National Park Service
U.S. Department of the Interior

South Florida Natural Resources Center
Everglades National Park

ECOLOGICAL MODEL REPORT

SFNRC Technical Series
2022:1

Generalized Additive Modeling of Alligator Nest Sighting for Resource Management in Everglades National Park

Generalized Additive Modeling of Alligator Nest Sighting for Resource Management in Everglades National Park

ECOLOGICAL MODEL REPORT
SFNRC Technical Series 2022:1

South Florida Natural Resources Center
Everglades National Park
Homestead, Florida

National Park Service
U.S. Department of the Interior

Cover photograph by Mark Parry, Everglades National Park

TABLE OF CONTENTS

CONTRIBUTING AUTHORS

Dilip Shinde[1], Leonard Pearlstine[1,] Amy Nail[2], and Mark Parry[1]

[1]South Florida Natural Resources Center, Everglades National Park, 950 N. Krome Avenue, Homestead, FL 33030-4443
[2]Honestat Statistics & Analytics, LLC, Cary, NC, 27513-2115

Comments and Questions:

Dilip_Shinde@nps.gov

ACKNOWLEDGMENTS

We sincerely thank all contributors who supported the development of this model, including the field researchers for their tireless efforts collecting and analyzing the data. The authors acknowledge the Everglades Depth Estimation Network (EDEN) project and the U.S. Geological Survey for providing the water depths for the purpose of this research and report. We especially thank all reviewers and contributors who provided feedback on this work including Donatto Surratt, Jimi Sadle, and Antonia Florio (Everglades National Park, National Park Service); Laura D'Acunto (U.S. Geological Survey); Laura A. Brandt (U.S. Fish and Wildlife Service); Christa Zweig (South Florida Water Management District); and Frank J. Mazzotti (University of Florida). Sincere thanks are extended to Troy Mullins for help with GIS data processing and Gregg Reynolds for providing help with R programming during this work (Everglades National Park, National Park Service). Michelle Collier (Everglades National Park, National Park Service) was editor and desktop publisher of the report.

THIS REPORT SHOULD BE CITED AS:

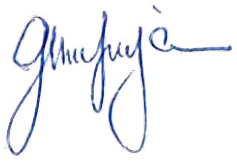
Shinde, D., L.G. Pearlstine, A. Nail, and M. Parry. 2022. Generalized additive modeling of alligator nest sighting for resource management in Everglades National Park. National Park Service, Everglades National Park, South Florida Natural Resources Center, Homestead, FL. Ecological Model Report. SFNRC Technical Series 2022:1. 49 pp.

FOREWORD

This report, "Generalized additive modeling of alligator nest sighting for resource management in Everglades National Park", describes a model designed to evaluate effects of changes in hydrology and land cover on nesting of alligators (Alligator mississippiensis). The American alligator is a keystone species within Everglades marsh systems whose activity structures the landscape creating dry season refugia and increasing the diversity of habitat and species. Alligators are dependent on spatial and temporal patterns of water fluctuations that affect courtship and mating, nesting, and habitat use. The Modified Water Deliveries Project and the Comprehensive Everglades Restoration Plan are programs for reversing past environmental degradation and restoring habitat for wildlife, such as the alligator. Ecological modeling tools that can guide the planning efforts and simulate the effects of restoration are of keen interest to natural resource managers.

The report describes the use of 24 years of alligator nest monitoring survey data in the development of a model of the probability that a nest will be built in a given grid cell within Everglades National Park in a given year. The major input requirements for the model include the habitat, hydrological, meteorological, and landscape data that provide the predictor variables. The model can evaluate the influence of alternative water management operations on alligator nesting.

Examination of the probability of alligator nesting spatially during a year provides insight to any limiting hydrologic conditions that contribute to a poor nesting probability, thus inhibiting successful alligator reproduction. This model enhances ENP's ability to preserve and protect natural resources while engaging in restoration efforts in southern Florida. The model and its application as shown in this report demonstrates our efforts to protect wildlife resources and restore the Everglades, while underscoring the need for continued monitoring of alligators.



G. Melodie Naja
Director
South Florida Natural Resources Center
Everglades National Park

September 2022

1 INTRODUCTION

Everglades restoration focuses on correcting hydrologic conditions in southern Florida after people diverted and drained the region for development starting in the late 1800s. The main beneficiaries of this enormous environmental restoration project are the varied and unique plants and animals that make up the Everglades ecosystem. The Modified Water Deliveries Project (MWDP) (USACE 1992) and the Comprehensive Everglades Restoration Plan (CERP) are two of the most significant Everglades restoration programs (https://evergladesrestoration.gov/, accessed May 2020). Projects in these programs range from large construction projects, such as canal removal and road reconstruction, to water regulation plans that consider the water needs of all stakeholders. The CERP is being implemented using an applied science strategy framework (Ogden and Davis 1999, Ogden et al. 2003) that links alternative plan evaluation with ecological models, monitoring, and research to provide more effective scientific support to Everglades restoration and allow for adaptive management. Ecological modeling tools can evaluate the effects of restoration on key components of the Everglades ecosystem, such as alligators (*Alligator mississippiensis*), as well as other keystone and indicator species.

