**R.A. van der Roest**

MSc Thesis

# Habitat suitability mapping for *Tursiops truncatus* in the Aegean Sea

| | |
|---|---|
| Date: | 01-03-2019 |
| Supervisor Wageningen University: | A.R. Bergsma |
| Responsible Professor Wageningen University: | R.J.A. van Lammeren |
| Supervisor Archipelagos Institute of Marine Conservation: | T.J. Grandjean |

**Colophon**

Geographical Information Management and Applications MSc Thesis
March 2019

This master thesis is commissioned by the University of Utrecht, University of Wageningen, University of Twente and the Delft University of Technology as part of the Geographical Information Management and Applications (GIMA) Master of Science.

Supervisor Wageningen University:
Ir. A.R. Bergsma

Responsible Professor Wageningen University:
Dr. Ir. R.J.A. van Lammeren

External Supervisor Archipelagos:
MSc T.J. Grandjean

**Contact information**

Ronja van der Roest
ronjavanderroest@gmail.com

UU Student number: 4093666

# Summary

It is a globally accepted principle that marine biodiversity needs to be conserved, there are clear links between the level of biodiversity of an ecosystem and its functioning. The common bottlenose dolphin is one of the animals keeping this biodiversity intact within the Mediterranean sea. Since 2006 this dolphin his qualified as vulnerable and got listed on the ‘International Union for Conservation of Nature and Natural Resources' (IUCN) red list of threatened animals (International Union for Conservation of Nature and Natural Resources, 2000). The main aim of this research is to indicate the favourable temporal habitat in the Aegean Sea of the common bottlenose dolphin, this is done by researching to what extend ecological and environmental factors can explain the favourable temporal habitat of this dolphin. When it is known what the favourable temporal habitat is throughout the year, it gets easier to put on specific and more effective conservation programs.

The program Maxent (a species distribution model, SDM, program) is used to model this favourable temporal habitat, the temporal aspect used in this research are the four seasons throughout the year. The choice for Maxent is based on a literature study, researching which program has the best fit with the data available for this research. Within Maxent ten different environmental variables (bathymetry, chlorophyll, distance from shore, nitrate, oxygen, phosphate, phytoplankton, salinity, slope and temperature) and occurrence points gathered around the island of Samos in the years 2016, 2017 and 2018 are analysed to find where the highest likelihood of occurrence per season is in the Aegean Sea and which variables contribute most to this likelihood of occurrence.

The results show that the likelihood of occurrence per season differs throughout the year, so every season shows a different favourable temporal habitat. The variable importance per season is comparable for the seasons winter, fall and summer in which fall and winter are extremely comparable, spring shows differences in variable importance. Bathymetry is considered as the main variable in all four seasons, in spring followed by the variables slope and second runner up temperature, the other three seasons have oxygen and phytoplankton as a second and third variable.

Concluding there can be said that according to the Maxent outputs of this research the common bottlenose dolphin has a different favourable temporal habitat per season and that ecological and environmental variables can explain this favourable temporal habitat. To really make a start with more specific conservation management more research is needed in this subject, but this research did make a proper beginning by indicating that there are seasonal differences in the favourable temporal habitat with both comparable and differing variables of importance within the seasons.

# Acknowledgements

I would like to thank all the people that were involved in this thesis for both, their support and contributions. At first, I would like to thank my supervisors;

**Aldo Bergsma;** Wageningen University, for his extremely positive feedback during the entire research process and guidance into the right direction. Furthermore, he was always available to quickly answer questions over email which helped to not get stuck in a problem. **Ron van Lammeren**; Wageningen University, for all his feedback on my thesis and the regular meetings. Next to this for the motivation and the trust in me to finish this project on time, which really helped! **Tim Grandjean**; Archipelagos Institute of Marine Conservation, for his guidance in this entire research. Not only he helped me set up this project, he was always ready to help when problems occurred and to give feedback on pieces I wrote. Furthermore, he has been extremely helpful in coming up with solutions when I was not sure anymore how to go further with my thesis.

Furthermore, I would like all the other people at Archipelagos that helped me in creating this thesis; Belen Quintana, Guido Pietroluongo, Karthik Ashok and Simone Antichi. Also a big thanks to all the interns and volunteers that have been going on boat surveys the past three years to gather all the occurrence data that I have used in this thesis. Without all this sighting effort this research would not have been possible.

I hope you will all enjoy reading this thesis, and that this research will both stimulate in and help in saving the common bottlenose dolphin in not only Aegean Sea but all over the world!

# List of Figures, Images, Maps and Tables

# Table of Contents

# 1. Introduction

## 1.1 The common bottlenose dolphin in Greece

The common bottlenose dolphin, *Tursiops truncatus,* has been studied intensively in numerous locations around the world, making this species one of the best-known cetaceans from the over 85 species that exist. In the Mediterranean Sea, relatively little is known about this dolphin species since modern cetacean field studies only started in the late 1980s in this area (Bearzi, Fortuna, & Reeves, 2009). It is, for example, unclear how many common bottlenose dolphins currently live in this area. One of the first attempts to assess their status was during a regional Red List workshop in March 2006. Participants agreed that the Mediterranean bottlenose dolphin had to be qualified as vulnerable and the species got listed on the 'International Union for Conservation of Nature and Natural Resources' (IUCN) red list of threatened animals (International Union for Conservation of Nature and Natural Resources, 2000; Reeves & Notabartolo Di Sciara, 2006). This classification was primarily based on direct kills and incidental mortality in fishing gear, the assessment also took ongoing threats from overfishing of prey populations and habitat degradation into account. These factors are not easily quantified but were suspected to be contributing to an overall decline of the common bottlenose dolphin in the Mediterranean region (Bearzi et al., 2009).

## 1.2 Conservation within the Mediterranean

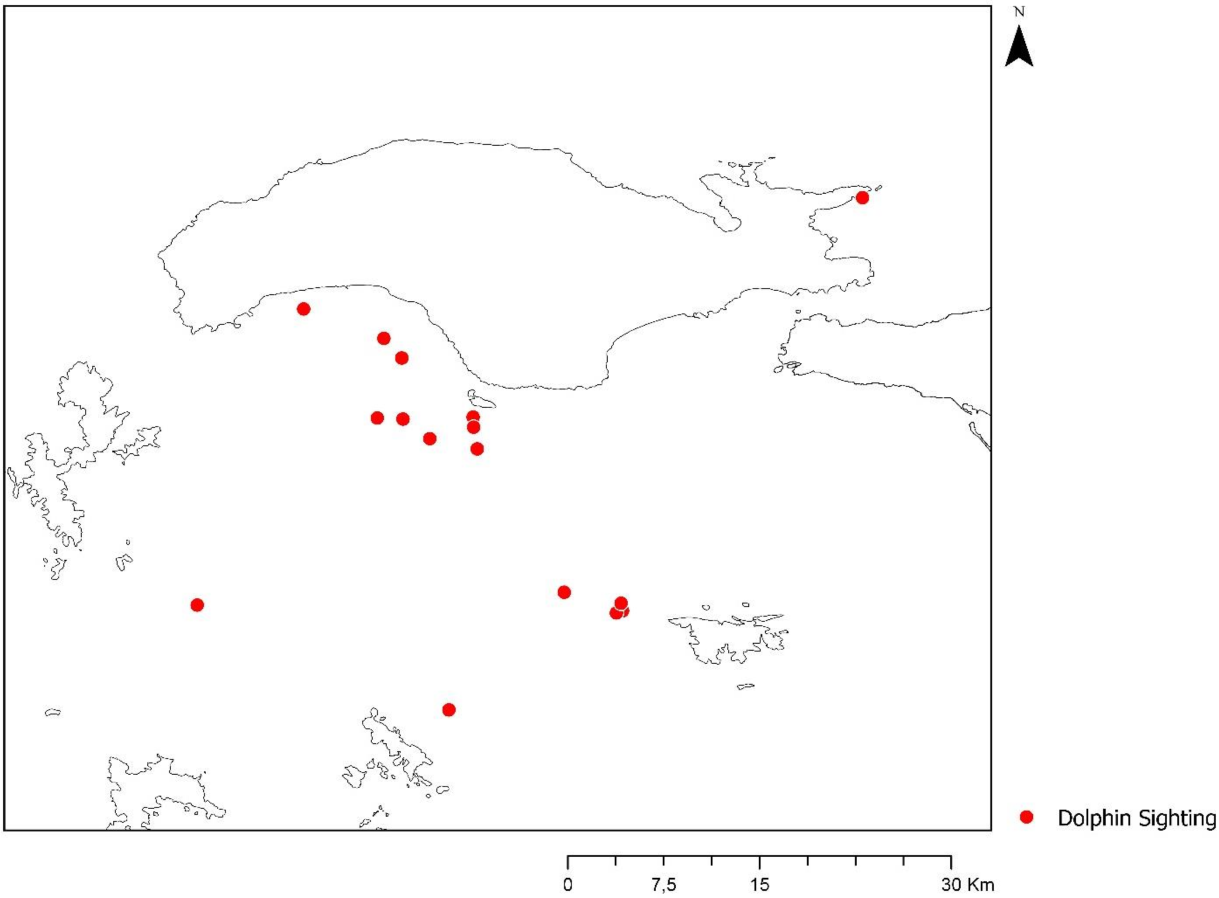
It is a globally accepted principle that marine biodiversity needs to be conserved, this is enshrined in several international agreements and conventions, from which most notably the Convention on Biological Diversity (CBD) (United Nations, 1992), it is in force since 1992 and consists of 193 parties (United Nations, 1992). Biodiversity within the seas is an important issue since there are clear links between the level of biodiversity in an ecosystem and the functioning of this ecosystem. Furthermore, protection of (among other cetaceans) the common bottlenose dolphin is important due to their intrinsic natural value (apex predator) and conservation actions for this dolphin may extend their benefits to other species and to the environment they are part of. Considering the high mobility of these dolphins conservation needs cooperation amongst all countries within the range of the dolphins. The dolphins do not simply stay in the waters of one single nation (Notarbartolo-Di-Sciara & Birkun, 2010). Protection of the biodiversity in the marine ecosystems is among others regulated within so-called Exclusive Economic Zones (EEZ) in which a country can exploit, conserve and manage the natural resources in the zone, among which the marine life (Katsanevakis et al., 2014). At the international level, articles 56(1)(b)(iii), 192 and 194 of the United Nations Convention on the Law of the Sea (UNCLOS) oblige the coastal countries to protect the marine environment in their own EEZ, one of these measures that are taken up in the law is the establishment of Marine Protected Areas (MPA). The regulations of these areas must be based on the rules set by the UNCLOS (Czybulka & Bosecke, 2006). The effectiveness of the MPAs and other protection programs such as the EU habitat directive have a more severe impact in protecting species that are highly site attached as for species that have a wide-ranging moving pattern, this due to the fact that the moving species are more vulnerable to fishing (Edgar, Russ, & Babcock, 2007; Notarbartolo-Di-Sciara & Birkun, 2010). You cannot simply build a fence in the sea to keep animals within the conservation area. Neither it is possible to protect the entire living habitat of wide-ranging species. Knowing what the favourable habitat of species is throughout the year will help setting up specific conservation programs. At this moment the favourable temporal habitat of the common bottlenose dolphin throughout the year within the Mediterranean is unknown. Firth, Wilson, Thompson, & Hammond, (1997) researched the seasonal migration of the common bottlenose dolphin in the Northern parts of Scotland. In their research it is stated the common bottlenose dolphin shows seasonable area use, this group of dolphins researched was one of the most northern groups researched at this time so, seasonable fluctuations were expected. They were found but not at such a high presence as predicted. The seasonal differences in favourable habitat depend in this area on several factors such as predator avoidance, the rearing of offspring, mating and foraging (Firth et al., 1997).

## 1.3 Research aim and area

This research attempts to make the first step towards more flexible conservation management by determining which factors make the favourable temporal habitat of the common bottlenose dolphin. So this brings us to the main objective which is finding this favourable temporal habitat. Within this research, a species distribution model (SDM) also called ecological niche model or habitat suitability model, is created to find the favourable temporal habitat of the common bottlenose dolphin within the study area, this will be explained more thoroughly in Chapter 3.

The research area of this research covers the entire Aegean Sea, the sightings of the common bottlenose dolphin used for this research are located around, especially, the islands of Samos, Lipsi, Patmos, Arkoi and Fourni in the Aegean Sea. The area in which the occurrence points are gathered is smaller but still considered sufficient for the larger research area since there are several different options in SDM models that are suitable for a small area or number of occurrence points (Jane Elith et al., 2006). The sightings of the common bottlenose in spring over 2016-2018 are depicted below in Map 1 as an example;

*Map 1; Sightings of T. truncatus in spring, 2016-2018*




## 1.4 Research questions

To reach the main objective of this research a main- and several sub-questions are conducted in which the first sub-question focusses on the literature, the second one on the methodology, the third and fourth one are control questions.

*To what extent can ecological and environmental factors explain the favourable temporal habitat of Tursiops truncatus in the Aegean Sea?*

- Which temporal habitat suitability model is suitable for mapping the favourable temporal habitat suitability of *T. truncatus*?
- Which environmental and ecological factors might have an influence on *T. truncatus* favourable temporal habitat?
- How strong are the links/correlations between the factors and the presence of *T. truncatus* within the research area?
- To what extent can these results be extrapolated and validated in other regions?

## 1.5 Research scope

This research is about showing the optional favourable temporal habitat of the common bottlenose dolphin within the East Aegean Sea. This research does not intend to pinpoint the exact area of where the common bottlenose dolphin will live during a season, whether the dolphins are really there or not within certain time frames lays outside the scope of this research.

## 1.6 Reading guide

In the next chapter, Chapter 2, different habitat suitability methods are discussed. Chapter 3 covers information about the common bottlenose dolphin itself and which environmental and ecological variables are of importance for this species. The datasets containing the variables of importance and the dataset containing the sightings of the common bottlenose dolphin are described in Chapter 4. Chapter 5 discusses the used methodology. Chapter 6 covers the results of the performance of the model, the likelihood of occurrence of the common bottlenose dolphin in the Aegean Sea and shows how important every environmental and ecological variable is concerning the likelihood of occurrence across the research area. Chapter 7 presents the conclusion of this thesis, followed by the discussion and recommendations in Chapter 8.

# 2. Habitat suitability modelling

This chapter gives an overview of several habitat suitability models. This is done by first introducing the term, a description of two presence-only models (BIOCLIM and DOMAIN), regression models (BRT and GLM & GAM), machine learning (MaxEnt and GARP) and there is concluded with a comparison.

## 2.1 Introduction to habitat suitability modelling

The question of how species are distributed on earth over time and space has a long history. Most modelling approaches, developed for predicting animal species distribution, have their roots in quantifying species-environment relationships (Guisan & Thuiller, 2005). Various names have been given to the studying of suitability analyses, including bioclimatic envelope, species niche and habitat suitability. Within this research, the term Species Distribution Model (SDM) will be used (Booth, Nix, Busby, & Hutchinson, 2014). A species distribution model (SDM) can be seen as an empirical model which relates field observations to environmental predictor variables, based on theoretically derived response surfaces (Guisan & Thuiller, 2005). They are used to tackle conservation issues, such as managing species distribution, assessing ecological impacts of various factors, risk of biological invasions, endangered species management and the understanding of ecological and evolutionary determinants of spatial patterns of diversity (Jane Elith et al., 2006; Hirzel, Le Lay, Helfer, Randin, & Guisan, 2006). Models statistically relate field observation to a set of environmental variables and produce spatial predictions indicating the suitability of locations for a target species. There are different types of modelling techniques to fit different types of recorded information to put in the model (Jane Elith et al., 2006).

SDMs are able to predict the distribution of species by combining known occurrence records with digital layers of environmental variables. The most common strategy for estimating the actual or potential geographic distribution of species is to characterise the environmental conditions that are suitable for the species and to then identify where suitable environments are distributed in space. The spatial distribution of environments that are suitable for the species can be estimated across a study area by running an SDM (Pearson, 2010). There are two approaches in creating an SDM, a mechanistic or a correlative approach. The mechanistic model aims to incorporate physiologically limiting mechanisms in a species' tolerance to environmental conditions. The correlative model aims to estimate the environmental conditions that are suitable for a species by associating known species' occurrence records with suites of environmental variables that can reasonably be expected to affect the species' physiology and the probability of persistence. In both models, the observed distribution of the species provides the information for the environmental requirements of that species (Pearson, 2010). The second consideration that needs to be made is choosing an explanatory model or predictive model. Explanatory models try to seek insight into several ecological processes by finding relationships between the data.

These relationships are often determined from statistical models that ascertain the strength of the statistical relationship between a response and a (combination of) one or more explanatory variables. Predictive models are seeking to provide the user with a statistical relationship between the response and a series of predictive variables for use in predicting the probability of species occurrence or estimating numbers of occurrences at new locations. These models have as goals a model that predicts the ecological attributes of interest from a restricted number of predictors (Guisan, Edwards, & Hastie, 2002).

### 2.1.1 Different types of habitat suitability models

The input data for habitat modelling can be split up in factors and location data of the mapped species. This data can be presence-only data, presence/absence data or only absence data, most research has focussed on data that is collected in a systematic manner. However, occurrence data for most species have been recorded without planned sampling schemes, and the great majority of these data consist of presence-only records form museum or herbarium collections that are increasingly accessible electronically. Environmental data layers of high spatial resolution, for example, derived from satellite images, are now more abundant and available than some time ago. There is a big amount of methods to model species' distribution and habitat suitability and they all vary in how they deal with the data and how they model the habitat suitability (Jane Elith et al., 2006). Elith et al. (2006) did a comparative research on several models to analyse habitat suitability. This research is used as an indication of the most common and acceptable models. The described models are split up in two types of models, of which the first group uses presence-only data and the second group describing the presence of other species, i.e. community-based techniques (Jane Elith et al., 2006).

