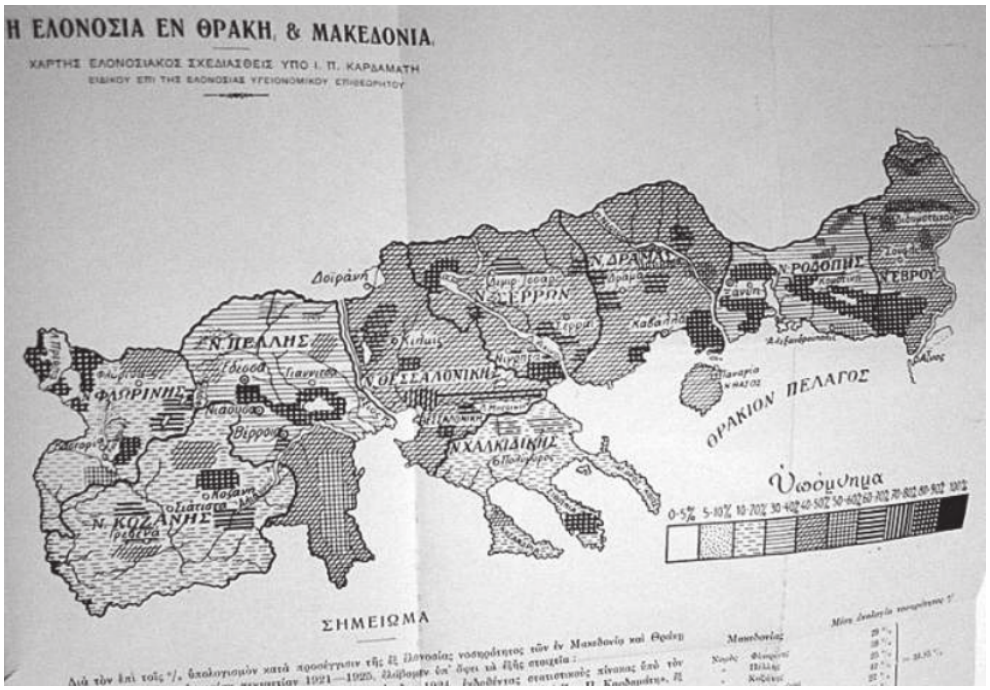


Figure 1.9: Swamps and malaria in Northern Greece, published by the Anti-Malaria League in 1924 (Mandyla et al., 2011)

1.4 Mosquito life cycle and dependence on climatic parameters

The understanding of the different developmental stages of the mosquito life cycle is crucial for decoding the various aspects of malaria transmission (Asare, 2016). All mosquito species proceed through four basic stages, that include the aquatic part (egg, larvae and pupae) and the terrestrial part, in which the adult mosquito emerges. All compartments of the development of a mosquito are directly or indirectly affected by a series of climatic factors, a fact that very vividly delineates the importance of a solid understanding of how climatology/meteorology affects the various aspects of the mosquitos biology, yielding to the characteristic ecology of the vector. The life cycle of the mosquito and its relation with the various climatic factors is presented in Figure 1.10.

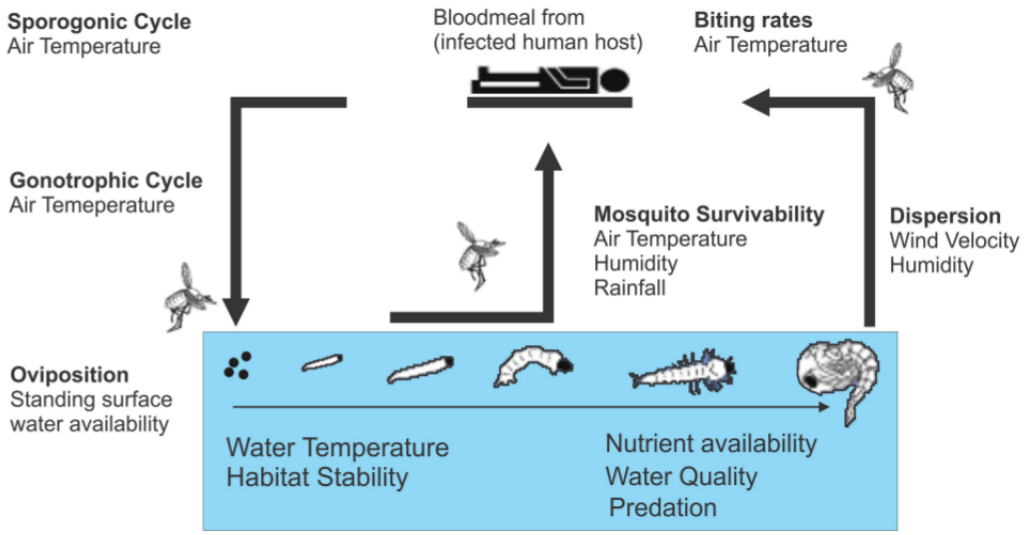


Figure 1.10: The life cycle of a mosquito and its relation with climatic/meteorological factors (Asare, 2016 adopted from Smith et al., 2013).

Egg stage

The availability of water-covered areas is essential for oviposition and hence, for the initiation of the mosquito life cycle. Female mosquitoes of all species lay their eggs on areas covered with water either in rafts (*Culex spp.*) or individually (*Anopheles spp.*). The preference of the type of water body on which the female mosquitoes lay their eggs varies from species to species (Asare, 2016). For instance, the female mosquitoes of the *An. gambiae* species prefer to lay their eggs on small and warm (sunlit) water bodies (Asare, 2016), while *An. sacharovi* species prefer waters with a restricted content of salinity. In addition, mosquitoes of the *Aedes albopictus* species can lay their eggs in any container on which there is available standing water. *Aedes albopictus* species are also termed as small container breeders because they can lay their eggs even in the cap of a

plastic bottle. Typically, oviposition includes the laying of 50 to 200 eggs.

Larvae stage

After oviposotion, all larvae lie in the water, some fully covered (*Culex spp., Aedes spp.*), while other species (*Anopheles*) lie parallel to the water surface (Figure 1.11). During this stage, high mortality is observed which can be due to abiotic reasons, such as flushing rainfall and increased water temperature, or due to biotic reasons, such as predation and competition (Asare, 2016).

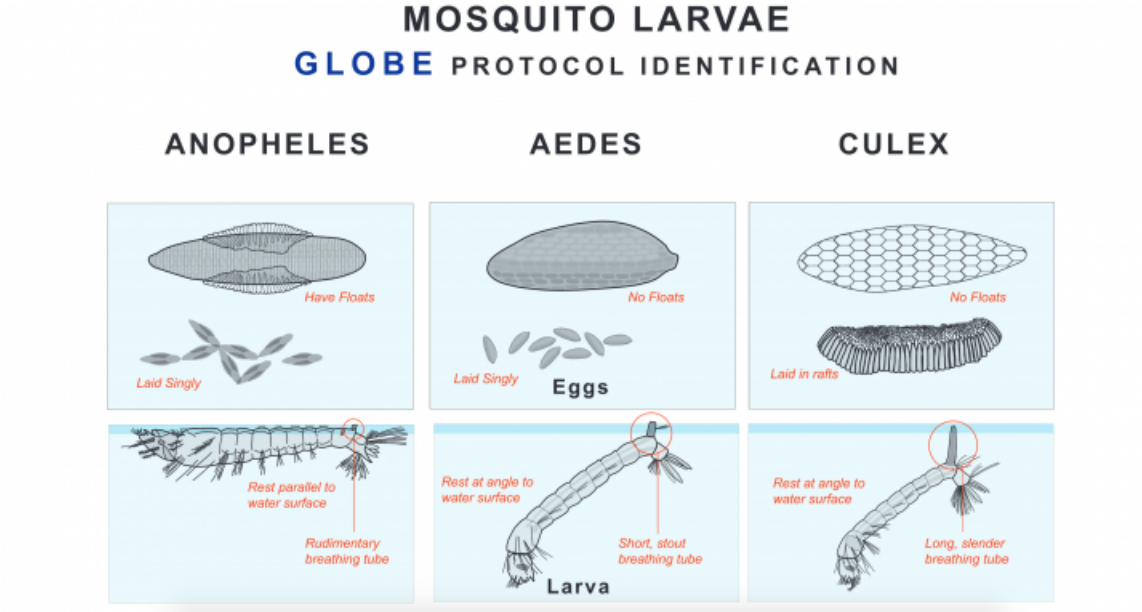


Figure 1.11: Eggs and larvae for *Anophels sp.*, *Aedes sp.* and *Culex sp.* Source: https://www.globe.gov/documents/11865/0dcf909a-b4b3-4793-969a-5f88c48fbf26

Pupae stage

The last stage of the aquatic part of the mosquito life cycle is the pupae stage. This part is very short and typically its duration is 2 days, but as all stages, this is also temperature dependent, and hence, there might be some variation from the 2 days normal period. At this stage metamorphosis takes place and the adult stage of the mosquito is initiated (Asare, 2016).

Adult stage

The adult mosquito that results from the metamorphosis of pupae, is capable of producing eggs after some days (Asare, 2016). At this stage the female vectors are in a blood seeking process and it is at this stage that biting and hence transmission, occur (in the case that the vector acquires the parasite from an infected person or an infected vector bites a susceptible human). After a blood meal is acquired, female mosquitoes have sufficient protein supplies in order to proceed to

their gonotrophic cycle. The completion of the gonotrophic cycle results to the laying of new eggs on available water surfaces. The activity of the vector, the length of the gonotrophic cycle, the frequency that a blood meal is required and the frequency of oviposition are all climate-sensitive parameters (Asare, 2016).

1.5 Malaria and *Anopheles* Mosquitoes

In the study of the dynamical evolution of an infectious disease, such as malaria, epidemiology employs two fundamental approaches in order to simulate, explain, treat and prevent the disease in concern. These two approaches consist of either the application of individual-based (or agent-based) models (Bomblies, 2009) or of the use of compartmental models (Mandal et al., 2011). Agent based models attempt to model in a very detailed way all the various individuals that constitute the community in focus, modeling explicitly all the specific actions and interactions that might possibly be related to the transmission of the disease (Arifin et al., 2013). Mainly, agent-based models attempt to model the spatio-temporal behavior of the host. Currently, the behavior of the vectors is not explicitly modeled. The application of agent-based models is possible only on a very localized scale due to its specificity and its requirement of very detailed local information. Such an attempt is made in the framework of the Open-Malaria project which is an open-source agent-based model, especially applicable in areas with high endemicity (Williamson et al., 2016).

Compartmental models treat the modeling of the disease dynamics in a more crude and generic way, by discriminating the population of the community in concern in certain compartments, that describe their position with regards to the transmission cycle (Chitnis et al., 2008). Depending on the disease, its evolution and dynamics, different kinds of compartmental models can be applied, in order for a more representative modeling of the disease to be achieved. The more complete form of a compartmental model is considered to be the SEIR model, which separates the initial population into four categories, which are: S: Susceptibles (the total number of the naive population that are the potential candidates for infection), E: Exposed (the compartment of the population that has been infected but is not infectious yet), I: Infectious (the number of people that can transmit the disease), R: Recovered (the number of people that have recovered from the disease). Depending on the disease, the individuals can remain in this class if they have acquired immunity, or otherwise return to the susceptible class (and re-enter the SEIR cycle). An SEIR model can be very detailed, discriminating subclasses inside every class, accounting for gender, age, occupation, demographical movement (births, deaths, migration) and any other characteristic that could be relevant to transmission. Nevertheless, the most generic form of an SEIR model is presented below:

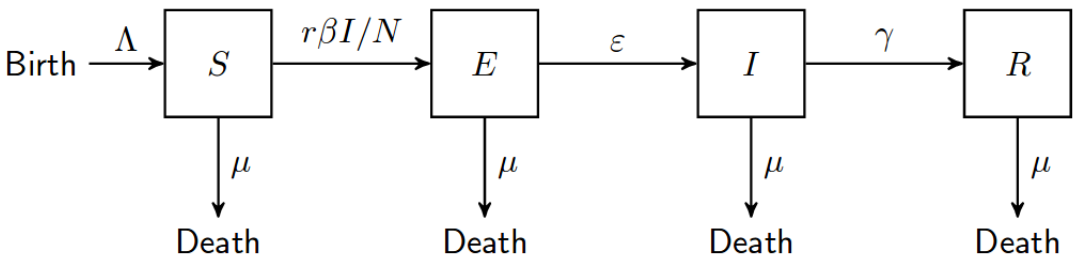


Figure 1.12: The flow diagram of an SEIR model (Chitnis, 2017 (from a presentation)).

In the diagram presented above, there is a constant flow from the S class towards the R class. Also, in every compartment there is a probability that certain portion of the initial population is removed from the class, due to death caused either by the disease, or by other demographical factors. The available pool of Susceptibles is given by Equation 1.1:

$$\frac{dS}{dt} = \Lambda - r\beta\sigma\frac{I}{N} - \mu\sigma \tag{1.1}$$

The constant recruitment rate is denoted by Λ and represents the rate at which people are added to the Susceptible class either by birth or migration. The number of contacts per unit time between a Susceptible and an Infectious individual is denoted by r. The definition of r is dependent on the nature of the disease (vector-borne diseases, air-borne diseases, sexually transmitted diseases etc.). In the case of malaria, r is the mosquito bite. The probability of disease transmission is represented by β. For malaria it is estimated that approximately 20-30 % of the contacts lead to a successful transmission. N is the total population of the community, while I is the part of the population that has already been infected and is capable of transmitting the disease onwards. Lastly, μ is the rate at which people are removed from every class. The progression of the Susceptible class to the Exposed class is described by Equation 1.2:

$$\frac{dE}{dt} = r\beta\sigma\frac{I}{N} - \epsilon E \tag{1.2}$$

In Equation 1.2, the per-capita rate of progression to the Infectious class is denoted by ϵ. The continuation in the disease progression leads to a certain part of the population becoming infected and hence, infectious. The infectious state is described by Equation 1.3, in which with γ the recovery rate (per-capita) is represented.

$$\frac{dI}{dt} = \epsilon E - \gamma I - \mu I \tag{1.3}$$

Finally, the Recovered class is described by Equation 1.4:

$$\frac{dR}{dt} = \gamma I - \mu R \tag{1.4}$$

Successful malaria transmission requires the co-existence of an *Anopheles* female mosquito and a host, one of which should belong to the Susceptible class and one of which should belong to the Infectious class. Transmission occurs through the bite of the female *Anopheles* mosquito, during its blood-feeding process (Asare, 2016). If the mosquito is infectious, it is able of transmitting *Plasmodium* sporozoites through its salivary fluids during its bloodacquiring bite. The mosquito injection infects the host and the life-cycle of *Plasmodium* inside the host is initiated. Approximately, one hour after the bite, sporozoites enter the hepatic cells, via the bloodstream (CDC, 2016). By the time sporozoites reach the hepatic cells, their asexual reproduction begins, resulting to the production of schizont and eventually, to the production of merozoites. During the time of reproduction inside the liver, the disease is in a latent state and the host is asymptomatic. Also, if the infection is caused by *Plasmodium vivax* or *Plasmodium ovale*, it is possible that the

sporozoites lie dormant inside the liver, forming hypnozoites, which are able of causing relapses after months or years (Cogswell, 1992).

After the individual acquires the *Plasmodium* sporozoites, there is a certain amount of time required for the maturing of sporozoites and their invasion to the blood stream. The incubation period is known to vary within different kinds of *Plasmodiums*. For instance, the *Plasmodium vivax* requires 6-8 days to develop, the *Plasmodium falciparum* requires 5-7 days, *Plasmodium ovale* requires 9 days, while *Plasmodium malariae* requires 12-16 days (Ngasala, 2010). The maturing of schizonts leads to the release of merozoites into the red blood cells. From this state onwards, the patient becomes symptomatic. In addition, it is possible that some merozoites, do not reproduce asexually successfully, and generate gametocyte cells. In the case that a susceptible vector bites an infectious host, the gamecytes are ingested by the mosquito, through the blood and the development of the *Plasmodium* inside the vector is initiated (sporogonic cycle). The duration of the sporogonic cycle is temperature dependent (Detinova, 1962). The sporogony inside the vector undergoes the stages of producing a zygote and ookinete cells that enter the female mosquitos oocyst and eventually, after the bursting of the cyst, the sporozoites reach the vector's salivary gland. From that stage, the vector remains infectious until its death. A schematic representation of the *Plasmodium* life cycle is given in Figure 1.13.

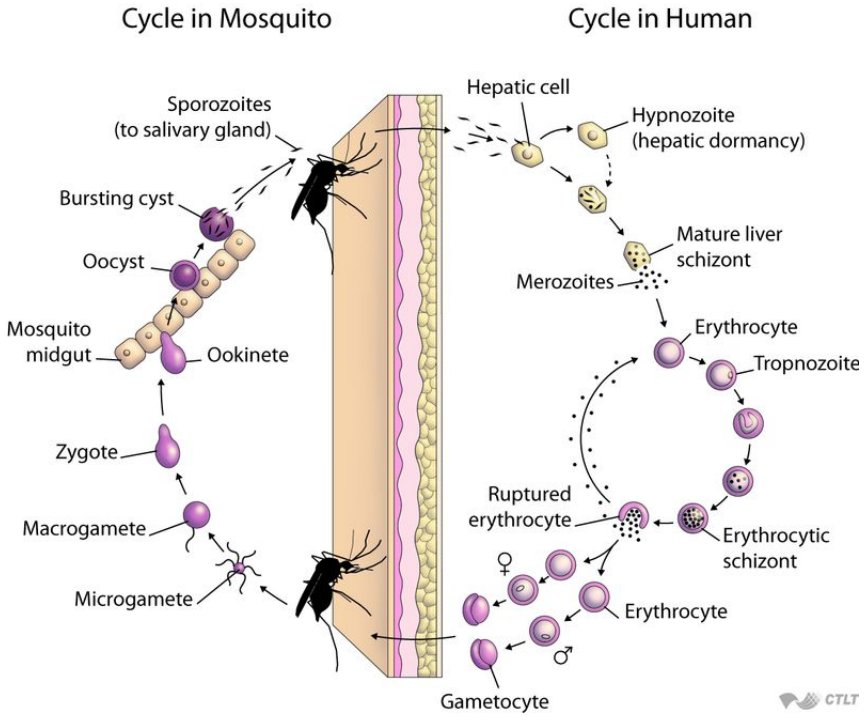


Figure 1.13: The cycle of the *Plasmodium* parasites inside the mosquito and inside the human (Source: malwest.gr)

A graphical representation of an SEIR model applied for malaria is given in Figure 1.14 as taken from Mandal et al. (2011). As it is shown, the Susceptible class is constituted by all the individuals of the community and is also affected by demographic movement (births, deaths, immigration, emigration). After the incubation period is completed, the infected individual becomes infectious

and can transmit the disease. If the patient recovers from the disease, then moves to the Recovered class, and thus, if immunity is not acquired, the person is re-located to the pool of Susceptibles. Also, if the infection was caused by *Plasmodium vivax* or *ovale*, there might be a possible connection from the Recovered class to the Exposed class, because these two Plasmodiums are known to cause relapses of the disease, due to their dormant lying in the liver (Cogswell, 1992).

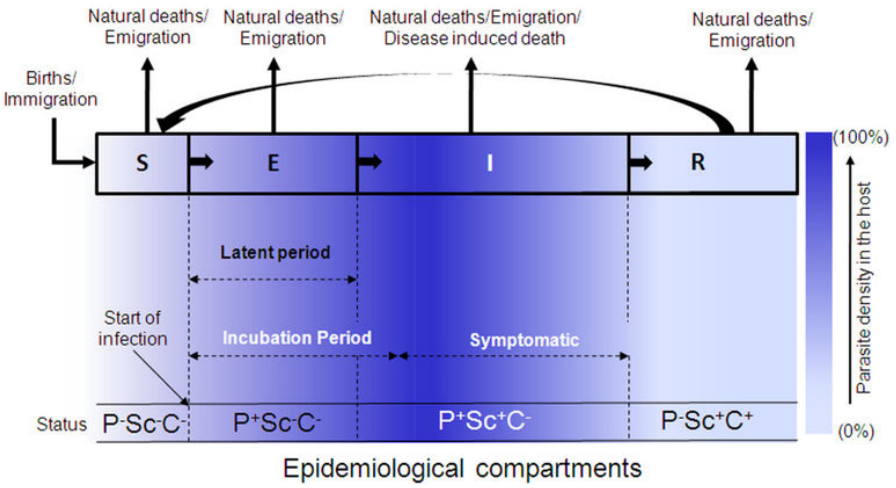


Figure 1.14: An SEIR model applied for malaria (taken from Mandal et al., 2011).

The modeling of the disease dynamics, requires that both groups are modeled as to their infectious state. In both groups (vectors and hosts) an SEI and an SEIR model could be applicable. Malaria transmission occurs either when an infected mosquito bites a susceptible human, or when a susceptible mosquito bites an infectious human. Also, a difference between the two populations is that an infectious vector remains so until its death, hence, there is no Recovered class in the mosquitos compartmental model.

1.6 Modeling the impacts of climate on malaria distribution

Modeling the climatic impacts on vector-borne diseases requires that all the various mechanisms of the disease are studied and modeled effectively. This is achievable by primarily two possible suite of models: the mechanistic models and the correlative models (Peterson et. al., 2017). In addition, studies related to the spatial and temporal distribution of vector-borne diseases are either focusing on the ecology of the pathogen itself or on the ecology of the vector responsible for transmitting the pathogen to the miscellaneous hosts, or both.

A great challenge in ecology is the modeling of the ecological niche of a species. The term ecological niche is used for the description of a multi-dimensional space that favors the viability of a species and the maintenance of their population (Hutchinson, 1957). The ecological niche models appear in the literature with different terms such as species distribution models, habitat distribution models or climatic envelope models. Silero (2011), proposes the use of the term ecological niche models, which better reflects the modeling of the suitable habitats, in contrary to the term species distribution models, which alludes to the per se modeling of the species distribution (which is extremely difficult and usually not possible). Sillero's proposed terminology is adopted throughout the whole extent of the current work.

The theoretical establishment of the ecological niche theory was laid on the foundations set by Joseph Grinnell in 1917 and by Charles Elton in 1927. In Grinnell's notion, the niche constitutes a subset of the habitat that is governed by environmental conditions that are conducive to the species survival and reproduction. Elton's perspective of the niche was rather systemic, as certain species constitute a node in a complex food network and are a functional part of a community. In Grinnell's opinion, the environmental conditions constitute a primary modulator in the formulation of the species spatial distribution, while in Elton's view the interspecies interactions is the dominant forcing factor of the species spatial distribution.

From a geographic perspective, the occurrence of a species over a region could be approximated as a point pattern. A fundamental question that arises from observing the clustering (or no clustering) of the occurrences (which translates to points, when a point pattern analysis is applied) has to do with whether the causes of a certain point pattern type lies in the environment in which the occurrences co-exist or whether it lies on the interactions between the occurrences (Hengl et al., 2009). If the spatial phenomenon in concern is governed by changes in the mean field of a broad scale environmental variable, then these interactions are classified as first order interactions. If the phenomenon in concern is modulated by the spatial interactions between neighboring occurrences of the same phenomenon, then these interactions are classified as second order interactions and they display high spatial autocorrelation. Placing the ecological niche theory into a geographic context, Grinnell's view can be interpreted as a first order interaction between species and the environment, while Elton's standpoint fits better to a second order interactions. In most cases, the reality amalgamates both kinds of interactions, hence it could be stated that the best ecological

niche approximation is a combination of both Grinnell's and Elton's views.

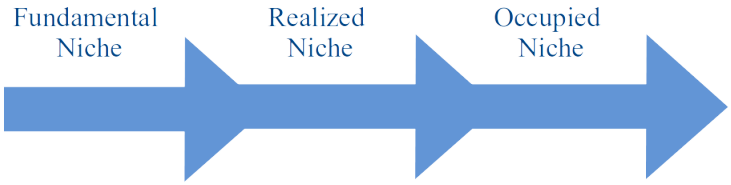


Figure 1.15: A gradient of the different types of niches, moving from the more generic (left) towards the more specific (right).

Types of Ecological Niches

An issue that has justifiably attracted a lot of attention in ecology and the study of species distribution, is the identification of the kind of niche that it is possible to be studied and modeled, so that the conclusions drawn based on the model outputs are in agreement with what it is actually modeled (Sillero, 2011). In scientific literature there seems to be an overall consensus on the existence of different kinds of niches, that could be possibly placed on a niche gradient (Figure 1.15). Moving from the more generic ones towards the more constrained ones, these are the niche typologies that arise from the relevant literature: the fundamental niche, the potential niche, the realized niche and the occupied niche.

The full geographic range of a region that is governed by environmental conditions suitable for the species viability and reproduction constitutes its fundamental niche. Nevertheless, there is a set of biotic and abiotic factors that hinder the species from occupying their full fundamental niche, hence, the species in concern may be absent from regions that carry suitable environmental conditions. The biotic and abiotic factors in action that oppose the full expansion of the species are primarily due to biotic interactions such as competition, predation, symbiosis and parasitism and due to geographical limitations such as the existence of mountainous barriers (Sillero, 2011) or the land use change that can disconnect patches that function as suitable species habitats (Wiens et al., 2009). The biotic constraints, such as species competition, limit the fundamental niche to the realized niche, while if abiotic factors are also included, then there is a jump from the realized niche

towards the occupied niche, which is even more constrained. The general notion of the occupied niche is that it reflects in a more realistic way the actual geographic area that a species can be found and most authors agree that the occupied niche is a subdivision of the realized niche, which in turn is a subset of the fundamental niche (Figure 1.16). Nevertheless, according to Pulliam (2000) it is possible that the realized niche can occupy a broader area than the fundamental niche, due to species immigration. This theory, under the name Source-Sink theory, claims that suitable habitats for a species can function as sources and unsuitable habitats as sinks and that there can be a constant flow from the source regions towards the sink regions.

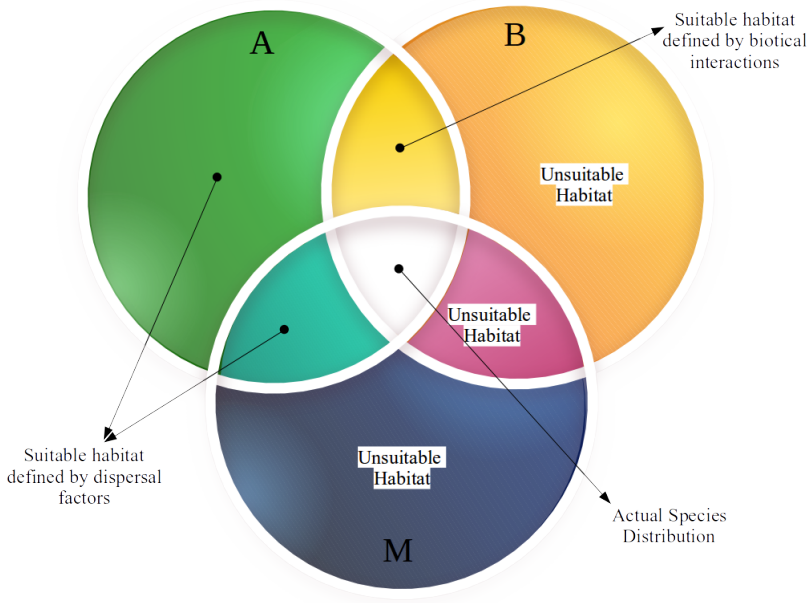


Figure 1.16: The BAM (biotic, abiotic, movement) diagram depicting what are the main reasons formulating the spatial distribution of a species and the niche that results from the three kinds (BAM) of niche modulators (adopted from Sillero, 2011).

1.6.1 Correlative Ecological Niche Models

Ecological niche models are either correlative or mechanistic models that exploit the information of a species observed location, in combination with information that characterizes the environmental profile in which the species are located. For correlative models the occurrence of a species is related to the environmental information by statistical means, while in the case of mechanistic models the relation built between species occurrences and environmental conditions is rather based on phenological laboratory experiments (Silero, 2011). A seemingly widely accepted fact is that correlative models are able to spatially predict the species realized niche (Silero, 2011). A species fundamental niche can only be approximated by using mechanistic models, that are able to identify and build a mathematical (rather than a statistical) relationship that inherently describes a causal relationship between the spatial distribution of a species and the environmental conditions engulfing the distribution of the species in concern. Such an approach would be unaffected by the number of data available (Silero, 2011), hence a mechanistic model is not empirically derived. As it was mentioned above, mechanistic models seek to establish a causal relationship between the components of the studied system, while correlative models seek to correlate the different components of the same system. The difference between the two aforementioned approaches lies on the fact that correlation does not necessarily translate to causality. Nevertheless, the observed correlation can assist and educate the procedure of establishing a mechanistic relationship.

Ecological niche models are useful tools for a variety of applications related to conservation and planning, spatial epidemiology and disease modeling, ecology, study of invasive species in new geographic domains (Phillips et al., 2006) and the study and modeling of the potential impact of climate change (Wiens et al., 2009) on species distribution and the risk that arises with respect to the spread of infectious diseases. Concerning the response of certain species to climate change, ecological niche models can provide an estimate of whether their current environment would be suitable in the future for their maintenance and reproduction and if not, what are the regions that would provide a suitable habitat that would enable the species to survive and have offspring. This approach is based on the assumption that the environmental requirements of a species will remain the same in the future and that the species will shift its ecological niche geographically, so as to find a favorable environment in which it could be established. However, the species might also be able to adjust in their current location to the new climatic conditions and not seek to shift their niche or they could also go extinct (Wiens et al., 2009).