The American alligator has been studied as part of the Everglades ecosystem for decades. It is a keystone species within Everglades marsh systems (Mazzotti and Brandt 1994) because its activity structures the landscape resulting in increased diversity of habitat that is critical to many wildlife populations for nesting, resting, or foraging sites (Craighead 1968, Kushlan 1974, Deitz and Jackson 1979, Kushlan and Kushlan 1980, Hall and Meier 1993). For example, Everglades fishes concentrated in the remaining dry season pools are readily available for wading bird forage, making the holes alligators excavate for themselves a critical driver of wading bird nesting success and spatial distribution (Frederick et al. 2009). Alligators may also serve as "nest protectors" for wading birds (Burtner and Fredrick 2017).

Alligators are dependent on spatial and temporal patterns of water fluctuations affecting courtship and mating, nesting, and habitat use. Alligator abundance, nesting effort, nest success, growth, survival, and body condition serve as indicators of the health of the Everglades marsh system. Because of this, the alligator population in Everglades National Park (ENP) has been monitored closely and park researchers have conducted annual Systematic Reconnaissance Flights (SRF) to monitor alligator nesting since the 1980s. Combining this rich dataset with that of the equally rich hydrologic dataset that exists for the park provides an opportunity to find relationships between the two.

Changes in water management have influenced the pattern of water levels in the southern Everglades, causing unnatural flooding of alligator nests (Kushlan and Jacobsen 1990). Hydrological alterations of the system have reduced prey availability corresponding to reduced growth, survival, and reproduction of alligators (Mazzotti et al. 2007). Increased drought frequency and depth of drying have reduced suitability of southern marl prairie and rocky glades habitats and the number of alligators occupying alligator holes (Mazzotti et al. 2009, Fujisaki et al. 2012), limiting important nesting resources. Reproduction is a vital contribution to the persistence of a species in a region. Given that successful alligator nesting is dependent on hydrologic conditions and that sufficient data on alligator reproduction and habitat use as well as hydrologic data exists for the Everglades region, an ecological model designed to test the effects of restoration alternatives on alligator nesting would be of keen interest to natural resource managers, restoration, and conservation planners.

Earlier modeling efforts such as the Alligator Production Suitability Index Model (APSI; Shinde et al. 2014) and other habitat suitability models of alligators (Rice et al. 2004, Palmer et al. 2004 and Newsom et al. 1987) based on expert knowledge, judgment, and some empirical data, provide a deterministic response (0-1 index) of productivity and habitat suitability over the spatial domain grid cells. Higher scores from these models indicate better conditions, but the relationship between a specific score and expected number of nests and their success has not been established (RECOVER 2014). The index provided by these models represents neither measurable quantities nor probabilities of clearly defined events, and model validations are consequently qualitative. As such, these models do not inform the uncertainty associated with the index or represent actual alligator production values.

We determined that the richest, most comprehensive dataset— the annual SRF alligator nest survey data— can be used to develop a statistical model of the probability of a nest being built as a function of variables describing hydrology, habitat, and meteorology. This statistical model could also be validated and tested by SRF data that were withheld from the model-building process and by SRF data recorded in years following the time span of the data used for model-development. Since the statistical model would provide a probability distribution, the uncertainty of the predicted probabilities of a nest being built would also be quantified.

Moreover, the statistical model of the probability of a nest being built could be a function of the same input variables used for the habitat, breeding potential, courtship and mating, and nest building indices in APSI. This means the results could represent cumulative effects of all variables and alligator breeding-cycle stages on the probability that a nest is built in a spatial cell during a year. Since this model could be a function of hydrological variables in the breeding

potential, courtship and mating, and nest-building periods, it would also accomplish the objective to be achieved by the APSI model: to allow natural resource managers in charge of programs such as the MWDP and the CERP to link alternative plans to their effects on alligator habitat suitability. Though both models are supported by expert opinion, this model represents an improvement over the APSI because: 1) all model decisions (e.g., functional forms) are based on rigorous analysis of the appropriate data, 2) as a statistical model, it naturally allows quantification of uncertainty, and 3) since the output is the probability of a specific event that is routinely observed, it can be validated using validation and test datasets and future-year data.

1.1 Objective

The objective of this report is to describe how data from the annual SRF nest surveys was used to develop a model of the probability that a nest will be built in a given grid cell within ENP in a given year, as a function of predictor variables similar to those used for the habitat, breeding potential, courtship and mating, and nest-building components of the APSI, as well as variables describing the structure of the landscape and meteorology.