## 2.2 Presence only models

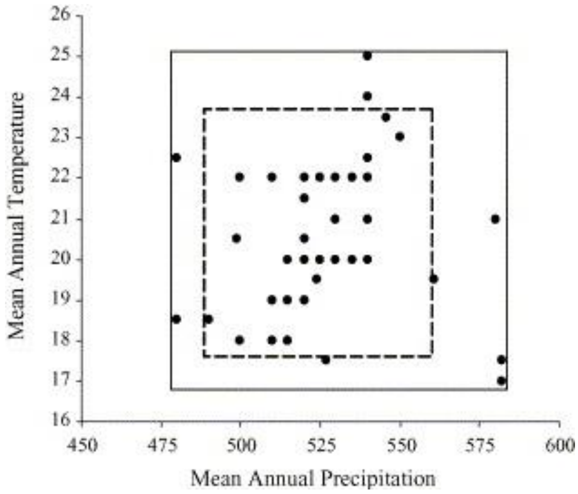
Presence only models can be categorised within three main categories; these are (1) those that use the species data without reference to any environmental data; (2) those that model a species-environment relationship in reference to the species presence data; and (3) those that model a species-environment relationship by characterising the 'background' environment across the region of interest, and modelling the species presence in comparison to this background. The first category of models is useful for estimation of ranges but not for more detailed maps of species distribution. The second category can be seen as a climate envelope method that maps habitat that is climatically suitable for the researched species, this category includes the models of Bioclim and Domain. The third category includes most other presence-only methods. One of the downsides of using presence-only data is the lack of available and broadly accepted methods for evaluating the predictive performance of these models (Wintle, Elith, & Potts, 2005). Within this subchapter the presence-only models Bioclim and Domain will be discussed.

### 2.2.1 BIOCLIM

BIOCLIM is a bioclimatic model, these assume that climate ultimately restricts species distributions. Bioclimatic models summarise a number of climatic variables within the known range of a species, thus generating a bioclimatic envelope (summary of the climate at the locations from where the species has been recorded) which on its point again can be used to find other suitable environment with the same bioclimatic variables (Beaumont, Hughes, & Poulsen, 2005). For many years Bioclim has been the leading SDM package and it remains widely used (Booth et al., 2014). Bioclim summarises up to 35 climatic parameters (available in four different spatial resolutions) throughout a species' known range and assesses the climate suitability of habitat under current and future climate scenarios. It is a correlative modelling tool that interpolates up to 35 climatic parameters for any location for which the latitude, longitude, and elevation are known. The modelling tool can be used for three tasks; (I) describing the environment in which the species has been recorded, (II) identifying other locations where the species may currently reside, and (III) identifying where the species may occur under alternate climate scenario's (Beaumont et al., 2005).

Bioclim is a range-based model that describes the species climatic envelope as a rectilinear volume depicted in figure 1 it suggests that species can tolerate locations where values of all climatic parameters fit within the extreme values determined by the set of known locations. The current potential distribution of a species is identified by interpolating the climate profile of the species within each grid cell of a Digital Elevation Model (DEM) and comparing it to the climatic profile of the species. When a location has all climatic variables will be classified as climatically suitable by Bioclim. Multiple levels of classification can be achieved by removing the extreme values of each parameter and identifying locations with climatic values that lie within different percentile limits (Beaumont et al., 2005). This can be seen in figure 1 depicted below. This figure shows the diagrammatic representation of a hypothetical two-dimensional bioclimatic envelope. Dots represent the values of mean annual temperature and mean annual precipitation at each known location of a hypothetical species. In predicting a species' potential distribution BIOCLIM classifies all locations with values within the extremes of the species' envelope, which is the unbroken line, as suitable (Beaumont et al., 2005).

*Figure 1; BIOCLIM (Beaumont et al., 2005)*




When predicting a species' potential distribution, Bioclim would classify all locations with values within the extremes of the species envelope, which is the unbroke line (Figure 1), as suitable. The dashed box represents those areas where climatic values outside of the 5-95$^{th}$ percentiles of the species envelope are excluded (Beaumont et al., 2005).

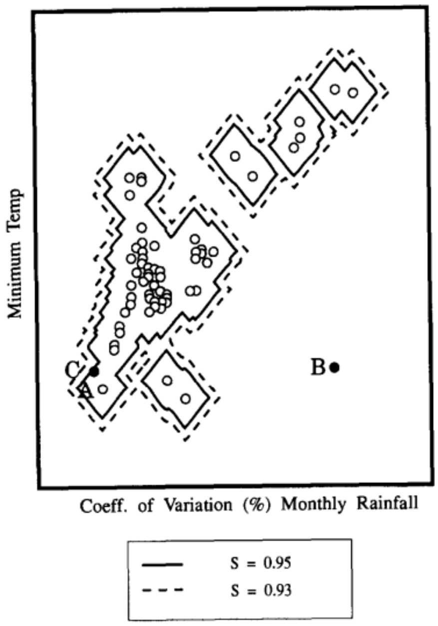
A criticism of BIOCLIM is that the use of all 35 parameters may lead to over-fitting of the model which in turn may result again in misrepresentations od species' potential ranges and to the loss of biological reality. (Beaumont et al., 2005). Furthermore, Bioclim treats each climatic axis independently which in some cases may lead to ecological unsound predictions (Carpenter, Gillison, & Winter, 1993). Most comparison articles conclude that Bioclim is the least effective method to use (Jane Elith et al., 2006; Hernandez, Graham, Master, & Albert, 2006).

### 2.2.2 Domain

Domain is a model working with multivariate distance and can be run in among others R and DIVA-GIS, which is a free online GIS program (Jane Elith et al., 2006). DOMAIN can be seen as one of the more simple environmental matching models, same as Bioclim (Jane Elith, Graham, & Elith, 2009). Domain uses a point-to-point similarity metric to assign a classification value to a candidate site based on the proximity in environmental space of the most similar record site. This can be seen in figure 2, depicted on the next page.

Domain does not define a discrete boundary for the climate envelope. All candidate points are assigned similar values and user-defined thresholds or contour intervals determining the actual ranges mapped. When there are absence records available they can be used to establish thresholds more objectively by determining the value which minimizes the classification error. The variable sensitivity is high in this model. It performs well with limited site data, gives similarity value to all sites and is easily implemented (Carpenter et al., 1993).

*Figure 2; DOMAIN (Carpenter et al., 1993)*




## 2.3 Regression models

Regression methods, have been a commonly applied method for modelling and predicting habitat (Guisan et al., 2002; Wintle et al., 2005). Within this subchapter several regression-based models will be discussed, started with Boosted Regression Trees (BRT) and followed by a subchapter about Generalised Linear models (GLMs) and Generalised Additive Models (GAMs).

### 2.3.1 BRT

BRT is a model and stands short for Boosted Regression Tree. BRT can be used within R (Jane Elith et al., 2009). Regression models are frequently used for quantifying the relationship between one variable and other variables upon which this first one depends. It is important for, among others, ecologists to have techniques that are flexible enough to express typical features on their data such as nonlinearities and interactions when using statistical models for explanation and prediction (J. Elith, Leathwick, & Hastie, 2008). BRT is an ensemble method for fitting statistical models that differs fundamentally from conventional techniques that aim to fit a single model. BRTs combine the strengths of both regression trees and boosting. Regression trees are models that relate a response to their predictors by recursive binary splits and boosting is an adaptive method for combining many simple models to give an improved predictive performance. The final BRT model can then be seen and understood as an additive regression model in which individual terms are simple trees, fitted in a forward, stage-wise fashion (J. Elith et al., 2008). Within this subchapter regression trees will first be discussed followed by a better explanation of boosting.

Tree-based models partition the predictor space into rectangles, using a series of rules to identify regions having the most homogeneous responses to predictors. They then fit a constant to each region, with classification trees fitting the most probable class as the constant, and regression trees fitting the mean response for observations in that region, assuming normally distributed errors.

The outcomes of the model are unaffected by monotone transformation of the data, same as differing scales of measurement among predictors and irrelevant predictors are actually never selected. Furthermore, are the trees not responsive to outliers and can accommodate missing data in predictor variables by using surrogates. The trees have a hierarchical structure which means that the response to one input variable is dependent on the values of inputs higher in the tree. This means according to Elith et al., (2008) that this method is as good as GLMs and GAMs. Regression trees do have difficulties in modelling smooth functions, also the tree structure depends on the sample of data, and small changes in training data can result in very different series of splits within the tree (J. Elith et al., 2008).

Boosting is a method for improving the accuracy of the used model. It is based on the idea that it is easier to find and average many rough rules of thumb, than to find a single, highly accurate prediction rule. Boosting is unique because of its sequential character, it is a forward, stage-wise procedure (J. Elith et al., 2008). It works by sequentially applying a classification algorithm to reweighted versions of the data and then taking a weighted majority vote of the sequence of classifiers thus produced. Weak classifiers are combined in such a way that together they produce a powerful 'committee' of classifiers (Friedman, Hastie, & Tibshirani, 2000). There are multiple boosting algorithms and they vary in how they quantify lack of fit and selecting settings for the next iterations.

The positive thing of working with boosted regression trees is that they incorporate the advantages of tree-based methods, handling different types of predictor variables and accommodating missing data. BRTs have no need for prior data transformation or elimination of outliers, they can fit complex nonlinear relationships and automatically handle interaction effects between predictors. Fitting multiple trees in BRT will overcome the biggest drawback of single tree models which is their relatively poor predictive performance. Although BRT models are complex, they can be summarized in ways that give powerful ecological insight, and their predictive performance is superior to most traditional modelling methods (J. Elith et al., 2008).

### 2.3.2 GLMs & GAMs

One of the advantages of using GLMs and GAMs in describing species-habitat relationships and predicting the spatial distribution of suitable habitats is that they are easily used on free available software (R) (J. Elith et al., 2008). Both approaches have been noticeably used within ecological research, partly due to the ability of the models to deal with the multitude of distributions that define ecological data and they blend in well with traditional practises used in linear modelling and analysis of variance (Guisan et al., 2002). Both models also increased our capacity to analyse data with nonnormally distributed errors (presence-absence and count data), and to model nonlinear relationships (J. Elith et al., 2008). Within this paragraph, a description of both models is given.

GML stands for generalized linear model and is the same as a multivariate regression model. The first thing that should be noted is the difference between a general linear model and a generalized linear model is a generalization of the concept of a linear model. Generalized linear models include linear models as a special case but also include logistic regression, exponential regression, and gamma regression as special cases. Log-linear models for multinomial data are closely related to GLMs (Christensen, 2011). GLMs are composed of a random component which is described by the assumed distribution of the observation data, a systematic component which specifies a linear combination of explanatory variables and a 'link' between the random and systematic components of the model that specifies how the mean response (the observation) relates to the explanatory variables in the linear prediction (Wintle et al., 2005).
One of the advantages of GLMs above normal regression models is that GLMs are mathematical extensions of the former and do not force data into unnatural scales, and thereby allow for non-linearity and non-constant variance structures in the data. They are based on the presumption that there is a link between the species looked at, and the other data that needs to predict the habitat. Used data can be of several probability distributions, these include normal, binominal, Poisson, negative binominal or gamma distributions, many of which better fit the non-normal error structures of most ecological data. GLMs are better suited for analysing ecological relationships and more flexible in use than classical Gaussian distributions (Guisan et al., 2002). GLMs have as an advantage to normal linear regression models that they have a better ability to handle larger classes of distributions for the response variable. Furthermore, they can accommodate more general qualitative and semi-quantitative response variables. They are able to constraint the predictions to be within a range of possible values for the response variable.. And, the model incorporates potential solutions to deal with overdispersion (Guisan et al., 2002).

Generalized additive models are a non-parametric, the data does need to fit a normal distribution, generalization of GLMs in which the relationship between both is defined by non-parametric smoothing functions, which create an approximating function that attempts to capture important patterns in the data. This means that the linear predictor that defines the relationship between the dependent and explanatory variables within GLMs is replaced by smoothing functions. When choosing a GAM the most crucial step is choosing this level of smoothing for a predictor (Guisan et al., 2002; Wintle et al., 2005). You can say that the different regression models are being nested within each other. Simple and multiple LS linear regression (SLR and MLR) being the most limiting cases and GAMs the most general (Guisan et al., 2002).

GAMs are more flexible what gives them a better position than GLMs. Also, GAMs are able to fit the data more closely for a given number of degrees of freedom because they are not constrained to fit predefined parametric shapes. It is easier to interpret GMLs, GAMs do not have a retrievable model formula in the classic sense and interpretation generally requires a plot of the fitted response curves. The software that is capable of doing this is R (Wintle et al., 2005).

## 2.4 Machine learning models

During the 1980s and 1990s, several machine learning methods were developed. According to Elith et al. (2008), these methods are used less frequently than regression methods within ecology. One of the reasons that are given for this is that these methods are considered less interpretable and therefore less open to investigation. Machine learning approaches differ from statistical approaches in that statistical methods start by assuming an appropriate data model, and the parameters for this model are then estimated from the data. While machine learning avoids starting with a data model and rather uses an algorithm to learn the relationship between the response and its predictors. Statistical approaches focus on questions such as what model will be proposed, how the response is distributed and whether observations are independent. The machine learning approach assumes that the data generating process (in ecology this can be seen as nature) is complex and unknown, and tries to learn the response by observing inputs and responses and finding dominant patterns. This places the emphasis on a model's ability to predict well and focuses on what is being predicted and how prediction success should be measured (J. Elith et al., 2008). Within this subchapter three machine learning methods will be discussed based on the article by Jane Elith et al., (2009), MaxEnt, GARP and RF.

### 2.4.1 MaxEnt

MaXent is a presence-only species data program for modelling species distribution. The predictive performance of MaXent is consistently competitive with the highest performing methods. The program has been used extensively for modelling species distributions which covers diverse aims across ecological, evolutionary, conservation and biosecurity applications (Jane Elith et al., 2011). It is one of the most popular tools for species distribution. Its popularity is mainly to be due to that MaXent typically outperforms other methods based on predictive accuracy and that the software is particularly easy to use (Merow, Smith, & Silander, 2013). Within its modelling MaXent makes use of a so-called background sample of environments in the region of interest. The data input in the model can be named Presence-Background (PB) data. Maxent applies its maximum-entropy or 'MaXent' principles for fitting the model so that the estimated species distribution differs from a uniform distribution as minimally as required to explain the observations (Guillera-Arroita, Lahoz-Monfort, & Elith, 2014).

MaXent focuses on fitting a probability distribution for the occurrence of the species to the set of pixels across the study region, based on the idea that the best explanation to unknown phenomena will maximize the entropy of the probability distribution, subject to the appropriate constraints. Within this research, the constraints consist of the values of those pixels at which the species has been detected. Normally, default parameters are used for MaXent models (Townsend Peterson, Papeş, & Eaton, 2007).

Originally MaXent was built to estimate the presence across a landscape. Density estimation implicitly assumes that individuals have been sampled randomly across the landscape so, the samples occur in proportion to population density. When the population size is known estimations can be made of the species per grid cell. The problem is that the occurrence isn't known most of the time so no precise estimation can be made. Relative occurrence estimations are possible which results in a relative occurrence rate (ROR). The ROR is the relative probability that a cell is contained in a collection of presence samples. The ROR corresponds to MaXent's raw output. MaXent can be used to predict the probability of presence only by using a transformation of the ROR, called logistic output. This does rely on strong assumptions which have been criticized. MaXent predicts RORs as a function of the environmental predictors at that location (Merow et al., 2013).

### 2.4.2 GARP

GARP models work with presence-only data, with absence information included via automated, random sampling of 1250 pseudo-absence points from the set of pixels at which the species has not been detected. So, this is no true absence data. GARP works with an iterative process of rule selection, testing and incorporation or rejection. At first, the method is chosen from a set of possible methods, these are; logistic regression, bioclimatic rules, range rules and negated range rules. The next step is applying this to the training data and a rule developed; rules may evolve by a number of means to maximize productivity. The predictive accuracy is then evaluated based on the 1250 points resampled from the intrinsic test data and the 1250 pseudo-absence points. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 1000 iterations of these processes or until convergence (Townsend Peterson et al., 2007).

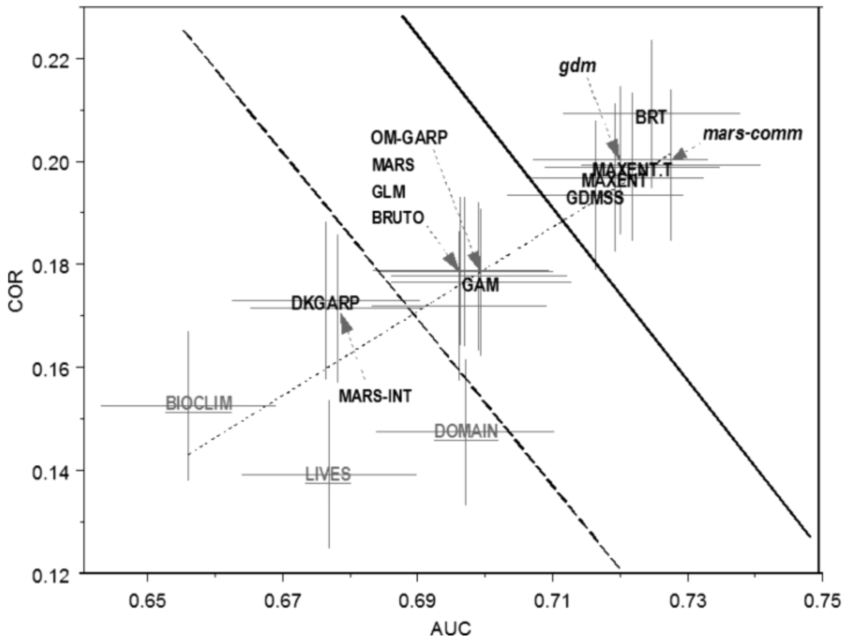
## 2.5 Comparison of the different habitat suitability models

It is common to compare different programs on their performance by using AUC, correlation (COR) and ROC curve. The AUC stands for Area Under the receiver operating characteristic (ROC) plot. It is a measurement of the discriminatory capacity of classification models. It is a widely used method for assessing the accuracy of species distribution models. It takes sensitivity (*Se)* into account which is the proportion of instances of presence correctly predicted as presence and specificity (*Sp)* into account which is the proportion of instances of absence correctly predicted as an absence. The ROC curve plots *Se* versus (1-*Sp)* across all possible thresholds between 0 and 1, this is the commission error and can be explained as the proportion of instances of absence wrongly predicted as presence. A model will be okay when the curve lies above the diagonal of no discrimination, in other words, if the AUC is higher than 0.5. When using only presence data with a background the ROC curve needs to be modified so that instead of plotting *Se* against (1-*Sp*) it is plotted against the proportion of the background locations predicted as presences for all possible thresholds (Jiménez-Valverde, 2012).

Within the study of Hernandez et al., (2006) Domain, Maxent, Garp and Bioclim are compared by using AUC and COR. The AUC of MaxEnt and Domain are the highest followed by Garp and lowest Bioclim. The prediction success is looked at using the ROC curve, MaxEnt scored highest, followed by Garp, Bioclim and last Domain. A side note is that when the sample size of the species became better Garp starts performing better as MaxEnt on the ROC curve. The evaluation discussed above is only done by using presence-only data (Hernandez et al., 2006).