A question of primary interest in ecology and evolution is related to the spatial distribution of species, that is strongly determined by the environmental regime in which the species in concern are able to establish and reproduce, delineating their ecological niche. Nevertheless, the observed niche that species occupy over a geographic region is usually a subset of the one that they could potentially occupy, thus, there has been a distinction between the potential niche and the realized niche. There is a plethora of factors contributing to the formulation and the differentiation between the various kinds of niches such as the species biology, the spatial resolution in which the analysis is carried out, the explanatory variables that are included in the modeling approach and the

modeling method that is employed for the approximation of the niche (Sillero, 2011), therefore, the differentiation of the various kinds of niches is not a straightforward procedure.

1.6.2 Dynamical-Mechanistic Models

The occurrence and evolution of malaria can be characterized as a highly dynamic system, as its status and progression may alter severely in the course of a short time. Also, the incidence and the change over time of malaria could be potentially described as a chaotic system, as it displays dependency on initial conditions. The initial conditions refer to the system constituted of the host and the vector population and provides information on the proportion of infected humans and/ or infected vectors, that contribute to the propagation of the disease. Furthermore, the evolution of the system can be importantly affected by intervention measures, such as the usage of bed nets and the application of vector control practices (spraying with insecticides etc.) (Tompkins and Ermert, 2013). Also, the response of a certain community to the emergence of malaria occurrence or to the malaria risk of occurrence, may affect severely the disease dynamics. In addition, the progression of malaria may also lie on factors inherently describing the characteristics of the host and vector populations. For instance, different vectors species may rank in different scales of anthropophilicity or zoophilicity, which is a metric describing vectors biting preferences in acquiring their blood meal, but is also directly influenced by the availability of either humans or livestock. Concerning the host population, not all people are equally attractive to vectors, yielding to a very differentiated spatial distribution of bites, inside a relatively small district that could be regarded as homogeneous. From the aforementioned it is evident that the way host and vector populations interact is highly variable and any attempt of modeling this system would unavoidably introduce a large degree of uncertainty.

The modeling of malaria dynamics is performed either by the development and application of statistical or dynamical models (Tompkins and Ermert, 2013). Statistical models rely on the regression of past cases to a suite of environmental covariates. Nevertheless, although statistical models are very widely-applied for modeling malaria transmission, they display certain limitations such as the fact that the model that is produced depends on a certain number of samples, which may hinder the successful application of the model when the training data available are limited. Also, models built for a very specific location, may not be applicable for locations with different characteristics and geographic attributes. In addition, statistical models may display some inadequacy when daily or sub-daily variations of the disease are required (Tompkins and Ermert, 2013). On the other hand, if the equations describing the evolution of a dynamic system (such as malaria transmission) are known, then they can efficiently be used in a dynamical model to simulate, model and forecast the transmission cycle (Tompkins and Ermert, 2013). The use of either models finds application in studies related to the construction of an early warning system or on applications related to the impact that climate and environmental change may have on the spatio-temporal dynamics of the disease. In both applications the goal is that a better intervention and response strategy may be designed and achieved. Despite the fact that both types of models carry their own suite of advantages and disadvantages, the dynamical models are comparatively less applied and studied. One dynamical model that was developed with the purpose of explicitly modeling

the malaria dynamics is the vector-borne disease community model of the International Center for Theoretical Physics, Trieste called VECTRI (Tompkins and Ermert, 2013).

VECTRI is a dynamical malaria model accounting for the impact of temperature, rainfall and population density on the transmission dynamics of the disease (Tompkins and Ermert, 2013). Some of the innovative applications of malaria modeling in VECTRI is that it treats and models the host-vector interaction as a unified system, instead of treating the host and vector populations as two different entities (Tompkins and Ermert, 2013). Also, the impact of climatic factors is explicitly modeled. More specifically, the impact of temperature is modeled using the concept of degree days that need to be completed in order for the *Anopheles* mosquito to proceed from one stage to the next in its developmental cycle. Rainfall has a direct impact on the surface hydrology model parameterized in VECTRI, that provides an estimate of the available temporary water bodies that form after a rain episode and provide the necessary aquatic environment for oviposition and the development of the aquatic stages of the mosquito. Also, the interaction between host and vector is achieved by providing information on the population density. This layer of information can provide a quantitative insight into what is the availability of blood meals for *Anopheles* mosquitoes. The ability of VECTRI to model the host-vector dynamics system, enables the distinction between malaria occurring in peri-urban and rural environments. The VECTRI model in its current set up is parameterized to model the transmission of the *Plasmodium falciparum* parasite that is transmitted through the *Anopheles gambiae* complex (Tompkins and Ermert, 2013). The current tunning of the model is suitable for malaria occurring in regions of sub-Saharan Africa, but it would require further tunning and calibration to model malaria that is caused by the *Plasmodium vivax* parasite and is transmitted by other *Anopheles species* (Sinka et al., 2012).

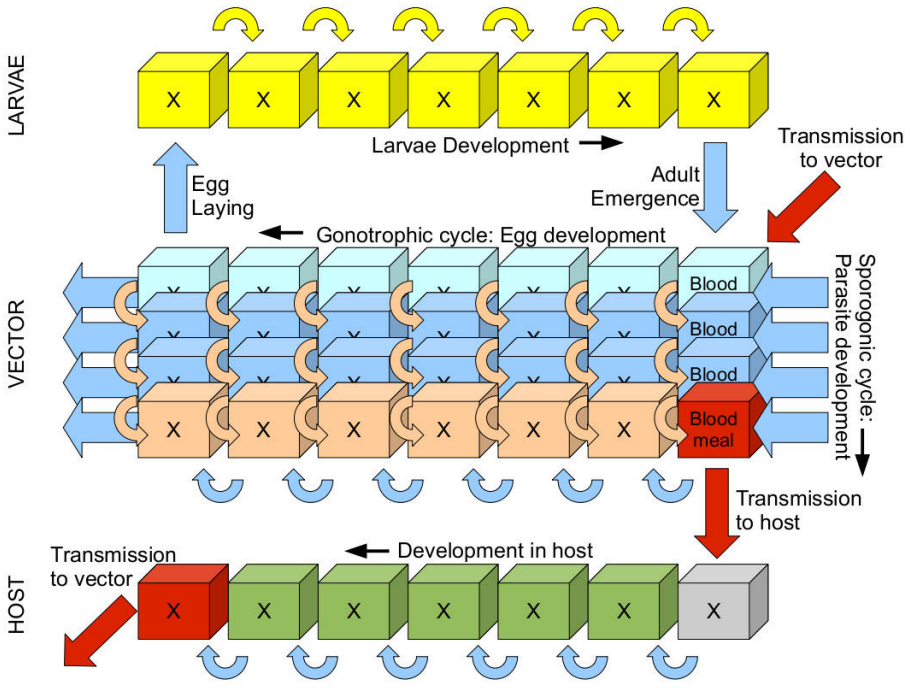


Figure 1.17: Schematic representation of the VECTRI model.

1.7 Dissertation Aim and Structure

The discussion concerning the potentiality of malaria transmission lies on several factors related to the possibility of introduction of the malaria parasite in a naive population, the environmental suitability for an effective malaria vector and the existence of socioeconomic factors that would nurture the emergence or even re-emergence of the disease. In order for malaria to emerge, all the three aforementioned aspects have to co-exist (Semenza, 2015). Such a research question is multi-sectoral (Danis et al., 2013) and requires input from the fields of medicine and epidemiology, climatology, geography and ecology. Each of the aforementioned sectors has to provide answers to certain questions, that they themselves will function as input in a decision support system that would enable the making of educated and data-driven operational choices (Semenza, 2015), in order to avert malaria risk.

For the case of Greece where malaria was endemic until the 1960s (Danis et al., 2013), the primary research goal would be to identify what are the various aspects in the malaria transmission cycle that would cause a potential revival of the disease. Once these aspects are successfully identified, then anti-malaria activities can be informed so as to break the transmission cycle. A study carried out under the presence of re-emergent cases should ideally respond to what is the measure of receptivity and vulnerability in the area of study (Danis et al., 2013). Receptivity refers to the presence of a vector that is capable of transferring the disease and thus, receptivity lies on climatic and ecological factors. On the other hand, vulnerability refers to the introduction or the presence of the parasite in a community, that could be potentially transferred through the biting activity of the vectors. In Danis et al. (2013) it is stated that:

There appears therefore ample space for the development of an integrated malaria vector control action plan adapted to the disease transmission patterns and the ecological setting of Greece.

Therefore, the four main question that the current dissertation aims to investigate are the following:

1. What is the ecological setting of Greece favoring the occurrence of *An. Sacharovi* (what is its ecological niche)?
2. What is the spatial outline of receptivity of malaria over Greece?
3. How would a dynamical malaria model reproduce malaria activity in Greece?
4. How could a dynamical malaria model be calibrated for *An. Sacharovi*?

In Chapter 2 of the current document, the data that were used in the current analysis are presented. Chapter 2 is divided in four main subsections. The first subsection is referring to the climatic data used, the second subsection is referring to all the various geographic data used and the third and fourth subsections are referring to the mosquito and malaria data used, respectively.

In Chapter 3 the methodological approach of the current analysis is presented. Chapter 3 is also divided in two subsections and these are: the methodology of the correlative ecological niche models and the methodology of the VECTRI model. In Chapter 4 the results are presented and in Chapter 5 an overall discussion and criticism of the analysis is presented and the next research questions are delineated.

Chapter 2

Data

2.1 Climatic Data

2.1.1 Regional Climate Models

The current state-of-the-art tools in climate science is the use of climate models, which are mathematical representations of the atmosphere, simulating the known dynamic and thermodynamic processes that govern the atmosphere in its entirety (Neelin, 2011). The fundamental question that climate models were initially trying to address, was how would the energy equilibrium of the climatic system be altered in a changing climate. The tools that were originally used for this purpose were the Global Circulation Models (GCMs) that provided a crude representation of the state of the atmosphere. Nevertheless, as time proceeded and the knowledge concerning climate change became more solid, the questions of the global climate research community became more specific and the requirements of the end-users became more refined, hence the need for a more detailed representation of the climate arose. This evolution in the knowledge and the mechanics of the climate system lead from the use of GCMs to the development of the Regional Climate Models (RCMs). RCMs allow for the use of climatic information in a great variety of impact studies, extending from agriculture and forestry, to renewable energy sources, sustainable development in urban planning and studies of health impact. Of course, RCMs do not replace GCMs, but they rather provide an added value in climate research.

The transition from GCMs to RCMs required the development of methodologies that would efficiently allow for the downscaling of the initial information contained in GCMs to a finer spatial resolution. For this purpose, two basic methodologies were developed and applied: the statistical downscaling and the dynamical downscaling techniques. The dynamical downscaling techniques translate the climatic conditions calculated and contained in a coarser resolution grid to an RCM that is correctly nested and located into the coarser resolution grid (Neelin, 2011). The RCM receives the information of the dynamic and thermodynamic state of the atmosphere from the parent domain and re-solves the fundamental equations describing motion and energy in the inner domain of concern. In the case of dynamical downscaling, the performance of the nested RCM is highly dependent on the quality of the lateral conditions received from the parent domain (Laprise et al., 2008, Laprise et al., 2012), and hence, a lot of effort is put on providing accurate and

detailed information as initial conditions to drive an RCM. The initial conditions can be provided by a GCM or by other products such as reanalysis datasets. Reanalysis datasets have undergone the process of data assimilation and are considered to be the most accurate representation of the past and current state of the atmosphere (Dee et al., 2011). When reanalysis data are used as forcing fields of a climate experiment over a limited area, then this experiment is regarded as an experiment performed in a perfect boundary setting (Giorgi and Gutowksi, 2015).

Perfect boundary experiments are used in hindcast simulations that aim to identify the bias that is due to the RCM used (and not the bias due to the forcing fields used in the experiment), so that the performance of the model can be assessed for a past period, for which the climate has already occurred and thus, observational datasets are already available. Once the performance of RCMs is assessed, then they can be used for future studies that investigate the spatio-temporal patterns of climate change. Also, the miscellaneous hindcast experiments can effectively inform the procedure of producing a model ensemble of many RCMs, so that the uncertainty of the final product can be minimized (Pavlidis, 2015). This procedure is especially useful when the use of climate information aims at guiding decision making and planning. From the aforementioned it is shown that dynamical downscaling experiments is a rapidly growing field of increasing importance and hence, there is a need for a coordinated application of such techniques.

This need in the field of climate research is met by the Coordinated Regional Climate Downscaling Experiment (CORDEX) which aims at organizing and regulating simulation experiments, downscaling climatic information to RCMs (Giorgi and Gutowksi, 2015). More specifically, the CORDEX initiative provides a common temporal and spatial context for downscaling experiments and also regulates the various methods through which the simulations are performed, along with the evaluation techniques employed (Giorgi et al., 2009). The spatial resolution of the CORDEX experiments are performed on grids of 0.11 degrees, 0.22 degrees or 0.44 degrees which corresponds to an approximately 12 km, 25 km and 50 km spatial resolutions respectively. Also, within the CORDEX experiment, both multi-model and multi-physics experiments are conducted (Kotlarshki et al., 2014, Katragkou et al., 2015).

For the current analysis, the output of 8 RCM hindcast model simulations was employed (Table 2.1), that were run for the EURO-CORDEX domain on a 0.11 degrees spatial resolution and covered the standard hindcast period from 1990 to 2008. The model runs were forced with the ERA-Interim reanalysis data and were made available on a daily timestep.

2.1.2 WorldClim Dataset

The WorldClim dataset is a global climatology product referring to the period 1950-2000. It is derived using station data assembled in large data repositories which are spatially interpolated using the splines interpolation method. The spatialization of the point data is assisted using the elevation derived from the SRTM DEM. The variables that are produced in the WorldClim project are minimum, maximum and mean monthly temperature and total monthly precipitation (Hijmans et al., 2005), while the second version of the WorldClim project also included solar

Table 2.1: The suite of Regional Climate Models employed in the current analysis, along with the information concerning the institutes that developed the models and the institutes that run the models.

Model	Model Run by:	Model Developed by:
ALADIN53.v1	Centre National de Recherches Mtorologiques	Centre National de Recherches Mtorologiques
CCLM4-8-17.v1	Climate Limited-area Modelling Community	Climate Limited-area Modelling Community
HIRHAM5.v1	Danish Meteorological Institute	Danish Meteorological Institute
RACMO22E.v1	Royal Netherlands Meteorological Institute	Royal Netherlands Meteorological Institute
RCA4.v1	Swedish Meteorological and Hydrological Institute	Swedish Meteorological and Hydrological Institute
RegCM4-2.v1	Meteorological and Hydrological Service of Croatia	International Centre for Theoretical Physics
REMO2009.v1	Helmholtz-Zentrum Geesthacht, Climate Service Center, Max Planck Institute for Meteorology	Max-Planck-Institut fr Meteorologie
WRF331F.v1	Institut Pierre Simon Laplace/Institut National de lEnvironnement Industriel et des Risques (IPSL- INERIS)	National Center for Atmospheric Research

radiation, vapour pressure and wind speed (Fick and Hijmans, 2017). The gridded surfaces are available on a 1 km spatial resolution. The WorldClim is a widely-used dataset in ecological niche modeling studies. The climatic variables of minimum, maximum monthly temperature and total monthly precipitation were also utilized (along with the ensemble mean of RCMs described in the previous section) for the calculation of the 19 bioclimatic variables, used as input in the ecological niche modeling by means of employing correlative models.

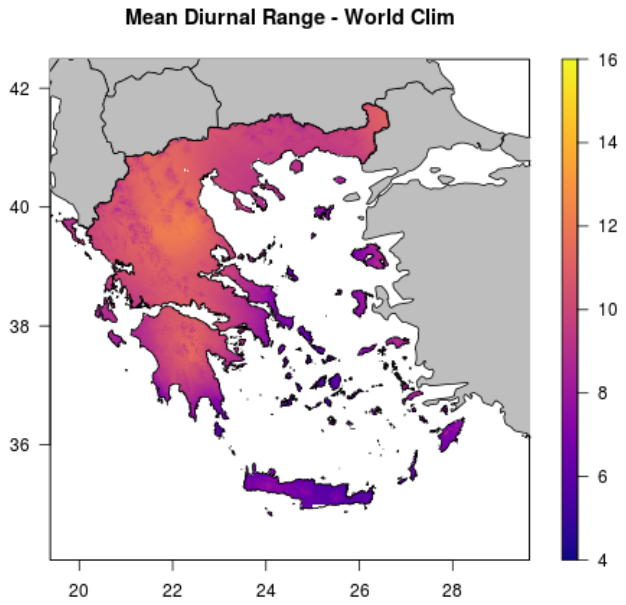
2.1.3 Bioclimatic Variables

The bioclimatic variables used in the current analysis were derived from the WorldClim database, where minimum and maximum mean monthly temperature, along with total monthly precipitation were utilized for their calculation. The worth of calculating bioclimatic variables, instead of using the initial climatic variables per se, lies on the fact that the bioclimatic variables obtain a biological-ecological-environmental meaning that allows their use in various studies concerning applied climatology.

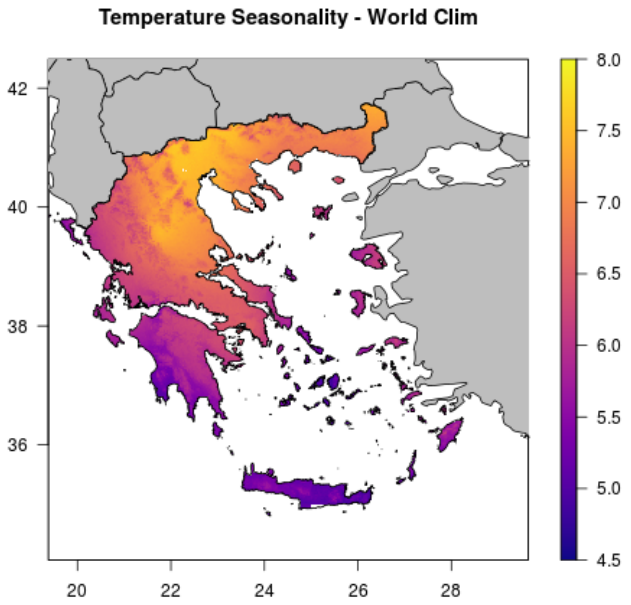
The whole suite of the available variables consists of 19 bioclimatic variables (available through the WorldClim website) that utilize temperature and precipitation in order to construct variables related to the mean, variance and extremes of temperature and precipitation. Furthermore, a

subset of the bioclimatic variables combine both temperature and precipitation in one variable (eg. mean temperature of wettest quarter). In addition, the bioclimatic variables refer either to yearly values (eg. annual precipitation) or to monthly values (eg. minimum temperature of warmest month) or to certain quarters (eg. mean temperature of coldest quarter - quarters refer to a three-month period). In the current analysis quarters were selected over singular months, because the available vector data usually cover the summer months. In the Figures below, the selected bioclimatic variables that were exploited in the ecological niche modeling by means of correlative models are presented.

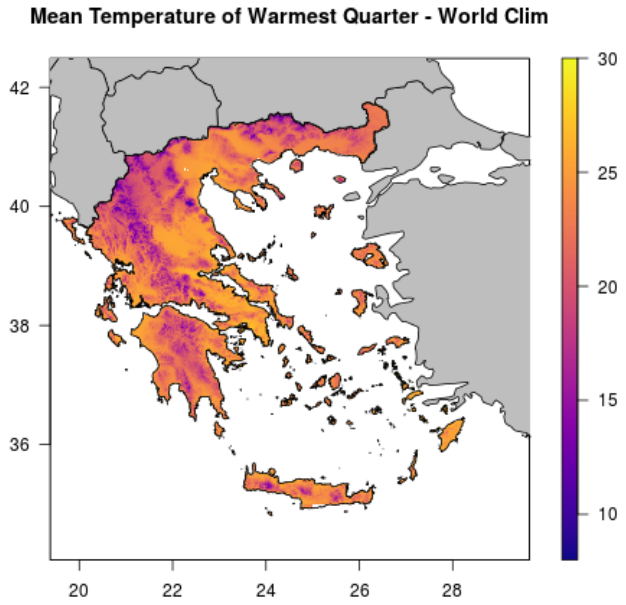
The bioclimatic variables were selected meeting both quantitative and qualitative criteria. Concerning the quantitative criteria, a detailed listing of those is presented in the Methodology Chapter following. Concerning the qualitative criteria, those were the product of discussion with individuals with expert opinion in the field of vector biology and control, concerning the parameters that are potentially driving the activity of Anopheles mosquitoes.



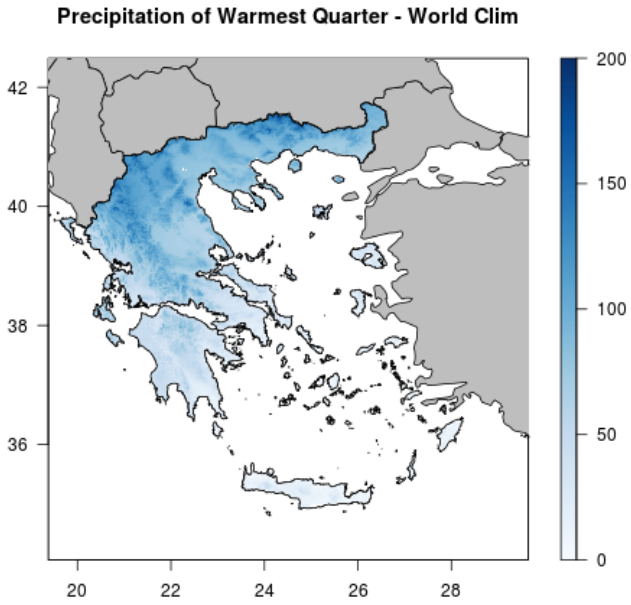
(a) Bio 2



(b) Bio 4



(c) Bio 10



(d) Bio 18

Figure 2.1: The bioclimatic variables used in the ecological niche modeling by means of correlative models (the variables are calculated using the WorldClim dataset, equivalently the same variables were also calculated using the RCM ensemble mean).

2.2 Geographic Data

2.2.1 Digital Elevation Model

The surface topographical characteristics are provided through a Digital Elevation Model (DEM) that contains values of topographical elevation on a grid. DEMs are usually derived from satellite products or from interpolation techniques, where points of known (measured) altitude are spatialized. The spatial resolution of such products is normally high, ranging from a couple of hundred of meters to a couple of tens of meters. In the current analysis, the Shuttle Radar Topography Mission (SRTM) is employed, with a spatial resolution of approximately 92 m in the area of the Equator and approximately 85 m in the area of interest (Greece) (Farr et al., 2007). The SRTM project was a collaboration between NASA, the National Geospatial Intelligence Agency, and the German and Italian Space Agencies. The Shuttle operation scanned the whole globe on the February of 2000 and the technique employed for data acquisition was the Interferometric Synthetic Aperture Radar (InSAR) where a dual polarization radar scanned the Earths surface in order to obtain terrestrial height (bathymetry estimations are not included) measurements (Farr et al., 2007). The SRTM dataset is widely-used among a plethora of applications and also used for the derivation of surface derivatives such as slope, aspect and curvature in standard GIS applications. The elevation map of Greece is depicted in Figure 2.2.

After remapping the DEM file to the WorldClim grid, the elevation values range from 0 to 2709 m. Also, as it is shown in Figure 2.3, the majority of pixels (60 %) belong to areas where the elevation is less than 500 m, while the remaining 40 % is constituted of pixels describing elevations higher than 500m, with a decreasing frequency as the elevation increases. For instance, 13 % of the pixels describe elevations greater than 1000 m, 3 % of the pixels belong to elevations greater than 1500 m and only 0.3 % of pixels characterize elevations greater than 2000 m. From Figures 2.2 and 2.3 it becomes evident that the area of interest is dominated by a topographical variety, which eventually has an impact on the climatic regimes and furthermore, it can function as a potential geographical barrier in the expansion of species.

The DEM file was also used for the derivation of the slope and aspect maps. Slope can be defined as the rate of change of the elevation and it is the first derivative of the DEM file. For its calculation, Horns algorithm was employed, which utilizes a 3x3 pixels neighborhood around every pixel in concern. The 3x3 neighborhood scans the whole study region as a moving window and eventually produces the slope file, measured in degrees. Concerning the aspect map, it uses slope as input and calculates the orientation of the slopes. The aspect map is outputted in a scale from 0 to 360 degrees, where 90 degrees is placed on the North, 180 degrees is placed on West, 270 degrees is placed on the South and 360 degrees is placed on the East. This scale represents aspect values starting from the east and progressing counterclockwise. Slope and aspect were calculated in GRASS GIS v7.0.

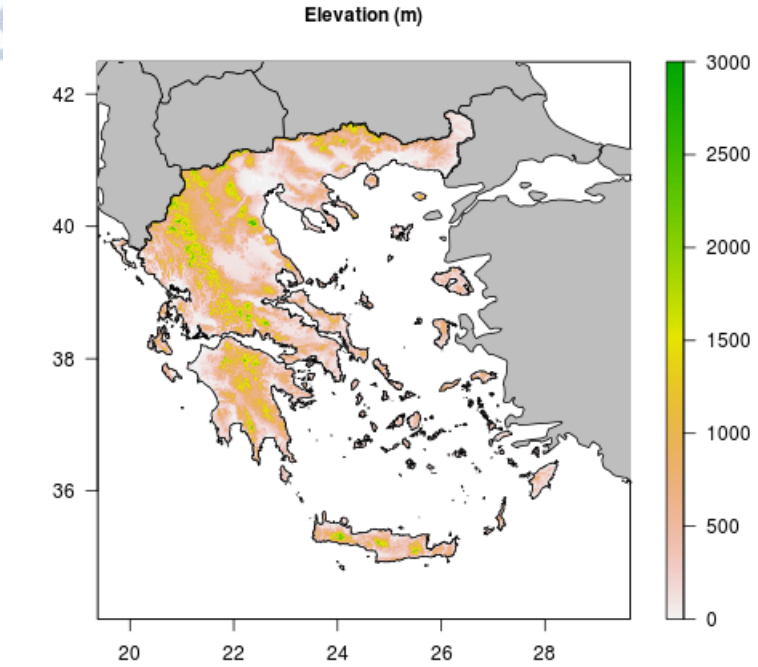


Figure 2.2: Elevation derived from the SRTM dataset over the study area.

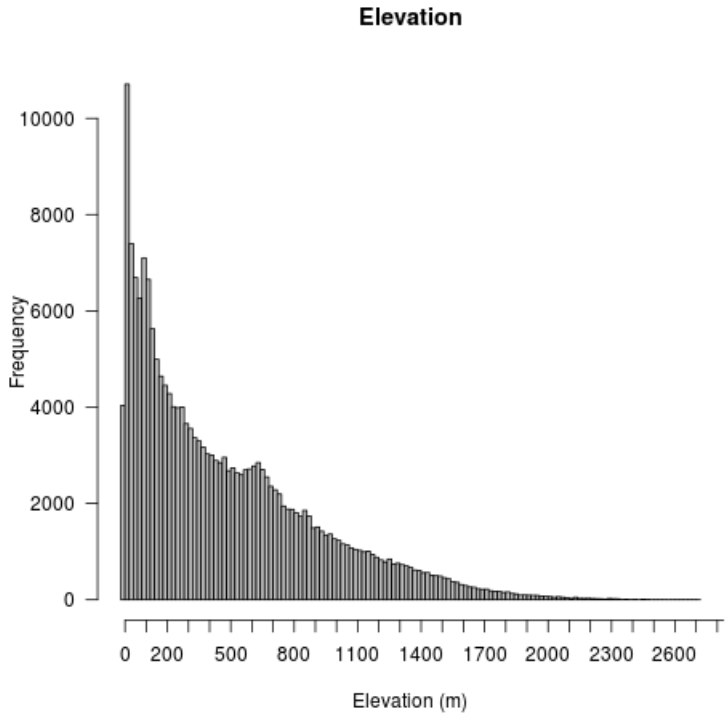


Figure 2.3: The histogram of the elevation file over the study area.

2.2.2 Population Dataset

Population density data display usefulness in many impact applications related to human presence. Population information is usually acquired through census campaigns that are administered by national governments and normally carried out every 10 years. Nevertheless, these data refer to administrative areas that are of various geographic expands and irregular shapes and sizes, which makes them inappropriate for spatial studies that require gridded data. The use of areal data on studies that require spatialized data is known to cause the ecological fallacy issue, which assigns the value of the corresponding areal polygon (eg: a polygon delineating an administrative unit) to all the underlying pixels that belong to each respective administrative unit, creating a spatial artifact.

A common practice for dealing with the aforementioned issue is the use of covariates (such as satellite products), in order to spatially disaggregate census data and estimate population density in a grid cell. Other approaches have employed the use of CLC 2012 urban classes, in order to spatialize census data (Gallego 2010). The population data used in the current study are taken from the Socioeconomic Data and Application Center (SEDAC) operated by NASA. The dataset is termed as the Global Rural-Urban Mapping Project (version 1) (GRUMPv1) and it is derived from the Gridded Population of the World (version 3) dataset (GPWv3) (Balk et al., 2005). The GRUMP dataset has a spatial resolution of approximately 5km in the Equator and refers to data corresponding to 2000. The dataset describes population density. For its construction, both census and satellite data were employed (Balk et al., 2005).