Section 2 describes the SRF nest survey data which provides the response variable in the model, as well as the habitat, hydrological, meteorological, and landscape data that provides the predictor variables. It also gives the rationale for considering each predictor variable as a candidate in the model. Section 3 serves the purpose of the typical "Methods" section of most scientific journals, describing the model-building methods. Section 4 serves the purpose of the "Results" section, giving the details of each model-building decision as the methods of Section 3 are applied. Section 5 gives an assessment of the final model selected using the withheld test-set data, and Section 6 uses the model to assess different scenarios.

2 DATA

2.1 Systematic Reconnaissance Flights for alligator nest sighting

Alligator nest systematic reconnaissance flights (SRFs) are carried out annually along well-established and defined latitudinal transects spaced two kilometers apart (Figure 1). The search pattern resulting from this method provides approximately 25% survey coverage of ENP's alligator nesting habitat. Flights are initiated when nest construction and egg deposition for most or all nests is expected to be complete in order to observe the maximum number of nests. The date for nest completion varies annually but SRFs usually commence the first week of July.

The alligator SRFs were initiated in 1985 as a cost-effective tool to detect landscape-level change in alligator reproduction effort and success within ENP, particularly in response to measurable hydrological change (Fleming 1991, Dalrymple 2001, Ugarte 2006, Parry and Bass 2009). The SRFs were expanded in 1992 to include all hydrological basins (adding East Slough, Rocky Glades, Taylor Slough, Long Pine Key, and Panhandle to Northeast-, Upper-, and Lower-Shark Slough basins; Figure 1) in ENP and currently comprise a continuous dataset. The SRFs provide a relative nesting effort yearly within ENP (i.e., to compare with previous years), but not an absolute number of nests occurring in ENP.

Flights are conducted according to standardized protocols as described by Fleming (1991) and Ugarte (2006). Doors are removed from the helicopter and one observer occupies the front left seat while the second observer occupies the right rear seat. The pilot maintains a flight speed of 50 knots, average elevation of 200-300 feet (60-90 m), and keeps the aircraft on the transect centerline (Figure 1). Observers look for nests or indications of likely nesting activity (e.g., heavily used trails or alligator ponds) within 250 m both north and south of the transect centerline, resulting in a 500 m strip width. When a likely nest location is identified, the pilot is directed to more closely investigate the areas as necessary, and then return to the centerline before proceeding. When a nest is observed, the aircraft hovers as near as possible without disturbing nest materials while the location is marked on a global positioning system and recorded on data sheets.

2.1.1 Space-time resolution and coverage

Each observation is a year-grid cell combination, or cell-year and is coded as 1, if at least 1 nest was observed in that cell that year, or as 0 if no nest was observed. In the SRF domain (Figure 1), there are 2,332 grid cells with dimension 400 m (horizontal: east-west) x 500 m (vertical: north-south). The data span the years 1992-2015, for a total of 24 years and 55,968 observations. The 500 m vertical width of the grid cells was chosen because SRF survey observers record nests 250 m to either side (north and south) of the aircraft as it flies along east-west transects. The 400 m horizontal width of the grid cell was chosen so that the east-west dimensions of the cells line up (Appendix: I) with the EDEN (Everglades Depth Estimation Network; http://sofia.usgs.gov/eden/index.php) hydrology data layers.