Elith et al., (2006) also did research in comparing several models. They describe that the most obvious differences between models are in the proportion of the region that appears to be predicted most suitable. Within their study, they also found clear indications that presence-only data can provide the basis for accurate predictions, but also marked variation in modelling success. According to them an assessment of modelling success using AUC and COR indicate that methods can be analysed in three groups depicted in figure 3.3 below. Within this figure, more models are depicted as described within this research but still gives a clear indication of the differences in the performance of the several models.

*Figure 3; comparison SDM Models (Jane Elith et al., 2006)*




The lowest performing groups are shown below the dotted line. As to be seen Bioclim scored worst from the discussed models in this research on AUC and Domain scored worst on COR. Both models fall in the lowest group. The middle group contains Garp, GLM and GAM, all scored on average. The highest performing group contains BRT and Maxent (Jane Elith et al., 2006).

So, concluding there can be said that both Maxent and BRT are best performing models considering COR and AUC.

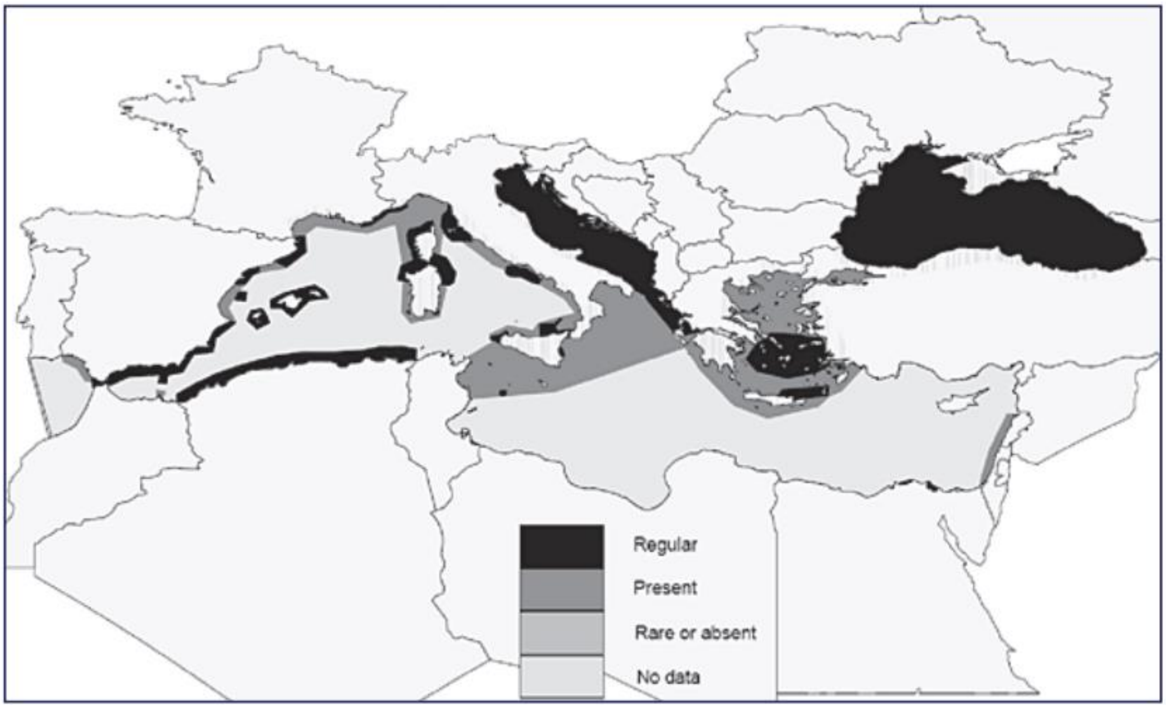
# 3. The common bottlenose dolphin and variables of importance

Within this chapter, a description of the common bottlenose dolphin will be given followed by a description of three environmental factors (slope, distance from shore and water depth) that are of importance for the favourable temporal habitat of the common bottlenose dolphin are discussed. Followed by the ecological factors that might be of influence for the distribution of this dolphin.

### 3.1.1 Introduction to the common bottlenose dolphin

Within the Mediterranean, the common bottlenose has been reliably reported in the waters of Albania, Algeria, Croatia, Cyprus, France, Gibraltar (United Kingdom), Greece, Israel, Italy, Montenegro, Morocco, Slovenia, Spain, Tunisia and Turkey (Bearzi et al., 2009). They occur regularly around offshore islands and archipelagos. The distribution and the density of the dolphins across this region vary greatly. These differences and variations are related to several factors among which habitat characteristics, local availability of suitable prey and the generally gregarious nature of bottlenose dolphin communities. Next to these factors, the effects of past extermination campaigns and the variety of ongoing threats distribute to the distribution of the bottlenose dolphins. The different zones can be characterized by ocean floor topography and by features such as surface salinity, productivity and temperature (Bearzi et al., 2009). Furthermore, the Turkish Straits system, the Strait of Gibraltar and the Suez Canal form natural borders for the dolphins in the Mediterranean (Bearzi et al., 2009; Reeves & Notabartolo Di Sciara, 2006). In map 2 the distribution in the Mediterranean can be seen, based on the IUCN in 2006.

*Map 2; distribution of T. truncatus in the Mediterranean (Reeves & Notabartolo Di Sciara, 2006)*



The common bottlenose dolphin is one of the best-known cetacean species in the world. It is to no surprise that this species is intensively researched in many areas around the world (Bearzi, Agazzi, Bonizzoni, Costa, & Azzellino, 2008; Bearzi et al., 2009; Natoli et al., 2008; Pabst, Rommel, McLellan, Williams, & Rowles, 1995). One of the reasons that this cetacean is so well known is due to its primary near coastal behaviour and the widespread use of this dolphin within oceanaria (Shane, Wells, & Wursig, 1986). They have some distinctive characteristics, varying in size from 1.9 to 3.8 meters in which males are bigger than females. The colour of the dolphins varies from light grey to nearly black on the back and sides, fading to white (sometimes with a pinkish hue) on the belly. The belly, as well as the lower sides, are sometimes spotted. From eye to flipper there is a dark stripe and there is a faint dorsal cape on the back, generally only visible at close range. They often have brushings of grey on multiple places on their body. On each jaw the dolphin has 18 to 26 pairs of robust teeth, these may be worn or missing (Jefferson, Leatherwood, & Webber, 1993). See image 1 on the next page for an impression of the common bottlenose dolphin as seen within the study area of the Aegean Sea.  Even though they look like each other, the Mediterranean bottlenose dolphin is genetically different from the species that live in contiguous areas such as the North Atlantic Ocean and Scottish waters. The differences between the dolphins coincide with the transitions between different habitat areas. These different zones can be characterised by ocean floor topography and by features such as surface salinity, productivity and temperature. The two populations that were identified within the Mediterranean Sea are the western- and eastern Mediterranean bottlenose dolphin. Within these two populations significant genetic differentiations where found (Bearzi et al., 2009). Research studying different species such as certain fish also noted this strong boundary and suggest that differences in hydrographic characteristics defined the different habitats in these basins. The western Mediterranean is more influenced by the Atlantic Ocean and the Eastern Mediterranean is characterized by water circulation limited to the Libico-Tunisian Gulf and by low activity in the rest of the basin (the Adriatic and the Aegean Sea) which are under influence of cool waters of low salinity (Natoli, Birkun, Aguilar, Lopez, & Hoelzel, 2005).

.

*Image 1; T. truncatus around the island of Samos (Archipelagos, 2018)*

As to be seen in map 2 big parts within the Mediterranean lack data. As is also stated by Reeves et al., (2006) surprisingly little is known about abundance, distribution and movement within the Mediterranean sea. The distribution and by which the migration of the bottlenose dolphins depends highly on the habitat in which the animals are living. According to Bearzi et al. (2009) especially feeding habitat shapes the behaviour of dolphins. Diet and foraging behaviour appear to vary widely depending on the area, season or tropic niche occupied by the local dolphins. Even groups that are geographically contiguous may show different foraging behaviour and prey preferences (Bearzi et al., 2009). As already stated in the introduction movement patterns and so the behaviour of the common bottlenose dolphin highly depend on predator avoidance, the raring of offspring, mating and foraging (Firth et al., 1997).

There is no clear number in average travel distance per day from the common bottlenose dolphin. One of the reasons for this is that the research groups studying the dolphins tend to go back and forward to the same port every day, so when a group of dolphins is not seen for a couple of days little is known about the whereabouts of the animals. Bottlenose dolphins in the Mediterranean often tend to stay in their home base region but travelling does occur. Dolphins that are living in more coastal areas tend to travel less than dolphins in more open and protected areas. There are several reasons for common bottlenose dolphins to travel, such as foraging and mating (Bearzi, Bonizzoni, & Gonzalvo, 2011). In several areas, research has been carried out to the study of migrating dolphins. Research, executed in the Pelagos sanctuary located on the left from Italy, found that the dolphins living there show a residential attitude with excursions that are on average 50 kilometres, usually always within a distance of 80 km (Gnone et al., 2011). Within the Mediterranean common bottlenose dolphins are once in a while sighted in areas deeper as 2000 meters and movements across pelagic waters can occur. It is known that the common bottlenose dolphins in the Mediterranean have a high site fidelity, but again research is limited and movements across geographical space are also common (Bearzi et al., 2011).

## 3.2 Environmental variables of importance

Within this chapter, a discussion about the important environmental factors is given. The factors discussed here are the distance from shore, see depth and slope.

### 3.2.1 Distance from shore

The common bottlenose dolphin is known to have two different ecotypes. One group of common bottlenose dolphins is known to be more of an offshore animal while the other group is more an inshore animal. Within the Mediterranean Sea, the common bottlenose dolphin relates to the inshore ecotype, although some the dolphins on the South of Spain have more connections with the Atlantic ecotype (Gnone et al., 2011). Furthermore, it is known that common bottlenose dolphins inhabit a wide variety of habitats including continental shelf waters, lagoons and enclosed seas and the waters surrounding islands and archipelagos. Bottlenose dolphins in the Mediterranean are often regarded as predominantly coastal or inshore. But, the animals can also be found on continental shelves and in shallow plateau waters at any given distance from shore. The only information on the behaviour of bottlenose dolphins in the Mediterranean comes from groups living close to the shore, so little is known and can be said about offshore groups (Bearzi et al., 2009; Jefferson et al., 1993).

### 3.2.2 Water depth

Common bottlenose dolphins prefer to stay in more shallow waters. According to the results of Gnone et al., (2011) bottlenose dolphins tend to prefer waters that are less than 100 meters deep. This research is conducted within the Pelagos Sanctuary, the largest MPA within the Mediterranean. The dolphins within this area also preferred to stay within close distance of the continental shelf (Gnone et al., 2011). Bearzi et al., (2009) also state that the dolphins prefer more shallow waters, they do write about some deeper water as described by Gnone et al., (2009), namely 200-600 meters deep, over steep slopes. Movements into and across pelagic waters may occur and sometimes they are found in waters deeper than 2000 meters (Bearzi et al., 2009). Reasons for the preference of the more shallow water could be related to the feeding habitats of this species, which is as already stated above mostly on benthic and demersal fishes (Gnone et al., 2011).

### 3.2.3 Slope

When looking at research conducted by de Stephanis et al., (2008) they found a positive correlation between steeper slopes and the presence of the common bottlenose dolphin using GAM’s in the street of Gibraltar. Bearzi et al., (2009) also state that the common bottlenose dolphins tend to be found over steep slopes, mainly in waters between 200 and 600 meters deep as stated above. But, again most research is conducted in nearshore areas so not much can be said yet about the areas located further from shore.

## 3.3 Ecological variables of importance

As stated before, research to the common bottlenose dolphin within the Mediterranean is quite scarce (Reeves & Notabartolo Di Sciara, 2006). So, there isn’t that much research yet into the ecological factors determining the favourable temporal habitat in this area. The study of Gomez & Cassini (2015), does describe that sea temperature, sea salinity, ocean productivity (the production of organic matter by phytoplankton), fishing effort and prey distribution, influence the favourable temporal habitat of dolphins. Within their research, an SDM is created to estimate the effects of environmental variables on habitat suitability of river dolphins *Pontoporia blainvillei* along their overall biogeographical distribution (Gomez & Cassini, 2015). Natoli, Birkun, Aguilar, Lopez, & Hoelzel, (2005) researched the habitat structure of the common bottlenose dolphin using among other factors the ecological factors chlorophyll and salinity the later also researched by Hornsby et al., (2017).

# 4. Data description

The aim of this research is to indicate the favourable temporal habitat of the common bottlenose dolphin. This will be done by creating a habitat suitability model for this dolphin with occurrence data of this dolphin gathered in the years 2016, 2017 and 2018 by Archipelagos institute of marine conservation. This model should be able to indicate the factors of importance for the favourable temporal habitat of the common bottlenose dolphin. Within this chapter, the available occurrence data and the to be used environmental and ecological data will be described. There will be concluded with a model selection based on the previous literature chapter and the available data description in this chapter.

## 4.1 Sighting data of *T. truncatus*

The occurrence data used for this research is retrieved by means of boat surveys organised by Archipelagos Institute of Marine Conservation. Within this subchapter, a description is given of the occurrence data by first describing how the data is gathered, followed by describing the type of data and finally the distribution and time.

### 4.1.1 Data gathering

The data is gathered by means of boat surveys on which dolphin sightings are done by a crew from Archipelagos. It is known where the common bottlenose dolphin tends to be within the research area (island of Samos). The collection of the occurrence data is based on these known presence sites, and the data is by, exception gathered, in other areas around the island. There is no systematic way of data collection so the data is totally random. Furthermore, the data is not gathered in a continuous matter, this due to the fact that the surveys are highly weather and equipment dependant (Archipelagos, 2018). The observation protocol of Archipelagos is widely used during cetacean research boat surveys and is an efficient method. When the boat is on survey 4 people are observing at all times, working on rotation with at least 8 crew-members, surveying for the maximum of one hour before having a 30 minutes break. The boat is seen as the centre of a clock, each person will take a quarter scanning the horizon in their area. The environmental characteristics are notated every 30 minutes when there are no sightings and every 3 minutes during a sighting, even as the coordinates and the time of the sighting. The angle and distance towards the sighting of the dolphins are evaluated with more or less accuracy by the crew. The datasheets used during the survey are digitalised after the survey and saved in one big google drive form. This dataset contains sighting information about the dolphins themselves such as location and numbers as well as data about the environment at the moment of the sighting. These include among other wind direction and sea state. Furthermore, sightings are only noted in case of 100% certainty (Archipelagos, 2018).

### 4.1.2 Data type

The gathered occurrence data is presence only, there is some absence data available that could be used, however, the absence records are not of good enough quality and the possibility of false absences exist which makes the data unreliable (Pearson, 2010). No surface recording doesn't mean the absence of a species in the researched water column, the survey typology is only able to collect presence data for the surface and not the rest of the water column. Furthermore, it is always possible for the observants to miss dolphins and misinterpret the sighting.

Wintle et al., (2005) describe that presence only data often suffers from the problems that observations are unplanned and tend to be biased towards places the species is common. They are often of dubious reliability and unspecified spatial accuracy, the variation in survey effort between different environments and geographical areas cannot be controlled or adjusted in model fitting. Nonetheless, presence-only modelling methods are widely applied due to the prevalence of presence-only data (Wintle et al., 2005).

There are three types of presence only category models that can be used:

1. Those that use the species data without reference to any environmental data;
2. Those that model a species-environment relationship in reference to any environmental data;
3. Those that model a species-environment relationship by characterising the 'background' environment across the region of interest, and modelling the species presence in comparison to this background.

Of these three types of models, the models falling in the third category tend to have higher predictive performance (Wintle et al., 2005).

Furthermore, Elith et al (2011) state that it does not matter that much, whether you have absent records or not to include in your final model. They describe that regardless absence is used in modelling the pattern in the presence records will suggest the area is unsuitable, and the model in place will again be affected by this patterning. Similarly, if the detectability of a particular species varies from site to site, this can result in false absences in presence-absence data which again can also affect the pattern in the presence only data. Formally there is said that numbers of distribution cannot be determined with presence-only data, this actually means that it cannot be determined exactly regardless of the sample size. But, since absence data has so many limitations of its own the question is if even presence-absence data can give a good estimation of distribution (Jane Elith et al., 2011).

Another limitation with presence-only data is that sample selection bias has a strong effect on presence-only models, for more explanation see Elith et al. (2011). Whether a species is present or absent is also dependent on the time frame of the data. It might be possible that a species is present during summer but is absent in winter. When using presence-absence data the definition of the response variable should naturally be consistent with the sampling method. E.g. if the available data are surveys of 1-m2 quadrats, then the response variable should correspond to the species being present in a 1-m2 quadrat. With using presence-only data most of the time this is not possible. A solution for this problem is to implicitly assume a sampling unit of size equal to the grain size of available environmental data (Jane Elith et al., 2011).

### 4.1.3 Distribution, density & time

The dolphin data is gathered in a random matter on a relatively small scale. The data is biased towards the places where the dolphins tend to be. In the used dataset this means the data collection is specially collected around the southern edge of the island of Samos. Furthermore, the data consists of “double” sightings. Since the data collection is targeted dolphins are followed when spotted which results in multiple sighting points at different locations of the same pot of dolphins. For this research, this should not be a problem due to the fact that the pot of dolphins which is followed still swims in the favourable temporal habitat. So, those points of the same pot actually enhance the data instead of setting it back.

The fact that the data is collected in a targeted way also enhances the density of the occurrence data in certain areas. This due to the fact that the dolphin occurrence is written down every three minutes during a sighting. This is a point that needs to be taken into account when using the data in the eventual to be chosen model. It should not be a problem since the data only consists of presence data. When a (grid cell) area has occurrence records that entire grid cell will be listed as presence. When the density of the occurrence data does influence the to be chosen model the data should be pre-processed in a correct manner.

According to Pearson (2010), it is possible that biased data in some cases may lead to a non-representative sampling of the available environmental conditions but it does not necessarily have to be the case. Biased data does have a stronger effect on presence-only models than presence-absence models (Jane Elith et al., 2011; Pearson, 2010). When there is more data available the eventual habitat suitability model will be more accurate, because the scale of observations for this research is relatively small it is important that the eventual model used is able to work with a small scale even should the used environmental datasets be available on a small scale.