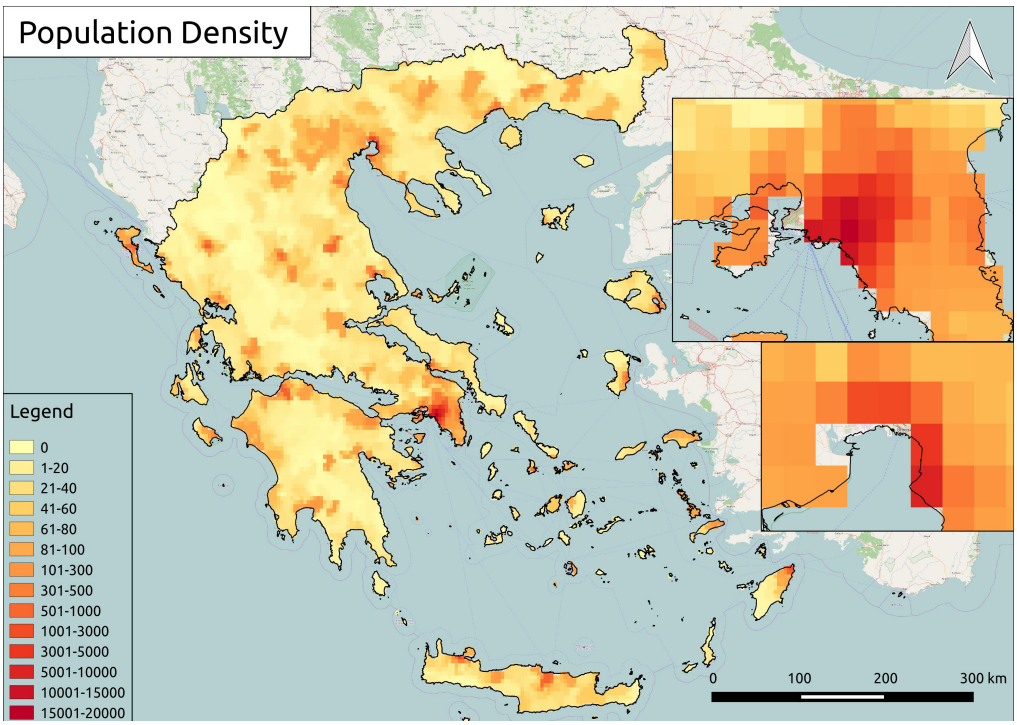


Figure 2.4: Population density over the study region.

2.2.3 Distance to Water-Covered Areas

As it has been extensively discussed in Chapter 1, the biology of the *Anopheles* mosquitoes includes three aquatic stages, hence, the presence of available aquatic environments is of vital importance for mosquito occurrence and survival. Nevertheless, the aquatic environment must occupy a relatively small expanse, in order to accommodate efficiently the eggs of *Anopheles* mosquitoes. Also, in limited area water bodies, there is absence of predators and thus, the survival probability of the eggs increases. From the aforementioned, the knowledge of where are the water-covered areas located can yield important information concerning the potential presence of mosquitoes and can serve as a strong determinant of environmental suitability. In the current analysis, this information was extracted from the Coordination of Information on the Environment (CORINE) Land Cover (CLC) database.

The CLC initiative is an effort of the European Union to map the European land cover landscape. It originated in 1985 under the auspices of the European Commision, until 1994 when the CLC program came under the administrative responsibility of the European Environmental Agency (EEA). CLC contains 44 land cover classes that are produced by the Eionet network National Reference Centers Land Cover (NRC/LC). Until the present day, there have been four CLC versions. The last CLC (termed as CLC 2012) is included in the Copernicus Land Monitoring Services. The CLC spatial resolution is defined using the Minimum Mapping Unit (MMU) which is equal to 25 hectares for areal data and 100 meters for linear data. The CLC database is mainly constructed by means of visual interpretation of satellite images with high resolution, however, there are some countries that apply semi-automatic classification methodologies. Whichever the classification methodology applied, the CLC database is a satellite bi-product and widely recognized as a land cover database. Nevertheless, the 44 classes contained in the database integrate both land cover and land use classes resulting to the mistaken interchangeable use of the terms (land use and land cover). Nonetheless, because CLC is a satellite-derived database, the term land cover is adopted and regarded as the correct term. Briefly, the CLC versions are presented in the following table:

Table 2.2: The Corine Land Cover missions and characteristics.

	CLC1990	CLC2000	CLC2006	CLC2012
Satellite data	Landsat-5 MSS/TM	Landsat-7 ETM	SPOT-4/5	IRS P6 LISS III
Time consistency	1986-1998	2000 +/- 1 year	2006+/- 1 year	2011-2012
Geometric accuracy	$\leq 50m$	$\leq 25m$	$\leq 25m$	$\leq 25m$
Min. mapping unit	25 ha/ 100m	25 ha/ 100m	25 ha/ 100m	25 ha/ 100m

The CLC 2012 dataset was retrieved from the website of the National Cadastre and Mapping Agency S.A. in a polygon vector file in the Greek Geodetic Reference System (EPSG: 2100), but was reprojected to the World Geodetic Reference System (EPSG: 4326) for consistency reasons with the other spatial covariates. The classes that were extracted from the initial set of the 44 classes were the following: Permanently irrigated land (code: 212), Rice fields (code: 213), Inland

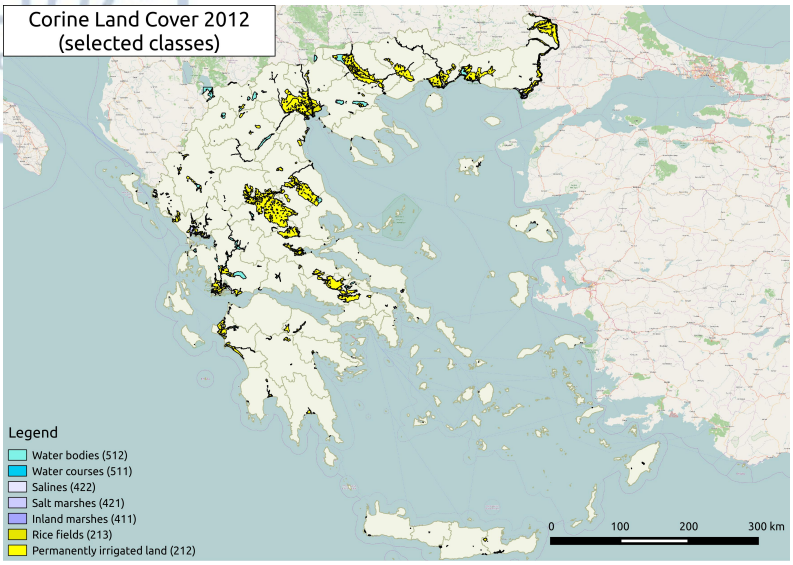


Figure 2.5: Water-covered areas extracted from CLC 2012.

marshes (code: 411), Peat bogs (code: 412), Salt marshes (code: 421), Salines (code: 422), Water courses (code: 511), Water bodies (code: 512).

From all the aforementioned classes, only the class with code 412 was not found to be present over Greece and was excluded from the analysis. In the current project, the water-covered areas were not used directly as input to the set of the covariates, but their distance to the rest of the geographic space was calculated and used instead. For this purpose, all the aforementioned classes (except of peat bogs) were spatially aggregated and were treated onwards as a single spatial feature (a polygon shapefile). The shapefile was rasterized in a 1km x 1km spatial resolution and the geodetic growing distance was calculated. Nevertheless, it is known that mosquitoes can fly up to a certain distance, hence it would be neither useful nor correct to introduce the geodetic distance directly into the ecological niche models. For this purpose, there was a scaling of the derived surface using an exponential function, so that from a threshold distance and beyond, the existence of water-covered areas would be insignificant. The relation that was used was the following:

$$ScaledDistance = e^{-\frac{distance}{\tau}} \tag{2.1}$$

In the relation given above, the τ parameter is set equal to 5 km, because it is considered that this is a reasonable median distance up to which mosquitoes can fly. However, when the ecological niche modeling was carried out using the EURO-CORDEX ensemble which is available in an approximately 12 km spatial resolution, the scaling to 5 km would yield distances that the EURO-CORDEX grid would be unable to capture efficiently. Hence, a second scaled distance is derived where the τ parameter is set equal to 15 km for the ecological niche modeling that uses bioclimatic variables derived using the EURO-CORDEX ensemble. The resulting surfaces are depicted in Figures 2.6 and 2.7. As it is shown, the scaling to 15 km results to a relatively more generalized pattern, nevertheless, in both scalings one can discern the shape and location of the water-covered areas as depicted in Figure 2.5.

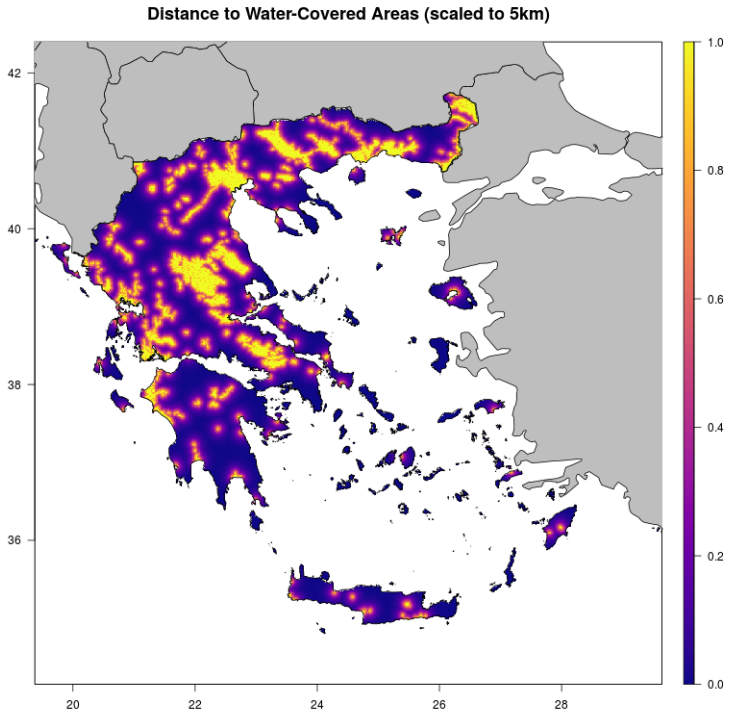


Figure 2.6: Distance to water-cover areas scaled to 5 km distance.

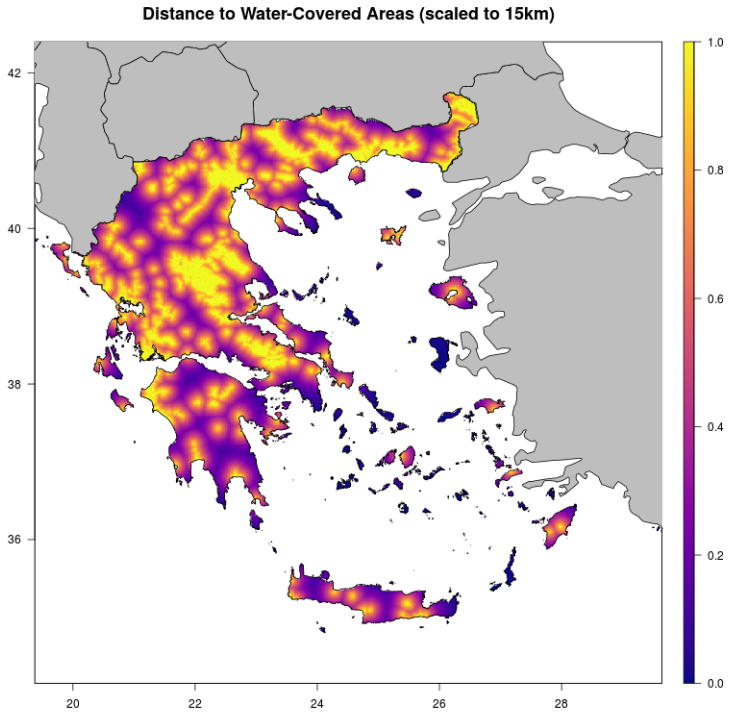


Figure 2.7: Distance to water-cover areas scaled to 15 km distance.

2.3 Mosquito Data

A fundamental requirement for ecological niche modeling by means of correlative models is the existence of presence/absence data. For the current analysis, these data were made available by Ecodevelopment S.A., a private company which is located in northern Greece, that operates in the area of mosquito control and precision agriculture. For the purpose of mosquito control applications, a wide network of mosquito traps is maintained through mainland Greece and measurements of mosquito abundance are usually carried out weekly or biweekly. The sampling campaigns usually extend through the whole period of vector activity, which is mainly from May until early October. The companys mosquito sampling network is primarily constituted of 4 sub-networks with different spatial coverages, and in some cases different sampling methodologies are applied (CO_2 traps and human bait traps). Also, in some networks the entomological analysis is carried out on a species level, while on other networks the analysis is performed on a genus level. For the present analysis, it was required that the spatial coverage of the stations had the largest possible extent and also, that there was available information on a species level. Also, because the original data described species abundance, it was required that abundance measurements were efficiently translated to presence/absence data. The original data comprised of 106 stations sampling from May of 2011 until September of 2011 and of 26 stations measuring from April of 2016 until September of the same year.

An issue that arose from the fact that the mosquito traps were normally sampled more than once throughout the course of the available time-series, is that certain stations recorded presence of the species in concern for some dates, while on other dates, no species was observed for the same trap (Figures 2.8 and 2.9). Nevertheless, the fact that even once a presence was recorded in a trap, it was regarded as a sufficient factor so that this trap would be classified as a "Presence" station. This approach constituted of characterizing a station as a "Presence" station if at least one measurement in the available time-series included a presence. Due to the spatial structuring of the initial network, in which there is an increased clustering of traps in North-central Greece, it was necessary that a second approach would be developed, that would allow for the "declustering" of the initial network. This approach made use of the stations that measured abundances $=> 1$. Stations that measured high abundance and were relatively the most remote ones were selected for the Presence network, while stations in small distances that measured small abundances were omitted from the network. The distance criterion is very important, because it provides an indication of how clustered the network is. For applications such as ecological niche modeling by means of correlative models, it is necessary that the presences are not highly clustered, because that would introduce a sampling bias in the analysis and the suitability of that specific region would erroneously be overestimated. On the other hand, not too many stations must be excluded from the analysis because that would have a direct impact on the validity of the model output. The network of presence and absence data that was eventually used for building the correlative models is presented in Figure 2.10.

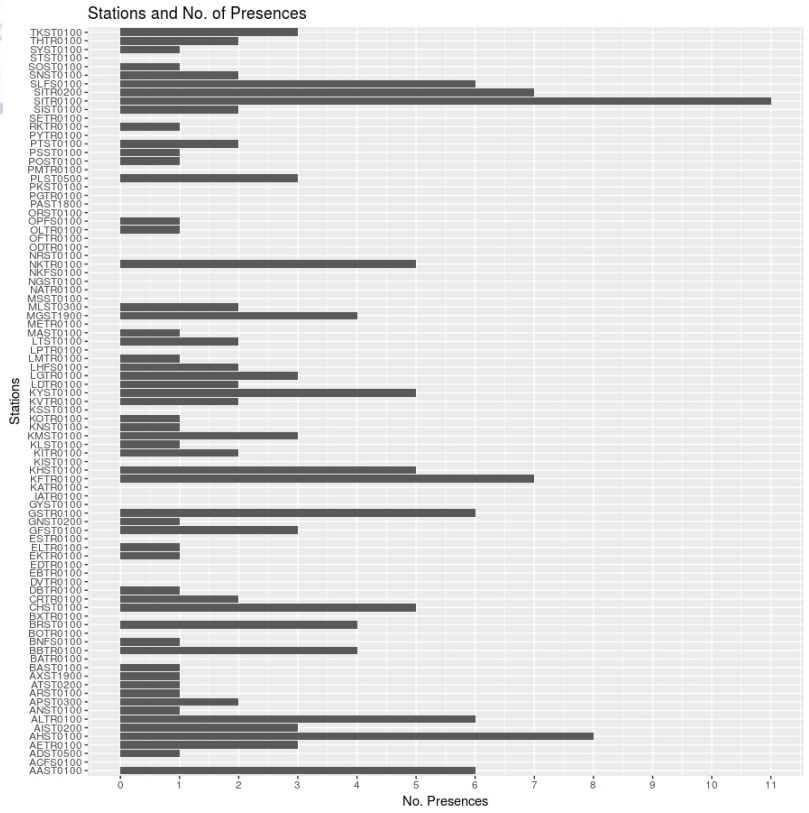


Figure 2.8: Number of presence records in every station.

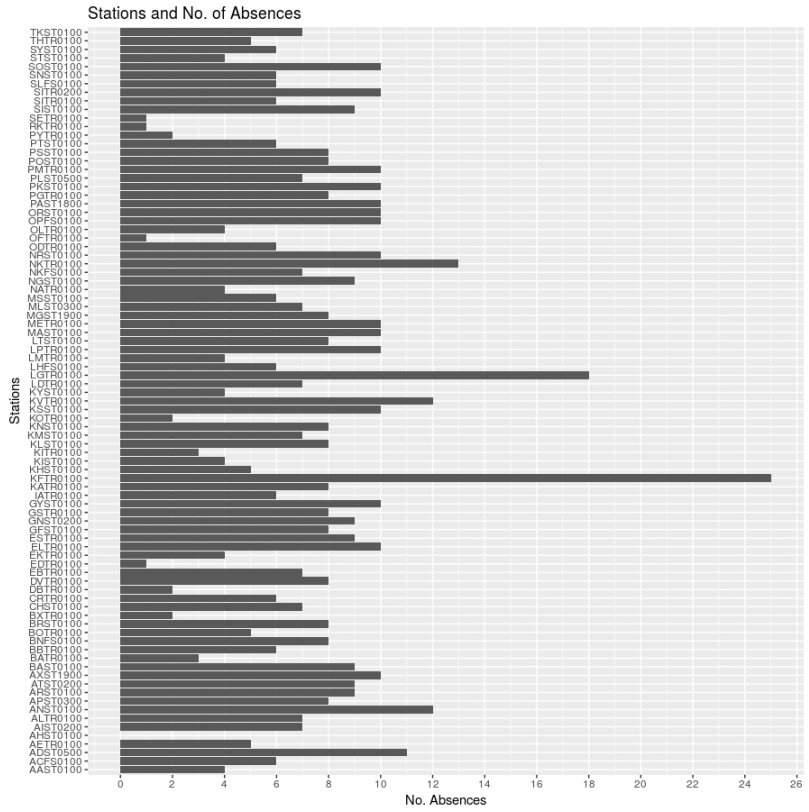


Figure 2.9: Number of absence records in every station.

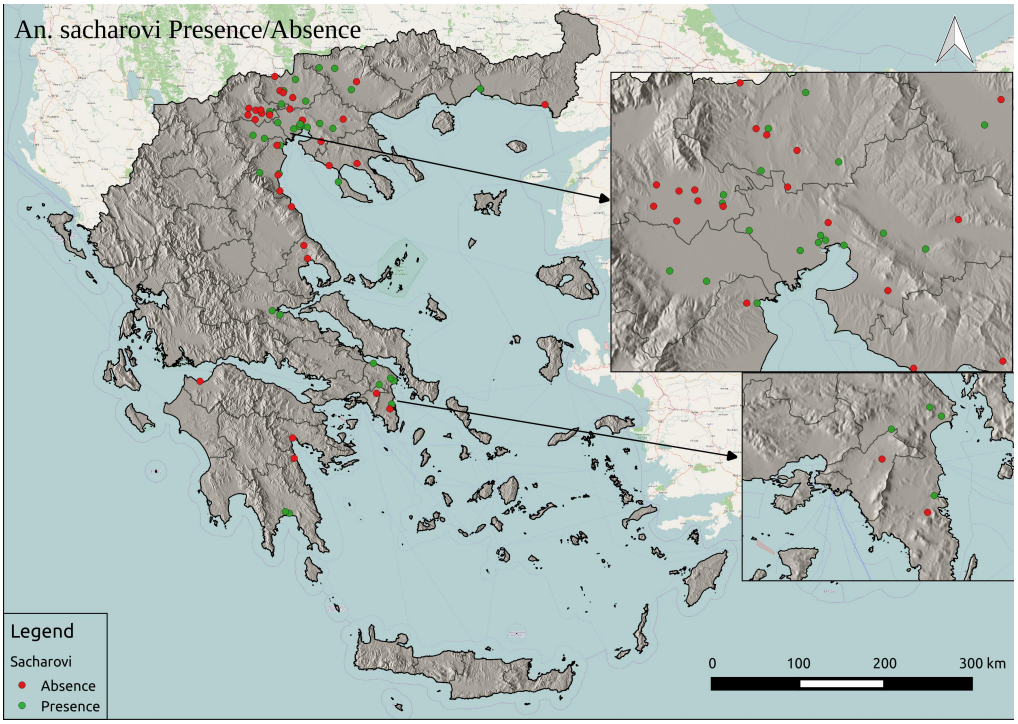


Figure 2.10: Number of absence records in every station.

2.4 Malaria Data

Malaria data were made available by the Hellenic Center for Disease Control and Prevention (HCDCP) and they refer to the period 2004-2015. The original data received, contained information concerning the Plasmodium kind, the date of the reported start of the symptoms, the date of the hospitalization (in the case that the patients were hospitalized), the date of the treatment start, the municipality of residence and a classification of whether the incident was imported or indigenous.

The primary *Plasmodium kind* of indigenous cases was *Plasmodium vivax*, while in 2010, a *Plasmodium falciparum* incident was reported (HCDCP report, 2011) in the area of Attiki (Marathonas). Although there is an observed malaria activity from 2004, the indigenous cases were observed from 2009 onwards 2.11. The greatest number of incidents was observed during 2011 (37 cases). The cumulative spatial distribution of all *Plasmodium kinds* for both imported and indigenous cases is presented in the Figures 2.12.

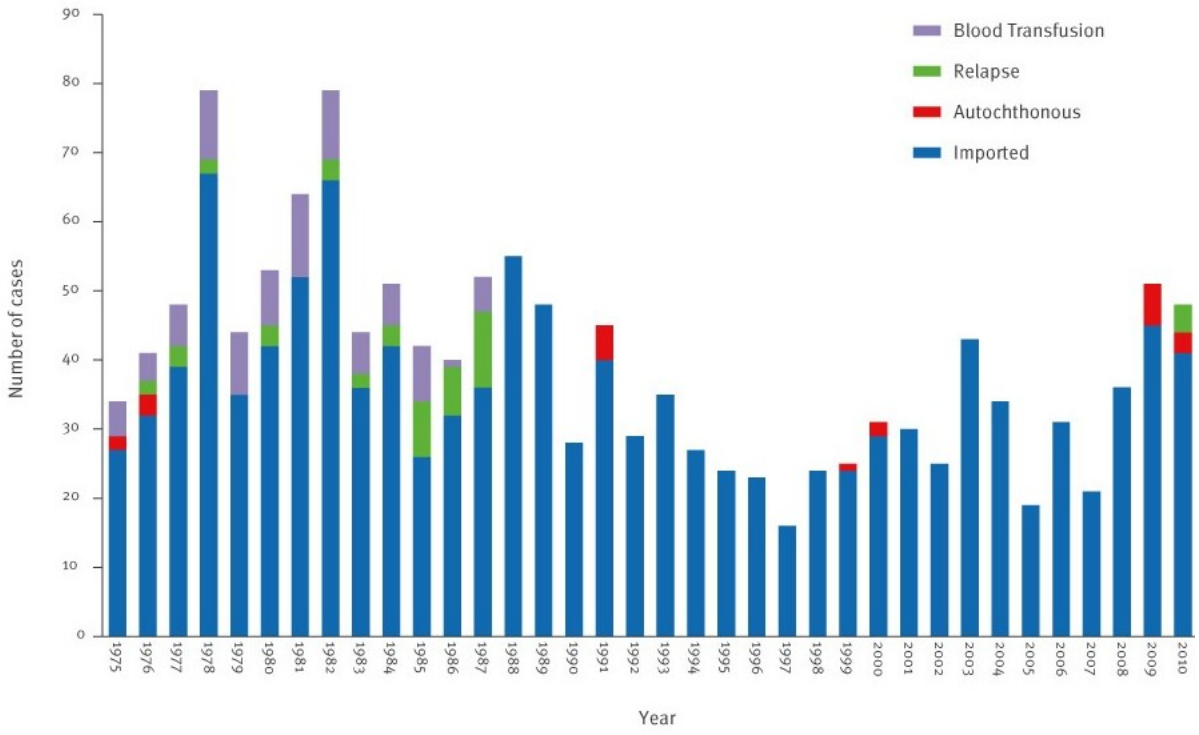
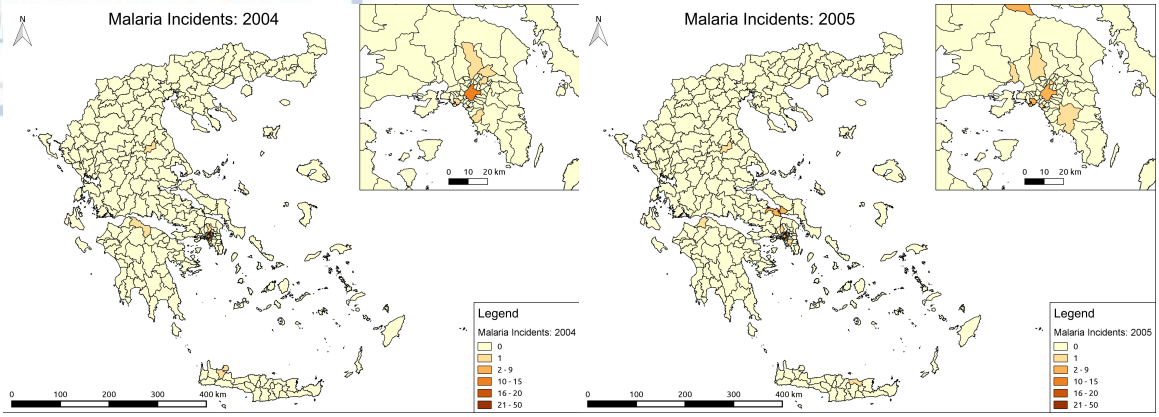
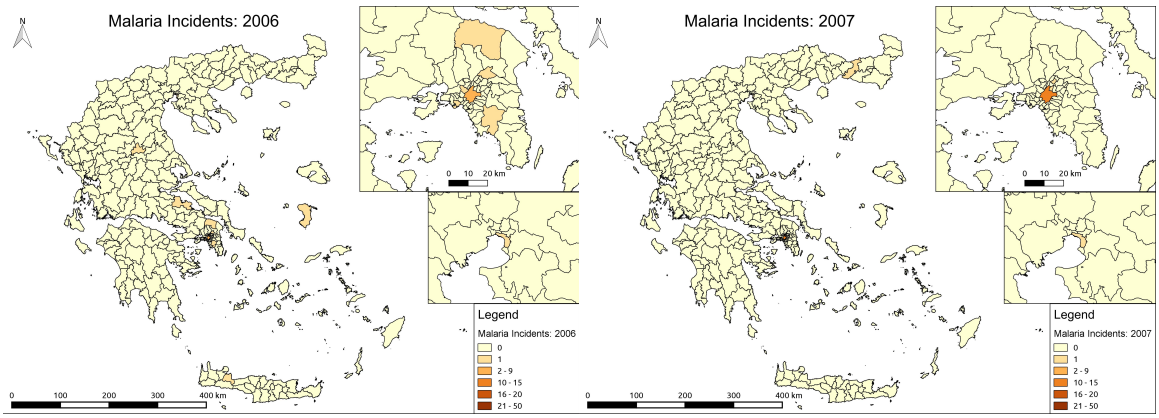


Figure 2.11: Number of malaria cases in Greece, for the period 19752010 (Vakali et al., 2012).