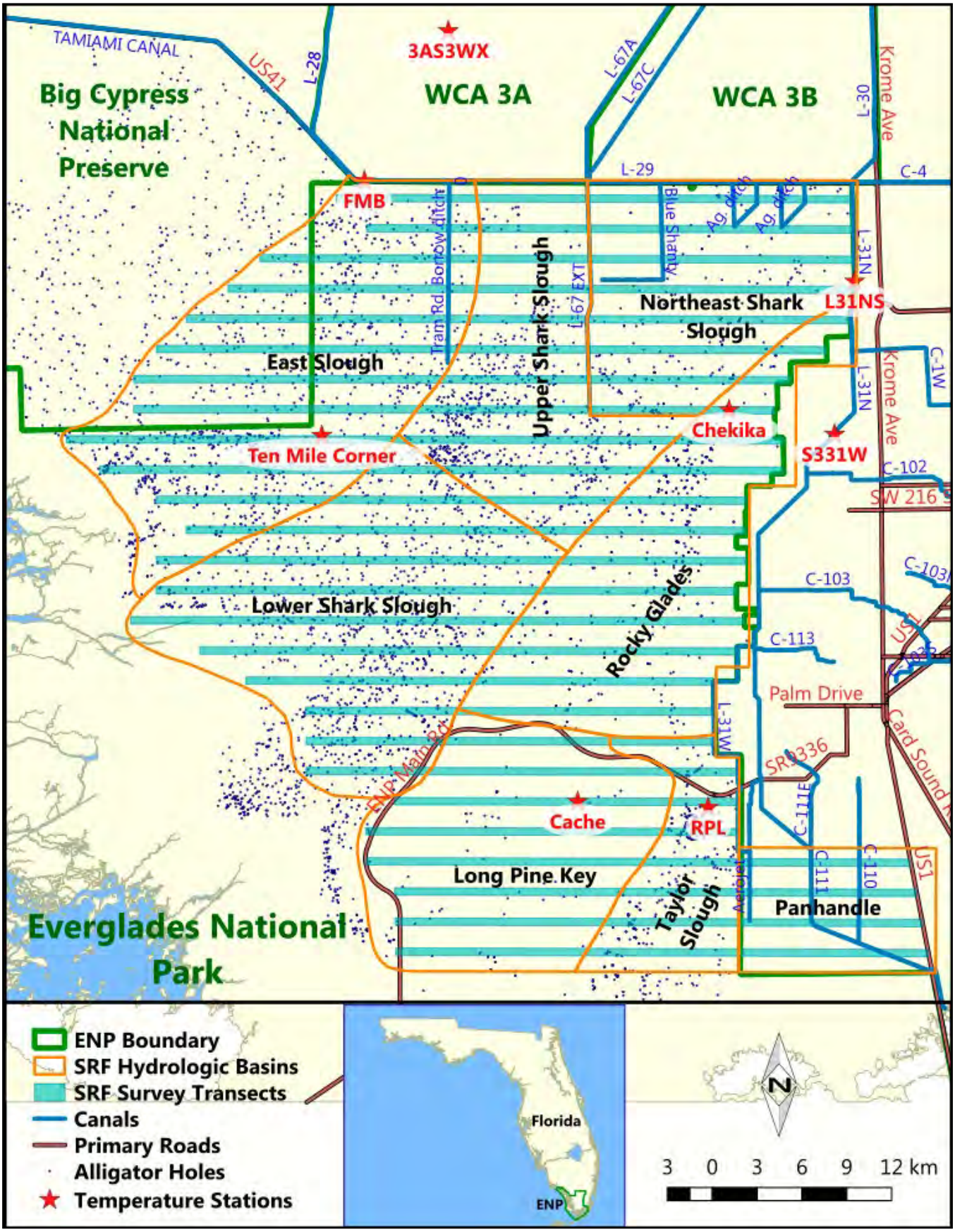


Figure 1. Location of SRF survey transects, canals, roads, alligator holes, and temperature stations.

2.1.2 Detectability

A nest sighting or non-sighting in the SRF survey may not exactly match nest presence or absence because a surveyor might not detect every nest that is present in his or her line of sight. Studies of alligator nest detectability have been conducted in other domains (Rice et al. 2000), including Arthur R. Marshall Loxahatchee National Wildlife Refuge (LNWR) (Brandt 2018; Graham 2004). Since ENP does not have extensive tree islands or any other form of vegetation that would block visibility from a helicopter, we assume that a nest sighting or non-sighting does correspond to a nest being present or absent, respectively. Since the survey observers check roughly 80% of SRF detected nests in follow-up nest visits each year and record when one is indeed old, we removed such nests from consideration.

2.2 Predictor variables

The predictor variables (Table 1) can be classified into alligator hole variables, space variables, distance variables, hydrological variables, meteorological variables, and habitat variables. The following sections discuss each variable in detail.

2.2.1 Alligator holes

Alligators breed in relatively deep, open water, and the suitability of an area as breeding habitat is influenced by the amount and type of open water. Bayous, canals, and deeper water areas of lakes and ponds are the preferred areas for breeding throughout the alligator's range (Newsom et al. 1987). Such conditions are easily provided by alligator holes and proximity to canals. Alligator holes become more important during periods of drought when they provide better feeding opportunities. Most nests are located adjacent to alligator holes or ponds (Fleming, 1990). In the Everglades, sloughs, alligator holes, and canals provide these deeper water areas. Deeper water is preferred because during mating, females must be mounted and forcefully submerged before they will engage in copulation (Fleming, 1990). Ideal nest locations are those where the eggs will be above the seasonal high-water level, but remain near enough to the water's edge to prevent desiccation, and with suitable nursery habitat for young (Mazzotti and Brandt 1994).

Location of alligator holes in the model domain is shown in Figure 1. Location data for alligator holes in ENP are from Rice and Mazzotti (2007). Substrate depth, or the depth to limestone, may serve an important role in the current and historic distribution of alligator holes. It is likely that