On average the dataset consists of at least one boat survey per week. There is a big difference between summer and winter as is depicted in figure 4 and 5 below, which are included as examples for 2017:

Figure 4; number of boat surveys, 2017 (Archipelagos, 2018)

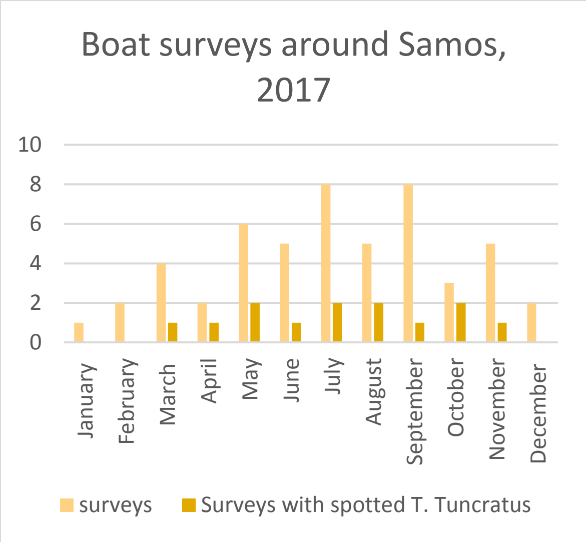
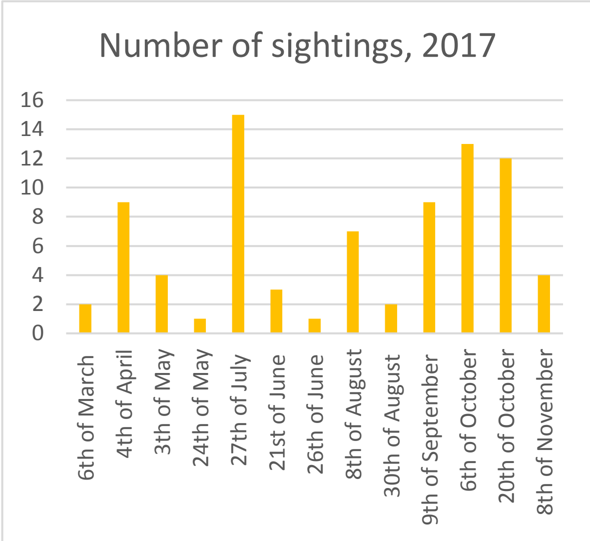



Figure 5; number of spotted T. truncatus, 2017 (Archipelagos, 2018)




As to be seen in the graph the amount of surveys is highest in the months of July and September. As also can be seen not every boat survey leads to common bottlenose dolphin sightings. This could be seen as absence data since the transect lines of (most of) the boat surveys are known but they won't be used due to previously discussed difficulties of using presence-absence data and needing to know for 100% sure whether the absence is really absence data. This is just impossible with just visual sightings and the nature of dolphins to spend most of their time underwater. As seen in figure 4 there are around one to two sightings of pots of the common bottlenose dolphin per month in 2017, except for the months of December-March. Per sighting, the number of points taken (every 3 minutes a recording is made, as described before) varies from 1 to 15 in 2017.

Due to the fact that pots of dolphins are followed while taking the points, many sighting locations are located rather close to each other. In the article by Elith et al., (2006) this was also the case. They fixed the problem by reducing the PO data to one sighting per used grid cell. This might be one of the options but is highly dependable on the to be chosen SDM model. Furthermore, the grid size of the environmental and ecological data that will be used plays a role in how to handle the grid size. It is of importance to choose the correct grid size because as is stated in Elith et al., (2011) the interplay of data quality is among others dependant on the scale of analyses (Jane Elith et al., 2011).

## 4.2 Description of the environmental and ecological variables

The main question of this research asks which ecological and environmental factors have an influence on the favourable temporal habitat of the common bottlenose dolphin. For this not only the occurrence data of this dolphin is needed, but there also needs to be a comparison with several ecological and environmental factors. Since very little is known about which ecological factors influence the favourable temporal habitat of the common bottlenose dolphin (see Chapter 3.3) and this study has more an explorative character there is decided to use some additional factors next to the factors that are already discussed in literature (salinity, temperature, phytoplankton) which are nutrients, chlorophyll, oxygen and phosphate. These datasets are all retrieved from Copernicus – marine environment monitor system, the environmental datasets are retrieved from EMODnet.

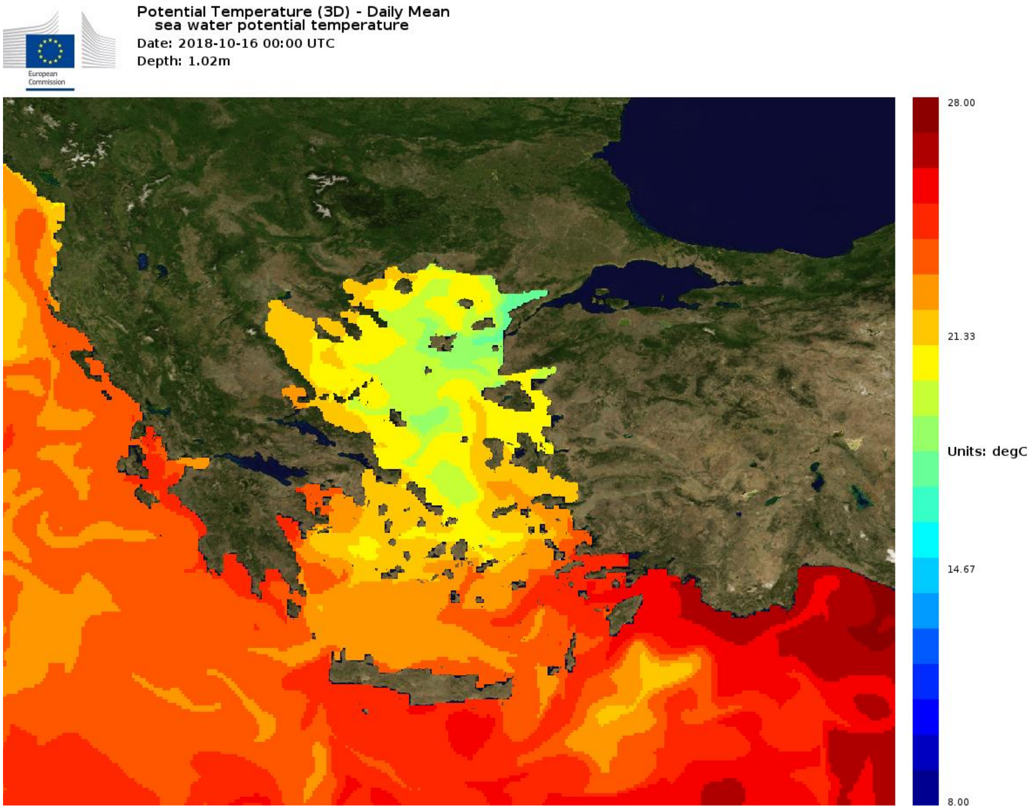
### 4.2.1 Ecological variables

The ecological variables used in this research are retrieved from Copernicus (Copernicus, 2018c). Copernicus is the European Programme for the establishment of a European capacity for earth observation and monitoring. The Copernicus program consists of three components, these are; space, in-situ and services. The space components include the Sentinel-satellites, which are currently still in development for the specific needs of the Copernicus programme. This component is together with several other missions regulated by organisation from several levels. The in situ component calibrates (among others) the satellite data. The service component addresses several themes among which the marine monitoring program. This ocean monitoring program delivers free to use raster datasets in netCDF format which provide the environmental and ecological data for this research (Copernicus, 2018a). As stated before in chapter 2 (introduction) there is a difference between explanatory models and predictive models. For this research, the focus lays more on a predictive model instead of an explanatory model. This due to the fact that the aim of this research is finding which factors contribute to the favourable temporal habitat of the common bottlenose dolphin instead of looking into the statistical relationship between a response and explanatory variable. Predictive models are seeking to provide the user with a statistical relationship between the response and a series of predictive variables for use in predicting the probability of species occurrence or estimating numbers of occurrences at new locations. These models have as goals a model that predicts the ecological attributes of interest from a restricted number of predictors (Guisan et al., 2002).

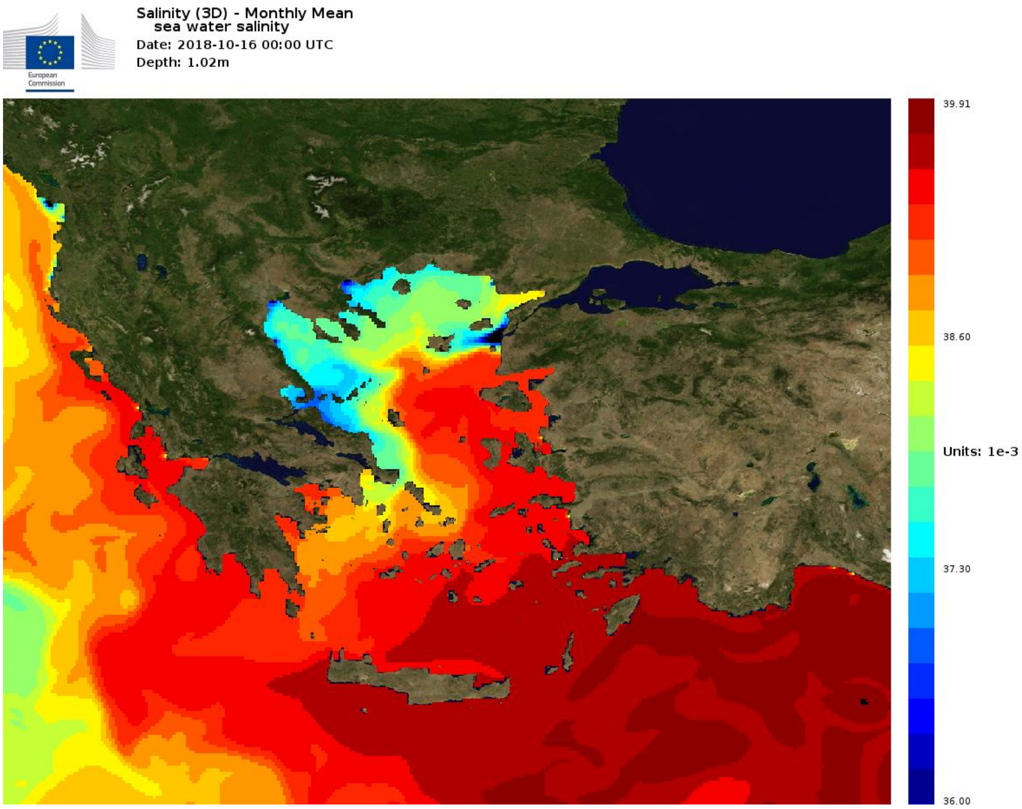
#### 4.2.1.1 Sea temperature and salinity

Two of the datasets used within this research are taken from the product; MEDSEA_ANALYSIS_FORECAST_PHY_006_013. This data is collected by a coupled hydrodynamic-wave model which is implemented over the entire Mediterranean basin. The horizontal grid of the model is circa 4 kilometres and it has 141 unevenly spaced vertical, which results in a cell size of approximately 4x4 kilometres (Copernicus, 2018b). For this research this scale is considered small enough, this since the value of the variables does not change significantly per for example one kilometre, taking a smaller cell size would not influence the results in a positive way, rather making the datasets bigger and slowing processing time. The factors retrieved from this product are salinity and temperature. When retrieving this data several parameters can be set of which depth and time frame are the most important. Depth is an important parameter to set since (almost) all factors will differ in numbers on for example a depth of 40 meters or on the surface area. Since the used occurrence data is at the surface area the surface area data will be retrieved. Since this research focusses on the favourable temporal habitat it is important to keep an eye on the time frame of the data.  For Salinity, there is among others a monthly mean available and for temperature, it is a daily mean. Maps 3 and 4 give an example of how this data looks like in a bigger research area, when zooming in to smaller more specific areas the scale will change showing more local differences.

*Map 3; potential temperature Aegean Sea, 2018-10-16 (Copernicus, 2018a)*




*Map 4; salinity Aegean sea, 2018-10-16 (Copernicus, 2018a)*



4.2.1.2 Chlorophyll, oxygen, phytoplankton and nutrients

These datasets are retrieved from the product MEDSEA_ANALYSIS_FORECAST_BIO006-014. This product is described as the biogeochemical analysis and forecasts for the Mediterranean Sea. This dataset is produced by means of the MedBFm model system which is run by OGS ( Oceanografia e Geofisica Sperimentale) and uses as physical forcing the outputs of the Med-currents products (managed by INGV (National Institute of Geophysics and Volcanology)). Seven days of analysis and ten days of forecast are bi-weekly produced on Wednesday and on Saturday, with the assimilation of surface chlorophyll concentration from satellite observation (produced by the CMEMS-OCTAC (Copernicus Marine Environment Monitoring Service – Ocean Colour Thematic Assembly Centre) (Copernicus, 2018c). The cell size of these factors are the same as the datasets temperature and salinity, so approximately 4x4 kilometre.

### 4.2.2 Environmental variables

The environmental variables used in this research are retrieved from EMODnet (EMODnet, 2018). The bathymetry of the research area is retrieved from EMODnet who offer the bathymetry datasets for free downloading. A complete EMODnet Digital Terrain Model (DTM) is generated for the European sea region from selected bathymetric survey data sets, composite DTMs, Satellite Derive Bathymetry (SDB) data products, parts with no data coverage are completed by integrating the GEBCO Digital Bathymetry, the global terrain models for ocean topography. The dataset is available on a grid resolution of 1/16 * 1/16 arc minutes which is approximately a grid resolution of 115 * 115 meters (EMODnet, 2018). This resolution is considered small enough for this research, dolphins move around during the day and will not stay at 115 x115 meters area. This dataset is used to show the distance from shore, slope and water depth.

## 4.3 Habitat suitability model selection

In previous chapters, there has been an explanation of the goal and background of this topic describing the common bottlenose dolphin and a short introduction towards the marine conservation and it needs in the EAS (chapter 1 and 2). Followed by a discussion of several SDM models in chapter three and an analysis of the occurrence data and a short introduction to the Copernicus data to be used. Within this chapter, a model will be selected using the information discussed in previous chapters.

When looked at the conclusion of Chapter 3, Boosted Regression Trees and MaxEnt are the best performing models. BRT is designed to work with both, presence and absence data where MaxEnt is designed to work with presence/background data. Both models are performing really well, but since this research is focusing on presence data and setting this off against environmental the goal is more in line with MaxEnt. This, in combination with the fact that BRT is designed for presence/absence data, gives the reason to work with MaxEnt in this research.

# 5. Methodology

Within this chapter, the research methodology will be outlined. This is done by first discussing the data pre-processing and followed by which settings are used in Maxent.

## 5.1 Data pre-processing for Maxent

### 5.1.1 Sighting data

The datasheets containing the sighting information needed to be pre-processed and several decisions needed to be made preparing the data in a useful way for MaxEnt.

When there is a sighting the information is noted down every 3 minutes. Since a lot of information needs to be noted down when the sighting happens (for example about the behaviour) the coordinates are only written down at the beginning of the sighting and when the location drastically changes. So, the coordinates are often missing on the data sheet at the moment the sighting takes a longer period of time. Due to the fact that the dolphins are followed on a slow speed (around 2 knots), the location will barely change but still, there is no 100% certainty of the sighting location of that moment. So, the decision was made to remove sighting data of which there is no longitude and latitude. In some cases, the number of animals is also written down just ones, at the beginning of the sighting. Since the boat is following the same group of dolphins we can assume that the number of dolphins is the same in the blank spaces.

So instead of not using this data, the numbers noted down at the beginning are copied down from the first notation. This may lead to some bias within the data but since MaxEnt assigns grid cells as a species being present or not this bias does not matter for this research. In some cases, it hasn't been clear how many animals there were exactly and there is noted down >3 or <10. Since with these signs, the column does not only exist of numbers anymore these signs are removed. So less than three becomes three. The sightings where the spotted species is unknown are deleted from the used dataset. To make the table useable in for example a program as ArcGIS a column stating the ObjectID is added. What remains is a dataset with the object ID, sighting number, date, Latitude, Longitude, species, time and number of animals. Maxent only uses longitude, latitude and species the other fields are ignored in this program but stay useful when the dataset is used in different programs or purposes.

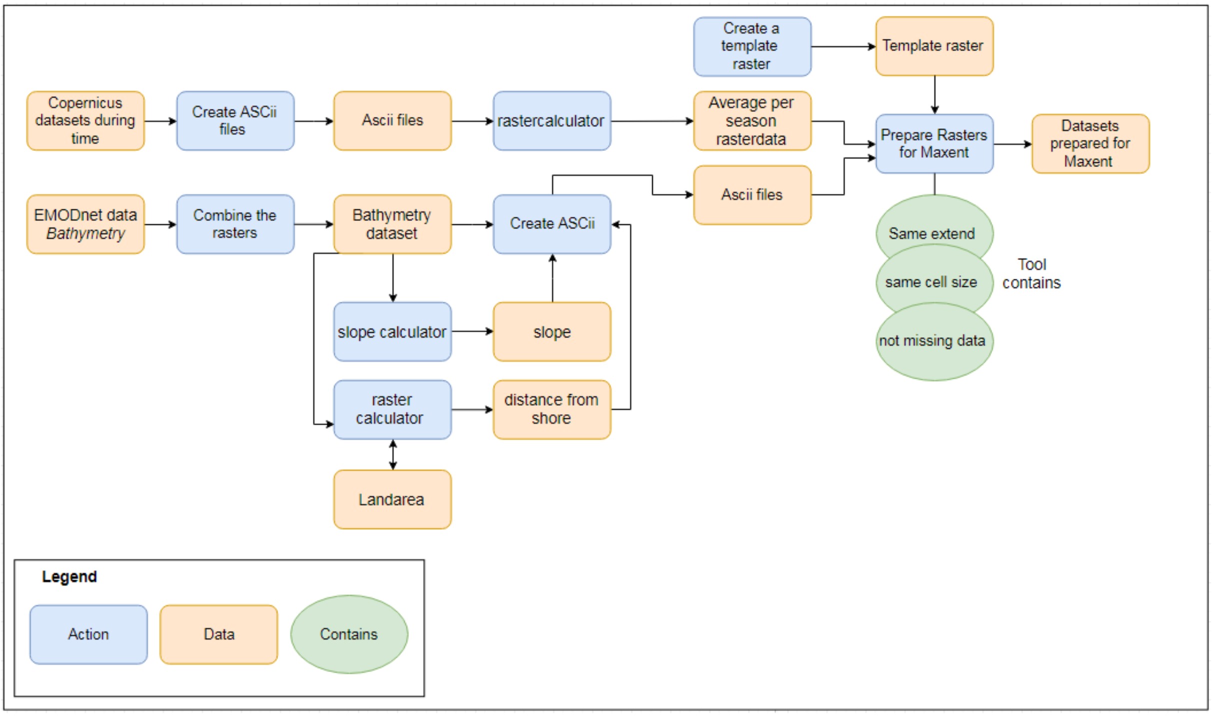
This research focusses on the favourable *temporal* habitat of the common bottlenose dolphin in the Aegean Sea. As described in Chapter 1 dolphins tend to show seasonal differences in favourable habitat (Firth et al., 1997). The research of Firth et al., (1997) showed those seasonable difference in the North of Scotland but since their research is comparable to this research, but in a different area and with a different SDM method, there is decided to keep this temporal division. So, to find out whether there are temporal differences between the different seasons  Maxent will be run in different temporal scenario's. These are the meteorological seasons, winter; December-February, spring; March-May, Summer; June-

August, Fall; September-November. The sightings, spread over the three years, are divided into these seasons resulting in one sighting .csv file per season.