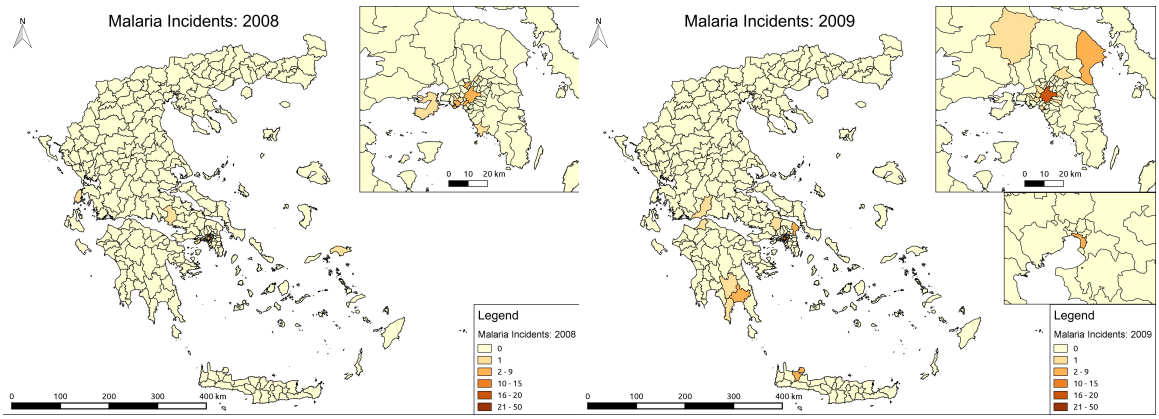
(a) 2004

(b) 2005



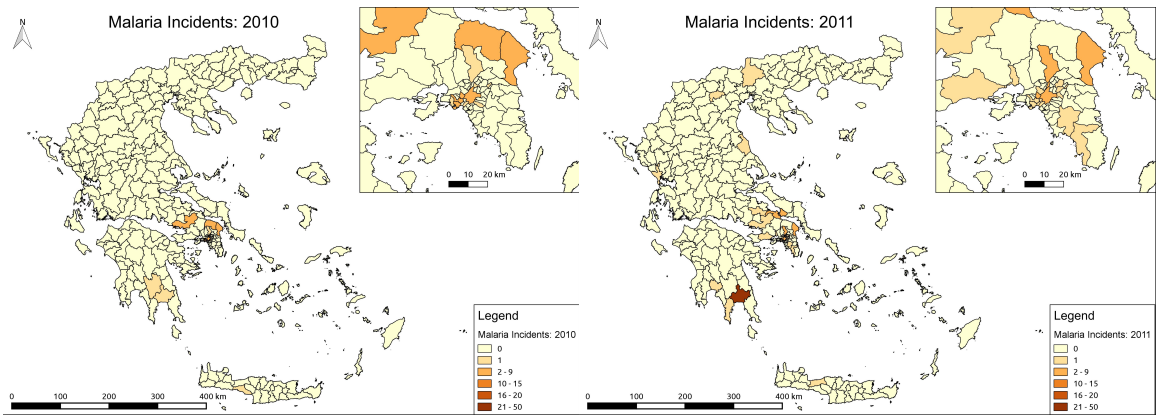
(c) 2006

(d) 2007



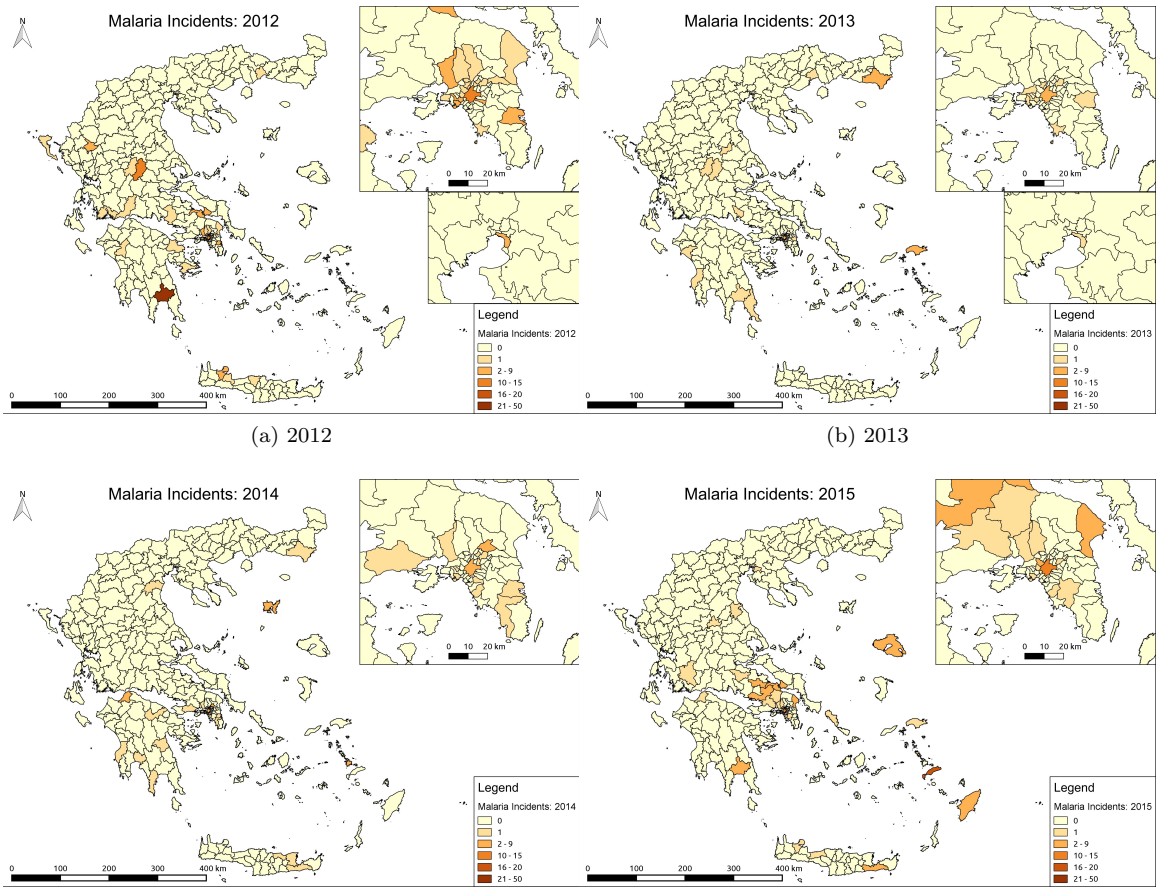
(e) 2008

(f) 2009



(g) 2010

(h) 2011



(a) 2012 (b) 2013

(c) 2014 (d) 2015

Figure 2.12: Malaria incidents reported for the period 2004-2015 by HCDCP (all Plasmodium kinds depicting the accumulated number for both indigenous and imported cases).

Chapter 3

Methodology

3.1 Construction of Bioclimatic Variables

The bioclimatic variables used in the current analysis were derived from the WorldClim database, where minimum and maximum mean monthly temperature, along with total monthly precipitation were utilized for their calculation. In conjuction, another set of bioclimatic variables was created, using as input the RCM model runs obtained from the ESGF database. The variables used were, likewise, minimum temperature (tasmin), maximum temperature (tasmax) and precipitation (pr). All RCM models were used in an ensemble and the ensemble mean of the three aforementioned variables was utilized for the construction of the 19 bioclimatic variables. A subset of the 19 bioclimatic variables was used as input in the ecological niche modeling by means of correlative models.

3.2 Variable Selection

An issue of primary concern when constructing correlative models is the cautious and meaningful selection of the suite of variables that will function as auxiliary parameters in the modeling process. Such variables need to have a physical correlation with the dependent variable that is modeled, so that the constructed model will have some explanatory capability. In addition, it is important that the variables used as explanatory will not be correlated among themselves (presence of collinearity). For this purpose in the current analysis, the Spearman correlation was calculated for the whole suite of variables (the whole suit of variables is constituted by the 19 bioclimatic variables contained in the WorldClim database, along with elevation, slope, aspect, population density and distance to water-covered areas). The range of values for Spearman correlation varies from -1 (for strong negative correlations) to 1 (for strong positive correlations). The Spearman correlation is given by Equation 3.1:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \tag{3.1}$$

where n stands for the number of observations. The Spearman correlation was selected instead of Pearson correlation, because the variables in concern were not linearly correlated among themselves.

For the six variables that were eventually selected, a limit of 0.8 (or -0.8) was considered as the threshold above (or below) which the variables were considered correlated. Of course, this threshold is relatively arbitrary, but nevertheless it secures that strongly correlated variables will not be used as explanatory variables. However, besides the quantitative criteria that were set, the expert knowledge of Ecodevelopment S.A. (the private company that provided the mosquito data) was also utilized, so that the selected variables would have a biological meaning on formulating the distribution of *An. sacharovi*. In Figure 3.1 the correlation matrix of the six variables that were eventually used is depicted. The correlations were calculated on a 0.95 confidence level. The matrix presented below is symmetric. The non-statistically significant correlations were crossed out.

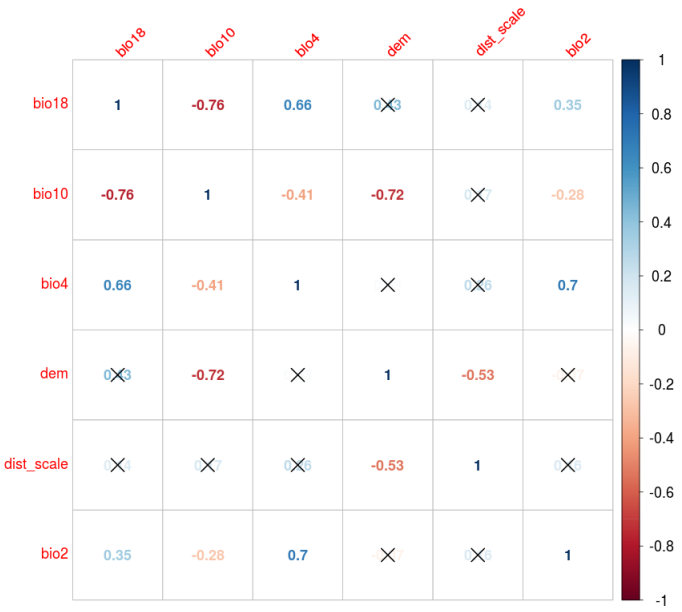


Figure 3.1: Spearman correlation for the 6 auxiliary variables used in the correlative ecological niche models.

3.3 Correlative Ecological Niche Modeling

Model Requirements

A basic requirement in ecological niche modeling is the availability of georeferenced data of where certain species have been observed. This data are widely termed as presence data. In addition, an important amount of information can be also inferred not only from where the species have been observed, but also from which locations the species was absent (hence, absence data). Both presence and absence data are complimentary one to the other and they co-function so as to provide a more coherent picture of what is the actual distribution of the species. Also, inherently in the presence and absence data lies the information of biotic and abiotic constraints that affect the observed species distribution, therefore, they produce an acceptably realistic view of the actual realized or occupied niche. Another important aspect, is that the number of presences and absences is able to modify severely the final outcome of the niche modeling procedure. Also, the spatial structure of the sampled locations can introduce a sampling bias in the niche modeling (Silero, 2011), highlighting the importance of a well-planned sampling methodology in the field, that is systematically designed so as to address very specific scientific questions.

Regardless of the importance of the absence data, in many cases this information is not available (Phillips et al., 2006). A proxy solution to this issue is the construction of a set of pseudo-absence data (pseudo from the greek word pseudo which means falsely or counterfeit), that represent areas where the species is most probably absent. The definition of the pseudo-absence data can be either determined by the random sampling of an n number of points on geographic space or by using a more sophisticated criterion that provides a higher level of confidence. The latter approach can be defined either by setting a minimum distance from observed presence locations, or by using the statistical distribution derived from presence data and using this same distribution to allocate pseudo-absence data in unsuitable locations. Nevertheless, no matter how sophisticated the pseudo-absence data definition is, there is always a risk that the pseudo-absences are placed over locations carrying suitable environmental conditions for the species in concern (Silero, 2011). In addition, there have been developed certain niche modeling methodologies that exploit only presence data and they have no use either for absence or pseudo-absence data. In the current analysis, 1000 pseudo-absence points were used and placed randomly at least 30 km away from presence locations.

Model Uncertainty

Ecological niche models, as all models, attempt to describe a realization of the reality and because reality is way too complex from what a model is able to capture, the modeling approach is possible only by means of applying simplifications and generalizations on what the model perceives as a reality. The assumptions and simplifications applied comprise an inevitable source of

uncertainty. If the output of the ecological niche models is to be used operationally for planning and decision making, it is important that the uncertainty inherent in the models is quantified, and also, that the exact source of uncertainty is identified (Wiens et al., 2009). The uncertainty introduced in ecological niche modeling can be classified into four main categories: Uncertainty in the explanatory variables used, uncertainty in the niche modeling algorithm applied, uncertainty in the field observations of species presences and absences and uncertainty related to the spatial scale in which the analysis is carried out (Phillips et al., 2006; Wiens et al., 2009).

3.3.1 Maximum Entropy

The Maximum Entropy (Maxent) algorithm is a machine learning methodology that aims at applying a robust statistical inference, based only on the available information, which is usually incomplete and fragmented (Phillips et al., 2006). Maxent was first applied in ecology in 2006 by Phillips et al., but before that, it was and still remains a data mining algorithm with various applications in astronomy, image processing, statistical physics etc. (Phillips et al., 2006). The basic concept of Maxent is that it utilizes the probability distribution of maximum entropy, in order to produce an estimation of the target probability which is unknown, due to the fact that the prior information is incomplete. The probability distribution of maximum entropy is influenced by certain constraints that provide a proxy for the incomplete knowledge that affects the application in concern (in the current case: the ecological niche of a certain species). These constraints are constituted by a set of environmental variables called features, that formulate and affect the target distribution. As an ecological niche modeling methodology, Maxent uses the locations of where the species in concern has been observed, accompanied by information that describes the environmental conditions observed in the same locations.

The big challenge in applying the Maxent algorithm is how can an unknown distribution be approximated. Jaynes (1957) provided the answer to the aforementioned problem, which eventually became the cornerstone of the Maxent approach: firstly, the unknown distribution can only be approximated by using information on what are the constraints imposed on that distribution and secondly, under those constraints the unknown distribution should acquire maximum entropy (Phillips et al., 2006). The entropy of the probability distribution is given by the following equation:

$$\eta(\widehat{\pi}) = -\sum_{x \in X} \widehat{\pi} \ln \widehat{\pi} x \tag{3.2}$$

where:

π : The unknown probability distribution

X: The set of pixels in the study area

The quantity H is non-negative and the maximum number it can acquire is the natural log of the number of pixels contained in the study area (X), hence, it can be seen that the geographic extent of the area in which the algorithm is applied, along with the spatial resolution of the constraint

features are able to greatly affect the model output. Generally, it can be stated that the higher the entropy, the more constrained is the process modeled. As stated by Jaynes (1990):

The fact that a certain probability distribution maximizes entropy subject to certain constraints representing our incomplete information, is the fundamental property which justifies the use of that distribution for inference; it agrees with everything that is known but carefully avoids assuming anything that is not known.

Maxent is a widely used methodology in ecological niche modeling and this is due to numerous reasons. Firstly, it requires presence-only data, a fact that makes Maxent easily applicable to all sampled datasets, since the lack of absence data is not uncommon and becomes a severe hindrance to the application of other ecological niche modeling methods that require absences (apart from presences). The set of presence data along with a set of spatialized environmental information (which can either be of continuous or categorical nature) are sufficient for the application of the algorithm. Secondly, there have been developed certain deterministic algorithms that secure the convergence to the maximum entropy probability distribution. Also, the maximum entropy distribution has a well-defined mathematical profile which is valuable in the analysis process. Thirdly, by the fine-tuning of the l1-regularizations one can avoid the over-fitting of the model to the data. Also, the model output is continuous and produces a scaled habitat suitability from 0 to 1, which is much more flexible than a binary output that results in a surface that is characterized either by 0 (unsuitable) or 1 (suitable) (Phillips et al., 2006). In addition, there are several ways one can apply the Maxent algorithm such as: The MaxEnt Software (v3.4.1), the biomod2, maxent or dismo packages in R, or by using the QSDM python plugin in QGIS.

Nevertheless, Maxent carries along some disadvantages which are mainly due to the fact that it is a generally new methodology (compared to regression models) and hence, the theory behind its correct application is limited (Phillips et al., 2006). Also, an important drawback in Maxent, arises from the fact that the user-friendly platforms that have been developed for its easy and quick application, have resulted to an extremely wide use of the algorithm using some default set values. It is common in recent studies that model the ecological niche of various species, that the default values are used with no further investigation of what their impact on the final model output is.

3.3.2 Generalized Linear Models

A very basic and common practice in statistics when there is a set of explanatory variables on which a random variable is dependent on, is the application of a regression model. There is a wide suite of regression models in statistics that range from very simplistic to more sophisticated models and their applicability depends upon the research question in concern. The main attributes that differentiate the various regression models with regards to their complexity is whether or not they allow for a non-gaussian distribution in the response variable, whether or not they allow for non-linearity in the explanatory variables and also, what are the main assumptions they make concerning the distribution of errors and their linkage to the mean.

One of the fundamental regression models in statistics is the General Linear Model, which assumes that a response variable y is a linear function of a set of predictor variables, so that:

$$y \sim N(\mu, \sigma^2) \tag{3.3}$$

$$\mu = b_1 x_1 + b_2 x_2 + ... + b_n x_n + b_0 \tag{3.4}$$

With y being the response variable, μ the mean field, σ^2 the variance, and the various x are the predictor variables. This kind of models are characterized as "General" because they allow for more that one predictor variables to be included in the model (in contrary to the simple linear models). In the General Linear Model described above it is assumed that the response variable y follows the normal distribution and that the errors are not correlated. In the present case, the response variable y is the habitat suitability for the occurrence of *An. sacharovi* mosquitoes and the covariates in the model do not bear a gaussian form, hence General Linear Models could be characterized as unfit for the current analysis, due to their assumptions.

The aforementioned deficiencies are dealt by the use of Generalized Linear Models (GLMs), which can be characterized as an extension of General Linear Models. One of the primary advantaged of GLMs is that they do not lie on the assumption that the response variable is normally distributed. GLMs also assume that the explanatory variables are linearly related to the response variable. In addition, a structural element of a GLM is a probability distribution describing the response variable. The probability distribution must belong to the exponential family of distributions, that have the following probability density function:

$$p(y_i, \lambda_i, \phi) = exp(\frac{y_i \lambda_i - b(\lambda_i)}{\phi}) + c(y_i, \phi) \tag{3.5}$$

The parameter λ_i is the canonical parameter that depends on the explanatory variables and ϕ is the dispersion parameter which is normally equal to 1 when the probability distribution used is either the Poisson or the Binomial. In the current analysis, since the response variable is of binary nature (presence-absence/pseudo-absence records) the most suitable probability distribution is the Binomial.

In the case of modeling binary data employing the binomial distribution, the response variable Yi can be approximated as:

$$Y_i \sim Binomial(n_i, p_i) \tag{3.6}$$

The property that it is desired to be modeled is $p_i = \frac{Y_i}{n_i}$ which is defined as the ration of the hits in n independent Bernoulli trials. The mean value and the variance of the p_i variable are defined as following:

$$E(\frac{Y_i}{n_i}) = p_i \tag{3.7}$$

for the mean and

$$var(\frac{Y_i}{n_i}) = \frac{1}{n_i} p_i(1 - p_i) \tag{3.8}$$

for the variance

Apart from the probability distribution, the second structural element that a GLM carries, is a monotonic link function that provides a description concerning how the mean relates to the linear predictor. For the binomial distribution, a typical link function is the logarithmic link function:

$$g(\mu_i) = logit(\mu_i) = log(\frac{\mu_i}{1 - \mu_i}) \tag{3.9}$$

where μ_i refers to the mean value of the response variable Y.

3.3.3 Generalized Additive Models

Generalized Additive Models (GAMs) can be considered as an extension of GLMs. The main characteristic that differentiates GAMs from GLMs is that GAMs allow for a non-linear relationship of the predictor variables, that basically utilizes a smooth function that is estimated from the data. The parameters calculating the smooth function are simultaneously estimated for all the predictors and then added (hence: additive). An issue that arises from using a smooth function is how this smooth function will be defined (what will the parameters of the spline be) and also, how smooth will it eventually be. GAMs can be characterized as models with high plasticity, which makes them suitable for fit to data with complex structure. Nevertheless, a basic drawback of such a case, is that GAMs are very easily over-fitted to the data. The general structure of a GAM model is the following:

$$g\mu_i = x_i\theta + f_1(x_{1i}) \tag{3.10}$$

where $\mu_i \equiv E(Y_i)$ and Yi is modeled using a distribution that belongs to the exponential family of distributions. Yi is the response variable, f_j are the smooth functions used for each predictor and x are the predictors that compose the model.

3.3.4 Classification Tree Analysis

Classification Tree Analysis is a machine learning algorithm that aims to provide a relationship between a response and various explanatory variables, in order to classify the variable in concern, according to certain constraints. The constraints are provided by the predictor variables that are used in the decision tree. The decision tree can either be categorical or a regression tree (if the variables contained are continuous).

3.3.5 Random Forest

The random forest method for ecological niche modeling is classified among the machine-learning algorithms and it comprises an extension of classification trees, in the sense that it uses many classification trees (instead of a single tree) and thus, creates a forest (hence the name random forest). Eventually, the final model combines the best predictions from all the trees (Cutler et al., 2007). The algorithm is based on randomly bootstrapping the original data, which in the case of ecological niche modeling are the predictor variables, for every node of the forest. The nodes of the forest, in the current analysis, are comprised by the set of absence records. The predictor variables that provide the best model for a certain node are retained.

The structure of a random forest consists of a root node that initiates the procedure using a large number of bootstrap samples and then proceeds to the interior nodes that use different randomly selected subsets of the predictor variables. In random forest, the randomly selected subsets are chosen in a procedure called bagging. Starting from the root nodes and proceeding to

the interior nodes, the random forest ends to the terminal node, that provide an estimate of the best fit model to the data. Random forest is a widely used classification method, and its advantages lie on the fact that it is a non-parametric approach, it can classify both continuous and categorical data, it can handle a large number of input data and the final result is independent of the different internal models that are developed, so that all models have an equal probability to be included in the final model. The basic functionality of a random forest is depicted in the diagram.

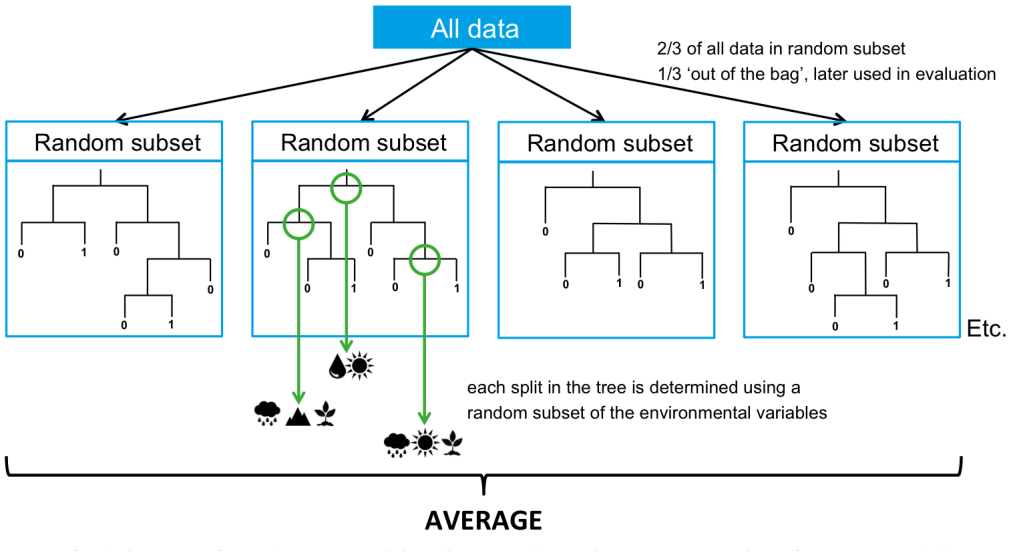


Figure 3.2: A representation of how the random forest algorithm works. Source: https://support.bccvl.org.au/support/solutions/articles/6000083217-random-forest

3.3.6 Ensemble Niche Modeling

There is a wide variety of correlative ecological niche modeling methodologies that provide a solution to the modeling of species suitable habitats. The whole suite of available statistical models could be classified into three main categories, which are: regression models, machine learning algorithms and classification - enveloping methodologies. Each approach is based on a different set of assumptions and parameters, resulting to different model outputs. The difference and disagreement in the modeled outputs, highlights the uncertainty inherently lying in the ecological models applied (Wiens et al., 2009) and the issue of equifinality that arises from the complex nature of the modeled phenomenon, which is approximated using statistical models that have too many parameters to be fine-tuned and the available knowledge and observation of the species distribution and the environmental conditions is not sufficient to provide a statistically robust estimate of these parameters, resulting to the issue of parameter non-identifiability (Dormann et al., 2012).

The aforementioned characteristics of the ecological modeling approach can be a great drawback when their output aims at better informing the decision making procedure (Wiens et al., 2009). A possible solution to this issue is the application of an ensemble of different niche models (Thuiller et. al., 2009), so that there can be a quantification of the uncertainty introduced. In the current analysis, the ensemble of the following ecological niche models was applied: Maximum Entropy, Generalized Linear Models, Generalized Additive Models, Random Forest and Classification Tree Analysis, that were described earlier.

As it was stated above, apart from the uncertainty introduced to the ecological niche model outputs due to the modeling algorithm used, there is an additional source of uncertainty that is imposed on the final output and it is introduced via the covariates. The environmental variables that are used as predictors are also model outputs and thus, are also subject to uncertainties. More specifically, climatic variables that result from Regional Climate Model (RCM) outputs can display a systematic bias, which needs to be addressed, especially when they are used as input in impact assessment studies. Despite the fact that RCM outputs are able to reproduce the basic climatic spatiotemporal patterns over Europe, it is also highlighted that they display certain deficiencies in their skill (Kotlarski et al., 2014). For instance, it is shown that there is a persistent warm and dry bias in summer over the region of Southeastern Europe (Kotlarski et al., 2014). A typical approach that is applied for addressing the aforementioned issue, is the calculation of a multimodel ensemble that uses the mean, weighted mean or median of a suite of RCM outputs (Kotlarski et al., 2014). This approach is also applied in the current work. The diagram above describes the procedure followed, so that both the uncertainty introduced due to the modeling algorithm used and due to the input variables used is effectively addressed.

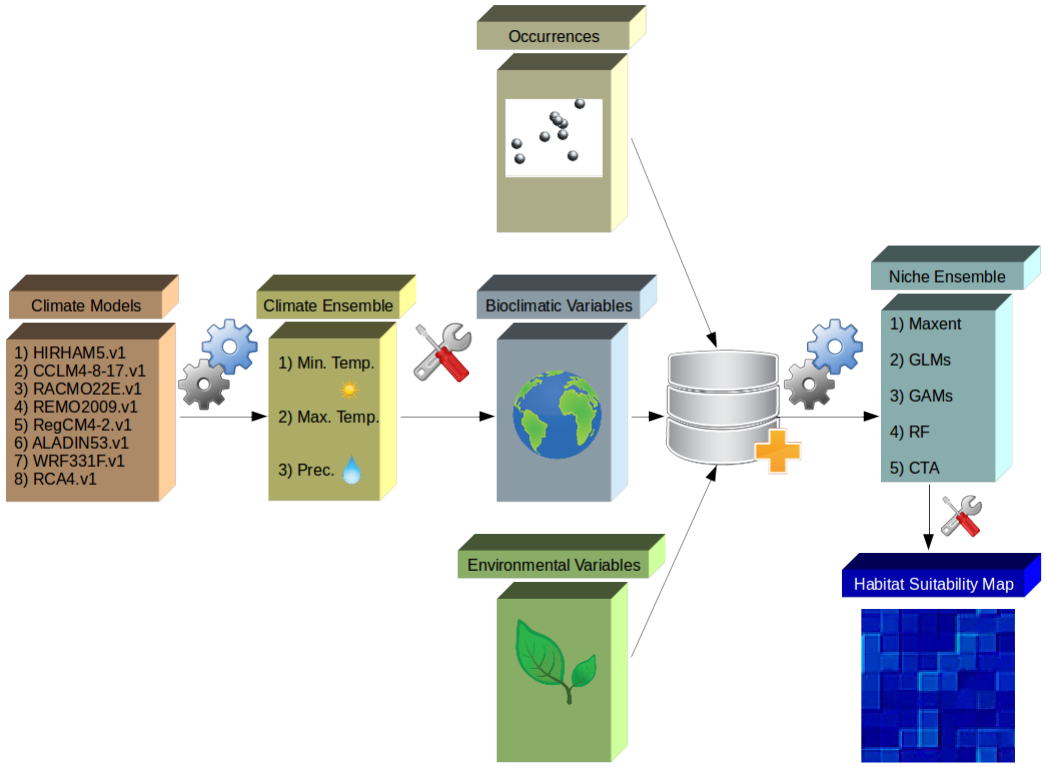


Figure 3.3: A schematic representation of the modeling procedure in the current analysis.

3.3.7 Ensemble Niche Model Evaluation

A very robust and solid question after applying a model of any kind, is how one evaluates the accuracy and validity of the model produced. Typically, there are two ways through which the validation procedure is performed. Firstly, the model output is compared pixel-by-pixel (once the model produces a spatial output) with a surface that is considered to be a valid realization of the reality. The comparison in such a case is performed by applying several statistical error metrics, that treat the true realization of the variable in concern as the observed value and the model output as the modeled value. In fact, the true value of the variable in concern in the whole study domain is actually unknown, but conventionally, this approach is applied in numerous evaluation studies and is considered as a comparative tool for the assessment of a models performance. Apart from the aforementioned method though, there is also another very widely-used approach for model validation and this is the splitting of the initial dataset into two separate sets, one used exclusively for training and one used exclusively for validation. There are different guidelines on how the initial dataset should be separated, but the most commonly applied method is the random splitting. In that, normally 60 % or 70 % of the initial data is used for training and the rest 40 % or 30 % is used for validation. In the current analysis, 70 % of the data is used for training and the remaining 30 % is used for validation.

In the case that the input model data are of binary nature, the model is required to have a good skill in identifying whether the explanatory variables used in the model are conducive to the model predicting a zero (0) or an one (1) (in the present case: 0 for species absence and 1 for species presence). This issue of discriminating and classifying a signal as true signal or a noise, was first observed in radar signal detection and processing and later on in medical research. From 2000 onwards it has found a wide applicability in ecological niche models, as they are normally trained on binary data (Jimenez-Valverde, 2012). The methodology applied for the successful identification of hits and misses of a certain procedure was performed by calculating the Area Under the Receiver Operating Characteristic (ROC) graph (Jimenez-Valverde, 2012). The area under the ROC curve utilizes the models sensitivity along with model's specificity, in order to assess how successful a model is in discriminating and correctly classifying presence and absence records, without mixing the two classes (e.g. classifying absences as presences (commission error) or classifying presences as absences (omission error)). More specifically, a models sensitivity is defined in the following way:

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{3.11}$$

while a models specificity is defined in the following way:

$$Specificity = \frac{TrueNegatives}{TrueNegatives + FalseNegatives} \tag{3.12}$$

In the current case, positives are regarded as the presences and negatives as the absences. The AUC plot uses Sensitivity (on the y axis) versus 1-Specificity (on the x axis). The diagonal of

this plot implies that the model has a random chance of discriminating correctly presences and absences (AUC=0.5).

For the case of Maxent, where pseudo-absences are used instead of real absences, the AUC plot uses in the x-axis the proportion of background points that were predicted by the model as presences (instead of using 1-Specificity) (Jimenez-Valverde, 2012). It is desired that the AUC value is as close to the unity as possible. Generally, when the AUC value is greater than 0.7, the model is considered skillful, although this is an arbitrarily set threshold. Nevertheless, despite its wide use and acceptance in assessing the accuracy of ecological niche models, the AUC metric has received some criticism that led to its revisiting and modification. This led to the creation of the True Skill Statistic (TSS), which is used in conjunction with AUC and it is considered to compensate the deficiencies of AUC.