Table 1. Variable class, definition, units, type, and time and space variation

Class	Variables	Time-varying?	Space-varying?	Type*	Units	Definition
Response	*nest*	yes	yes	Cat/Bin	indicator	Is = 1 if the grid cell has a nest, = 0 otherwise.
Alligator hole	*hole*	no	yes	Cat/Bin	indicator	Is = 1 if the grid cell has a hole, = 0 otherwise.
Alligator hole	*hole_count*	no	yes	Disc	holes	Number of alligator gator holes in the grid cell.
Space	*xCentroid, yCentroid*	no	yes	Cont	utm meters	X and Y-coordinate of the grid cell centroid.
Distance	*dist_AH* *dist_canals* *dist_ENPrds*	no	yes	Cont	m km km	Perpendicular distance between the grid cell centroid and the nearest of the alligator holes, canals, and roads, respectively.
Hydrological	*depth_bp, depth_cm, depth_nb, depth_wy*	yes	yes	Cont	cm	Mean water depth at each grid cell during the breeding potential, courtship and mating, nest building, and whole year periods, respectively.
Hydrological	*depth_max, depth_min*	yes	yes	Cont	cm	Maximum and minimum water depth at each grid cell during the breeding potential period, respectively.
Hydrological	*Hydroperiod, drydays_ max*	yes	yes	Cont	days	Number of days with water depth >15cm and maximum number of continuous days with water depth <15cm during the breeding potential period, respectively.
Meteorological	*rain_bp, rain_cm, rain_nb*	yes	yes	Cont	cm	Average rainfall at each grid cell during the breeding potential, courtship and mating, nest building, and whole year periods, respectively.
Meteorological	*temp_bp, temp_cm, temp_nb*	yes	no	Cont	°C	Average temperature park-wide during the breeding potential, courtship and mating, nest building, and whole year periods, respectively.
Habitat	*canal, edge, excluded, marsh, upland*	no	yes	Cat/Bin	indicator	Is = 1 if at least one of the 80 sub-cells of the given grid cell is labeled canal, marsh-upland edge, excluded, marsh, and upland, respectively, = 0 otherwise.
Habitat	*canal_pcent, edge_pcent, excluded_pcent, marsh_ pcent, upland_pcent*	no	yes	Cont	%	The percent of the 80 sub-cells in the grid cell that are labeled canal, marsh-upland edge, excluded, marsh, and upland, respectively.

* Type: Cont = continuous, Disc = discrete, Cat = categorical, Bin = binary (aka dichotomous)
‡ Variable suffix: *bp* = breeding potential period, *cm* = courtship and mating period, *nb* = nest building period, *wy* = whole year period; Figure 2.

alligators build new alligator holes and old ones may get filled up or abandoned, but, for the current model, we assume that this information does not change with time.

2.2.2 Distance variables

The distance variables quantify the perpendicular distance between the grid cell in question and alligator holes, different canals, or roads in and around the SRFs domain (Figure 1). Roads and canals are not part of the natural habitat of alligators, but they may influence breeding and nesting cycles. Distance metrics of canals and roads were explored to identify which one explains the influence of anthropogenic drivers.

2.2.3 Hydrological variables

Hydrological conditions have a great influence on alligator survival and production. To successfully produce young, alligators need (Shinde et al. 2014): 1) suitable habitat, 2) to have experienced environmental conditions prior to mating that are conducive to breeding (breeding potential), 3) conditions that allow them to mate (courtship and mating), 4) suitable nest sites (nest building), and 5) to not have their nests flooded. The crucial periods that correspond to these conditions are shown in Figure 2.

The *depth_bp* (breeding potential period- *bp*) provides an estimate of the conditions within each grid cell that may influence alligators to breed in the current year based on the hydrological conditions that existed during the preceding year (Figure 2). Water depths in the preceding year influence adult body condition (Dalrymple 1996a, 1996b and Barr 1997), which then influences successful breeding. In addition, water depths <15 cm limit the ability of alligators to move easily around the marsh (Frank Mazzotti and Laura Brandt, personal communication; Shinde et al. 2014), decreasing access to both food and mates (Rice et al. 2004). The relevance of *depth_cm* (courtship and mating period- *cm*) and *depth_nb* (nest-building period- *nb*) is described in Section 2.2.1. The *depth_wy* (whole year- *wy*) encompasses the other three periods in Figure 2 and is explored as a substitute for the other three periods. More details on their estimation are provided in Appendix: I.

2.2.4 Meteorological variables

The meteorological variables including rain (*rain_bp, rain_cm, and rain_nb*) and temperature (*temp_bp, temp_cm,* and *temp_nb*) correspond to the same periods as shown in Figure 2. Rain directly influences the local water depths. Daily rainfall data on a 3.22 km x 3.22 km grid was obtained from the South Florida Water Management District (used in SFWMM and RSM models extended until 2015, SFWMD; personal communication Jan. 2017: Walter M. Wilcox, Hydrologic and Environmental Systems Modeling and M. Clay Brown, Hydrology and Hydraulics Bureau). We interpolated this dataset to the SRF's grid for this work.