### 5.1.2 Environmental and ecological variables

To get the environmental data ready to use within Maxent some steps have to be taken, of which the main important things are aligning the cells(size), extend and converting everything to ASCII. For this, there is a difference between the data retrieved from the Copernicus program and the data that is retrieved from EMODnet. In figure 6 the workflow that needs to be followed for the data preparation is depicted.

*Figure 4; flowchart of preparing data for Maxent within ArcGIS*




Since this research focusses on the favourable temporal habitat, divided into the four seasons the environmental and ecological factors need to be divided accordingly. Bathymetry, slope and distance from shore do not show significant changes during the seasons. The variables chlorophyll, nitrate, oxygen, phosphate, phytoplankton, salinity and temperature do show seasonal and even daily differences. To do come with one seasonal file spread over the three years (2016-2018, since the occurrence points are gathered in this timeframe) the middle date is taken from each month from each year to download, resulting in 9 datasets per variable per season. These datasets are used to calculate the seasonal average by using the raster calculator (Esri, 2019h) in ArcGIS; (date 1 + date 2 …… + date 9)/9, resulting in the seasonal average in three years.

As to be seen in figure 6 it is needed in this research to create an input template raster. This is created by using one of the environmental layers, seven out of nine of the used variables are retrieved from

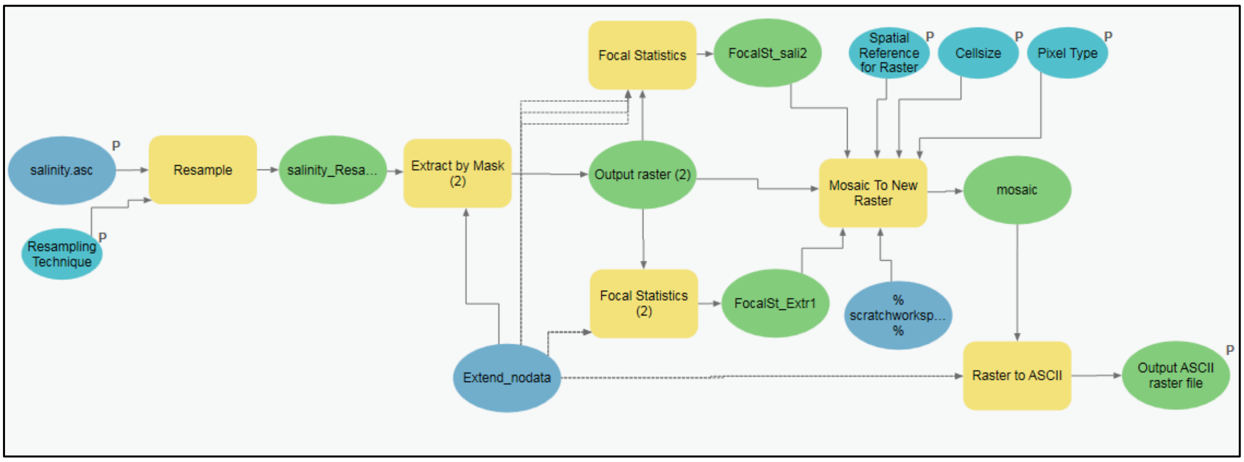
Copernicus, to make sure to keep most of the data as 'pure' as possible during pre-processing there is decided to keep the cell size of the Copernicus datasets as template. All of them have the same cell size so it does not matter which of these is used for the creation of the template raster file. The reclassify tool (Esri, 2019j) in ArcGIS is used to set all the values within this raster to zero, from now on it can be used as template raster.

Furthermore, a raster dataset is created containing only the land areas as numbers and the sea with NoData, this is done by using polygon to raster and then reclassifying the sea areas to NoData (Esri, 2019g, 2019j). This raster will be necessary for later stages of the research. By reclassifying the other way around, NoData to sea areas and land areas to NoData a file containing the sea areas is created.

As to be seen in Figure 6, all the dataset are available to download except for slope and distance from shore. These have to be created by using the bathymetry dataset. The distance from shore is created by using Euclidean distance with the following inputs; input raster or feature is a land cover shapefile with output cell size and extend from the template raster. The tool calculates for each cell the Euclidian distance to the closest source (Esri, 2019b). The Euclidean distance is calculated from the centre of the source cell to the centre of each of the surrounding cells (Esri, 2019m). Since dolphins do not follow straight gridlines this tool is considered the best option since it calculates the shortest distance from the centre of a cell to the nearest source, in this case land. The slope is calculated with the bathymetry dataset using the tool 'slope' within ArcGIS Pro. The input raster is the bathymetry dataset, output measurement is set on degree, the method is planar and the z factor is set on 1. This tool identifies the slope (gradient or steepness) from each cell of a raster (Esri, 2019l).

As to be seen in figure 6 there are actually two different paths for preparing the data to use in MaxEnt. The flow for Copernicus data and the flow for the bathymetry data. The flows come together within at the action prepare rasters for MaxEnt which is done within ArcGIS. The data that is used within MaxEnt needs to have the same extend, cell size and no missing data. The model that is used to properly prepare the data for MaxEnt is depicted below in Figure 7 and an explanation of the different tools can be found in Table 1.

*Figure 5; prepare rasters for Maxent*



Within this model, the data that needs to be used within MaxEnt is prepared for the program. The input datasets consist of an input template raster (named research_extend) and the input datasets (in this case only salinity.asc) that need to be prepared for MaxEnt. The output is an ASC raster file with exactly the same extend and cell size as the input template raster, without any missing data, reaching until the shoreline of the islands and with the 'land' data filtered out (obviously dolphins will not choose their favourable temporal habitat to be on land). The tools that are used within this model are described in table 1 below;

*Table 1; prepare rasters for Maxent (Figure 7) explanation*

| Tool | Explanation |
|---|---|
| Resample | The resample function changes the spatial resolution of the raster dataset and sets rules for aggregating or interpolating values across the new pixel sizes. The resampling size is set on nearest within this model since nearest returns the nearest cell value of the original raster. The other options (bilinear or cubic) will average values which might result in a loss of data or result in no data values, majority will only return the dominant value. The settings are left on default, except for the cell size which is based on the input template raster (Esri, 2019k). So, resampling is used to transform the cell size from the input raster(s). |
| Extract by mask | Extract by mask is used to extract the cells of the input raster(s) that corresponds to the area's defined by the input template raster (the mask). Extract by mask is used to remove land cells (islands) (Esri, 2019c). So, this tool makes sure the input raster(s) get the same extend as the template raster and filters out the data on land. This is the case for the datasets distance from shore, bathymetry and slope since these also have values on land. |
| Focal statistics | This tool fills up cells without data (Esri, 2019d). This is of importance for this research since the environmental data does not reach fully until the shore. Since the common bottlenose dolphin is mainly an inshore animal it is of importance that these areas are taken into account when using MaxEnt, since they will probably show a good fit for the dolphins.<br><br>Focal statistics calculates for each input cell location a statistic of the values within a specified neighbourhood around it (Esri, 2019d). The rectangle neighbourhood is used within this model, with the width being 3 and 5 (for focal statistics 2), the units type is set on cell and the statistics type is set on mean. There is decided to use mean as statistics type since this takes the |

| | |
|---|---|
| | average of the cells in the neighbourhood. The ignore NoData in calculations box is tapped on. Focal statistics is used twice, first with a search window fo 3x3 and the second one with a search window of 5x5. This allows to fill up smaller gaps with a less 'averaged' value than larger gaps. So, focal statistic is used to fill up cells with NoData. This makes sure the data reaches at least until the shoreline.<br><br>Since the cell size of the datasets in this research is approximately 4x4 km the focal statistics tool will address certain cells with a value that are actually partly on land. This is not considered as an issue since the data belonging to the on land cell is already filtered out previously by extract by mask. So, by executing these focal statistics the gaps between the datasets and the shoreline are filled up ensuring every bit of sea is used in the modelling, seeing the cells that are partly on land as sea instead of land. |
| Mosaic to new raster | Mosaic to new raster is executed to merge the three raster datasets into one new raster (Esri, 2019f). The order that is used for this tool is first the resampled file followed by the smaller focal statistics tool and last the bigger focal statistics tool. Within the mosaic operator the input rasters are; 1) resampled file, 2) focal statistics (3), 3) focal statistics (5). The Mosaic operator is set on first even as the mosaic colourmap mode. Cell size is set on 0, spatial reference is to be set, in this case, WGS 1984 and the pixel type is set to match the input raster. It is of importance that the pixel type is set to match the existing input raster dataset, if not a default is used which might give an incorrect output. In the case of this research, it is 32-bit float. |
| Raster to ASCII | This tool converts the datasets to an ASCII file representing the raster data (Esri, 2019i). Within this tool, it is of importance to put .asc in the output file since otherwise the file is saved as a .txt file. |

## 5.2 Settings within MaxEnt

Within MaxEnt there are many different settings that can be used and will give (slightly) different outputs within the results.

### 5.2.1 Regularization

Regularization is used to reduce overfitting of the Maxent model. Overfitting means that the model is overly fit to the specific input, and will not give any correct output when extrapolated. So regularization makes sure that the model is not fit too precisely. As described by Merow et al., 2013 it is possible to set the regularization coefficient to zero when you want your model to fit with all feature classes, but this should only be done when the number of features is small relative to the number of presences (Merow et al., 2013). Since this is not the case in this research the regularization parameter is left, as by default, at 1.0. Making it smaller will give a more localised (chance on more overfitting) output while making it bigger as 1.0 will give a more spread out and less localised output (Phillips, 2017).

### 5.2.2 Sampling bias

Maxent models are fit to the data assuming that every location, so every cell, within the research area is equally likely to be sampled (Merow et al., 2013). Within this research, this is not the case since the sampling took place around the island of Samos. When accounting for sampling bias the null hypothesis states the following; individuals are uniformly distributed in geographic space, the only reason they have been observed in particular locations is because those are the only places that were sampled, thus the prior distribution for the species occurrence is the sampling distribution (Merow et al., 2013). Since the search effort of the used presence datasets in this research is known, it gives the possibility to formulate the sampling bias models in geographical space. Typically there are two ways to deal with biased data, the first one is the biased prior method which is most widely used. In here the user provides an estimate of the relative search effort in each location of the landscape. This output reflects the assumption that the probability of observing an individual in a given location is based on the search effort there. There is also a biased background approach which is mainly used when it is not clear what the search effort has been for the occurrence data. This is not applicable for this research, for more information about this method see Merow et al., (2013) (Merow et al., 2013).

Within this research, there is decided to make use of a bias file indicating the area in which there has been surveyed for dolphins and in which it is known that the dolphins occur. This is done by laying a 50-kilometre buffer around the occurrence point per season. There is decided to use a 50-kilometre buffer since this is the approximate range the common bottlenose dolphin swims in the Mediterranean, based on the research by Gnone et al., (2011). The created bias grid will say that in this area the search effort is 1, and Maxent will only select background samples from this area (Fourcade, Engler, Rödder, & Secondi, 2014).

This file is created by first applying minimum bounding geometry on the occurrence points to create a polygon based on the occurrence points. The second step is to buffer (Esri, 2019a) this area with 50 kilometres as described above, after this the polygon is converted to a new raster (Esri, 2019g). It is important to know which areas are the biased areas in this raster and which areas are land and sea. This is done by mosaicking the newly created raster with the sea area raster and the land area raster which are described earlier in Chapter 5.1 (Esri, 2019f). The input is first sea, then the bias file and then land with the mosaic operator set on last. After this, you can classify this new raster in three classes (sea, land

and bias) and reclassifying sea and land to NODATA resulting in a file with only the biased area with cell value 1 (Esri, 2019j).

So, this file is indicating where there has been searched for dolphins and indicates where they at least live in the research area. By using this file in Maxent, Maxent will select the background point from this area, instead of assuming that all areas are sampled evenly.

### 5.2.3 Output types

The last important setting choice for Maxent is the type of output. Within Maxent there are three types of output, raw, cumulative and logistic. The outputs are related monotonically, so rank-based metrics for model fit will be identical to each other, if in one file something changes, this will change in all the outputs. The main point in difference within the three maps is the scaling, and with the scaling comes the visualisation (Merow et al., 2013). The raw output values are probabilities, between 0 and 1 such that the sum over all cells used during training is 1, a characteristic of this is that vales get really small, such as 0.00000037 which is quite difficult to work with. For the cumulative output, the value at a grid cell is the sum of all the probabilities of all grid cells with no higher probability than the grid cell, times 100. (Phillips, Dudík, & Schapire, n.d.). The logistic output transforms the raw output, this looks at the presence at 'average' presence locations and that logistic output can be interpreted as the probability of presence (Merow et al., 2013). Merow et al., (2013) advice to use the raw output since this output does not rely on post-processing assumptions. Cumulative output can be used when interpretations relate to omission rates when for example drawing boundaries. Cumulative values can be problematic when small differences exist between a large subset of cells because the cells will be ranked from highest to lowest in spite of potentially negligible differences. And to avoid logistic output whenever possible since this output heavily relies on again assumptions that are made post-processing. This since you have to state the probability of presence at average presence locations as a number in Maxent, which is by default 0.5. This can have big consequences on the predicted probabilities assigned to each location (Merow et al., 2013). Within this research there is decided to do use the cumulative output since this format is more easily interpreted showing the likelihood of occurrence and the problem proposed above does not apply to this research due to the size of the research area and the goal of the research, it is not intended to show small areal difference but more to indicate the wider patterns in likelihood of occurrence.

### 5.2.4 Other settings

Some other settings in Maxent are that the test data percentage is set on 30% which is a widely used default within MaxEnt and the repetitions is set on 50, this to make sure that if there are any outliers or strange outputs, these are smoothened out.

Since there is no sighting data available for the winter season, this due to the simple reason of harsh weather, a projection is done with the winter data variables projected on the summer occurrence data.

## 5.3 Data processing after Maxent

Maxent creates an output report by itself which results in the fact that this output can be seen as the direct results. No processing of any data needed anymore. The only thing that does need some additional processing is the comparison between the output maps per season. To compare the eventual likelihood of occurrence output maps that will be generated by Maxent the Minus tool in ArcGIS pro is used. This tool subtracts the value of the second input raster from the value of the first input raster on a cell by cell basis (Esri, 2019e). This results in an output raster showing the differences per season, did the likelihood of occurrence increase or decrease?

# 6. Results

Within this chapter, the summarised results of the 50 runs per season in Maxent for the common bottlenose dolphin are discussed. The results of the different seasons are discussed following the three subchapters in this chapter; model performance, likelihood of occurrence and variable importance.

## 6.1 Performance of the model

The model performance is measured by means of the Area Under the Curve (AUC), see Chapter 2.5 for more information. An AUC value of 0.5 indicates that the performance of the model is no better than random, values closer to 1.0 indicate better model performance. As Manna, Ronchetti, & Sar, (2016) describe in their article for interpretation of AUC values, models with values from 0.7 and higher are considered with good discrimination ability, in which 0.7-0.8 is of moderate discrimination, 0.8-0.9 is of good discrimination and 0.9 to 1 is of excellent discrimination (Manna et al., 2016).

### 6.1.1 Model performance results

Table 2 shows the AUC of the four different seasons. The Maxent model for summer scores best with an AUC of 0.902, followed by spring with an AUC of 0.803 and as lower two winter with 0.794 and fall as the lowest is fall with 0.780.

Table 2; area under the curve per season

| **Season** | **AUC** |
|---|---|
| **Spring** | 0.803 |
| **Summer** | 0.902 |
| **Fall** | 0.780 |
| **Winter** | 0.794 |

## 6.2 Likelihood of occurrence of the common bottlenose dolphin

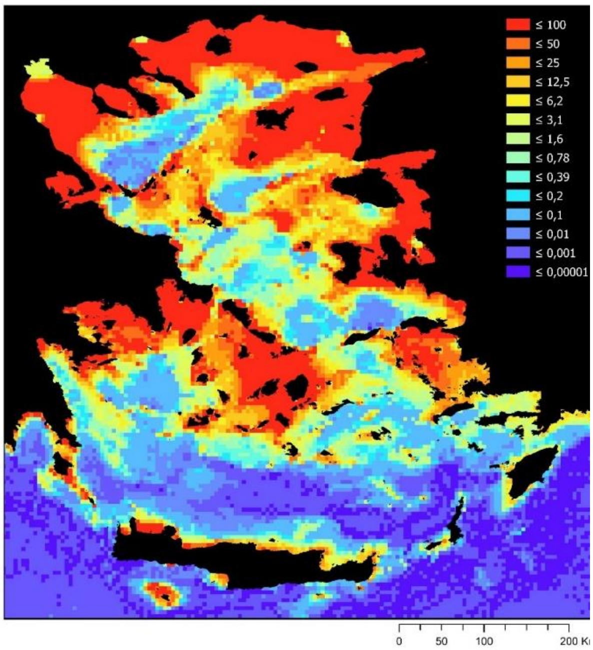
Map 5-8, showing the output of Maxent are cumulative, as described in Chapter 6.2.3. The cumulative output assigns, at a location, the sum of all raw values less than or equal to the raw value for that location and rescales this to lie between 0 and 100 (Merow et al., 2013). The output can be interpreted as suitable above a threshold in the approximate range of 1-20, this is depending on the level of predicted omission that is acceptable for the application. It is also possible to look at the predicted omission rate: when setting a cumulative threshold of *c*, the resulting binary prediction would have omission rate c% on samples drawn from the Maxent distribution itself, and we can predict a similar omission rate for samples drawn from the species distribution. So, when putting the threshold on for example 30 to predict a presence/absence surface will omit approximately 30% of presences (Merow et al., 2013; Phillips, 2017)

### 6.2.1 Likelihood of occurrence results

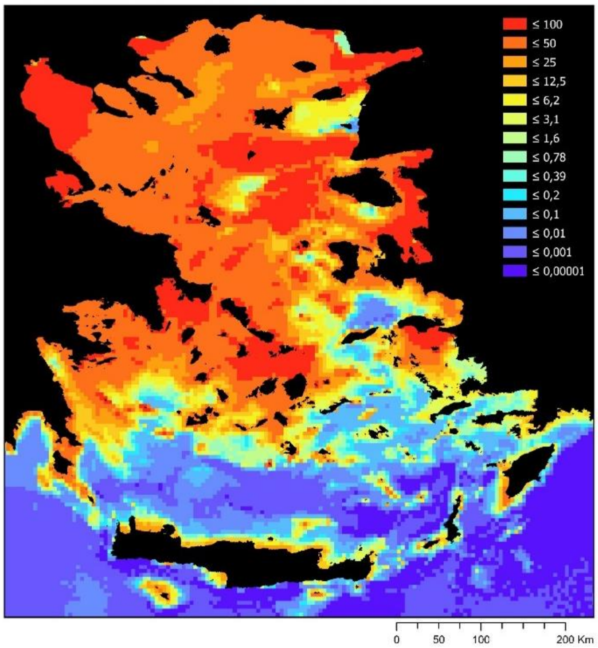
Maps 5 to 8 display the likelihood of occurrence of the common bottlenose dolphin per season. The redder the colours are on this map the higher the likelihood of dolphin occurrence, while the bluer the colours become the less likelihood of occurrence. This output can be interpreted as suitable above a threshold in the approximate range of 1-20. So, orange to red is considered suitable. The logarithmic scale is used to also show the differences in the less suitable classes (Phillips, 2017). To emphasise the differences between the four seasons maps 9-12 depict a comparison of the likelihood of occurrence per two seasons. The red colours mean that the dolphin likelihood increased and the bluer the colours become indicate that the likelihood of occurrence decreased. These are calculated by subtracting season 2 from season 1. A value of -100 in the output means that in season 1 there must be a likelihood of occurrence of 100 and in season two a likelihood of occurrence of 0, so the likelihood of occurrence decreased with 100. A value of +100 indicates the previous but exactly the other way around, so the likelihood of occurrence increased with 100.