The TSS score, also known as Hanssen and Kuipers discriminant, utilizes sensitivity and specificity scores, is independent of the size of the dataset and also, is independent of prevalence (Allouche et al., 2006) when certain assumptions are met and the size of the sample is large (Somodi et al., 2016). The values of TSS range from +1 to -1 with +1 describing a fully skillful model and -1 a completely unskillful model. Values close to 0 describe a model that can produce correct predictions only by chance (Allouche et al., 2006). The TSS metric is defined as:

$$TrueSkillStatistic = Sensitivity + Specificity - 1 \tag{3.13}$$

In the current analysis, both AUC (ROC) and TSS metrics are used for assessing the models skill.

3.4 VECTRI Malaria Model

VECTRI is a compartmental model that explicitly represents the progression of the various Anopheles development stages and the development of the parasite inside the vector (sporogonic cycle) and inside the host. As it was discussed in Chapter 1, the development of the Anopheles mosquitoes proceeds from the egg stage, to the larvae stage and to the pupae stage and eventually, if the conditions are conducive, the adult mosquito emerges. VECTRI employs the degree day concept in order to account for this progression. Also, the gonotrophic and the sporogonic cycles are explicitly resolved using an array of bins, representing the various stages of the malaria transmission cycle (Tompkins and Ermert, 2013). The progression from one stage to the other is performed using the advection equation:

$$\frac{dL}{dt} = R_L \frac{dL}{df} \tag{3.14}$$

where L is the larvae life cycle, f is the fractional growth stage and corresponds to the fractional growth rate (Tompkins and Ermert, 2013).

In addition, the vector life-cycle is also analyzed using the bin approach. The two attributes that directly influence the dynamics of malaria transmission is the development of the parasite inside the vector (sporogonic cycle) and the development of the egg inside the female Anopheles mosquito (gonotrophic cycle), which eventually has an impact on what rate new eggs are added to the larvae cycle and how often mosquitoes are in a blood-meal searching mode, which translates in bites (Tompkins and Ermert, 2013). Thus, the sporogonic and gonotrophic cycles are also included in VECTRI in a two-dimensional array V(Ngono, Nsporo). When the vectors reach the last stage of the sporogonic cycle they are considered to be infective to humans and the disease can transmit to the host at any possible bite (Tompkins and Ermert, 2013). It is noteworthy though, that not all bites from an infective vector to a susceptible host can result to a new infection. The percentage of successful transmissions from an infectious vector to a susceptible human or the acquiring of a blood meal from an infected host to a susceptible vector is estimated to be approximately 20-30 % (Ermert et al., 2010).

The basic concept on which the developmental stages in VECTRI are progressed, lies on the concept of degree days (Tompkins and Ermert, 2013). Degree days is a common practice in biology and ecology, yielding information on when the temperature criteria are met in order for a certain species to proceed from one stage of its development cycle to the next (Gu and Novak, 2005). Degree day studies normally include the observation of the species in concern in constant temperatures in laboratory experiments, but the output of this approach may yield very different results than what is normally happening in the field, due to a series of other reasons, such as the temperature variability, the quality of the environment in which the species is required to survive, the possible presence of other members of the same species (for instance, larvae density in the same pond), which would increase competition for the same resources etc. (Gu and Novak, 2005). In order for a degree day approach to be defined, it is important that the minimum (or threshold)

and maximum temperatures are identified. The degree day criterion requires that the average temperature of a day exceeds the threshold temperature by at least 1 degree (Teng and Apperson, 2000). If that requirement is met, then one degree day is completed. In order for the species to proceed in its developmental stages, it is necessary that a certain amount of degree days are completed, thus, every degree day is accumulated to the next, until the total degree day number is achieved.

3.4.1 Modeling of the Larvae Life Cycle

The life-cycle of larvae contains the transition from egg to larvae and from larvae to pupae. The progression from one stage to the next can efficiently be approximated using the degree day concept (Detinova, 1962). The larvae life-cycle contains the aquatic stages of the mosquitos life-cycle, hence, the variable that primarily governs its growth rate is water temperature T_{wat} which should be above a threshold temperature TL_{min} so that the larvaes survival is possible. The degree day concept included in VECTRI and describing the larvae growth rate is given by the following equation:

$$R_L = \frac{T_{wat} - TL_{min}}{K_L} \tag{3.15}$$

A parameter that requires investigation and its setting may be debated is the K_L parameter which is the rate coefficient, indicating how many degree days are required in order for the larvae to progress through its developmental stages. This parameter differentiates between different Anopheles species. Also, although the aquatic mosquito stages contain the egg-larvae-pupae cycle, VECTRI actually allocates one day for eggs progressing to larvaes and one day for larvaes progressing to pupae.

In addition, apart from the growth rate of the larvae, an other aspect that may greatly influence the transmission dynamics is the larvae mortality, which also displays strong dependency on temperature (Tompkins and Ermert, 2013). Apart from temperature though, other biotic factors may influence larvae mortality, such as the presence of predators or the increased larvae density. The effect of biotic and abiotic factors on larvae mortality is incorporated in VECTRI by employing the following relation:

$$PL_{surv} = (1 - \frac{M_L}{wML_{max}})K_{flush}PL_{surv0} \tag{3.16}$$

The term M_L describes the total larvae biomass per unit surface area of a water body, while w describes the fraction coverage of a grid cell by locations that could potentially function as breeding sites. The term ML_{max} describes the total carrying capacity of the water pond. The term K_{flush} incorporates the impact of heavy rainfall on larvae mortality and has a greater impact on young stage larvae than on older stage larvae (Paaijmans et al., 2007). The effect of flushing rainfall is incorporated in VECTRI using the following equation:

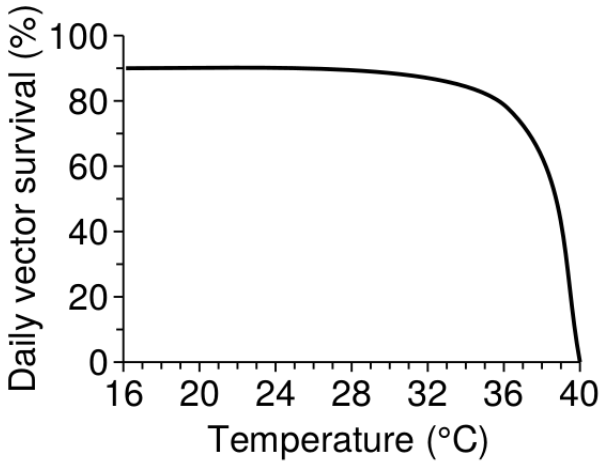


Figure 3.4: Daily vector survival (%) as a function of temperature for *An. gambiae* (Craig et al., 1999).

$$K_{flush} = L_f + (1 - L_f)((1 - K_{flush}, \infty))e^{\frac{-R_d}{\tau_{flush}}} + K_{flush}, \infty) \tag{3.17}$$

The R_d parameter describes the rainfall rate in mm per day, K_{flush}, ∞ is the maximum value of K_{flush} for newly hatched young larvae, and τ_{flush} describes how the effect of flushing increases on larvae mortality as a function of the rainfall rate.

3.4.2 Modeling of the Vector Life Cycle

As it is expected, the effect of temperature on the mosquito life-cycle extends beyond the larvae stage onto the adult stage (Tompkins and Ermert, 2013). High temperature is known to have a negative effect on adult mosquito survival (Craig et al., 1999; Mordecai et al., 2017), nevertheless, the exact thresholds of the maximum temperature that mosquitoes can tolerate is highly uncertain, and varies between different genuses and species. This uncertainty lies primarily on the fact that few samples are available for such high temperatures. Usually, the abrupt downward decrease of such curves is due to the fact that a statistical model is forced to pass through those few samples (Mordecai et al., 2017).

The effect of temperature on vector survival is incorporated in the VECTRI model using two available schemes that are described below, either as a quadratic function of temperature, according to Martens et al. (1995):

$$PV_{surv1} = K_{mar1,0} + K_{mar1,1}T_{2m} + K_{mar1,2}T_{2m}^2 \tag{3.18}$$

or as an exponential function of temperature, according to Martens et al. (1997) and Craig et al. (1999).

$$PV_{surv2} = exp(\frac{-1.0}{K_{mar2,0} + K_{mar2,1}T_{2m} + K_{mar2,2}T_{2m}^2}) \tag{3.19}$$

3.4.3 Gonotrophic and Sporogonic Cycles

As it was discussed above, the gonotrophic cycle describes the egg development inside the female mosquitoes. In order for the egg development to progress, it is necessary that a blood meal is acquired. Currently, in the VECTRI model it is assumed that female mosquitoes acquire their necessary blood meal during the first night of searching (Tompkins and Ermert, 2013), which might not always be the case. The achievement of a blood meal is a direct function of the proximity to populated areas, either by humans or by animals. The rate at which the egg development proceeds inside the female mosquito is also a function of temperature and is incorporated in VECTRI employing the degree day approach, as described below:

$$R_{gono} = \frac{T_{2m} - T_{mingono}}{K_{gono}} \tag{3.20}$$

The quantity in concern is the resulting number of female eggs that are laid by the female mosquito (the male eggs are indifferent in the transmission cycle). The number of eggs laid by a female mosquito is highly variable between the various species (Tompkins and Ermert, 2013).

The sporogonic cycle refers to the development of the malaria parasite inside the adult mosquito. During every blood meal there is a possibility that the parasite is transmitted from the host to a vector (or the opposite), depending on which of the two is infected and which is the susceptible (Tompkins and Ermert, 2013). During every time step of the model run in VECTRI (currently: daily time step) a certain portion of the susceptible mosquitoes move to the infected class, meaning that they have acquired the malaria parasite. From that moment on, the development of the parasite is occurring, until they reach the state when the mosquitoes themselves are infectious and thus, can potentially transmit malaria in any of their next blood-meals. The rate at which the parasite develops inside the mosquito is also dependent on temperature and is approximated in VECTRI using the degree day concept, as given below:

$$R_{sporo} = \frac{T_{2m} - T_{minsporo}}{K_{sporo}} \tag{3.21}$$

3.4.4 Host Community

One of the modeling innovations introduced in VECTRI is the representation of the population density on a spatialized grid, allowing for the inclusion of the ratio of potentially biting vectors to potential hosts for every grid cell and the calculation of the human biting rate (hbr), which describes the number of bites that individuals may potentially receive (Tompkins and Ermert, 2013). The human biting rate is given by the following equation:

$$h\bar{b}r = (1 - e^{\frac{-H}{\tau\infty}}) \frac{\sum_{j=1}^{N_{sporo}} V(1,j)}{H} \tag{3.22}$$

The term $1 - e^{\frac{-H}{\tau\infty}}$ indicates what is the amount of zoophily for the Anopheles species in concern. The zoophily parameter as given in the equation above gains greater weight in areas where population density is relatively low and the only sources of available blood meals are animals (Tompkins and Ermert, 2013). The term H represents the population density.

3.4.5 Hydrology in VECTRI

The input concerning surface hydrology is of primary importance in a model of malaria transmission, as it provides input related to the availability of potential breeding sites for vectors (Tompkins and Ermert, 2013) and captures the impact of rainfall on the disease dynamics, allowing for an investigation of the lag existing between rainfall events, adult mosquitoes emergence and malaria occurrence, which could be ideally used in an early-warning system. Potential breeding sites consist primarily of temporary water ponds that are found in the fringes of permanent water bodies, such as lakes and rivers, or over areas where the surface attributes, such as slope, curvature and soil type allow for the assemblance of the rainfall water in ponds. The latter, provides a more suitable environment for oviposition, as there is generally lack of predators and the waters are relatively calm (Tompkins and Ermert, 2013).

In the VECTRI model hydrology is not represented explicitly but a parametrization scheme is used instead (Tompkins and Ermert, 2013). The VECTRI model is designed to operate and simulate malaria transmission in a regional scale and it has been shown that an explicit representation of hydrology does not necessarily improve the model performance for regional scale studies (Asare et al., 2016). According to Asare et. al. (2016) the improved surface hydrology is given by the following equation:

$$\frac{dw_{pond}}{dt} = \frac{2}{ph_{ref}}\left(\frac{W_{ref}}{W_{pond}}\right)^{\frac{P}{2}}((Q(W_{max} - W_{pond}) + PW_{pond})(1-f) - W_{pond}(E + fl_{max}) \quad (3.23)$$

In the equation presented above, W_{pond} represents the fractional coverage of water for every grid cell and W_{max} is the temporary pond coverage area, E is the evaporation rate, I represents the infiltration rate and P is the precipitation rate. In addition, p is a parameter representing the shape factor of the pond, h_{ref} is the total depth of the reference pond, W_{ref} is the reference fractional coverage equated to K_w, Q is the runoff and f is equal to $\frac{1}{4}\frac{W_{pond}}{W_{max}}$.

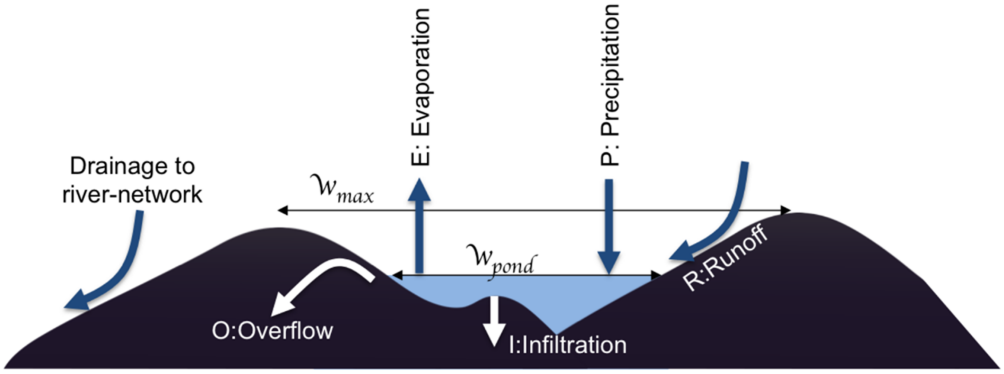


Figure 3.5: A schematic representation of the hydrology in VECTRI (Asare et al., 2016).

3.4.6 Model Requirements

In order for the VECTRI model to produce an estimate of the malaria transmission dynamics, it is required that precipitation and temperature fields are provided, and renewed in every time-

step of the model, yielding an estimate of what is the status of transmission at any given time. In addition, population density information is also required as a static input field. Nevertheless, apart from the variables that are necessary to be provided for a successful model run, there is also a suite of constants that need to be tuned, in order for the model to produce reliable and realistic estimates of the parameters related to the vector and to the dynamics of malaria transmission. Currently, these parameters are fine tuned for modeling the activity of *Anopheles gambiae* and the dynamics of *Plasmodium falciparum* (Tompkins and Ermert, 2013), but fine tuning of these constants is required for applying VECTRI to other geographic regions where other Anopheles species are found and other plasmodium kinds are observed.

In the current analysis, the input temperature and rainfall data consisted of the same RCM ensemble mean outputs that were used for calculating the bioclimatic variables that were employed in the correlative models. The input data covered the period 1990-2008 and had a daily timestep.

3.4.7 Model Outputs

After a successful run of VECTRI, there are several outputs describing the dynamics of malaria transmission. These outputs are:

Entomological Inoculation Rate (EIR): The number of infective bites that an individual receives after a period of time.

Detectable Parasite Ratio (PRd): Number of people with malaria that is detectable after 10 days.

Human Biting Ration (HBR): Number of people that are bitten daily.

Cirmu sporozoite Protein Rate (CSPR): Fraction of EIR/HBR

Vector Density: Mosquito density measured per m^2.

Larvae Density: Density of the mosquito larvae measured per m^2.

Cases: Number of new malaria cases emerging.

3.5 An attempt to calibrate VECTRI

As it was discussed above, VECTRI is currently fit for modeling malaria transmission over areas where the primary vector is *Anopheles gambiae*. Nevertheless, *Anopheles gambiae* is absent from the study region and the species in concern is *Anopheles sacharovi*, hence, it is desired that the VECTRI parameters currently fit for *An. gambiae*, would be modified so as to better describe *An. sacharovi*. The whole suite of parameters on which VECTRI is dependent on can be found in Tompkins and Ermert (2013) and they can be classified into the following categories: parameters related to larvae, parameters related to hydrology, parameters related to intervention measures and to mosquito biting characteristics. Also, VECTRI assumes parameters related to immunity and transmission, parameters related to the population and finally, parameters related to the sporogonic and gonotrophic cycles. Most of these parameters in the current VECTRI version are set using figures obtained from the literature.

The modifying of those parameters could either be performed by employing the same method of reviewing the literature related to the species in concern or by applying an intelligent machine-learning algorithm. By following the latter approach, model parameter setting would be treated as an optimization problem that explores efficiently an enormously big search space. However, before applying more intensive methods in model calibration, it is necessary that the available literature would be reviewed and certain runs with the modified parameters would be performed. The method of reviewing the literature was applied in the current analysis.

In the current attempt of calibrating VECTRI for *An. sacharovi*, only larvae parameters and parameters related to the gonotrophic cycle were modified and also, from those parameters, only parameters related to temperature were calibrated, with an exception to the larvae survival rate due to non-climatic factors. The reason for modifying parameters related to the gonotrophic cycle and larvae is because the VECTRI model output that is investigated in the current analysis is the vector density.

Table 3.1: The default larvae and gonotrophic cycle parameters related to temperature in VECTRI and the modified parameters taken from the literature.

Parameter	Default VECTRI value:	Modified value
Threshold temp. for egg development in vector (`rtgono`)	7.7	9.9
Degree days for egg development in vector (`dgono`)	37.1	36.7
Min. temp. for larvae survival (`rlarv_tmin`)	12.16	15.5
Max. temp. for larvae survival (`rlarv_tmax`)	38	35
Survival rate due to non-climatic factors (`rlarvsurv`)	0.987	0.85

Concerning the setting of parameters related to the gonotrophic cycle, the values were taken from Detinova (1962). However, there are other studies expressing the duration of gonotrophy in constant temperature, but do not provide information concerning the threshold temperatures. For instance, Tavanol and Caglar (2008), refer to gonotrophy's duration equal to 6.3 days under 27 °C,

while Kasap et al. (1989) refer to the gonotrophic cycle being equal to 13.2 under 24 °C and equal to 9.2 under 28 °C. In addition, the minimum temperature for larvae survival was set equal to 15.5 according to Kampen et al. (2003), however for maximum temperature values ranged in the literature from 30 - 40°C, thus this parameter was set equal to 35 °C. Also, according to Tavanol and Caglar (2008) the survival rate due to non-climatic factors was set equal to 0.85.

The VECTRI run using the default values is further on referred to as the "pre-calibration" run and the VECTRI run with the new parameters is referred to as the "post-calibration" run. Hence, firstly, the monthly and seasonal means for both runs were calculated and their differences were also calculated and investigated. Furthermore, for specific locations of known vector activity the time-series of the "pre-calibration" and the "post-calibration" runs were extracted. Also, for both runs, the vector density during the summer months was compared to the output of the ecological niche models using the run exploiting the RCM data. The covariance between "pre-calibration" and "post-calibration" to ecological suitability was also calculated.

Chapter 4

Results

4.1 Suitability maps

4.1.1 Bioclimatic variables computed with the WorldClim dataset

The environmental suitability map that results from the application of the five statistical models (GAM, GLM, RF, CTA and MAXENT) using the bioclimatic variables computed with the WorldClim dataset is presented in Figure 4.1. As it is shown, high suitability values are located over the region of North-Central Greece and sporadic suitable locations are additionally located over restricted areas in mainland Greece, over Attica, South Peloponnese and over the Eastern islands of the Aegean Sea.

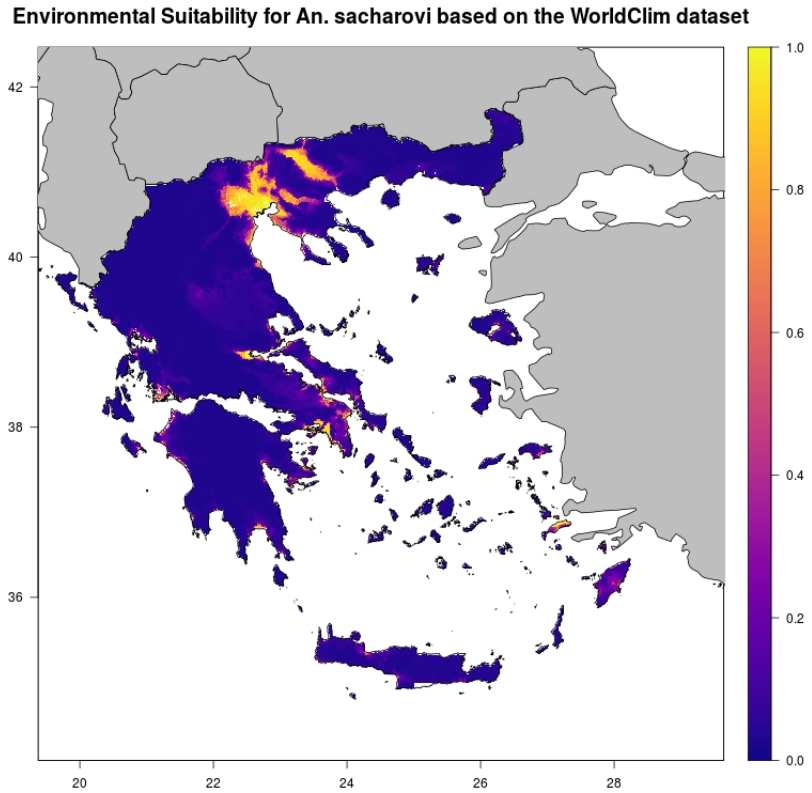


Figure 4.1: Environmental suitability of the ensemble niche modeling, where the bioclimatic variables were computed with the WorldClim dataset.

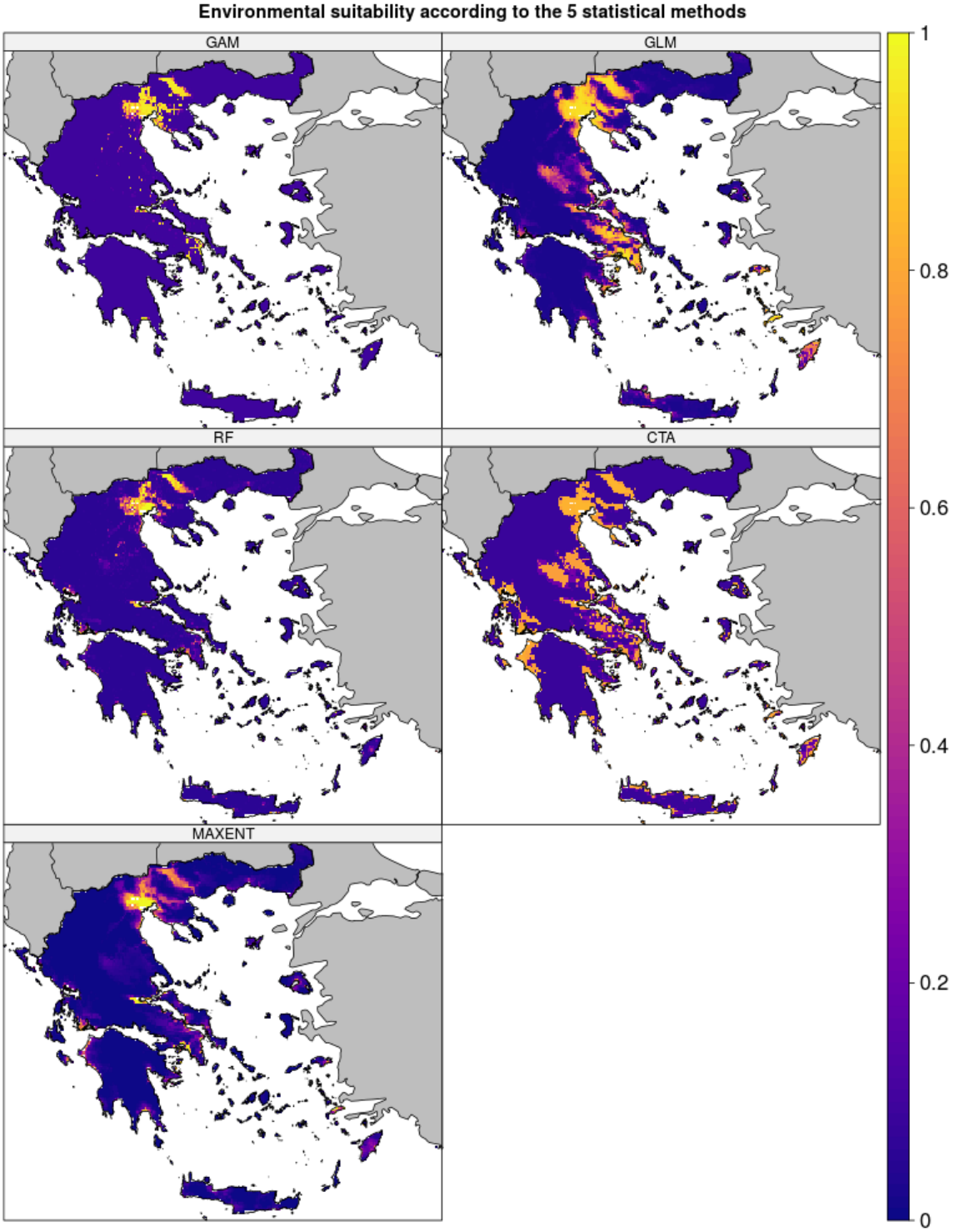


Figure 4.2: Environmental suitability according to the 5 statistical methods used, where the bioclimatic variables were computed with the WorldClim dataset.

The map in Figure 4.1 resulted from the calculation of the ensemble median of the individual maps presented in Figure 4.2. The ensemble median was chosen over the ensemble mean, because the mean is known to be highly influenced by outliers. The differences observed in the maps of Figure 4.2 is an aftereffect of the different mechanics, assumptions, limitations and capabilities of each method applied. The miscellaneous patterns of suitability that result from the various methods, highlight the need and the value of applying an ensemble methodology in correlative ecological niche models. Applying an ensemble median, is considered as a reliable and convenient option for filtering out the uncertainty introduced by each statistical method separately. As it is shown in Figure 4.2 the spatial patterns yielded by GAM, RF and MAXENT display significant similarities, while GLM and CTA yield a quite modified result that identifies suitable environments for *An. sacharovi* in more extended regions. Such an observation flows from the fact that GAM, RF and MAXENT are more sensitive to overfitting to the training dataset and thus their result is highly driven by observations. As it is displayed in Figure 4.2, GAM, RF and MAXENT predict high suitability over regions where there is an increased concentration of traps, while GLM and CTA are less influenced by the clustering of traps.

For each of the five statistical models 100 evaluation runs were performed and the ROC and TSS metrics were calculated for every run. The models that eventually contributed to the construction of the ensemble median are the ones that scored a value of TSS greater than 0.7. The ROC metrics of every statistical model of each run is presented in Figure 4.3. As it is shown, RF and MAXENT display the best performance with regards to the ROC metric, followed by GLM. More specifically, all RF and MAXENT runs yield ROC metrics greater than 0.75, while GLM yield to ROC metrics slightly less than 0.75 for certain runs, but its overall performance is similar to that of RF and MAXENT. GAM and CTA display a less good performance. With regards to the TSS metric, all models display a relatively weaker performance (compared to the ROC metric). Nevertheless, MAXENT seems to outperform the rest of the methods, followed by GLM, RF, CTA and eventually, GAM. For MAXENT the lowest performance is greater than 0.6, while for RF the lowest TSS score is equal to 0.6. For GLM the lowest TSS metric is less than 0.5 while for GAM and CTA it approximates 0.

The statistical means of the ROC and TSS metrics for each statistical method separately for all 100 runs are presented in Table 4.1. The highest ROC metric is achieved by MAXENT and is equal to 0.94 followed by GLM and RF both yielding a ROC metric equal to 0.92. Lastly, CTA and GAM result to a ROC metric equal to 0.79. With regards to the TSS metric, as it was obvious from the figures above, the overall performance drops, but the pattern of performance is similar between the five statistical methods. More specifically, MAXENT marks the highest TSS metric equal to 0.86, GLM follow with a TSS equal to 0.83, followed by RF, CTA and GAM with TSS metrics equal to 0.8, 0.6 and 0.58 respectively.