Spatially averaged daily temperatures (*temp_bp, temp_cm,* and *temp_nb*) for the whole SRF area were estimated using data collected at monitoring stations (henceforth referred to as "stations") FMB, Ten Mile Corner, L31NS, Chekika, Cache, and RPL (Figure 1). Missing data were estimated through regression with the highest correlated nearby stations. Other nearby stations used for estimating missing data were: S331W, 3AS3WX, JBTS, and MLRF1. These temperature data were accessed using the DBHYDRO Browser, SFWMD (FMB, JBTS, S331W, RPL, L31NS, 3AS3WX), Jan. 2017 site accessed, https://www.sfwmd.gov/science-data/dbhydro/, from Western Regional Climate Center (Ten Mile Corner, Chekika, Cache), Jan. 2017 site accessed, https://wrcc.dri.edu/, and from National Data Buoy Center (MLRF1), Jan. 2017 site accessed, http://www.ndbc.noaa.gov/.

South Florida is characterized by consistently high and equable temperatures compared to other parts of the alligator's range (Bugbee 2008). Howarter (1999) and

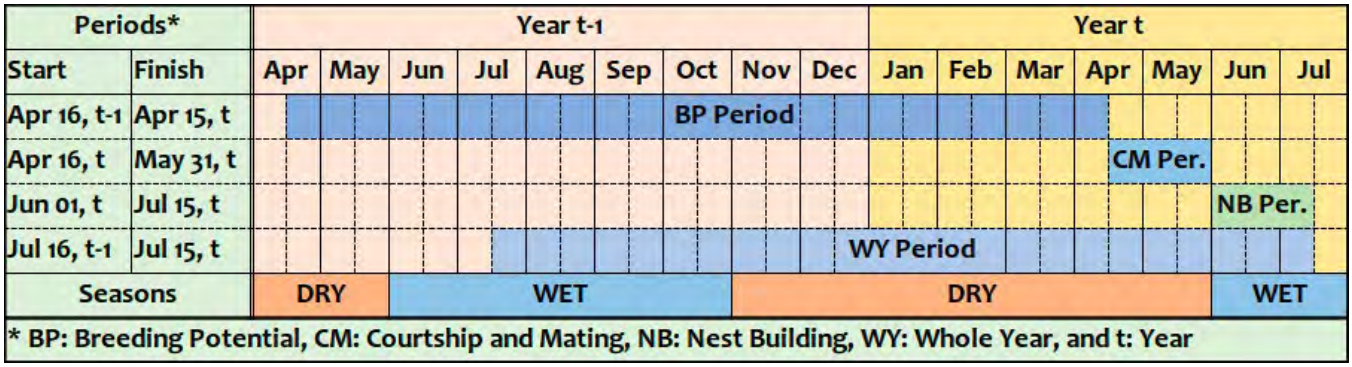


Figure 2. American alligator breeding cycle time periods (adapted from Shinde et al. 2014).

Percival et al. (2000) assert that the warm climate in southern Florida may result in high metabolic costs for alligators. Seasonal and daily temperature variations may also influence alligator movement and home range size (Chabreck 1965; Goodwin and Marion 1978; Joanen and McNease 1970, 1972; McNease and Joanen 1974; Morea 1999; Rootes and Chabreck 1993, Taylor 1984). Such increased mobility can be assumed to provide greater feeding opportunities and increased mating opportunity with more females.

2.2.5 Habitat variables

The habitat variables (*marsh, marsh_pcent, edge, edge_pcent, canal, canal_pcent, upland, upland_pcent, excluded, excluded_pcent*) describe land cover type in each model cell to support alligator growth, survival, and breeding. Habitat is grouped into five categories and their percentages in a model cell:

- Marsh is freshwater marsh, the primary habitat for alligators.
- Edge is potential upland nesting habitat immediately adjacent to freshwater marsh.
- Canals are considered unnatural areas and potentially ecological sinks for alligators even though alligators are found abundantly in canals adjoining marshes (Chopp 2003).
- Uplands have higher elevations than the marsh.
- Excluded are land cover that are not marsh, edge, upland, or canal such as lake, salt marsh, beach, levee, or road.

The habitat variables are an aggregation of the habitat classification used in Pearlstine et al. (2011). The aggregation crosswalk is presented in Shinde et al. (2014, Table 2). The habitat classification has a spatial grid cell resolution of 50 m x 50 m, so there are 80 habitat sub-cells within each 500 m x 400 m SRF transect grid cell. The percent habitat variable is the number of given 50 m x 50 m habitat sub-cells (X) in a given SRF grid-cell divided by the total number of sub-cells (80) in the SRF grid-cell multiplied by 100 (percent habitat=100*X/80).

Variables such as *marsh_pcent* and *edge_pcent* influence nesting for reasons described in Section 2.2.1. The *edge_pcent* associated with the presence of alligator holes influences nesting by providing access to higher elevation over water to build nests. Presence of edge in a grid cell indicates upland that will act as a suitable site close to water for nest building. Having a good proportion (~30%; Section 4.5.4) of grid cell with marsh conditions (*marsh_pcent*) indicates suitable habitat and increases the probability of a nest.