In map 5 the output of the Maxent model for spring is depicted, in this map the main parts with higher likelihood of occurrence are located in the middle of the Aegean Sea between the far point of Athens until approximately the island of Naxos, around the entire Northern coastline stretching out towards the entire Eastern coastline and some bigger parts around Crete. Map 6 shows the output of the likelihood of occurrence for summer. In this map, the main parts with higher occurrence are located in the middle of the Aegean Sea again between the far point of Athens until approximately the island of Naxos, a bit up towards the more eastern (middle) coast and around the northern parts of the Aegean Sea. Next to the areas described above with the likelihood in the 50-100 category, almost the entire inner part of the Aegean Sea has quite a high likelihood of occurrence falling in the 25-50 likelihood of occurrence rate.

*Map 5; likelihood of occurrence of T. truncatus, spring 2016-2018*

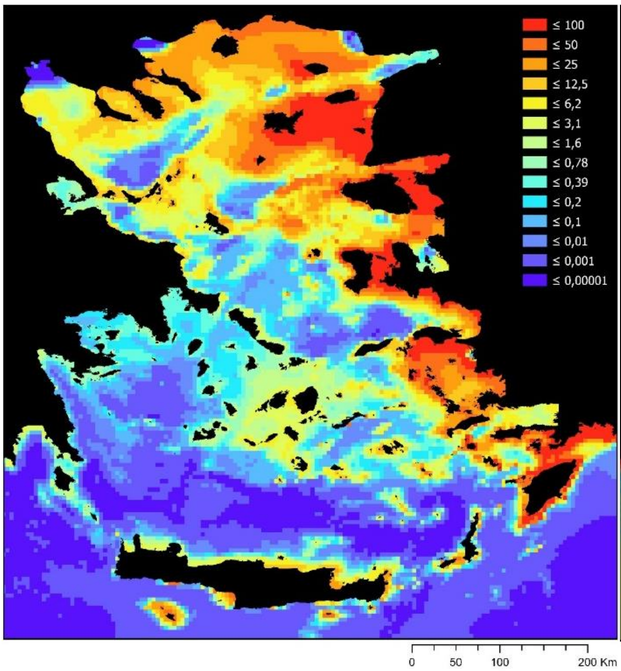



*Map 6; likelihood of occurrence of T. truncatus, summer 2016-2018*

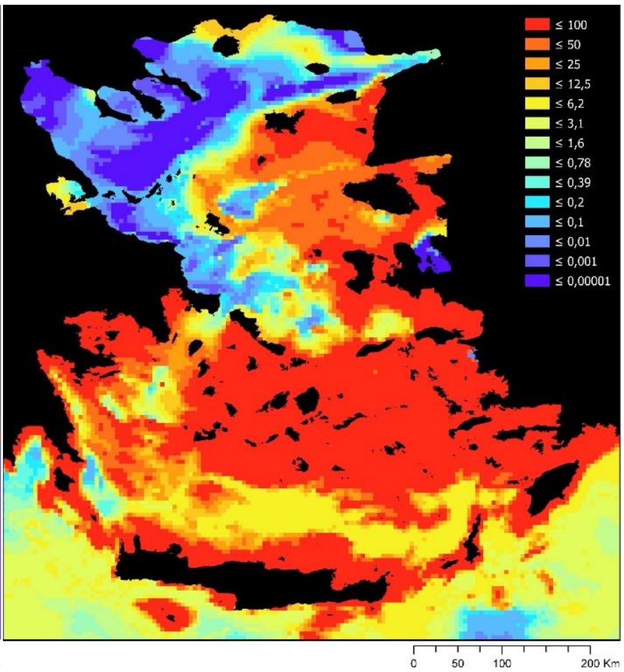


Fall is depicted in map 7, the area with a high likelihood of occurrence is located primarily along the eastern coast of the Aegean Sea stretching out towards the Northern part. Map 8 shows winter, the areas with a higher likelihood of occurrence are located throughout the entire middle part of the Aegean Sea, stretching from Athens until the coast of Turkey, going a bit more North along the Turkish coast are also big parts with a high likelihood of occurrence. Around the island of Crete the likelihood of occurrence is also considered high.

*Map 3; likelihood of occurrence of T. truncatus, fall 2016-2018*

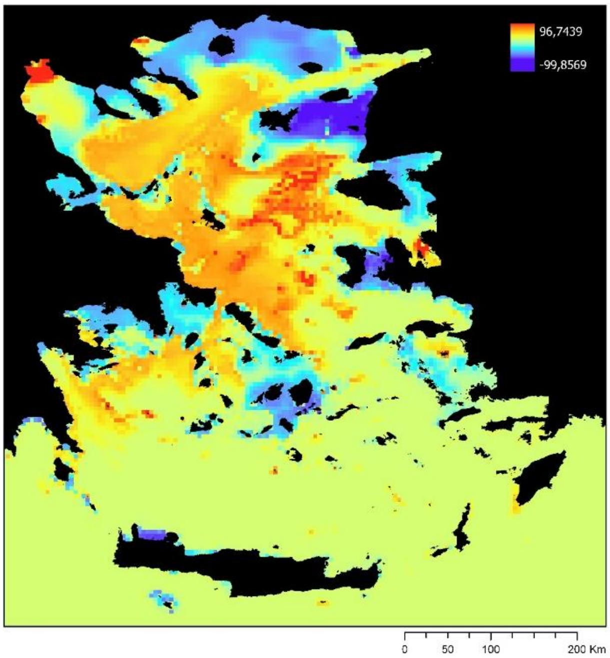



*Map 8; likelihood of occurrence of T. truncatus, winter 2016-2018*

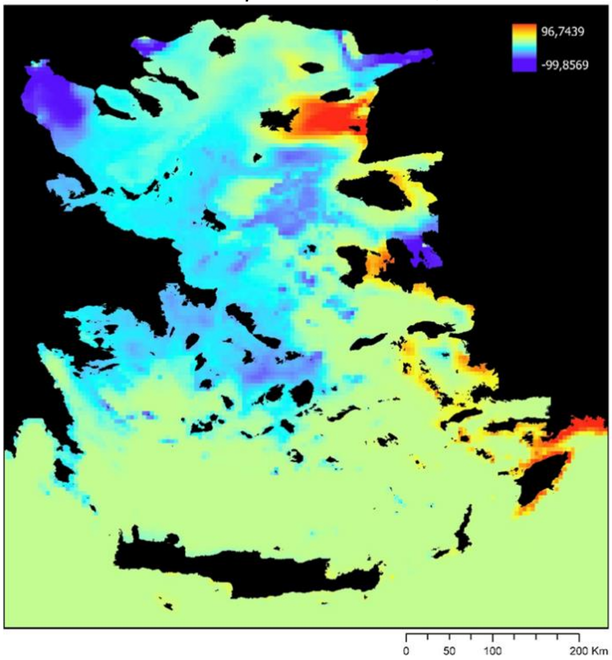



When looking at the comparison between summer and spring (map 9) it gets clear that the northern and eastern border areas of the Aegean Sea decreased in occurrence likelihood, while the middle part of the Aegean Sea increased significantly in occurrence likelihood. Map 10 depicts the comparison between fall and summer, almost the entire Aegean Sea shows a blue colour which means that the likelihood of dolphins decreased or did not change that much (the more greener colours), except for some areas along the Turkish coast which show a significant increase in the likelihood of occurrence. In Map 11 the comparison between winter and fall is depicted in which almost the entire middle stroke and areas around the island of Crete show an increased likelihood of occurrence. Parts in the Northern part of the Aegean Sea show a decrease in the likelihood of occurrence. The last comparison is between spring and winter (map 12), especially the more Eastern middle area shows a big decrease in the likelihood of occurrence, the more Northern parts show an increase.

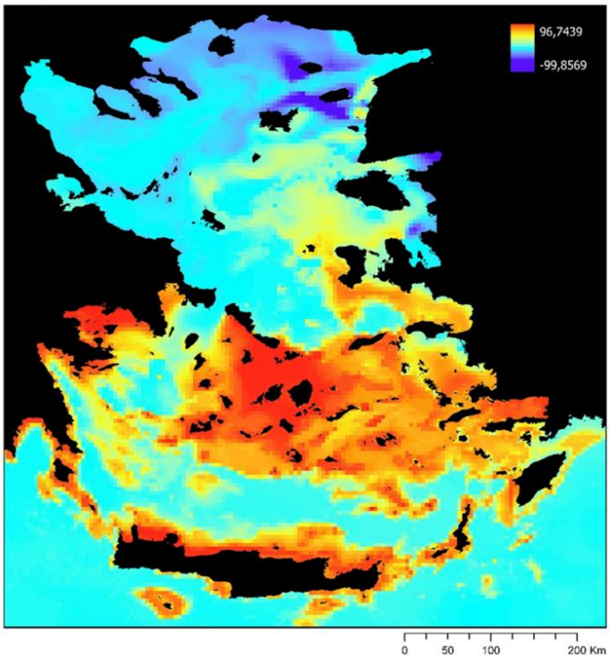
*Map 9; comparison of the likelihood of occurrence between summer and spring, 2016-2018*



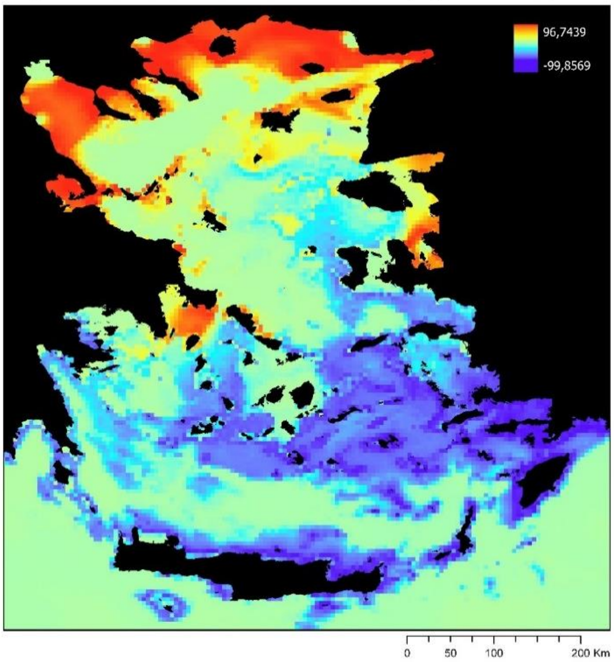

*Map 10; comparison of the likelihood of occurrence between fall and summer, 2016-2018*




*Map 11; comparison of the likelihood of occurrence between winter and fall, 2016-2018*




*Map 12; comparison of the likelihood of occurrence between spring and winter, 2016-2018*



## 6.3 Variable importance for the likelihood of occurrence

The variable importance will be described by means of the response curves, the percentual contribution per variable and the jackknife test of variable importance.

### 6.3.1 Response curves of the variables

Within this research, ten different variables are used to explain and predict the favourable temporal habitat in the Aegean Sea. The response curves in Figures 7 and 8 on the next two pages show how each environmental variable affects the Maxent prediction. The curves show how the predicted probability of presence changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. The curves show the marginal effect of changing exactly one variable whereas the model may take advantage of sets of variables changing together. The curves show the mean response of the 50 replicates Maxent runs and the mean +/- one standard deviation (this is the blue colour), only the red line will be described.

The vertical axe is the cumulative output going from 0 to 100. The horizontal axe shows a range of the corresponding variable in the study area, for temperature in spring this is for example 17.9 to 18.8 degrees. Some of the response curves of the variables are changing barely over the seasons, this is especially the case with the bathymetry. The response curve of chlorophyll is the same in spring fall and winter but differs in summer. Whit very low nitrate values spring differs from fall and winter and stays stable in summer. The response curve of phosphate in summer differs a lot from the other seasons. When looking at oxygen spring and summer have quite the same response curve, same goes for fall and winter but these two do differ from each other. The response curves from fall and winter stay low around the starting value of oxygen and start going upward during the middle value of oxygen while the response curves of spring and summer stay more stable during the entire value of oxygen. This is the same case for the response curves of salinity and temperature, although at temperature it is the other way around, fall and winter stay stable and summer and spring have a high response curve at the colder temperature and the curve goes down when the temperature gets warmer. The response curve of phytoplankton stays quite stable during spring, going slightly down when the phytoplankton goes up. In Summer the response curve goes from 100 to 0 around in the middle, fall and winter go slightly up when phytoplankton increases. The response curves for slope during the seasons stays stable, except for spring where the response curve goes from 100 to 0 in the very beginning of the graph.

Overall the response curves of fall and winter do very much overlap, while there is quite some difference between spring/summer/fall+winter.