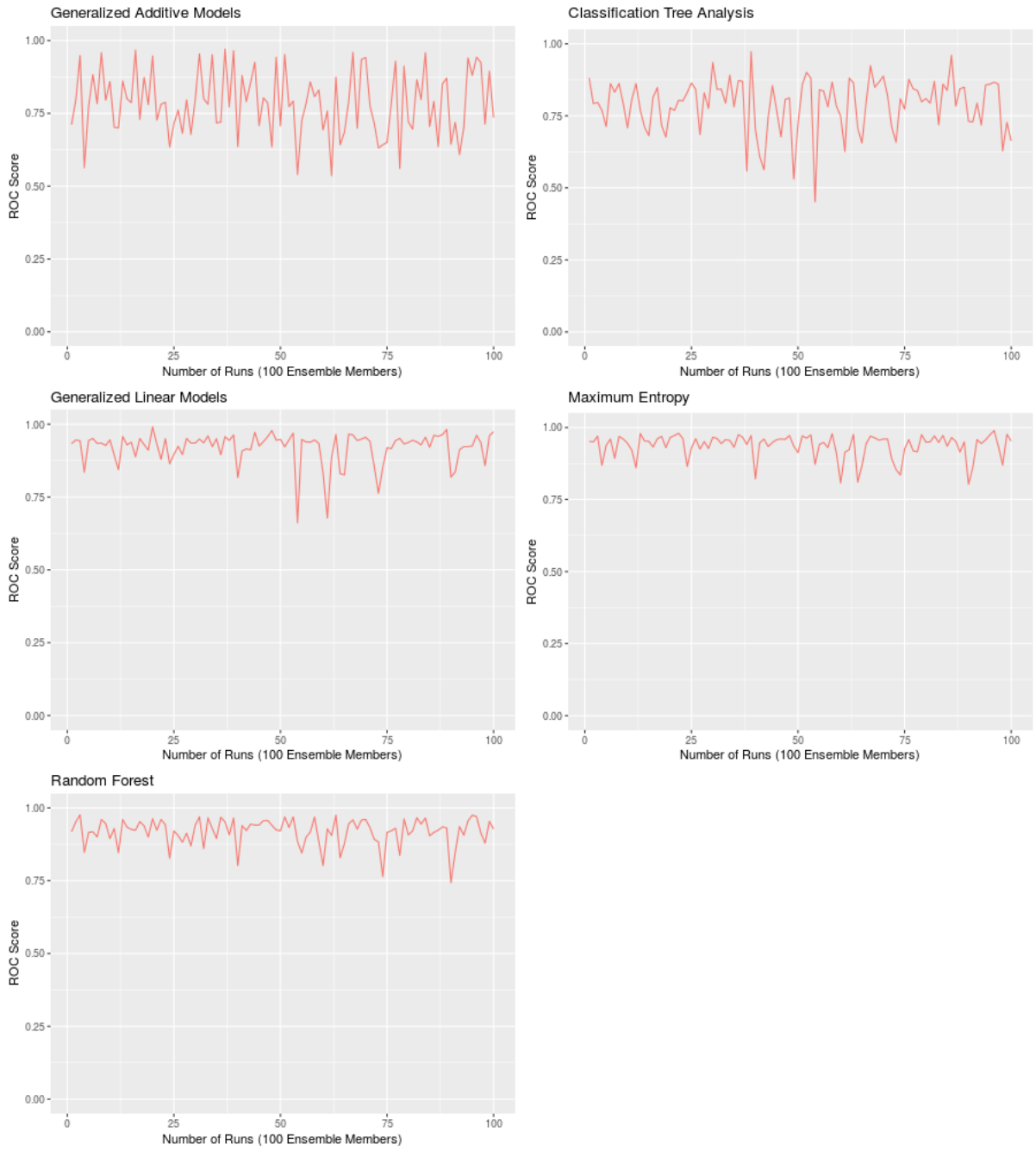


Figure 4.3: The ROC metric for all 5 statistical methods, for all 100 runs performed for each statistical method, run with bioclimatic variables computed with the WorldClim dataset.

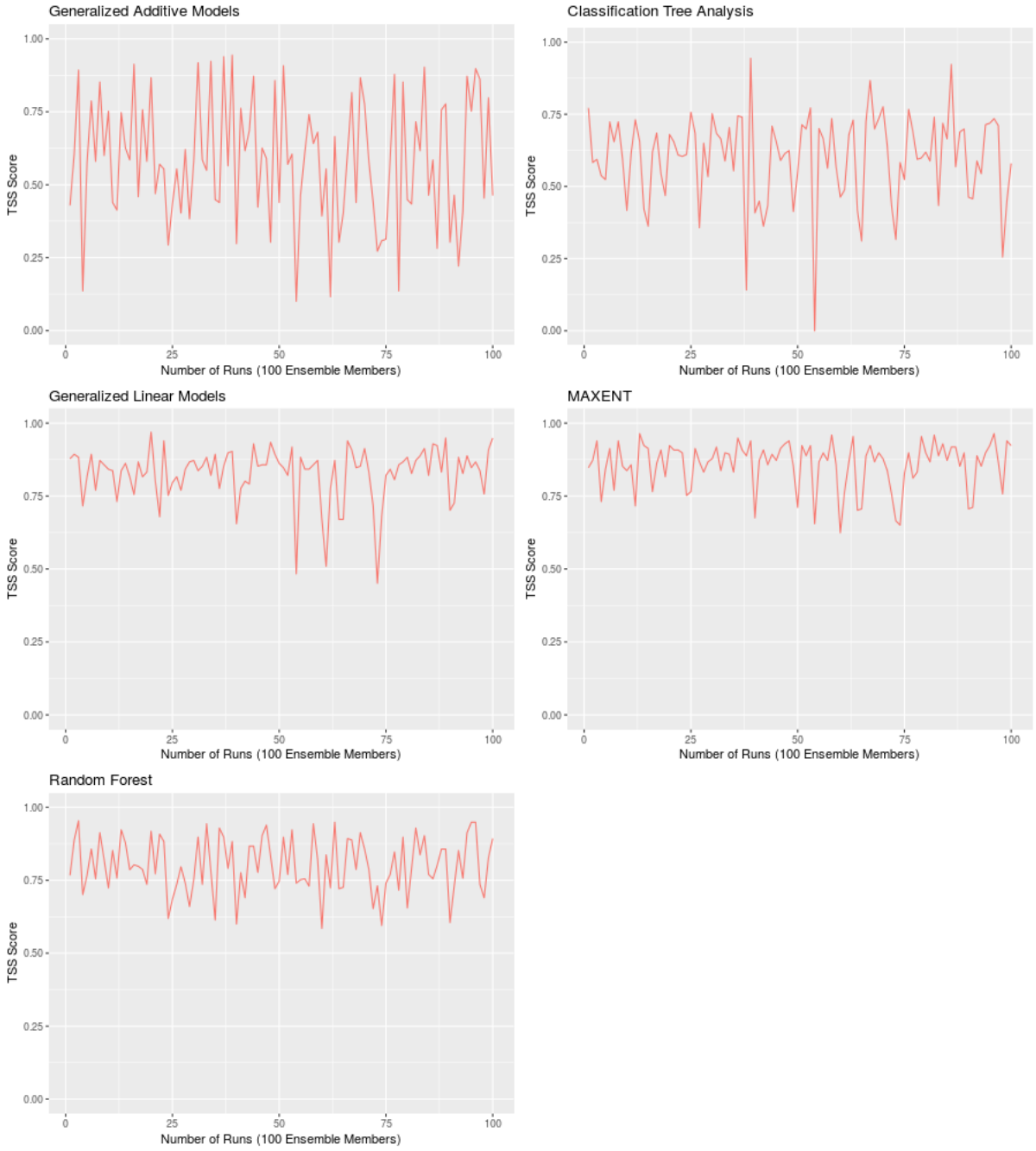


Figure 4.4: The TSS metric for all 5 statistical methods, for all 100 runs performed for each statistical method, run with bioclimatic variables computed with the WorldClim dataset.

Table 4.1: The mean values of the ROC and TSS metrics for all 100 runs performed with bioclimatic variables that were computed with the WorldClim dataset.

	GAM	GLM	RF	CTA	MAXENT
ROC	0.79	0.92	0.92	0.79	0.94
TSS	0.58	0.83	0.8	0.6	0.86

The results yielded from the aforementioned run in which the WorldClim data were employed, are displaying a substantial dependency on the presence records and they create model outputs that regard the departure from this assumption as erroneous and hence, classify those models as less successful. Although there is not a robust independent model output that it is known from the bibliography to be representative of the distribution of *An. sacharovi* over Greece, expert opinion of individuals being active in the field of vector control over Greece (Ecodevelopment S.A. company), state that the distribution of *An. sacharovi* is underrepresented from the map displayed in Figure 4.1. One of the purposes of the current work was to investigate what is the impact of the explanatory variables driving the statistical niche models. This investigation is presented in the next section.

4.1.2 Bioclimatic variables computed with the RCM ensemble mean dataset

The resulting environmental suitability map for *An. sacharovi*, exploiting the ensemble mean of RCM for the calculation of the bioclimatic variables is presented in Figure 4.5. The spatial resolution of the composed map is significantly reduced (equal to 12 km), but the produced result approximates in relatively better way the expected result (expected according to the expert knowledge provided by Ecodevelopment S.A.). The hot-spot of high environmental suitability of the region of North-Central Greece is maintained, but furthermore new areas are characterized as suitable, that they either had a much more constrained extent in the run using the WorldClim dataset or were completely obscured.

The additional areas that are characterized as suitable in the current run, are located over the plain of Thessaly in central Greece, over the coastal areas of northern Greece (Kavala and Alexandroupoli) and over the plain of Serres, also in northern Greece. Additionally, suitable areas are identified over coastal areas in southern mainland Greece, Attica, Peloponnese and over the eastern islands of the Aegean Sea. As it was described in the previous section, the final suitability map is composed using the result of 100 runs for every statistical method. Eventually though, only the "best" runs contribute to the construction of the ensemble (Figure 4.5). In the current run, the models composing the ensemble are presented in Figure 4.6.

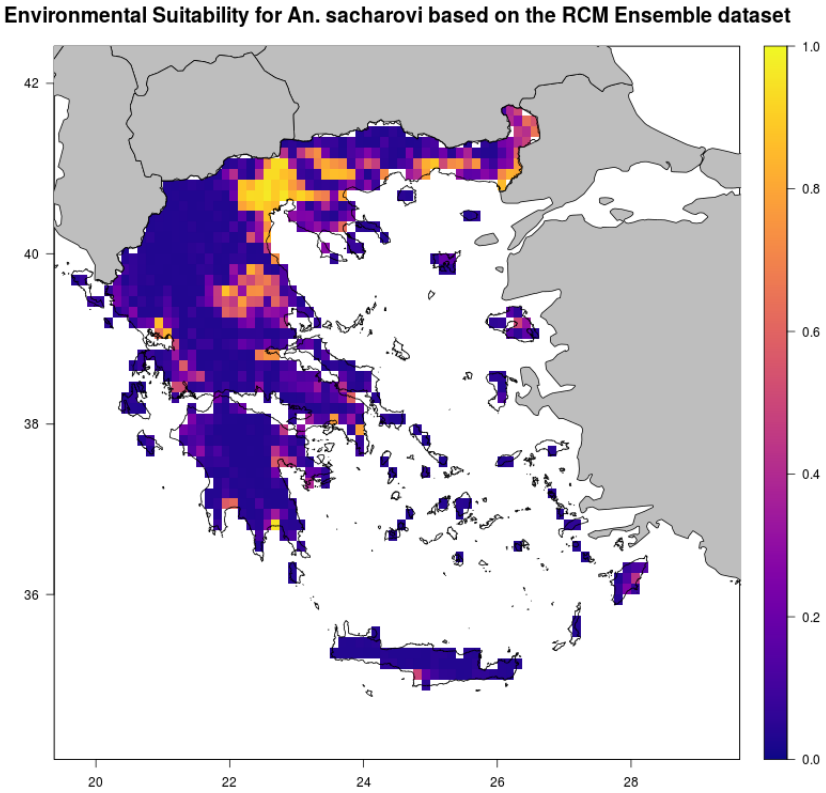


Figure 4.5: Environmental suitability of the ensemble niche modeling, where the bioclimatic variables were computed with the RCM ensemble mean dataset.

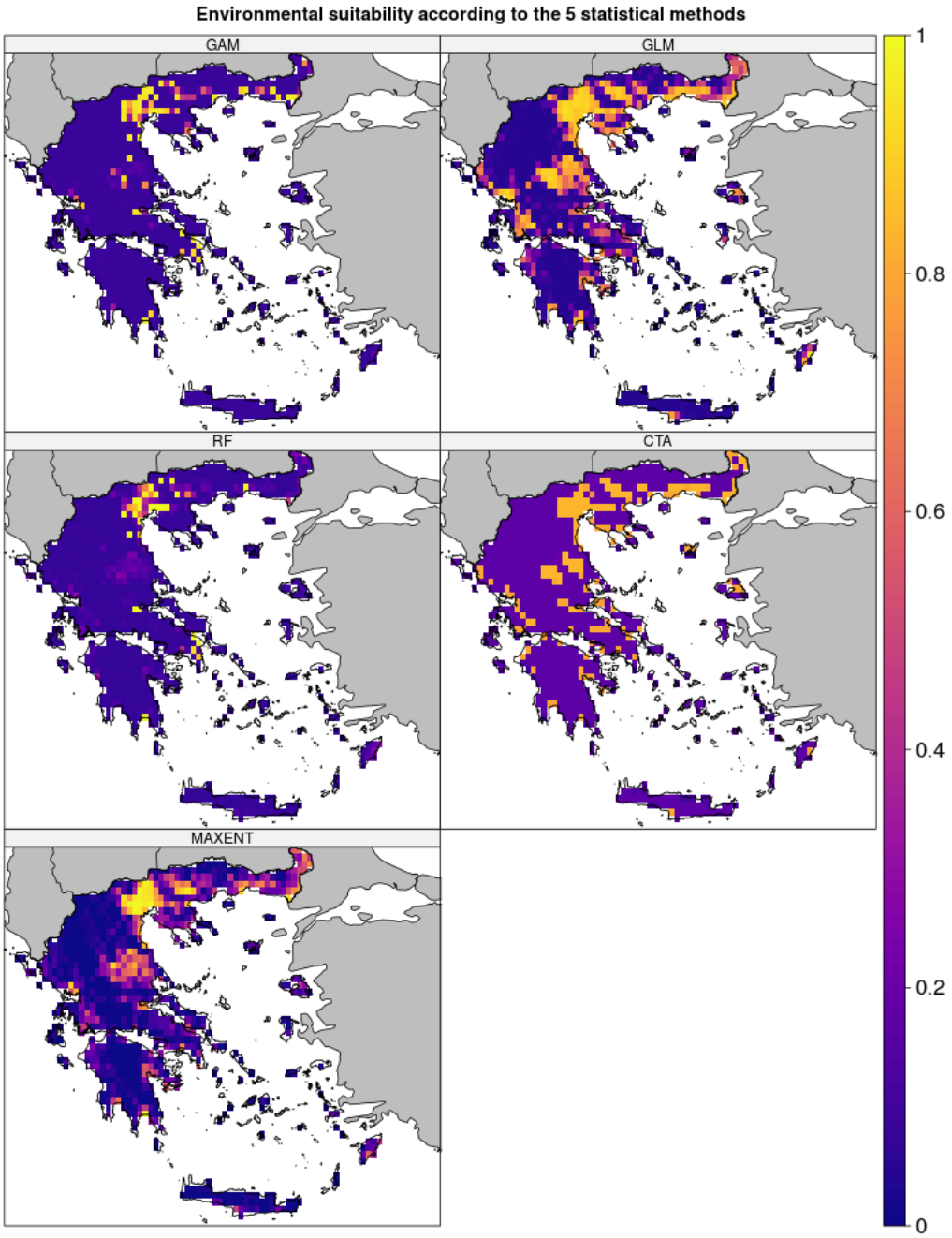


Figure 4.6: Environmental suitability according to the 5 statistical methods used, where the bioclimatic variables were computed with the RCM ensemble mean dataset.

As it is depicted, there seems to be a significant agreement between the predictions produced by GLM, CTA and MAXENT, although CTA is not capable of capturing the magnitude of suitability, but it rather captures its spatial pattern. In the current run, GAM and RF appear to be sensitive to the spatial structure of the training locations (traps) and they yield a rather localized result, due to overfitting. The fact that three models produce a more spatially extended result of environmental suitability (from the whole suit of five models) affects the calculation of the ensemble median, which subsequently affects the final composite suitability map. With regards to the reason why the run employing the ensemble mean of RCM produces a differentiated result to the run employing the WorldClim dataset, could possibly be attributed to the difference in the spatial resolution of the fields used as explanatory variables. In the case of employing the RCM ensemble mean, where the spatial resolution of the fields is approximately 12 km, it is possible that in areas with high trap density, more than one traps are located inside a pixel. This results in many traps being "filtered out" and hence, the mosquito trap network is, somehow, "de-clusterized".

Concerning the performance of the current run, there appears that the overall performance is relatively lower compared to the run using WorldClim. Based on Figure 4.7, the best performing models appear to be GLM and RF, followed by MAXENT, GAM and CTA. The lowest performance of GAM and CTA is slightly less than 0.5. With regards to the TSS metric presented in Figure 4.8, the performance is lower, compared to the ROC metric. GLM and RF yield as the best performing models, with their lowest values being greater than 0.25. The lowest performances of MAXENT, GAM and CTA approximates 0, with the difference that the bad performing runs in MAXENT have a reduced density. In the current application of the correlative niche models, the TSS metric displays higher variability across the 100 runs performed for each model.

Table 4.2: The mean values of the ROC and TSS metrics for all 100 runs performed with bioclimatic variables that were computed with the RCM ensemble mean dataset.

	GAM	GLM	RF	CTA	MAXENT
ROC	0.79	0.92	0.92	0.79	0.94
TSS	0.58	0.83	0.8	0.6	0.86

The statistical means of the ROC and TSS metrics for each statistical method separately for all 100 runs are presented in Table 4.2. The highest ROC metric is achieved by MAXENT and GLM and is equal to 0.85 followed by RF and GAM, yielding a ROC metric equal to 0.82 and 0.77 respectively. Lastly, CTA result to a ROC metric equal to 0.76. Concerning the TSS metric, GLM marks the highest TSS score equal to 0.68, MAXENT follows with a TSS equal to 0.67, followed by RF with a score equal to 0.61 and GAM and CTA with TSS metrics equal to 0.56 and 0.53 respectively.

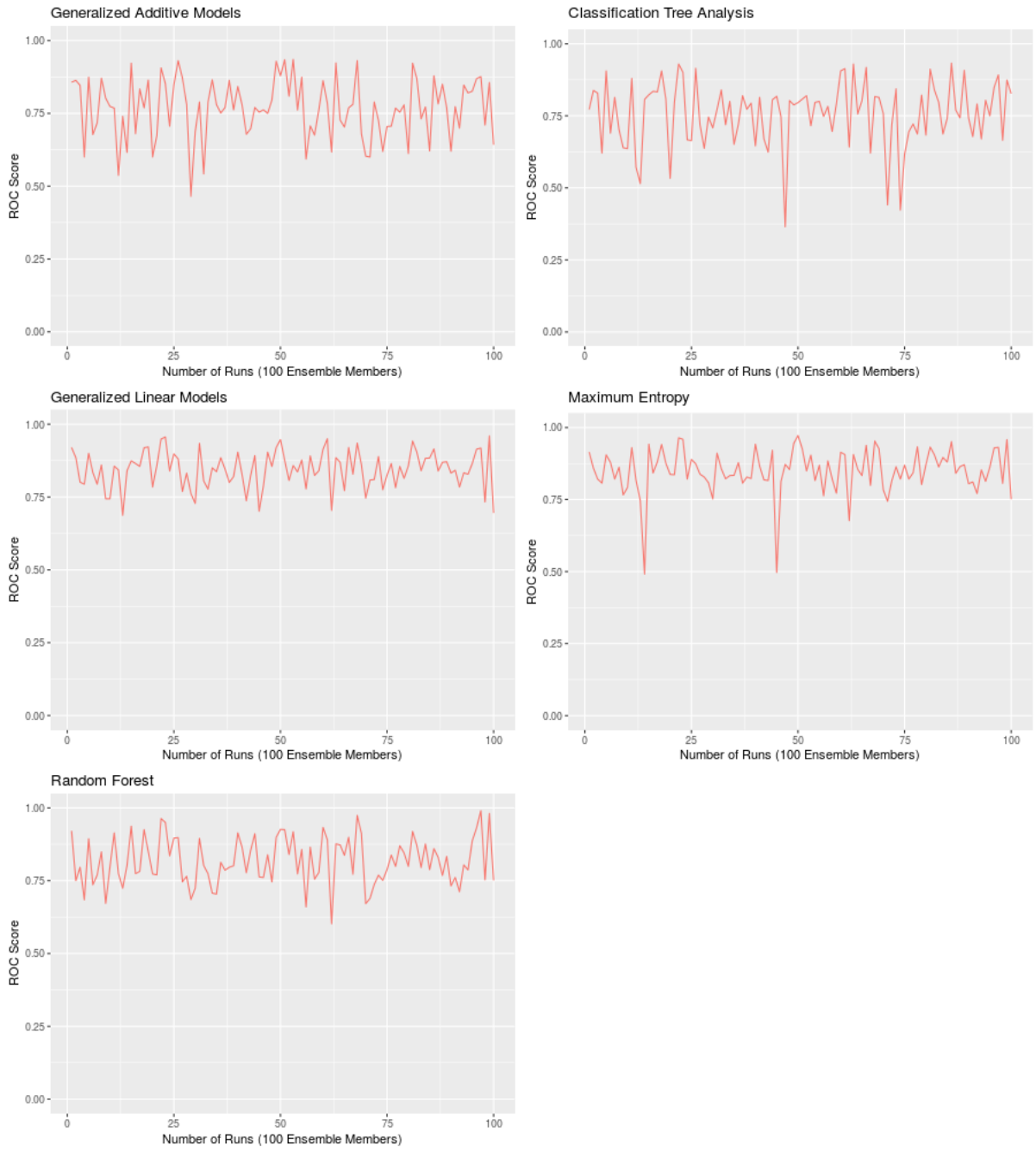


Figure 4.7: The ROC metric for all 5 statistical methods, for all 100 runs performed for each statistical method, run with bioclimatic variables computed with the RCM ensemble mean dataset.

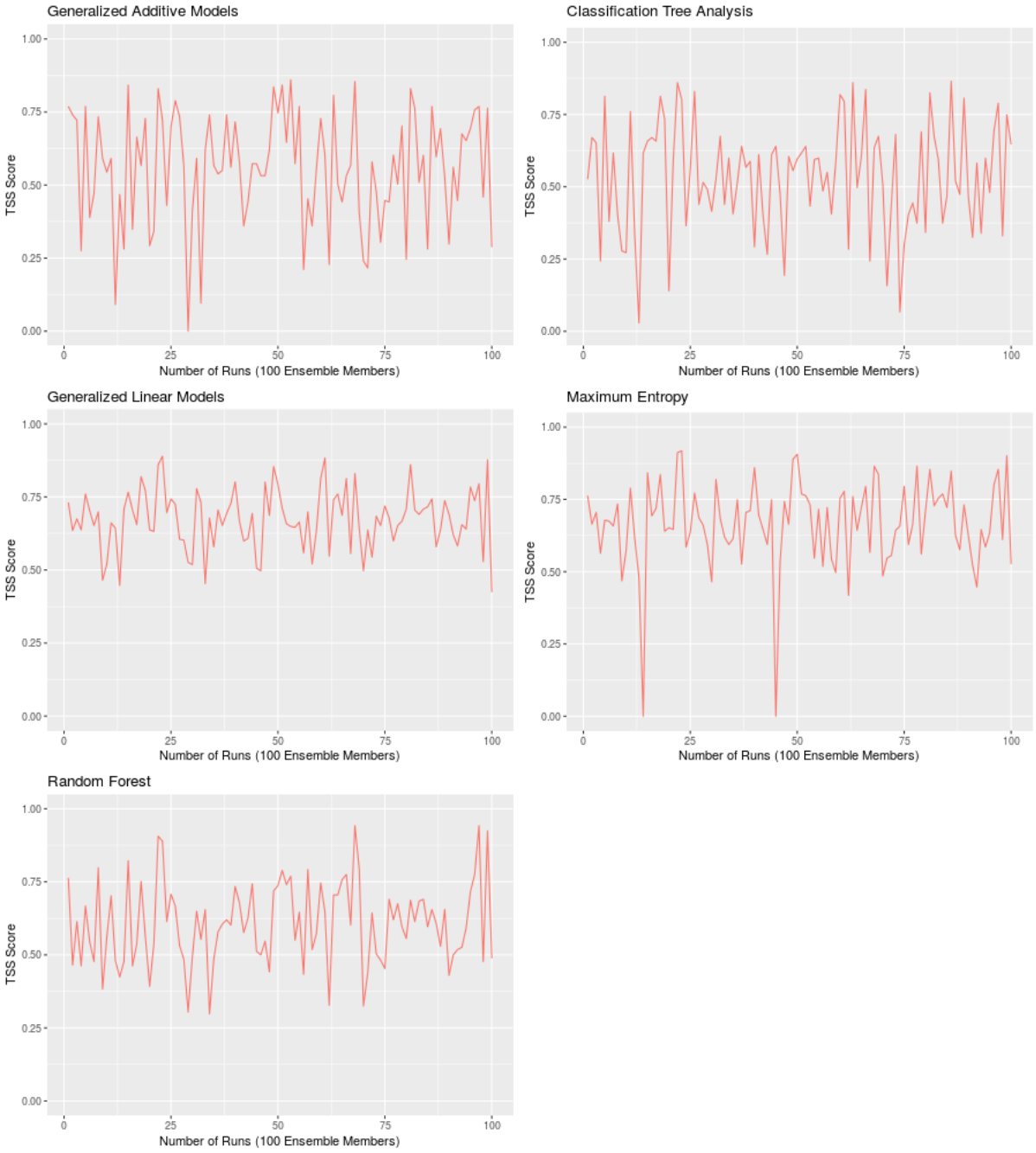


Figure 4.8: The TSS metric for all 5 statistical methods, for all 100 runs performed for each statistical method, run with bioclimatic variables computed with the RCM ensemble mean dataset.

4.2 Vector density output from VECTRI

4.2.1 Vector density before calibration

The VECTRI output of concern in the current analysis is vector density, expressed in m^{-2}. The "before calibration" runs refer to running VECTRI using its default parameters, describing the biology of *An. gambiae*. The first step towards calibrating VECTRI for *An. sacharovi* was to assess what does the model produce as output with its current setting, over the region in concern. The output of vector density with respect to its seasonal values is displayed in Figure 4.9. As it is shown, vector density is 0 during winter and spring and displays its highest values during summer, while during autumn vector density declines. During summer, vector density displays increased values throughout the whole domain of the study region, with an exception to the high elevation areas. Also, the highest values of vector density are observed over the plain of the Central Macedonia area in northern Greece, the plain of Thessaly and sporadic location over southern mainland Greece and over the Peloponesse. Also, high density values are observed in western Greece, a pattern that is substantially maintained also in autumn.

Analyzing the same output, but calculating its monthly values instead, yields the result presented in Figure 4.10. From what it is shown, vector's activity onset takes place during June, proceeds on to July, peaks during August and declines thereafter until November, when eventually it disappears. From what it is known through the available literature, vector activity is initiated early in spring . Also, the spatial extend of vector density is well expanded over areas with high elevation that it is known that provide an unwelcoming environment for mosquitoes.

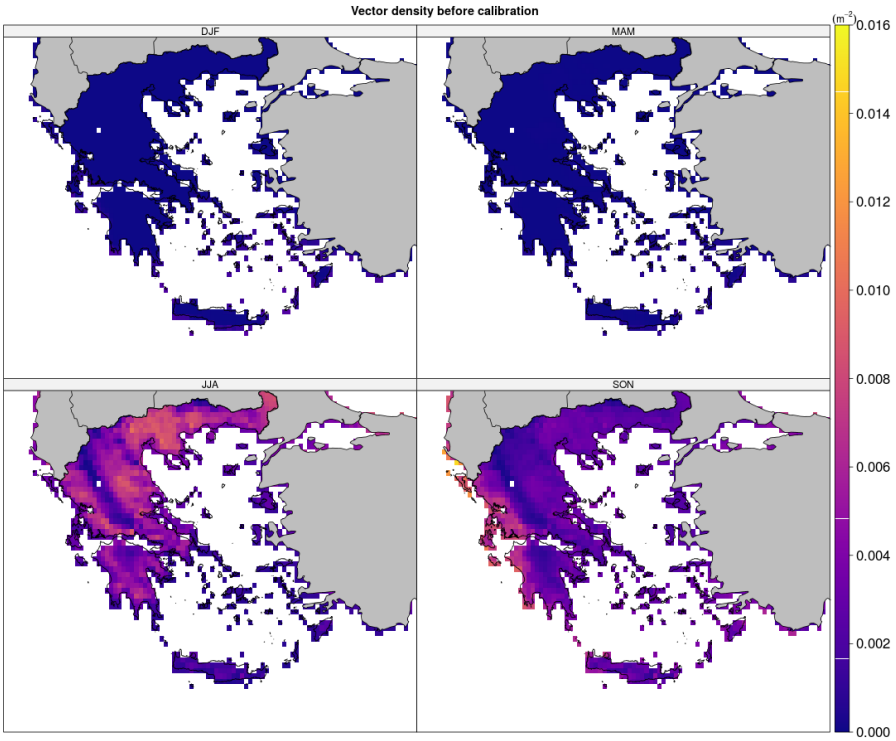


Figure 4.9: Vector density output from VECTRI before calibration. The maps correspond to the statistical means for the period 1991-2008.

The aforementioned observations are possibly due to the fact that the current temperature thresholds provide a wide temperature window for *Anopheles* mosquitoes and thus, provides the opportunity to them to be present over such extended regions with varying temperature regimes. This observation provides a hint to the successful calibration of VECTRI, so that it would restrict to a more confined range the temperature limits tolerated by *An. sacharovi*. This attempt is presented in the following section.