3 MODEL-BUILDING METHODS

The response variable nest was given a value of 1 if a nest was sighted in that grid cell that year, and it was given a value of 0 otherwise. Since the response variable is binary, a logistic regression generalized additive model (GAM) was used. Such a model will predict the probability surface of a nest sighting as a function of the predictor variables (i.e., how probable it is that there will be a nest in a given grid cell based on predictor variables such as temperature). If a prediction of the binary response variable nest was needed, then the predicted probability combined with a probability cutoff value could be used to classify cell-years as having value = 0 or = 1 for nest. In this application, however, the probability surface is enough for assessing the effects of water-management decisions.

Section 3.1 describes how and why we used data-splitting and the AUC (Area Under the receiver operating Curve) statistic as part of our model selection and assessment approach. The overall model-building process was to consider each of the categories of variables one at a time and Section 3.2 gives the rationale for the sequence in which we considered them. Section 3.3 describes how, within each category, we determined which variables to use and what functional form they should take. We built the model by adding terms and interactions from each category considered, but to keep model run times within a practical limit and computing limits (computer processing power and memory), we also eliminated some terms from the model at each stage, as described in Section 3.4. The measures we took to account for spatial correlation are described in Section 3.5, and practical details regarding the implementation of the model in R language and environment for statistical computing (R Core Team 2019) are given in Section 3.6.

3.1 Model selection and assessment

In the book *Elements of Statistical Learning*, Hastie, et al. (2001, Section 7.2), note that there are two major objectives regarding building models to be used for prediction: model selection and model assessment. Model selection is "estimating the performance of different models to choose the (approximate) best one." It is important to note that different people building models will make different decisions, and there are usually several different models of a system that have equivalent and optimal performance. Model assessment is described as, "having chosen a final model, estimating its prediction error (generalization error) on new data."

Hastie, et al. (2001) also note that when plenty of data are available, the best way to accomplish the objectives of model selection and assessment is to split the data into three parts,

which they refer to as the training-set, validation-set, and test-set.

> "The training data is used to fit the models; the validation set is used to estimate prediction error for model selection; the test set is used for assessment of the generalization error of the final chosen model. Ideally, the test set should be kept in a "vault" and be brought out only at the end of the data analysis," (Hastie, et al. 2001, p. 196).

We did not use information-theoretic approaches (e.g., AIC and BIC) or resampling approaches (i.e., cross validation and bootstrapping) in model selection because they were designed to be used when there is insufficient data to be split into three parts. Such approaches approximate the use of the validation-set to estimate model performance for model selection, and use the test-set to assess final model performance (Hastie et al. 2001, p. 196).

We have enough data so the full dataset containing the 55,656 cell-year observations was split into three subsets. A random sample of nearly half of the observations (27,840) was selected from the full dataset without replacement using a uniform distribution, and this sample was designated as the training-set. Of the remaining 27,816 observations, a random sample of nearly half of the observations (13,920; or nearly one quarter of the original set) was selected without replacement and designated as the validation-set. The remaining observations (13,896) were designated as the test-set.

For all our model-building decisions, the model was fit to the training-set. To avoid overfitting the training-set data, the model was then used to obtain predicted probabilities of *nest* for the observations in the validation-set data. These predictions were then compared to the observed values of *nest* in the validation-set data. The predicted probabilities are values between 0 and 1, but the observed values of *nest* are equal to either 0 or 1. There are different ways to compare model predictions to observed values in this situation.

The first is to choose a probability cutoff, 0.05 for example, and to set the predicted value of *nest* = 1 when the probability that *nest* = 1 falls above this cut-off, and to 0 otherwise. When such a cutoff is chosen, there are three quantities of interest: accuracy, sensitivity, and specificity. Accuracy is the proportion of cell-years correctly classified. Sensitivity is the proportion of true positives; that is, for all the cell-years having true classification of *nest* = 1, sensitivity is the proportion that were predicted to have *nest* = 1. Specificity is the proportion of true negatives. That is, for all the cell-years having true classification of *nest* = 0, specificity is the proportion that were predicted to have *nest* = 0. In other contexts, such as the use of medical tests, sensitivity and specificity have important meaning and should be examined carefully. In this analysis, however, the probability surface is of more importance than actual classification.

The second way to compare model predictions to observed values is to consider the area under the receiver operating curve (ROC). For all possible probability cut-off values, the sensitivity and specificity are calculated. The ROC is the curve for which sensitivity is on the vertical axis, and 1-specificity is on the horizontal axis, as in Section 5. When comparing two models, the one for which the area under the ROC (i.e., AUC) is greater, has greater sensitivity and specificity for a wider range of cut-off values (Hastie et al. 2001, p. 277-78). Since our interest is in the probability surface rather than in actual classification rates for a given cut-off value, we use the AUC calculated on the validation-set data to compare competing models for some model-selection decisions.