Figure 6; variable importance spring to winter

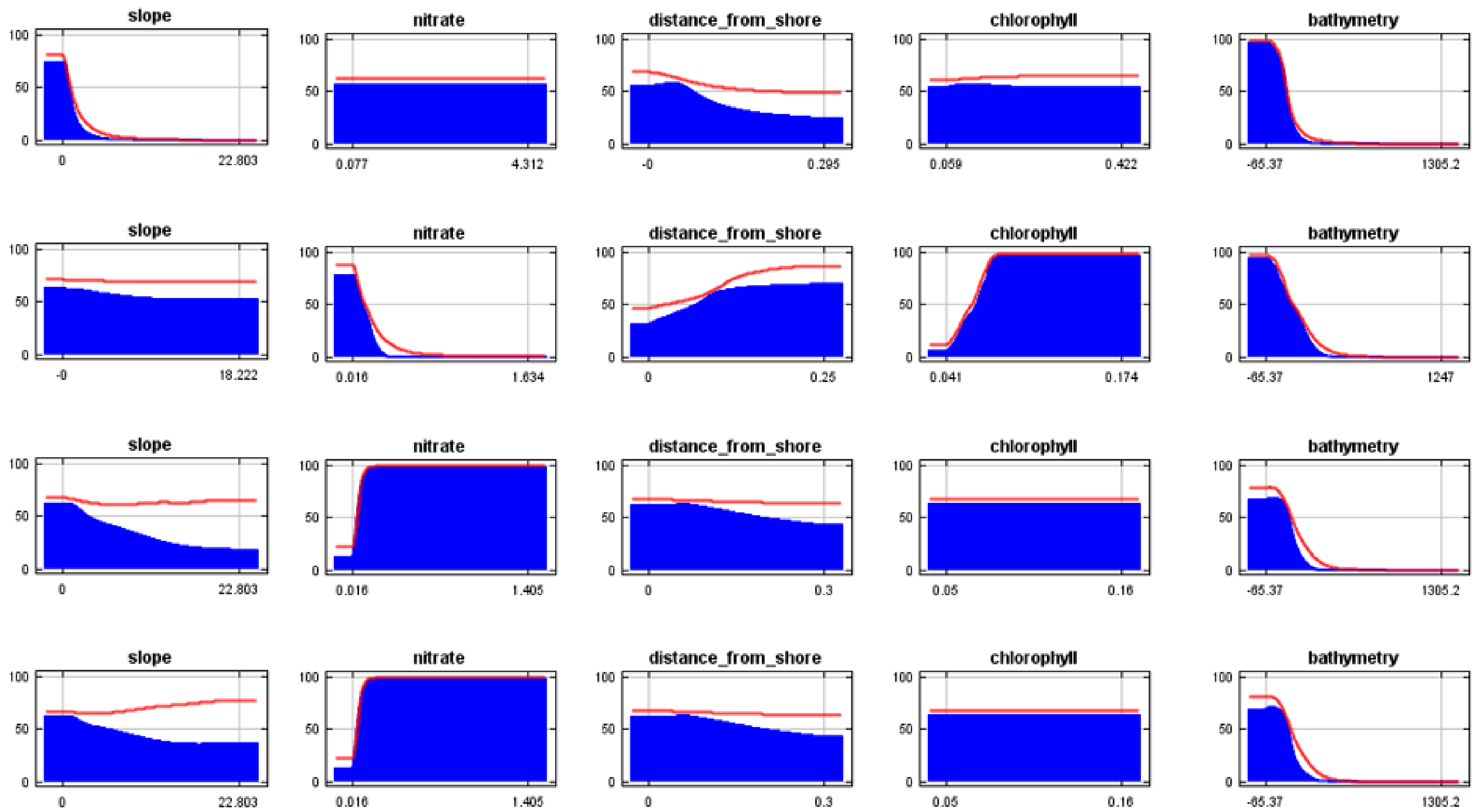
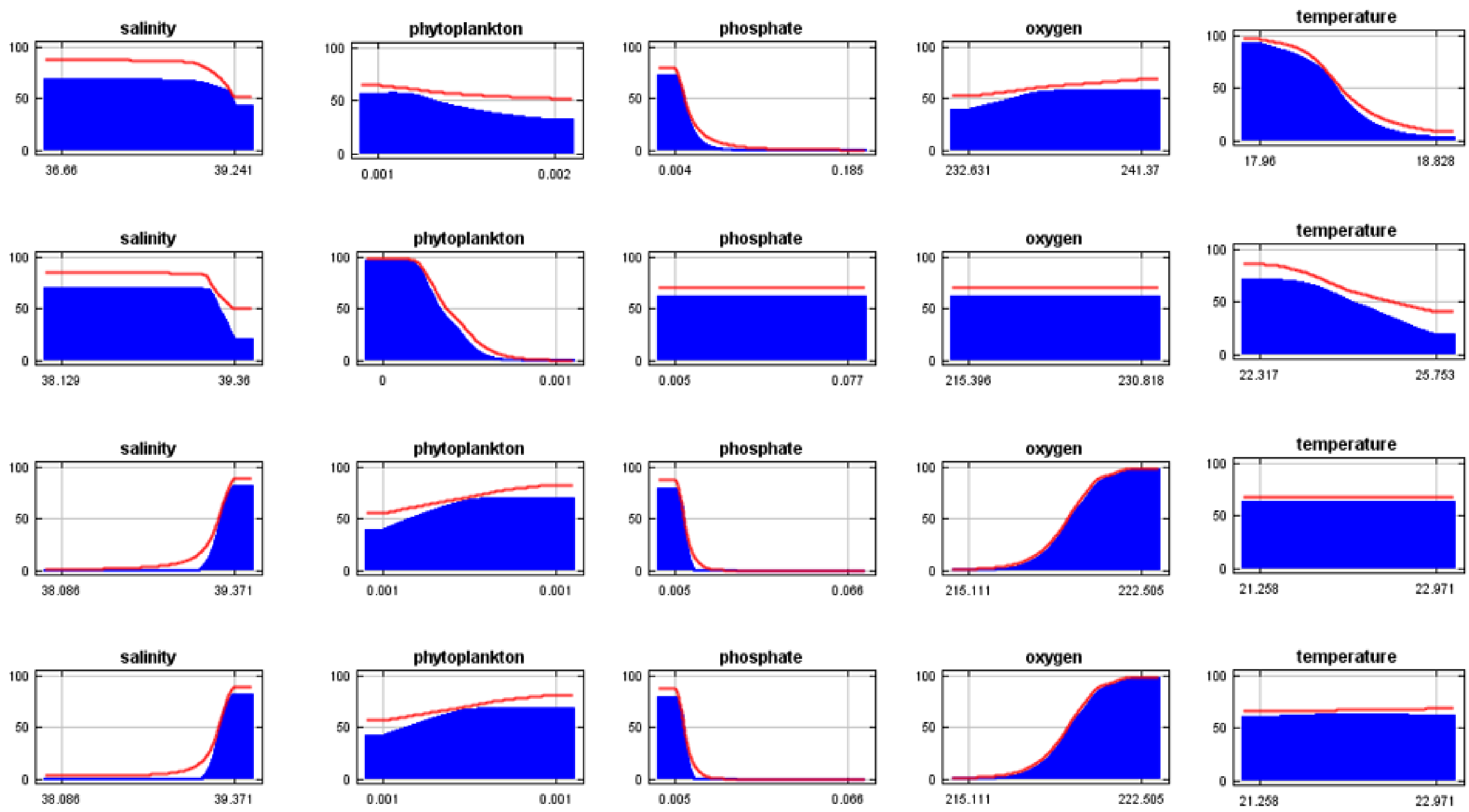


*Figure 7; variable importance spring to winter*



### 6.3.2 Percentual contribution of the variables

Table 3 gives estimates of the relative contribution of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted (change the order). The model is reevaluated on the permuted data, and the resulting drop in training AUC's shown in the table normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated. Values shown are averages over replicate runs. These values depend on the particular path that the Maxent code uses to get the optimal solution, a different algorithm could get to the same solution via a different path, resulting in different percentual contributions (Phillips, 2017).

Bathymetry is in every season the most important variable with in spring, fall and winter values over 60% and in summer a value of 35,7%. The second and third most important values are for summer fall and winter the same, oxygen (summer 23,5%, fall 11,7% and winter 12,4%) followed by phytoplankton as the third variable in these seasons (summer 16,6%, fall 10,9% and winter 10,8%). In spring the second variable is slope with a contribution of 16,9% and the third one is temperature with 4,8%, as described before this differs from the other seasons. In summer the 4$^{th}$ variable is also still of importance, chlorophyll with 9,1%, in the other seasons chlorophyll is of 0% importance. The other variables are spread out over the lower percentages and can be viewed in table 3.

*Table 3; variable importance over the seasons; percent contribution*

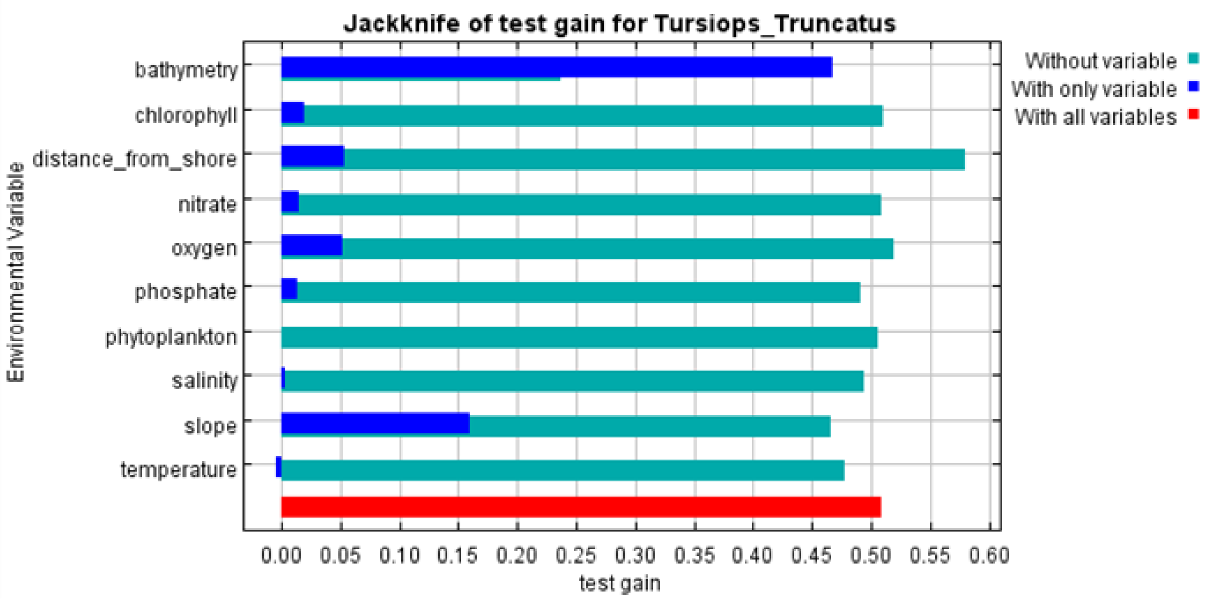
| ***Spring*** | | ***Summer*** | | ***Fall*** | | ***Winter*** | |
|---|---|---|---|---|---|---|---|
| ***Variable*** | ***%*** | ***Variable*** | ***%*** | ***Variable*** | ***%*** | ***Variable*** | ***%*** |
| *Bathymetry* | *65,8* | *Bathymetry* | *35,7* | *Bathymetry* | *62,3* | *Bathymetry* | *61,7* |
| *Slope* | *16,9* | *Oxygen* | *23,5* | *Oxygen* | *11,7* | *Oxygen* | *12,4* |
| *Temperature* | *4,8* | *Phytoplankton* | *16,6* | *Phytoplankton* | *10,9* | *Phytoplankton* | *10,8* |
| *Oxygen* | *4,3* | *Chlorophyll* | *9,1* | *Nitrate* | *4,1* | *Nitrate* | *4,1* |
| *Distance from shore* | *4,2* | *Distance from shore* | *6,3* | *Phosphate* | *3,3* | *Salinity* | *3,8* |
| *Phosphate* | *3,5* | *Nitrate* | *4,9* | *Salinity* | *3,2* | *Phosphate* | *3,2* |
| *Salinity* | *0,3* | *Temperature* | *2,9* | *Slope* | *2,5* | *Distance from shore* | *1,9* |
| *Phytoplankton* | *0,2* | *Salinity* | *0,7* | *Distance from Shore* | *1,8* | *Slope* | *1,8* |
| *Nitrate* | *0* | *Slope* | *0,5* | *Temperature* | *0,2* | *Temperature* | *0,4* |
| *Chlorophyll* | *0* | *Phosphate* | *0* | *Chlorophyll* | *0* | *Chlorophyll* | *0* |

### 6.3.3 Jackknife test of variable importance

A different way to look at which variables are important within the model is looking at the jackknife test of importance. Within this test each variable is excluded in turn and a model is made with the remaining variables. A model is created with each variable on itself and a model is made again using all variables as done previously (Phillips, 2017). This way does not have the problem with the values being dependent on the specific path Maxent chooses to use. Figures 10-13 display the jackknife test of variable importance over the four seasons. The darker blue line shows the environmental variable with the highest gain when used in isolation. The light blue line shows which variables decreases the gain the most when omitted.

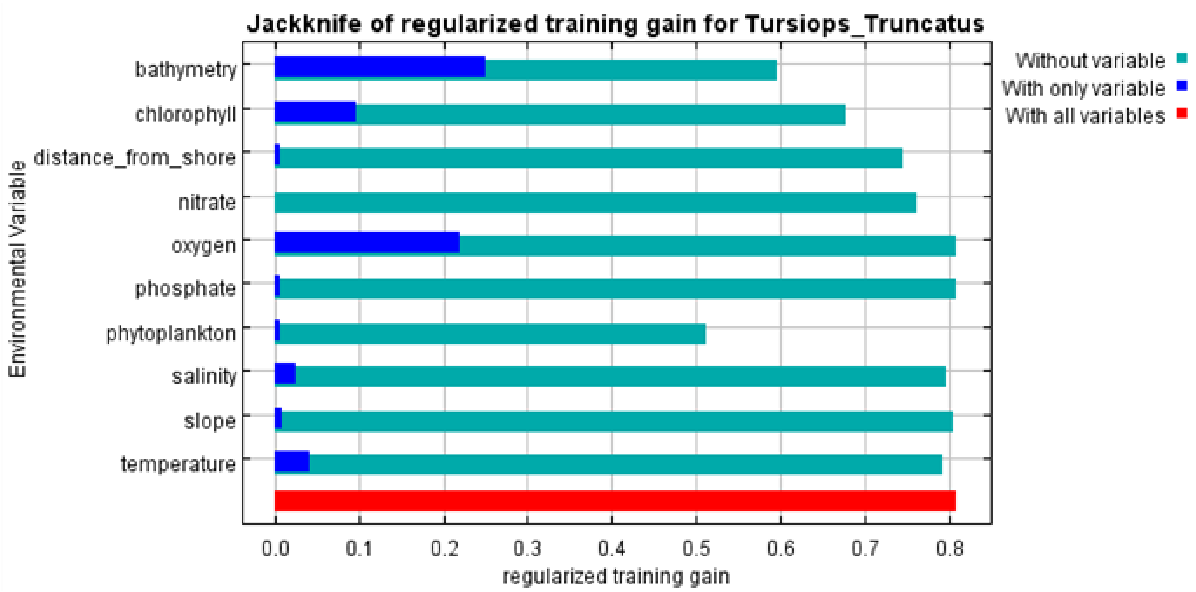
Within all the four seasons bathymetry is the variable that has the highest gain when used in isolation. It is also the variable that decreases the gain the most when omitted which can be seen by means of the light blue line. As to be seen in figure 10 in spring slope also has a big gain when used in isolation, the other variables stay quite low in this season.

*Figure 8; jackknife test spring*



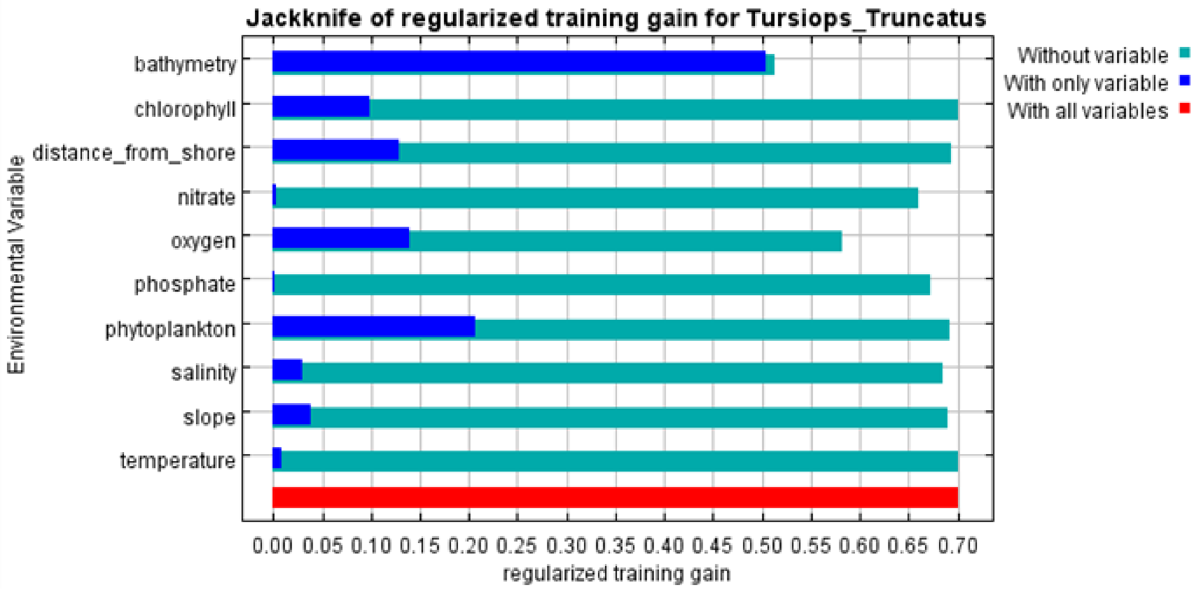

In summer, next to bathymetry, oxygen and chlorophyll both have quite a high gain compared to the others. The variable in summer which decreases the gain the most is phytoplankton. This can be seen in Figure 11.

*Figure 9; jackknife test summer*



In fall (Figure 12) phytoplankton is the variable with the highest gain after bathymetry, followed by oxygen, distance from shore and chlorophyll. The variable that decreases the gain the most when omitted is bathymetry.

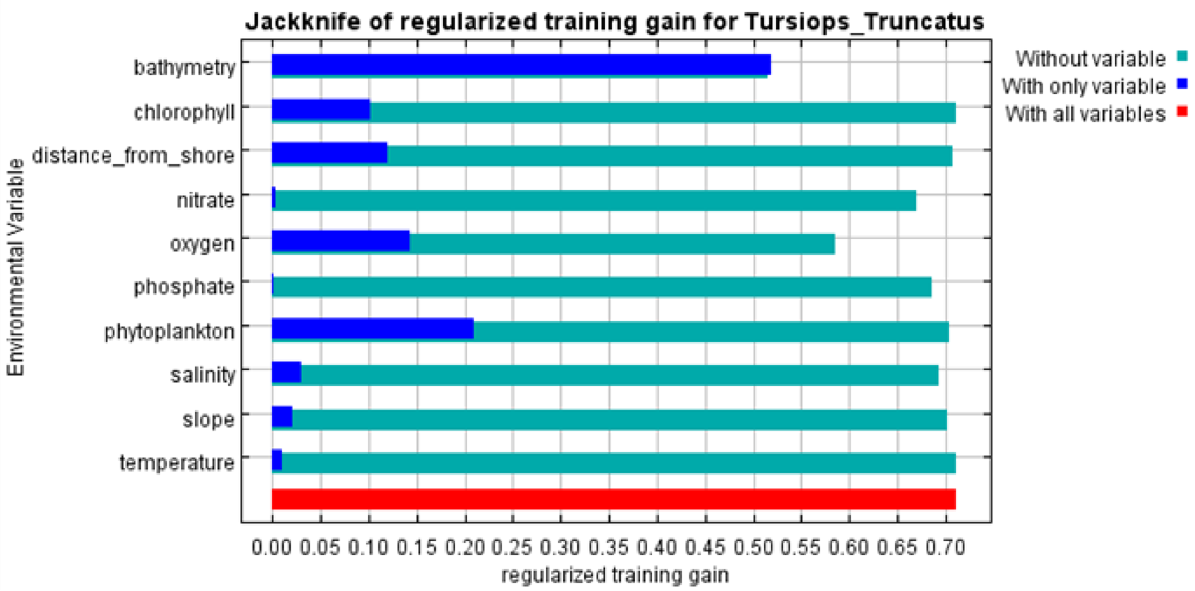
*Figure 10; jackknife test fall*




In winter it is also bathymetry followed by phytoplankton, oxygen, distance from shore and chlorophyll with the highest gain, also here, the variable that decreases the gain the most is bathymetry, Figure 13.

.

*Figure 11; jackknife test winter*



# 7. Conclusion

The main research objective of this thesis was to indicate the favourable temporal habitat of the common bottlenose dolphin, to reach this main objective the following research question was used;

*To what extent can ecological and environmental factors explain the favourable temporal habitat of Tursiops truncatus in the Aegean Sea?*

Maxent has been used to model the favourable temporal habitat of the common bottlenose dolphin based on occurrence points, gathered during the years 2016 to 2018, and by using ten different variables. These variables are (ordered alphabetically); bathymetry, chlorophyll, distance from shore, nitrate, oxygen, phosphate, phytoplankton, salinity, slope and temperature. In means of variable importance, it was expected that slope, distance from shore and bathymetry would be the most important variables since the common bottlenose dolphin is known to be an inshore animal. When looking at the results bathymetry is in all four seasons the most important variable. But the distance from shore and slope score relatively low in all seasons, except for spring in which slope is the second main contributor to the likelihood of occurrence of the common bottlenose dolphin in the Aegean Sea. Distance from shore scores low in all four seasons percentual contributions but when looking at the likelihood of occurrence maps the areas with the higher likelihood of occurrence are located for all the four seasons, among other areas, around the coastlines of the Aegean Sea. As stated above it got clear in the results that bathymetry, in every season, is the most important factor to determine the likelihood of occurrence for the dolphins across the Aegean Sea. Furthermore mainly oxygen and phytoplankton are considered as factors with a high contribution to the likelihood of occurrence of the dolphins. An exception is the spring season in which slope is the second most contributing factor. A likely explanation for this is the seasonal difference between the used values except for bathymetry, slope and distance from shore. Logically these do not differ in value during the different seasons, the other values do and form a likely explanation for the seasonal differences in the likelihood of occurrence. The seasons of fall and winter are very comparable in variable importance. There are more differences between the other seasons. Since there were no occurrence points available in winter a projection with fall is used for the winter months, probably resulting in the comparable season.