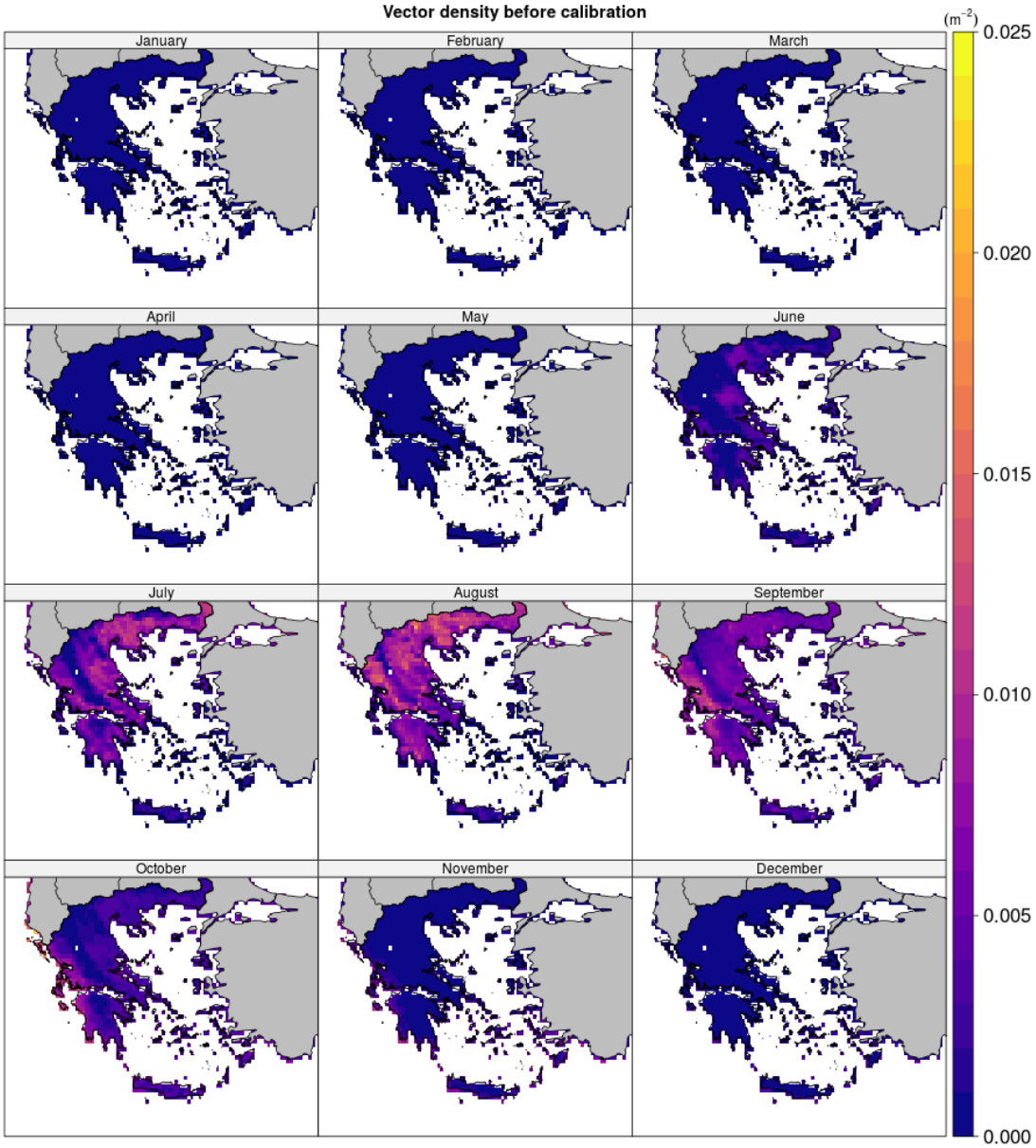


Figure 4.10: Vector density output from VECTRI before calibration. The maps correspond to the statistical means for the period 1991-2008.

4.2.2 Vector density after calibration

The result of modifying the parameters in VECTRI related to the gonotrophic cycle and the larvae characteristics of the vector are presented in Figure 4.11. The purpose of the current re-adjustment of the VECTRI parameters was to calibrate the model for *An. sacharovi*, instead of *An. gambiae* as it presently is. As it is shown, vector activity is 0 during winter and spring, but well-established during summer and autumn. In addition, the magnitude of vector density is severely affected and more specifically, it is severely reduced. Nevertheless, the spatial pattern of vector density acquired, impressively resembles the environmental suitability map for *An. sacharovi* that was discussed earlier in this Chapter. The difference between summer and autumn is minor, however, when looking at the monthly means of the same output in Figure 4.11, that is not the case.

As it was observed in the "pre-calibration" runs, the vector activity onset is missed, with the current run being even more delayed as vector activity is predicted to have its onset during July. The highest vector density is observed during August and September, a fact that it is in agreement with the literature. However, vector activity declines rapidly in October. Although there are certain questions that need to be addressed concerning the delayed vector onset and its decreased density values, it is encouraging that the spatial pattern of the current run comes in agreement with the suitability map for *An. sacharovi*.

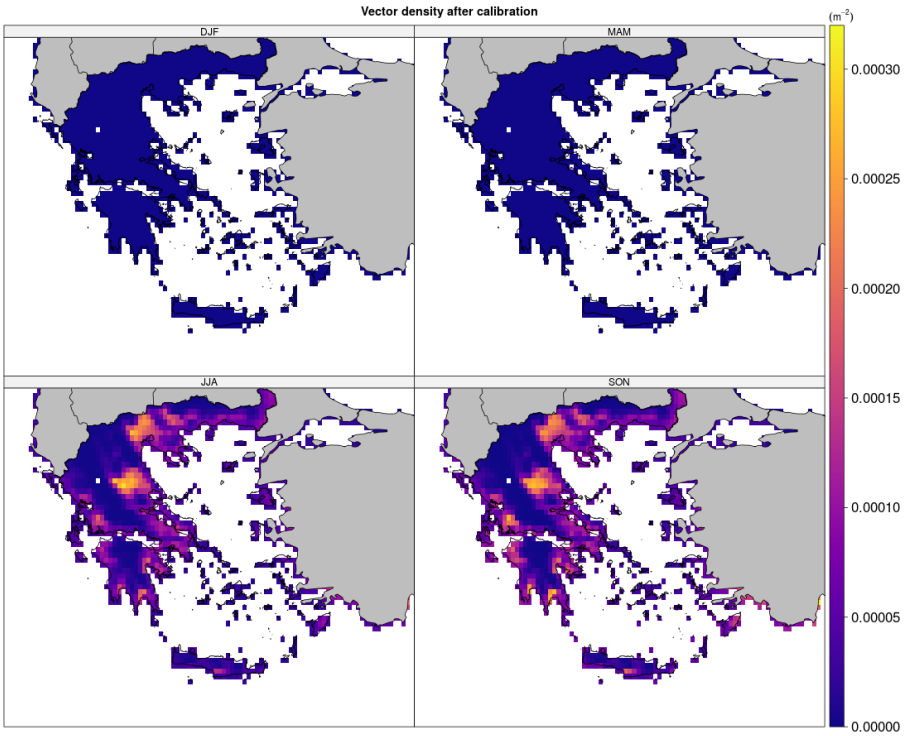


Figure 4.11: Vector density output from VECTRI after calibration. The maps correspond to the statistical means for the period 1991-2008.

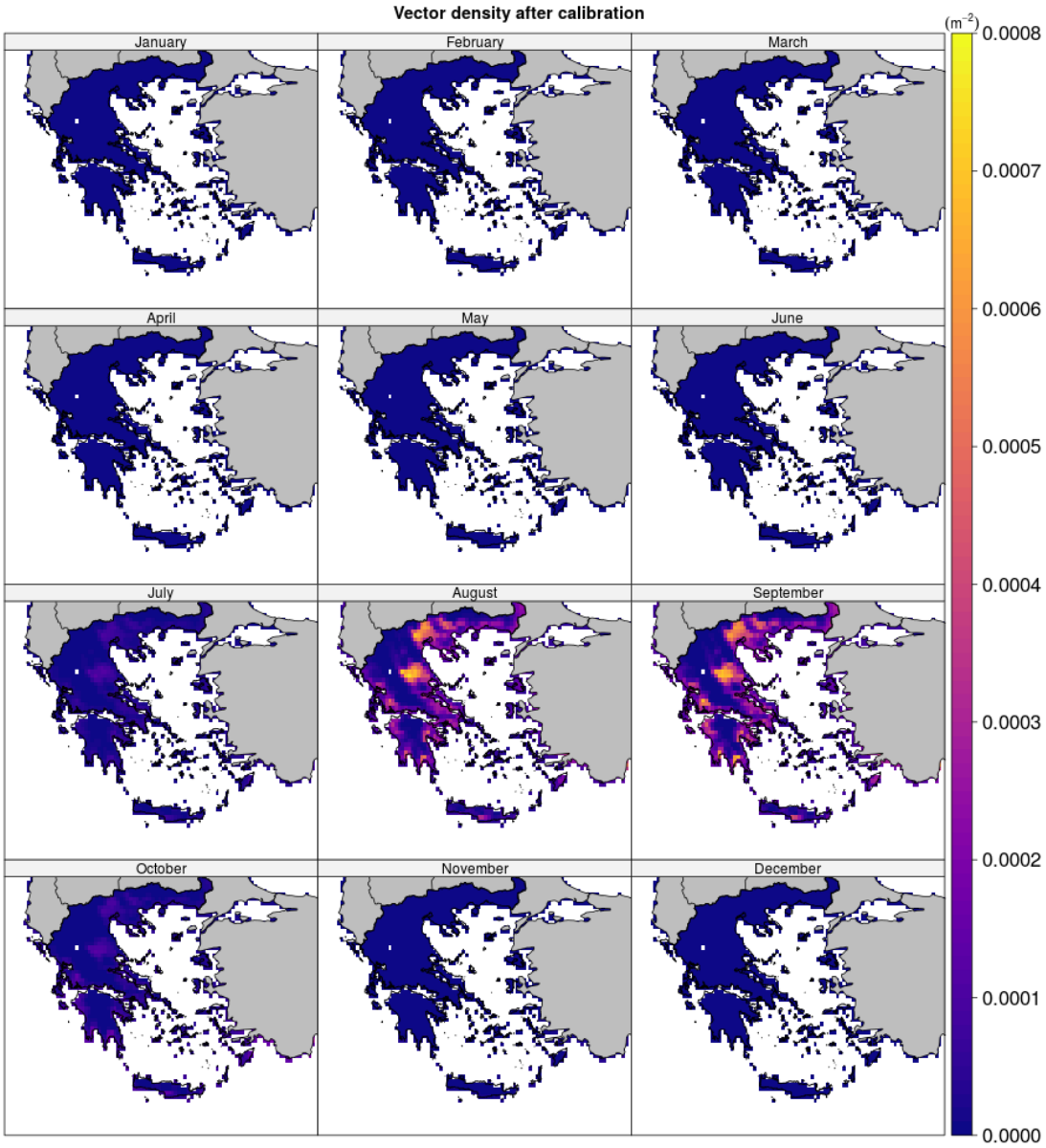


Figure 4.12: Vector density output from VECTRI after calibration. The maps correspond to the statistical means for the period 1991-2008.

4.2.3 Assessment of the performance of VECTRI calibration

Vector density time-series

Apart form the spatial characteristics of the VECTRI output, it is worth noting that vector density displays a well-established seasonality also. For seven locations (Figure 4.13) where vector activity was intense, the time-series of vector density were extracted and plotted over the time-period in concern. In Figures 4.14 and 4.15, both time-series as extracted from the pre-calibration and the post-calibration runs are plotted together for comparison. What is remarkable, is the difference in magnitude between the two runs. What additionally emerges as an interesting fact, is that although the temporal pattern is consistent, every year develops its own characteristics, a fact that is related to the climatic conditions prevailing during that particular year. For instance, years 1999, 2002 and 2005 display density values that exceed significantly the average values and also, after 2005 until 2008 a systematic decline is observed through all the locations located in mainland and northern Greece. The time-series in Geraki and Laerma do not strictly follow the aforementioned pattern and that is mainly due to the fact that the prevailing climatic conditions are much differentiated from the ones prevailing in mainland and northern Greece.

Concerning the time-series extracted from the the post-calibration run, these are also plotted separately in Figures 4.16 and 4.17. As it can be seen, vector density gradually increases from 1991 until 1995 and the following year it is very abruptly reduced. From 1995 until 2003, vector density is steadily increasing until it reaches its maximum during 2003 and the declines abruptly the following year. During years 2004, 2005 and 2006 vector density again displays progressive increase until lastly, it drops during 2008. The aforementioned pattern is observed for all locations with an exception to Laerma (in the island of Rhodes) that instead displays an almost static vector density with limited variability across the years 1991-1994, but displays an impressive increase during 1995, which the following year, however, returns to the values observed during 1991-1994. For the same location, values display a substantial increase throughout the period 2000-2008, with an exception to the year 2006.

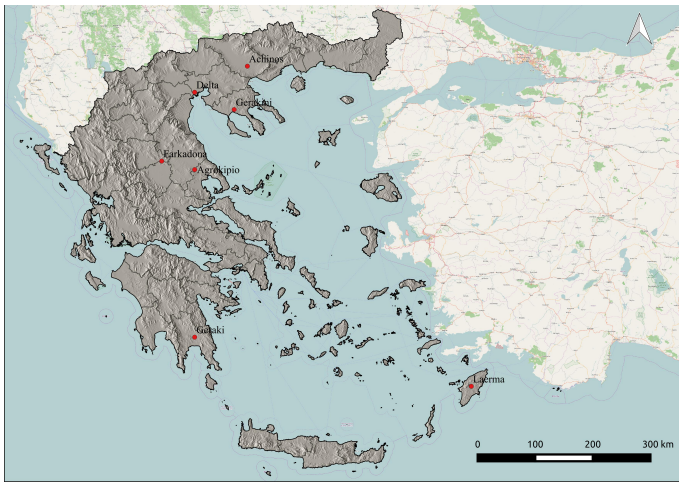


Figure 4.13: Random locations for which the following time-series were extracted.

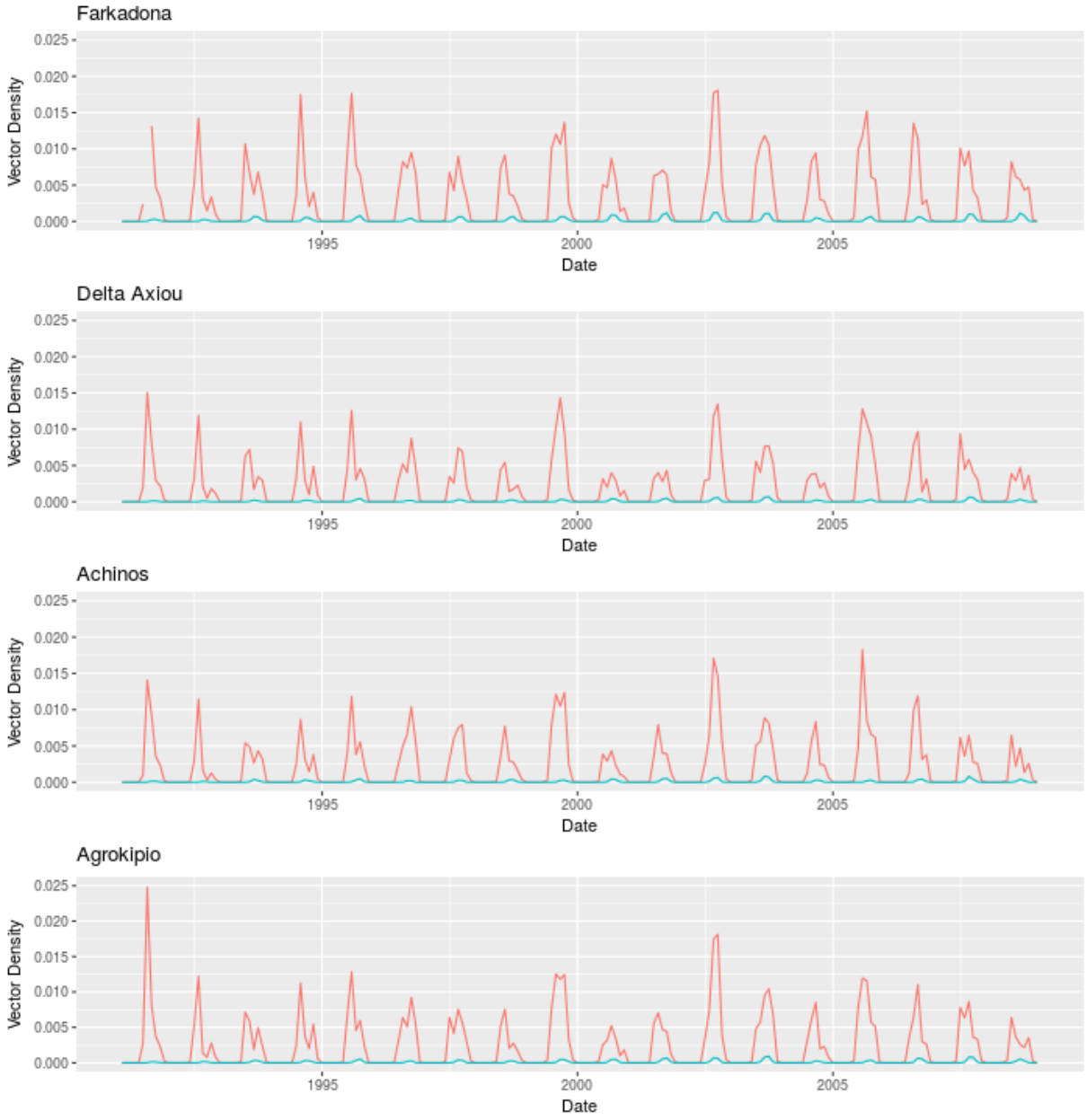


Figure 4.14: Time-series extracted from the VECTRI runs before calibration (in red) and after calibration (in blue).

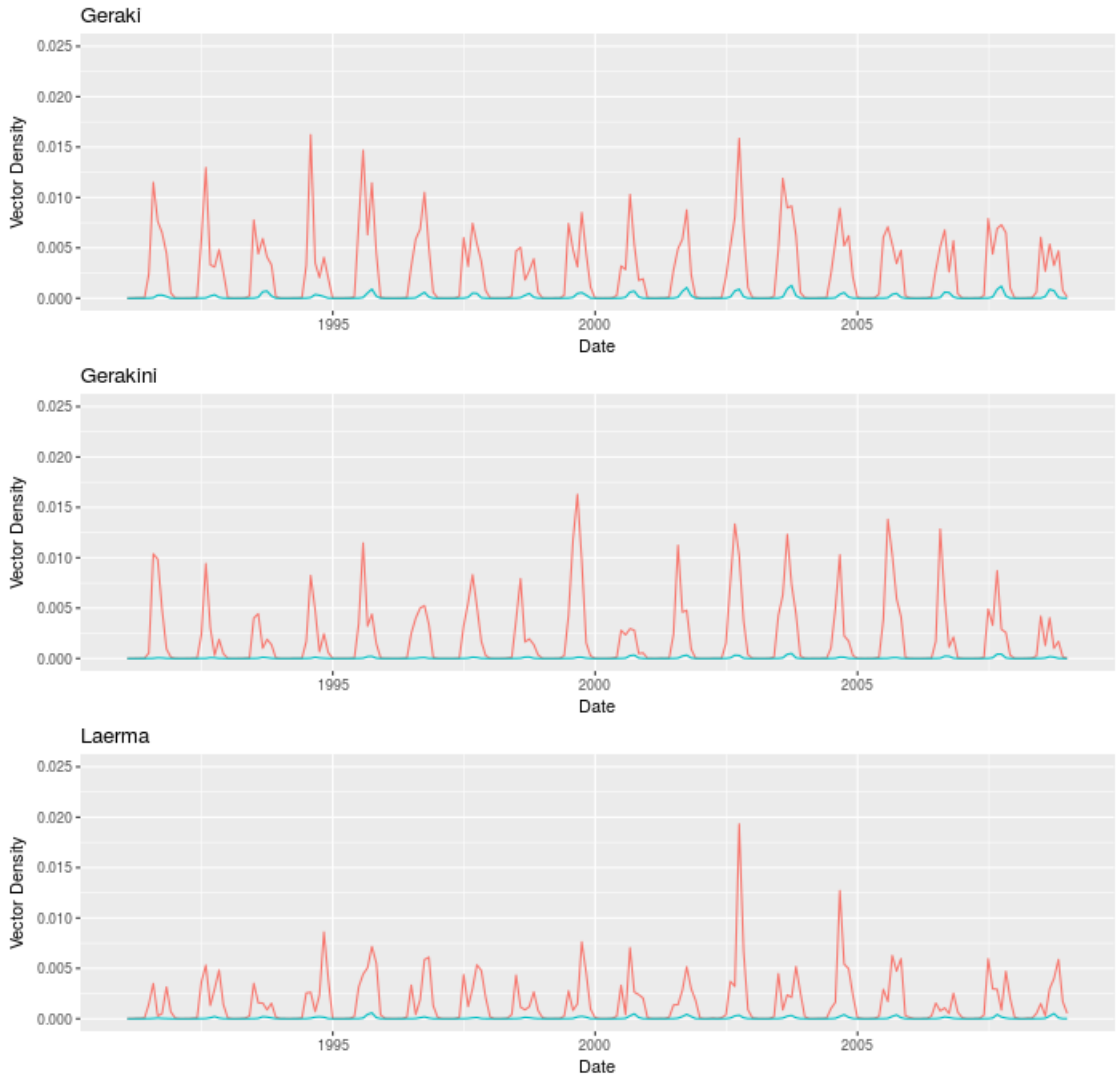


Figure 4.15: Time-series extracted from the VECTRI runs before calibration (in red) and after calibration (in blue).

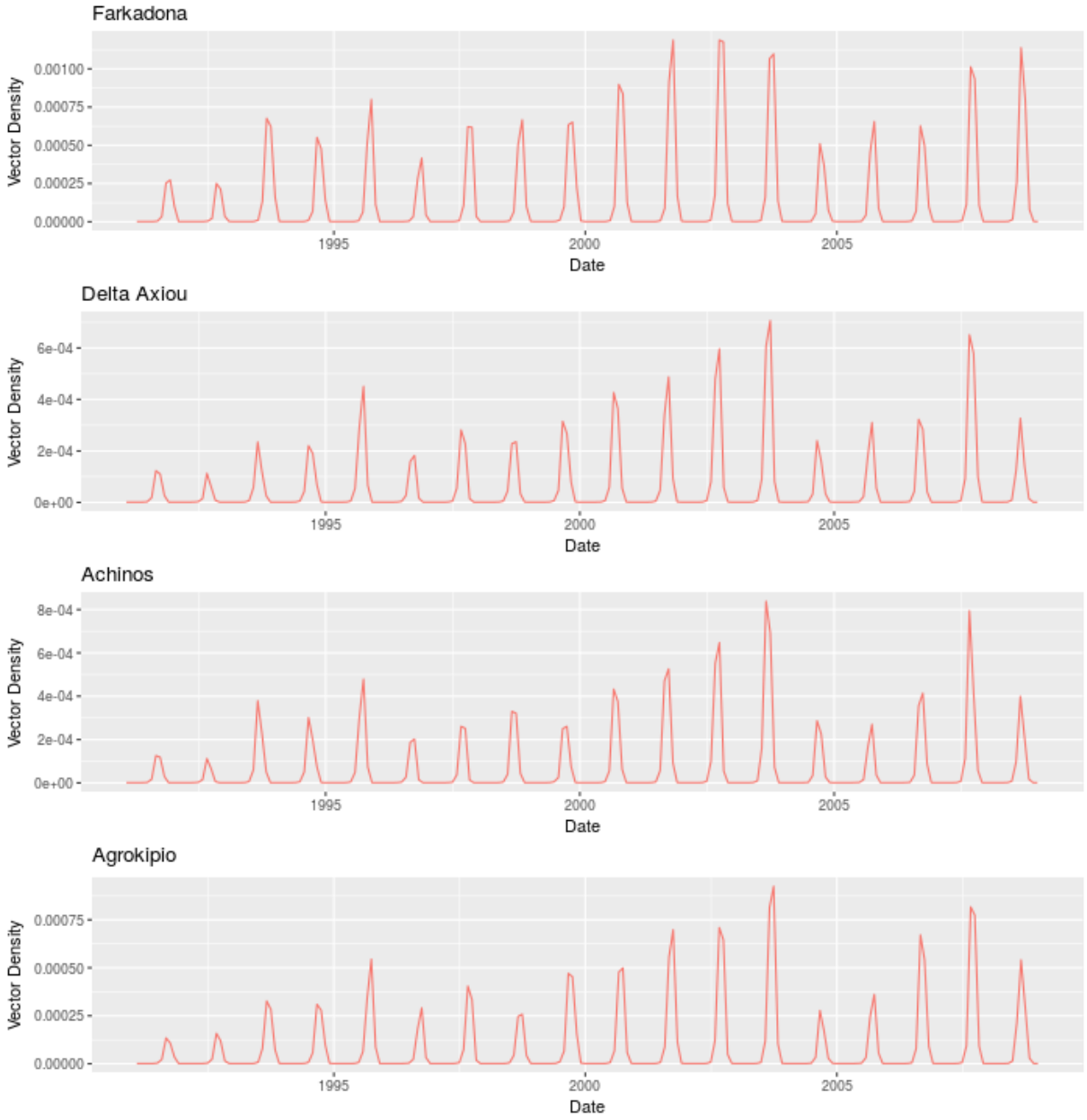


Figure 4.16: Time-series extracted from the VECTRI run after calibration.

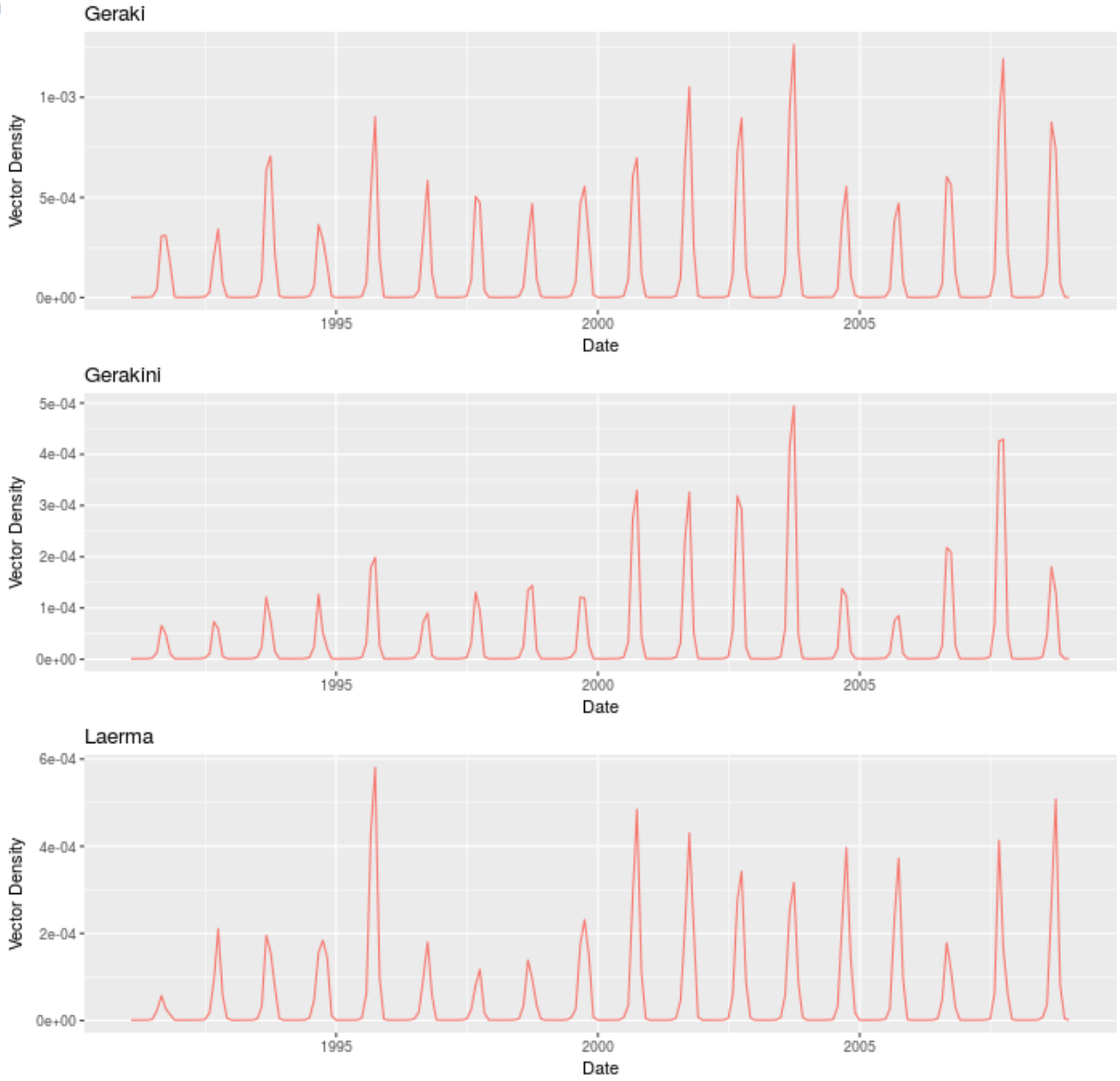
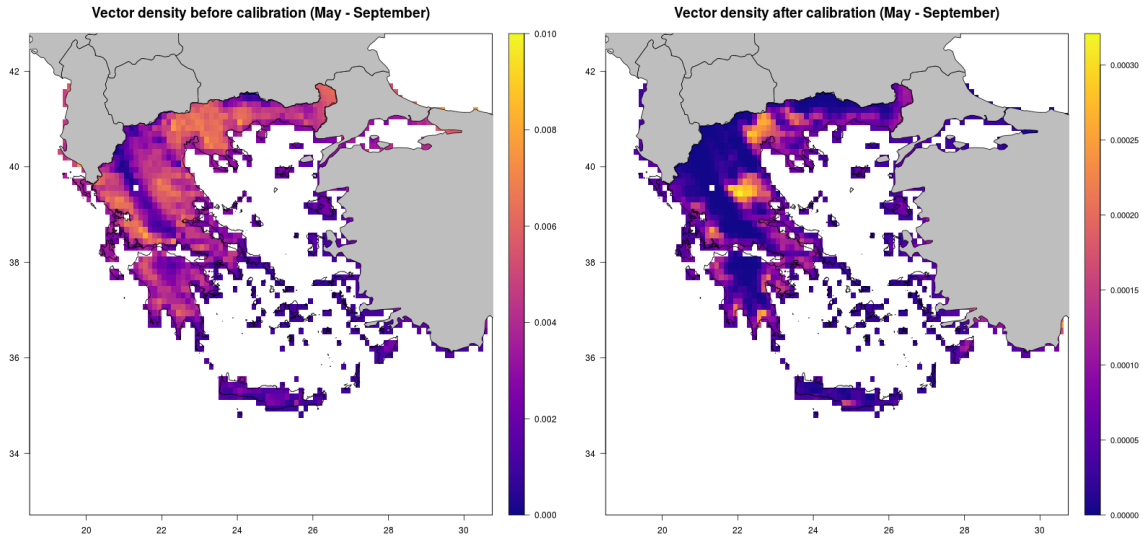


Figure 4.17: Time-series extracted from the VECTRI run after calibration.