What is interesting to see is that there are seasonal differences in the likelihood of occurrence of the dolphins. The World Nature Conservation Union (IUCN) stated that the common bottlenose dolphin has its regular habitat in the middle of the Aegean Sea in between the smaller islands that are common there. The seasonal output of this research shows that especially winter has this same likelihood of occurrence in between the smaller islands, matching almost entirely with the results of the IUCN research (Reeves & Notabartolo Di Sciara, 2006). The other three seasons show also quite a high likelihood of occurrence between the islands in the middle but also indicate different places with a higher likelihood of occurrence such as the more northern parts of the Aegean Sea and the border area with Turkey.

Concluding it can be said that ecological and environmental factors can explain the favourable temporal habitat of the common bottlenose dolphin. As described above, within this research the main important factor is the bathymetry, the favourable habitat differs significantly per season in which every season has its own distribution of how important the different variables are. Fall and winter are in these very comparable where the other seasons differ more in variable response.  Furthermore, the main aim of this research has been achieved since there is a clear identification of the likelihood of occurrence of the common bottlenose dolphin per season.

# 8. Discussion and Recommendations

The main aim and purpose of this study was to make a first step towards more flexible conservation management by determining which factors make the favourable temporal habitat of the common bottlenose dolphin in the Aegean Sea and indicating where this favourable temporal habitat is during the year (seasons). To reach this main aim there has been made use of a habitat suitability model, named Maxent. The decision for this program is based on a literature research wherein Maxent shows the best fit with the aim and the available data (Jane Elith et al., 2006). Maxent is used to analyse several environmental and ecological background datasets together with occurrence points of the common bottlenose dolphin gathered in three years (2016-2018). The main findings of this study are that there are seasonal areal differences in the likelihood of occurrence between the four seasons, varying across the Aegean Sea. Fall and winter are in variable importance very comparable as seasons, while summer and spring differ more from the others, except in that bathymetry is the most important factor in all, as stated above. From this research, we can say the ecological and environmental factors can explain the favourable temporal habitat of the common bottlenose dolphin, although additional research is needed to test the results in this research.

For this research the decision has been made to use Maxent to analyse the favourable temporal habitat, although Maxent is one of the top programs in SDMs (Jane Elith et al., 2006) the program has its limitations, and especially different setting might and probably will result in a different output (Merow et al., 2013). As described by Merow et al., 2013 there are several settings that will influence the results of Maxent. The first one is the used background data, the results would, of course, be different when using other environmental variables. This research has also been quite explorative in seeing which variables are of influence on the favourable seasonal habitat of the common bottlenose dolphin. Using other factors will most certainly give other (interesting) results. Furthermore, there are inputs and options for feature classes, regularization, sampling bias, types of output and evaluation (Merow et al., 2013; Phillips, 2017). All of these will influence the output results of Maxent. The output format for this research is cumulative since this enhances the visibility in showing the likelihood of occurrence, but the raw output does give a more ‘pure’ output form and would probably give less biased results since it least depends on pre-processing decisions that are made (Merow et al., 2013). Furthermore, there can be decided to change the random test percentage, replicates and replicated run type. The results probably won’t differ significantly but they’ll still be presented differently than the case is now. Also, the maxent algorithm can choose a different path to come to the same results so the percentual variable importance will differ when running the program again with the same settings and input data.

The biggest limitation of this research is the limited amount of occurrence points, the entire winter period does not have any occurrence points so there needed to be made use of a projection over the fall with the winter environmental factors. This is still considered as a proper way of analysing also the winter month but with actual occurrence points, it would have been a better and more reliable result (Phillips, 2017). Furthermore, there is no availability of absence data for the common bottlenose dolphin in the research area, which would really enhance the research when 100% certain absence is absence (Jane Elith et al., 2011; Pearson, Raxworthy, Nakamura, & Peterson, 2007; Wintle et al., 2005). This is extremely difficult to research within marine life due to the simple reason of our visibility underwater and the wide movement of the dolphins themselves.

The area in which the sightings of this research are carried out is relatively small in comparison to the Aegean Sea since the sightings are only carried out around mainly the south of Samos, Lipsi and Fourni. Even though Maxent is able to work with smaller sample sizes and areas (Fourcade et al., 2014) a bigger more spread out sample size would be preferable. It would be interesting to test Maxent on a smaller area between and around these islands, for now, this was outside the timeframe of this research but it most certainly is a recommendation for further research.

As discussed above one of the settings that can be chosen in Maxent is the bias file (Merow et al., 2013) Since Maxent evaluates each cell as evenly sighted there needed to be made use of a bias file which fixed the extremely biased area in which the sightings are done. But, the creation of this bias file is done by creating a buffer of 50 km around the point of sighting. This is based on several articles in which the daily travel range is researched, but none of these articles researched the dolphins in specifically this area (Bearzi et al., 2011; Gnone et al., 2011). Since the main travel distance of the dolphins in the other areas was around 50 kilometres this is also considered as the average travel distance of these dolphins. Another limitation in this way of thinking is that the buffer around the dolphin is a circle, in the real situation the dolphin might not travel to all different directions, or around some of the islands, this led to a biased bias file. Furthermore, there are several other options to deal with bias in the used data (Fourcade et al., 2014; Merow et al., 2013). And which even can be combined. It would improve the research to make more use of the several options to deal with bias.

The used data from Copernicus also has its limitations. Interpolating of the data was necessary since the environmental datasets used from Copernicus do not stretch exactly until the coastline. Since the common bottlenose is considered an inshore animal it is of importance that these datasets do reach until the shoreline, so interpolating was necessary. Another interpolation technique might give different results and we are never certain of the exact values in the interpolated area. Also, the cell size is of importance. Since the cell size of the environmental variables is quite big (4-kilometre x 4-kilometre) and the occurrence points are all located quite close together. The background points are sampled in a relatively small area. This problem is fixed with the 50-kilometre bias file discussed above. Another point is that the Copernicus data that is used is averaged per month. The data is available per day and using daily data and averaging that over the seasons would improve the research. Since the timeframe of this research was limited and the scope of this research is not to biologically determine which values of a certain variable are preferred by the dolphins, the monthly mean is considered sufficient. What would be interesting is to see on which days the sighting took place and use background points from these days and then extrapolate/project it within the season. What would be perfect is a model per day indicating the favourable habitat, so the shift during the year gets clear.

The recommendations for future research are already discussed in the limitations of the current research above. Some other recommendations for future research are; creating a Maxent model for the fish the common bottlenose dolphin eats the most, and compare/use this as environmental/ecological factor in the Maxent model used in this research. This will probably be a big factor contributing to the likelihood of occurrence of the common bottlenose dolphin.

An addition to this is including the fishing vessels as a factor in this model, they dump the fish remains in the seas which attract dolphins and is of influence on where they are. Also, other disturbing or constraining factors can be mapped within Maxent when they are known. It is also interesting to look at each individual factor more closely and analyse the relationship between occurrence and variable with other statistical programs. This would contribute in a very positive way to this research. It will also be relevant to know which specific range per value the dolphins prefer. When this is known a more specified model could be created. The best factor to add to the model is the z value of the dolphin and analyse this on different depth levels. This could be done with for example tracking some dolphins with sensors and analyse their movement and behaviour in relation to depth, location and time spend on location/depth.

# 9. Bibliography

Archipelagos. (2018). About us - Archipelagos. Retrieved September 4, 2018, from http://archipelago.gr/en/about-us/

Bearzi, G., Agazzi, S., Bonizzoni, S., Costa, M., & Azzellino, A. (2008). Dolphins in a bottle: abundance, residency patterns and conservation of bottlenose dolphinsTursiops truncatus in the semi-closed eutrophic Amvrakikos Gulf, Greece. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*(2), 130–146. https://doi.org/10.1002/aqc.843

Bearzi, G., Bonizzoni, S., & Gonzalvo, J. (2011). Mid-distance movements of common bottlenose dolphins in the coastal waters of Greece. *Journal of Ethology*, *29*(2), 369–374. https://doi.org/10.1007/s10164-010-0245-x

Bearzi, G., Fortuna, C. M., & Reeves, R. R. (2009). Ecology and conservation of common bottlenose dolphins Tursiops truncatus in the Mediterranean Sea. *Mammal Review*, *39*(2), 92–123. https://doi.org/10.1111/j.1365-2907.2008.00133.x

Beaumont, L. J., Hughes, L., & Poulsen, M. (2005). Prediciting species distributions:use of climatic parameters in BIOCLIM and its impact on predictions of species’ current and future distributions. *Ecological Modelling*, *186*(2), 251–270. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2005.01.030

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). bioclim : the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, *20*(1), 1–9. https://doi.org/10.1111/ddi.12144

Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, *2*(6), 667–680. https://doi.org/10.1007/BF00051966

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models - Ronald Christensen - Google Boeken*. Springer Science & Business Media. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=rQg2sgPo2gQC&oi=fnd&pg=PR3&dq=Christensen,+Ronald+(2002).+Plane+Answers+to+Complex+Questions:+The+Theory+of+Linear+Models+(Third+ed.).+New+York:+Springer.+ISBN+0-387-95361-2.&ots=25UwlrC5et&sig=wPSNhe-WoOe2jokX4C

Copernicus. (2018c). Copernicus - Marine environment monitoring service. Retrieved December 3, 2018, from http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_BIO_006_014

Copernicus. (2018a). Copernicus - Marine environment monitoring service. Retrieved November 30, 2018, from http://marine.copernicus.eu/about-us/about-eu-copernicus/

Copernicus. (2018b). Copernicus - Marine environment monitoring service. Retrieved December 3, 2018, from http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_PHY_006_013

Czybulka, D., & Bosecke, T. (2006). Marine Protected Areas in the EEZ in light of international and European Community law — Legal basis and aspects of implementation. In *Progress in Marine Conservation in Europe* (pp. 27–46). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-33291-X_3

de Stephanis, R., Cornulier, T., Verborgh, P., Salazar Sierra, J., Gimeno, N., & Guinet, C. (2008). Summer spatial distribution of cetaceans in the Strait of Gibraltar in relation to the oceanographic context. *Marine Ecology Progress Series*, *353*, 275–288. https://doi.org/10.3354/meps07164

Edgar, G. J., Russ, G. R., & Babcock, R. C. (2007). *Marine Ecology*. (S. D. Connell & B. M. Gillanders, Eds.). Oxford University Press.

Elith, J., Graham, C. H., & Elith, J. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, *32*. https://doi.org/10.1111/j.1600-0587.2008.05505.x

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., … E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

EMODnet. (2018). Data products - EMODnet Bathymetry. Retrieved December 3, 2018, from http://www.emodnet-bathymetry.eu/data-products

Esri. (2019a). Buffer—Help | ArcGIS Desktop. Retrieved February 28, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/analysis/buffer.htm

Esri. (2019b). Euclidean Distance—Help | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/euclidean-distance.htm

Esri. (2019c). Extract by Mask—Help | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/extract-by-mask.htm

Esri. (2019d). Focal Statistics—Help | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/focal-statistics.htm

Esri. (2019e). Minus—Help | ArcGIS for Desktop. Retrieved February 28, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/minus.htm

Esri. (2019f). Mosaic To New Raster—Help | ArcGIS for Desktop. Retrieved February 27, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/mosaic-to-new-raster.htm

Esri. (2019g). Polygon to Raster—Conversion toolbox | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/conversion/polygon-to-raster.htm

Esri. (2019h). Raster Calculator—Help | ArcGIS for Desktop. Retrieved February 27, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/raster-calculator.htm

Esri. (2019i). Raster to ASCII—Help | ArcGIS for Desktop. Retrieved February 27, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/conversion-toolbox/raster-to-ascii.htm

Esri. (2019j). Reclassify—Help | ArcGIS for Desktop. Retrieved February 27, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/reclassify.htm

Esri. (2019k). Resample function—ArcGIS Pro | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/help/data/imagery/resample-function.htm

Esri. (2019l). Slope—Help | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/slope.htm

Esri. (2019m). Understanding Euclidean distance analysis—Help | ArcGIS Desktop. Retrieved February 27, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/understanding-euclidean-distance-analysis.htm

Firth, M., Wilson, B., Thompson, P. M., & Hammond, P. S. (1997). *Habitat Use by Bottlenose Dolphins: Seasonal Distribution and Stratified Movement Patterns in the*. *Source: Journal of Applied Ecology* (Vol. 34). Retrieved from https://www-jstor-org.proxy.library.uu.nl/stable/pdf/2405254.pdf?refreqid=excelsior%3A93b89f3e7583a5b720ed031e4fc3d1b3

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS One*, *9*(5), e97122.

https://doi.org/10.1371/journal.pone.0097122

Friedman, J., Hastie, T., & Tibshirani, R. (2000). *ADDITIVE LOGISTIC REGRESSION: A STATISTICAL VIEW OF BOOSTING*. *The Annals of Statistics* (Vol. 28). Retrieved from https://projecteuclid-org.proxy.library.uu.nl/download/pdf_1/euclid.aos/1016218223

Gnone, G., Bellingeri, M., Dhermain, F., Dupraz, F., Nuti, S., Bedocchi, D., … Wurtz, M. (2011). Distribution, abundance, and movements of the bottlenose dolphin (Tursiops truncatus) in the Pelagos Sanctuary MPA (north-west Mediterranean Sea). *Aquatic Conservation: Marine and Freshwater Ecosystems*, *21*(4), 372–388. https://doi.org/10.1002/aqc.1191

Gomez, J. J., & Cassini, M. H. (2015). Environmental predictors of habitat suitability and biogeographical range of Franciscana dolphins (Pontoporia blainvillei). *Global Ecology and Conservation*, *3*, 90–99. https://doi.org/10.1016/J.GECCO.2014.11.007

Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2014). Maxent is not a presence-absence method: a comment on Thibaud *et al* . *Methods in Ecology and Evolution*, *5*(11), 1192–1197. https://doi.org/10.1111/2041-210X.12252

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, *157*(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*(5), 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*(2), 142–152. https://doi.org/10.1016/J.ECOLMODEL.2006.05.017

Hornsby, F., McDonald, T., Balmer, B., Speakman, T., Mullin, K., Rosel, P., … Schwacke, L. H. (2017). Using salinity to identify common bottlenose dolphin habitat in Barataria Bay, Louisiana, USA. *Endangered Species Research*, *33*, 181–192. https://doi.org/10.3354/esr00807

International Union for Conservation of Nature and Natural Resources. (2000). The IUCN red list of threatened species. Retrieved August 29, 2018, from http://www.iucnredlist.org/search

Jefferson, M. T., Leatherwood, S., & Webber, M. A. (1993). FAO species identification guide. Marine mammals of the world., 320.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498–507. https://doi.org/10.1111/j.1466-8238.2011.00683.x

Katsanevakis, S., Levin, N., Coll, M., Giakoumi, S., Shkedi, D., Mackelworth, P., … Hanadiv Rothschild, Y. (2014). Marine conservation challenges in an era of economic crisis and geopolitical instability: The case of the Mediterranean Sea. https://doi.org/10.1016/j.marpol.2014.07.013

Manna, G. La, Ronchetti, F., & Sar, G. (2016). Predicting common bottlenose dolphin habitat preference to dynamically adapt management measures from a Marine Spatial Planning perspective. https://doi.org/10.1016/j.ocecoaman.2016.07.004

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Natoli, A., Birkun, A., Aguilar, A., Lopez, A., & Hoelzel, A. R. (2005). Habitat structure and the dispersal of male and female bottlenose dolphins (Tursiops truncatus). *Proceedings. Biological Sciences*, *272*(1569), 1217–1226. https://doi.org/10.1098/rspb.2005.3076

Natoli, A., Cañadas, A., Vaquero, C., Politi, E., Fernandez-Navarro, P., & Hoelzel, A. R. (2008). Conservation genetics of the short-beaked common dolphin (Delphinus delphis) in the Mediterranean Sea and in the eastern North Atlantic Ocean. *Conservation Genetics*, *9*(6), 1479–1487. https://doi.org/10.1007/s10592-007-9481-1

Notarbartolo-Di-Sciara, G., & Birkun, A. (2010). *Conseving whales, dolphins and porpoises in the Mediterranean and Black Seas*.

Pabst, D. A., Rommel, S. A., McLellan, W. A., Williams, T. M., & Rowles, T. K. (1995). Thermoregulation of the intra-abdominal testes of the bottlenose dolphin (Tursiops truncatus) during exercise. *The Journal of Experimental Biology*, *198*(Pt 1), 221–226. https://doi.org/10.1242/jeb.01559

Pearson, R. G. (2010). *Network of Conservation Educators & Practitioners Species' Distribution Modeling for Conservation Educators and Practitioners. in Conservation* (Vol. 3). Retrieved from http://ncep.amnh.org/linc

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*, 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Phillips, S. J. (2017). *A Brief Tutorial on Maxent*. Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/.

Phillips, S. J., Dudík, M., & Schapire, R. E. (n.d.). Maxent software for modeling species niches and distributions (version 3.4.1).

Reeves, R., & Notabartolo Di Sciara, G. (2006). *The status and distribution of cetaceans in the Black Sea and Mediterranean Sea. IUCN Centre for Mediterranean Cooperation*. Retrieved from https://portals.iucn.org/library/sites/library/files/documents/2006-068.pdf

Shane, S. H., Wells, R. S., & Wursig, B. (1986). Ecology, behaviour and social organisation of the bottlenose dolphin: a review. *Marine Mammal Science*, *2*(1), 34–63. https://doi.org/10.1111/j.1748-7692.1986.tb00026.x

Townsend Peterson, A., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, *30*(4), 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x

United Nations. (1992). *Convention on biological diversity*. Retrieved from https://www.cbd.int/doc/legal/cbd-en.pdf

Wintle, B. A., Elith, J., & Potts, J. M. (2005). Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology*, *30*(7), 719–738. https://doi.org/10.1111/j.1442-9993.2005.01514.x