Agreement with the suitability maps

The environmental suitability map that was discussed at the beginning of Chapter 4, was constructed using vector data ranging from May to September, so for this reason, the VECTRI outputs in the current step were averaged over the same period and compared to the map provided by the correlative ecological niche models. he vector density outputs from VECTRI before and after calibration averaged over months May through to September are presented below in Figures 4.18. The scatterplots of the two VECTRI runs to environmental suitability are presented in Figures 4.19 and 4.20. As it can be seen, the r^2 value of the linear model applied is equal to 0.202 before the calibration and increases to 0.561 after the calibration. This increase is indicative of the growth in resemblance between the VECTRI post-calibration run and the environmental suitability map. Of course, this does not state that the post-calibration run is successful and free of errors, but it rather states that calibrating VECTRI can yield a useful resource and tool.



(a) pre-calibration (b) post-calibration

Figure 4.18: Pre-calibration and post-calibration vector density.

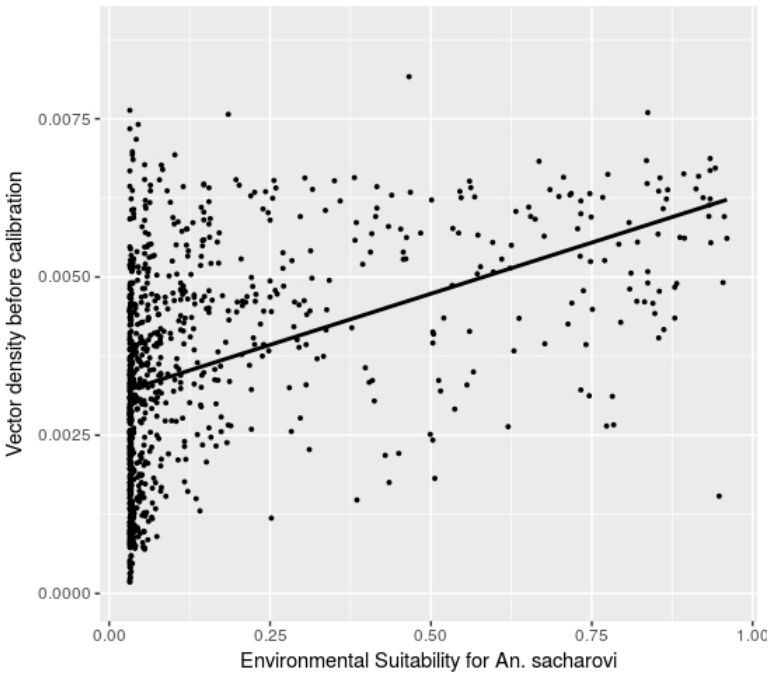


Figure 4.19: Scatterplot between vector density before calibration and environmental suitability for *An. sacharovi*. The r^2 of the linear model applied is 0.202. The equation that describes the linear model is the following: y=0.0031 + 0.032x.

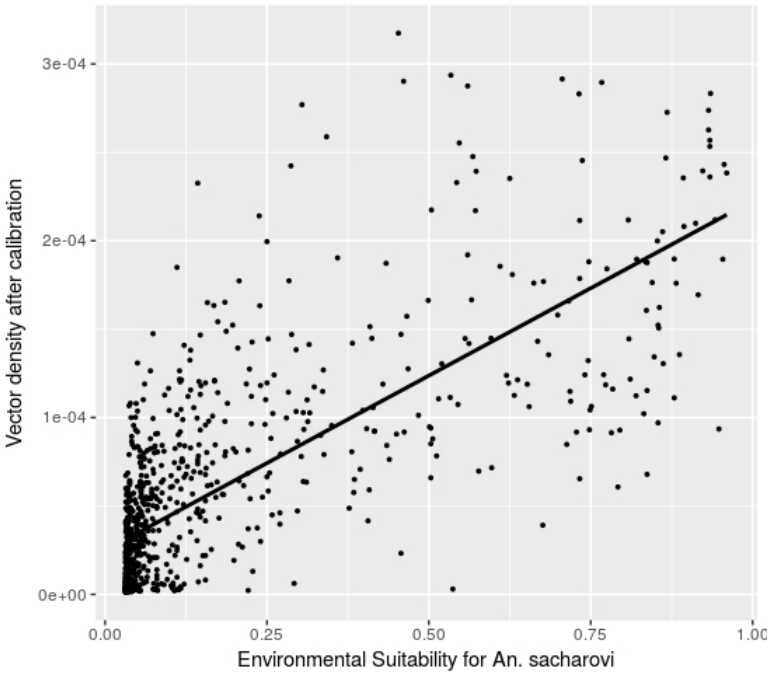


Figure 4.20: Scatterplot between vector density before calibration and environmental suitability for *An. sacharovi*. The r^2 of the linear model applied is 0.561. The equation that describes the linear model is the following: y=(2.5e-05) + (2e-0.4)x.

Chapter 5

Discussion

One of the objectives of the current analysis was to identify the ecological niche of *An. sacharovi* and thus further identify regions were there is a potential risk of malaria occurrence. For this reason, both correlative and dynamic models were employed, as both display different kinds of strengths and weaknesses. Through this practice, both models can contribute with a unique way in the modeling of the ecological niche of *An. sacharovi*.

Concerning the application of correlative models, a suite of Presence and Absence data was used, along with a suit of environmental variables. A common source of uncertainty in bibliography with regards to the accuracy of correlative ecological niche models, results from the fact that the covariates that are used as predictors contain errors and uncertainties themselves, that are consequently transferred into the niche models. A possible answer to such an issue is the use of an ensemble of environmental predictors, so that the final product of the modeling process could be reduced. In the current analysis, two runs of correlative ecological niche models were performed, one using the WorldClim dataset and one using the ensemble mean of seven RCM models ran in the context of the EURO-CORDEX domain. The results of the first run yielded to better error metrics, but the spatial patterns of the second run produced a suitability map that approximates better the expected reality. In this point, the expected reality was provided in a qualitative manner by Ecodevelopment S.A.

5.1 Occurrence records

Through the application of correlative ecological niche models it becomes evident that the predictions produced display high dependency on the spatial structure of the input data. This fact highlights the need for a thorough study concerning the optimal location of the initial network. The location of mosquito traps over a region is a form of spatial sampling and thus, there must be a very specific strategy regarding how this spatial sampling is going to be performed. A guideline in such issues is the identification of the research question (or questions) that need to be addressed. The spatial sampling needs to be performed in such a way so that it can effectively address the research question and can meet the requirements of the analysis that is to be applied. It can not

be stressed enough that not all spatial arrangements of spatial samples are suitable for all kinds of analysis. The correspondence of the sampling strategy and the analysis that is to follow are highly linked to the success and accuracy of the analysis.

In the current case of building a correlative model in order to assess the ecological niche of a species, two significant distinctions need to be made that are related to the nature of the species in concern. Normally, occurrence data are collected by geolocating the geographic points of where the species (or their nests) were found. However, in the current case, the presences were not extracted from the actual coordinates of where mosquitoes were found, but were rather inferred from the coordinates of the traps that captured the mosquitoes. As a result, the presence points used in the current work can be potentially considered as an underestimation of the actual presences of the species in concern over the study region. Additionally, as it was discussed in Chapter 2 the traps measured abundances of mosquitoes and therefore, a transformation from abundances to presences/absences was necessary. This transformation was an additional source of uncertainty to the results of the current analysis. A possible way of addressing how severe is the uncertainty introduced due to the aforementioned issue, would be the application of a sensitivity analysis, where various methodologies of transforming abundances to presences/absences is applied. Furthermore, the measurement of the abundances of mosquitoes in traps can be proven to be a challenge in the case that the niches of more than one mosquito species is studied. In that case, the transformation of the abundances to presence/absence records should be performed in such a way, so that the differences in occurrences between the miscellaneous species are emphasized.

5.2 Sensitivity to the spatial resolution of the covariates

As it became evident from the current analysis, the spatial resolution of the covariate fields that were used, generated substantial differences in the predicted environmental suitability of *An. sacharovi*. As it was discussed, this can be primarily attributed to the spatial arrangement and density of the mosquito traps, but it would be also important for the impact of the spatial scale of the covariate fields per se to be assessed. This could be attained by exploiting a high resolution climatic dataset and gradually upscaling it to lower spatial resolutions (bigger grid cells) and each time quantifying the changes in the resulting ecological niche. It is expected that there is a relation between the coverage of the area in concern, the proximity metrics of the occurrence network and the spatial resolution of the covariate fields. The challenging need would be to identify the mathematical form of such a relation, so that future studies of ecological niche modeling could be performed using covariates bearing the most suitable spatial resolution.

5.3 Fine tuning in parameters of correlative models

It is very common in the relevant bibliography that researchers applying various statistical models in ensemble niche applications, use the default set parameters of the respective softwares

or programming packages. Although this tactic secures the fact that the model outputs are comparable among themselves, with respect to the mechanics and parameters of each model, it does not secure that each model is used in its full capacity. The use of the default parameters of the "biomod 2" package in R were also used in the current analysis, but by no means this is considered a scientifically robust practice. The suitability maps produced are inseperably joint to the model parameters that generated the model outputs and thus the validity and accuracy of the final product is dependent on the model parameters. Another issue that needs to be highlighted is that the various statistical models are not disconnected among themselves. For instance, under certain parameterization a GAM can behave as a GLM (so it would no longer be a GAM) or MAXENT can behave as a logistic regression. Thus, the fine tuning needs to be performed under caution, so that the set of parameters that differentiate the miscellaneous models are not obscured, but rather strengthened.

5.4 Vector "reanalysis"

As it is extremely common in atmospheric sciences, the accuracy of a model output is usually assessed by comparing the desired surface to a dataset that it is relatively considered as the closest representation of reality. This practice proves to be very useful and provides a point of reference between atmospheric scientists. The products that are considered to be the most accurate and up-to-date are the climatic reanalysis products that combine both model output data, observational data and satellite products.

Likewise, it would be useful if an equivalent vector "reanalysis" product could be generated, so that there is a point of reference when assessing the accuracy of the predicted ecological niche of a vector species. One way that the vector "reanalysis" could be obtained is by spatially interpolating the vector abundances measured in every trap, employing multivariate geostatistical methods (regression kriging). The result of such a procedure would be to have an estimate of the number of mosquitoes for every grid cell of the study region.

5.5 Concerning future work

As a step forward, it would be interesting to find out how are the temporal patterns of environmental suitability going to change, under certain Representative Concentration Pathways in the future and identify locations that used to bare suitable conditions and are estimated to no longer do so in the future and vice versa. Concerning the calibration of VECTRI, more test run need to be performed and a robust sensitivity analysis needs to be carried out, concerning the impact of all the in-lying parameters on which the biology of *An. sacharovi* is dependent on. Furthermore, the calibration of VECTRI could be considered as an optimization problem, so its approximation and solution could be obtained through the implementation of an optimization algorithm, that would provide an acceptable value for each VECTRI parameter, based on a certain set of criteria.

Such an optimization algorithm is the Genetic Algorithms that mimic the mechanism of natural selection. Lastly, *An. sacharovi* is not the only effective malaria vector over the study region, so it would be a much desired goal to extend the current work to other Anopheles species, such as *An. hyrcanus* and *An. Pseudopictus*.

Bibliography

Abebe, A., Gemeda, A., Wondewossen, T., Lemu, G., 2011. Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia. Parasit. Vectors 4.

Alemu, A., Abebe, G., Tsegaye, W., Golassa, L., 2011. Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia. Parasit. Vectors 4, 30.

Alimi, T.O., Fuller, D.O., Qualls, W.A., Herrera, S.V., Arevalo-Herrera, M., Quinones, M.L., Lacerda, M.V.G., Beier, J.C., 2015. Predicting potential ranges of primary malaria vectors and malaria in northern South America based on projected changes in climate, land cover and human population. Parasit. Vectors 8, 431.

Allan, R.P., 2017. Global energy budget: Elusive origin of warming slowdown. Nat. Clim. Change 7, 316317.

Arifin, S.N., Madey, G.R., Collins, F.H., 2013. Examining the impact of larval source management and insecticide-treated nets using a spatial agent-based model of Anopheles gambiae and a landscape generator tool. Malar. J. 12, 290.

Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.-P., Awono-Ambene, P., Fontenille, D., Simard, F., 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar. J. 8, 307.

Asare, E.O., 2016. Development and evaluation of temperature and surface hydrology schemes for dynamical vector-borne disease models. Kwame Nkrumah University of Science and Technology.

Baeza, A., Santos-Vega, M., Dobson, A.P., Pascual, M., 2017. The rise and fall of malaria under land-use change in frontier regions 1, 0108.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., Nelson, A., 2004. The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents.

Bomblies, A., Duchemin, J.-B., and Eltahir, E. (2009). A mechanistic approach for accurate simulation of village scale malaria transmission. Malaria Journal, 8(1):223

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A., Dawson, T.P., Emerson, B.C., Ferrier, S., Gilbert, M.T.P., Herold, M., Jones, L., Leendertz, F.H., Matthews, L., Millington, J.D.A., Olson, J.R., Ovaskainen, O., Raffaelli, D., Reeve, R., Rdel, M.-O., Rodgers, T.W., Snape, S., Visseren-Hamakers, I., Vogler, A.P., White, P.C.L., Wooster, M.J., Yu, D.W., 2017. Connecting Earth observation to high-throughput biodiversity data. Nat. Ecol. Evol. 1.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Coln-Gonzlez, F.J., Stenlund, H., Martens, P., Lloyd, S.J., 2014. Impact of climate change on global malaria distribution. Proc. Natl. Acad. Sci. 111, 32863291.

Chitnis, N., Hyman, J.M., Cushing, J.M., 2008. Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. Bull. Math. Biol. 70, 12721296.

Chitnis, N., 2017. Introduction to SEIR Models lecture given during the Workshop on Mathematical Models of Climate Variability, Environmental Change and Infectious Diseases, Trieste, Italy, 8 May 2017.

Cogswell, F.B., 1992. The hypnozoite and relapse in primate malaria. Clin. Microbiol. Rev. 5, 2635.

Conley, A.K., Fuller, D.O., Haddad, N., Hassan, A.N., Gad, A.M., Beier, J.C., 2014. Modeling the distribution of the West Nile and Rift Valley Fever vector Culex pipiens in arid and semi-arid regions of the Middle East and North Africa. Parasit. Vectors 7, 289.

CORINE Land Cover Copernicus Land Monitoring Service [WWW Document], URL http://land.copernicus.eu/pan-european/corine-land-cover (accessed 7.16.17).

Cunze, S., Koch, L.K., Kochmann, J., Klimpel, S., 2016. Aedes albopictus and Aedes japonicus - two invasive mosquito species with different temperature niches in Europe. Parasit. Vectors 9.

Danis, K., Lenglet, A., Tseroni, M., Baka, A., Tsiodras, S., Bonovas, S., 2013. Malaria in Greece: historical and current reflections on a re-emerging vector borne disease. Travel Med. Infect. Dis. 11, 814.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hlm, E. V., Isaksen, L., Kllberg, P., Khler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thpaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc. 137, 553597.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Rev. Geophys. 45, RG2004.

Fick, S.E., Hijmans, R.J., n.d. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol.

Giorgi, F., Jones, C., Asrar, G.R., 2009. Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin 58, 175-183.

Giorgi, F., Gutowski, W.J.Jr., 2015. Regional Dynamical Downscaling and the CORDEX Initiative. Annu. Rev. Environ. Resour. 40, 467490.

Gallego, F.J., 2010. A population density grid of the European Union. Popul. Environ. 31, 460473.

Gething, P.W., Smith, D.L., Patil, A.P., Tatem, A.J., Snow, R.W., Hay, S.I., 2010. Climate change and the global malaria recession. Nature 465, 342345.

Gething, P.W., Van Boeckel, T.P., Smith, D.L., Guerra, C.A., Patil, A.P., Snow, R.W., Hay, S.I., 2011. Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax. Parasit. Vectors 4, 92.

HCDCP [WWW Document], URL http://www.keelpno.gr/en-us/home.aspx (accessed 7.16.17).

Hengl, T., Sierdsema, H., Radovi, A., Dilo, A., 2009. Spatial prediction of species distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecol. Model., Selected Papers on Spatially Explicit Landscape Modelling: Current practices and challenges 220, 34993511.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 19651978.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob. Change Biol. 12, 22722281.

Horton, R., Lo, S., 2015. Planetary health: a new science for exceptional action. The Lancet 386, 19211922.

How Aspect worksHelp — ArcGIS for Desktop [WWW Document], URL http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-aspect-works.htm (accessed 7.16.17).

How Slope worksHelp — ArcGIS for Desktop [WWW Document], n.d. URL http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-slope-works.htm (accessed 7.16.17).

Huang, F., Zhou, S., Zhang, S., Wang, H., Tang, L., 2011. Temporal correlation analysis between malaria and meteorological factors in Motuo County, Tibet. Malar. J. 10, 54.

Hutchinson, G.E., 1957. Concluding remarks. In: Cold Spring Harbour Symposium on Quantitative Biology, vol. 22 , pp. 415427.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Katragkou, E., Garca-Dez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R.M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P.M.M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., Jacob, D., 2015. Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. Geosci Model Dev 8, 603618.

Koch, L.K., Cunze, S., Werblow, A., Kochmann, J., Drge, D.D., Mehlhorn, H., Klimpel, S., 2016. Modeling the habitat suitability for the arbovirus vector Aedes albopictus (Diptera: Culicidae) in Germany. Parasitol. Res. 115, 957964.

Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Dqu, M., Gobiet, A., Goergen, K., Jacob, D., Lthi, D., van Meijgaard, E., Nikulin, G., Schr, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., 2014. Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7, 12971333.

Kousoulis, A.A., Chatzigeorgiou, K.-S., Danis, K., Tsoucalas, G., Vakalis, N., Bonovas, S., Tsiodras, S., 2013. Malaria in Laconia, Greece, then and now: a 2500-year-old pattern. Int. J. Infect. Dis. 17, 811.

Krefis, A.C., Schwarz, N.G., Krger, A., Fobil, J., Nkrumah, B., Acquah, S., Loag, W., Sarpong, N., Adu-Sarkodie, Y., Ranft, U., May, J., 2011. Modeling the relationship between precipitation and malaria incidence in children from a holoendemic area in Ghana. Am. J. Trop. Med. Hyg. 84, 285291.

Kulkarni, M.A., Desrochers, R.E., Kerr, J.T., 2010. High Resolution Niche Models of Malaria Vectors in Northern Tanzania: A New Capacity to Predict Malaria Risk? PLOS ONE 5, e9396.

Laprise, R., de Ela, R., Caya, D., Biner, S., Lucas-Picher, P., Diaconescu, E., Leduc, M., Alexandru, A., Separovic, L., 2008. Challenging some tenets of Regional Climate Modelling, Met. and Atm. Physics 100, 3-22.

Laprise, R., Kornic, D., Rapai, M., eparovi, L., Leduc, M., Nikiema, O., Luca, A.D., Diaconescu, E., Alexandru, A., Lucas-Picher, P., de Ela, R., Caya, D., Biner, S., 2012 Considerations of domain size and large-scale driving for nested regional climate models: Impact on internal variability and ability at developing small-scale details. Climate Change: Inferences from Paleoclimate and Regional Aspects, 181-199.

Leedale, J., Tompkins, A.M., Caminade, C., Jones, A.E., Nikulin, G., Morse, A.P., 2016. Projecting malaria hazard from climate change in eastern Africa using large ensembles to estimate uncertainty. Geospatial Health 11, 393.

Lyons, C.L., Coetzee, M., Terblanche, J.S., Chown, S.L., 2012. Thermal limits of wild and laboratory strains of two African malaria vector species, Anopheles arabiensis and Anopheles funestus. Malar. J. 11, 226.

Mandal, S., Sarkar, R.R., Sinha, S., 2011. Mathematical models of malaria - a review. Malar. J. 10, 202.

Mandyla, M., Tsiamis, C., Kousoumis, A., Petridou, E., 2011. Pioneers in the anti-malaria battle in Greece (19001930). Gesnerus 68, 180-197.

Millet, J.-P., Montalvo, T., Bueno-Mar, R., Romero-Tamarit, A., Prats-Uribe, A., Fernndez, L., Camprub, E., del Bao, L., Peracho, V., Figuerola, J., Sulleiro, E., Martnez, M.J., Cayl, J.A., Barcelona, Z.W.G. in, lamo-Junquera, D., de Andrs, A., Avellans, I., Gonzlez, R., Gorrindo, P., Sents, A., Simn, P., Bartumeus, F., Busquets, N., Alejo, I., Gascn, J., Muoz, J., Oliveira, I., Jess Pinazo, M., Rodriguez, N., Bocanegra, C., Espasa, M., Molina, I., Pou, D., Salvador, F., Snchez-Montalv, A., Pumarola, T., Rando, A., Serre, N., Soriano-Arandes, A., Trevio, B., 2017. Imported Zika Virus in a European City: How to Prevent Local Transmission? Front. Microbiol. 8.

Moffett, A., Shackelford, N., Sarkar, S., 2007. Malaria in Africa: Vector Species Niche Models and Relative Risk Maps. PLOS ONE 2, e824.

Mora, C., Dousset, B., Caldwell, I.R., Powell, F.E., Geronimo, R.C., Bielecki, C.R., Counsell, C.W.W., Dietrich, B.S., Johnston, E.T., Louis, L.V., Lucas, M.P., McKenzie, M.M., Shea, A.G., Tseng, H., Giambelluca, T.W., Leon, L.R., Hawkins, E., Trauernicht, C., 2017. Global risk of deadly heat. Nat. Clim. Change 7.

Mughini-Gras, L., Mulatti, P., Severini, F., Boccolini, D., Romi, R., Bongiorno, G., Khoury, C., Bianchi, R., Montarsi, F., Patregnani, T., Bonfanti, L., Rezza, G., Capelli, G., Busani, L., 2014. Ecological niche modelling of potential West Nile virus vector mosquito species and their geographical association with equine epizootics in Italy. EcoHealth 11, 120132.

NCMA S.A. [WWW Document], URL http://www.ktimatologio.gr/sites/en/Pages/Default.aspx (accessed 7.16.17).

Neelin, D.J., 2011. Climate Change and Climate Modelling, University of California, Cambridge University Press.

Ngasala, B.E., 2010. Improved malaria case management in under-fives in the era of Artemisini-based combination therapy in Tanzania, Karolinska Institutet.

Ngonghala, C.N., Pluciski, M.M., Murray, M.B., Farmer, P.E., Barrett, C.B., Keenan, D.C., Bonds, M.H., 2014. Poverty, Disease, and the Ecology of Complex Systems. PLOS Biol. 12.

Parham, P.E., Michael, E., 2010. Modeling the Effects of Weather and Climate Change on Malaria Transmission. Environ. Health Perspect. 118, 620626.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 3742.

Pascual, M., Cazelles, B., Bouma, M.J., Chaves, L.F., Koelle, K., 2008. Shifting patterns: malaria dynamics and rainfall variability in an African highland. Proc. R. Soc. Lond. B Biol. Sci. 275, 123132.

Pavlidis, V., 2015. Evaluation of climatic simulations of climate models for Europe for the period 1990-2008, MSc Dissertation, Aristotle University of Thessaloniki.

Peterson, A.T., Papes, M., Soberon, J., 2015. Mechanistic and Correlative Models of Ecological Niches, Eur. J. Ecol. 1, 28-38.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231259.

Prevention, C.-C. for D.C. and CDC - Malaria - About Malaria - Biology [WWW Document]. URL https://www.cdc.gov/malaria/about/biology/ (accessed 7.11.17).

Pulliam, H. r., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349361.

Sallam, M.F., Xue, R.-D., Pereira, R.M., Koehler, P.G., 2016. Ecological niche modeling of mosquito vectors of West Nile virus in St. Johns County, Florida, USA. Parasit. Vectors 9, 371.

Samson, J., Berteaux, D., McGill, B.J., Humphries, M.M., 2011. Geographic disparities and moral hazards in the predicted impacts of climate change on human populations. Glob. Ecol. Biogeogr. 20, 532544.

Semenza, J.C., 2015. Prototype early warning systems for vector-borne diseases in Europe. Int. J. Environ. Res. Public. Health 12, 63336351.

Sillero, N., 2011. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. Ecol. Model. 222, 13431346.

Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, Charles M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W. Burkot, T.R., Harbach, R.E., Hay, S.I., 2012. A global map of dominant malaria vectors. Parasit Vectors. 5, 1-69.

Sudre, B., Rossi, M., Van Bortel, W., Danis, K., Baka, A., Vakalis, N., Semenza, J.C., 2013. Mapping Environmental Suitability for Malaria Transmission, Greece. Emerging Infectious Diseases, 19, 784786.

Sweeney, A.W., Beebe, N.W., Cooper, R.D., 2007. Analysis of environmental factors influencing the range of anopheline mosquitoes in northern Australia using a genetic algorithm and data mining methods. Ecol. Model. 203, 375386.

Thom, D., Rammer, W., Dirnbck, T., Mller, J., Kobler, J., Katzensteiner, K., Helm, N., Seidl, R., 2017. The impacts of climate change and disturbance on spatio-temporal trajectories of biodiversity in a temperate forest landscape. J. Appl. Ecol. 54, 2838.

Tompkins, A.M., Di Giuseppe, F., 2014. Potential Predictability of Malaria in Africa Using ECMWF Monthly and Seasonal Climate Forecasts. J. Appl. Meteorol. Climatol. 54, 521540.

Tompkins, A.M., Ermert. V., 2013. A regional-scale, high resolution dynamical malaria model that accounts for population density, climate and surface hydrology. Malar. J. 12.

Tonnang, H.E., Kangalawe, R.Y., Yanda, P.Z., 2010. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. Malar. J. 9, 111.

Tseroni, M., Baka, A., Kapizioni, C., Snounou, G., Tsiodras, S., Charvalakou, M., Georgitsou, M., Panoutsakou, M., Psinaki, I., Tsoromokou, M., Karakitsos, G., Pervanidou, D., Vakali, A., Mouchtouri, V., Georgakopoulou, T., Mamuris, Z., Papadopoulos, N., Koliopoulos, G., Badieritakis, E., Diamantopoulos, V., Tsakris, A., Kremastinou, J., Hadjichristodoulou, C., 2015. Prevention of Malaria Resurgence in Greece through the Association of Mass Drug Administration (MDA) to Immigrants from Malaria-Endemic Regions and Standard Control Measures. PLoS Negl. Trop. Dis. 9.

Tsiamis, C., Piperaki, E.T., Tsakris, A., 2013. The history of the Greek Anti-Malaria League and the influence of the Italian School of Malariology. Infez Med 21, 60-75.

Vakali, A., Patsoula, E., Spanakos, G., Danis, K., Vassalou, E., Tegos, N., Economopoulou, A., Baka, A., Pavli, A., Koutis, C., Hadjichristodoulou, C., Kremastinou, T., 2012. Malaria in Greece, 1975 to 2010. Euro Surveill. 17.

Watts, N., Adger, W.N., Ayeb-Karlsson, S., Bai, Y., Byass, P., Campbell-Lendrum, D., Colbourn, T., Cox, P., Davies, M., Depledge, M., Depoux, A., Dominguez-Salas, P., Drummond, P., Ekins, P., Flahault, A., Grace, D., Graham, H., Haines, A., Hamilton, I., Johnson, A., Kelman, I., Kovats, S., Liang, L., Lott, M., Lowe, R., Luo, Y., Mace, G., Maslin, M., Morrissey, K., Murray, K., Neville, T., Nilsson, M., Oreszczyn, T., Parthemore, C., Pencheon, D., Robinson, E., Schtte, S., Shumake-Guillemot, J., Vineis, P., Wilkinson, P., Wheeler, N., Xu, B., Yang, J., Yin, Y., Yu, C., Gong, P., Montgomery, H., Costello, A., 2017. The Lancet Countdown: tracking progress on health and climate change. The Lancet 389, 11511164.

Weiss, D.J., Bhatt, S., Mappin, B., Van Boeckel, T.P., Smith, D.L., Hay, S.I., Gething, P.W., 2014. Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. Malar. J. 13, 171.

WHO — World Health Statistics 2016: Monitoring health for the SDGs [WWW Document], URL http://www.who.int/gho/publications/world_health_statistics/2016/en/ (accessed 7.10.17).

WHO/WMO. (2016) Climate Services for Health: Improving public health decision-making in a new climate. Eds. J.Shumake-Guillemot and L.Fernandez-Montoya. Geneva.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 1972919736.

Williamson, A.E., Ylioja, P.M., Robertson, M.N., Antonova-Koch, Y., Avery, V., Baell, J.B., Batchu, H., Batra, S., Burrows, J.N., Bhattacharyya, S., Calderon, F., Charman, S.A., Clark, J., Crespo, B., Dean, M., Debbert, S.L., Delves, M., Dennis, A.S.M., Deroose, F., Duffy, S., Fletcher, S., Giaever, G., Hallyburton, I., Gamo, F.-J., Gebbia, M., Guy, R.K., Hungerford, Z., Kirk, K., Lafuente-Monasterio, M.J., Lee, A., Meister, S., Nislow, C., Overington, J.P., Papadatos, G., Patiny, L., Pham, J., Ralph, S.A., Ruecker, A., Ryan, E., Southan, C., Srivastava, K., Swain, C., Tarnowski, M.J., Thomson, P.,

Turner, P., Wallace, I.M., Wells, T.N.C., White, K., White, L., Willis, P., Winzeler, E.A., Wittlin, S., Todd, M.H., 2016. Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. ACS Cent. Sci. 2, 687701.