3.2 Sequence of variables considered

The overall model-building process was to consider each of the categories of variable one at a time, roughly in the order given in Section 2.2: alligator hole variables; space and distance variables; hydrological variables; meteorological variables; and, finally, habitat variables.

Alligator holes were considered first as a predictor variable because of their importance in the life cycle of alligators (Section 2.2.1). Exploratory analyses and fit statistics calculated from single-variable models (SVMs) indicated that alligator holes were strongly predictive of nest presence, thus presence of alligator holes was almost certain to be needed in the final model. When a categorical variable is such a strong predictor, it is good to consider interactions between that variable and other variables, and since such interactions can double or triple the number of columns in the model matrix, we examined the need for such interactions first.

Of the remaining categories of variables, we considered the spatial coordinates and distance variables next because the spatial coordinates were being used to account for both landscape effects and spatial correlation. We added a spatial correlation component in the model early on to ensure that the effects of other variables being added to the model were strong enough to detect after spatial effects were considered. We also needed to determine the extent to which the distance variables and the spatial coordinates might be confounded with one another and use the distance variables that were most independent of the spatial coordinates.

Of the remaining three categories of variable— hydrological, meteorological, and habitat— the values of the AUC for the single variable models fit to the variables in each category indicated that as a group, the hydrological variables

demonstrated better predictive performance, followed by the meteorological, and then the habitat variables.

3.3 Functional-forms decisions

For some variables, the functional-form decision was a choice between a categorical variable or a continuous variable, as in the case of the habitat variables for which an indicator variable (e.g., present, absent) and a percent variable (e.g., 70% coverage) were available. This decision was made based on examining graphs that showed the level of continuity of the percent variables available in the data. For categorical variables, the functional-form decision was how many categories to use or if collapsing categories is warranted. For a continuous variable, the functional-form decision was how many basis-function columns to use in the model matrix.

Thus, to choose the number of columns to allow for a continuous variable, a model can be fit with more columns than could reasonably be needed, and then the effective degrees of freedom (*edf*) can be used to get an idea of how many columns were really needed. We used graphs of the empirical probability that *nest* = 1 as a function of each variable to get an idea of how wiggly the function might need to be. We then compared SVMs with $k = 5$ and $k = 20$ (term k, is equivalent to the decision about how many columns to use for that variable) and the *edf* that resulted from each to determine the value of k to use for each variable.

3.4 Variable elimination

Variables from a given category were added progressively to the model containing the variables selected from the previously considered categories, along with interactions among the main effect terms in the new category and those in the previous categories. This approach results in the number of columns in the model matrix growing very quickly, which unchecked could result in impractically long model run times. So before going on to the next category of variables, we also eliminated interaction terms that did not contribute to the overall model fit.

The R mgcv (Wood 2017) package offers automatic variable selection options. We used cubic splines with shrinkage, which resulted in calculating additional penalties for smooth terms that can result in reducing the *edf* to a very small (near zero) value. In our output tables, we noticed that the interaction terms for which the *edf* was close to zero had very small p-values.

Thus, we decided to eliminate any interaction terms that were not significant at the $\alpha = 0.001$ level. When using GAMs with penalty parameters, p-values are approximate, and with tens of thousands of observations, statistical significance at the typical $\alpha = 0.05$ level is far too easily attained. We set the p-value cutoff for our backward elimination to $\alpha = 0.001$ to protect against multiple testing, the potential for inadequately modeled spatial or temporal autocorrelation, and the fact that p-values calculated using the methods in mgcv are approximate. We used p-values only for back-elimination of interaction terms in the model-building process and at the end of the model-building process to eliminate some main effect terms.

Main effect terms were not considered for elimination until after all variables from all categories had been added to the model because sometimes a predictor variable affects a response variable not as a main effect, but as an interaction with another variable. This procedure allowed for all pairwise interactions to be considered.

Finally, after the category of habitat variables was considered, backward elimination of insignificant terms was performed using the rules enumerated below.

1. First set of passes:
    - 1.1 Eliminate any main effect that is not statistically significant unless there is an interaction term with that effect in it.
    - 1.2 Eliminate any interaction that is not statistically significant unless it is a 2-way interaction, and there is a 3-way interaction containing both terms in it. (Example: *ti(edge_pcent, xCentroid)* is not significant, but it is not eliminated because *ti(edge_pcent, xCentroid, yCentroid)* is significant).
    - 1.3 Continue to apply Rule 1.1 and Rule 1.2 until there is nothing more to eliminate.
2. Second set of passes:
    - 2.1 At this stage, if all insignificant interactions have been eliminated, then eliminate any 2-way interactions that are not significant even if there is a 3-way interaction term with that effect in it.
    - 2.2 If any 3-way interactions become insignificant, eliminate them per Rule 1.2.
3. Final set of passes:
    - 3.1 At this stage, if all insignificant interactions have been eliminated, then eliminate any main effect that is not statistically significant even if there is an interaction with that term in it.
    - 3.2 If any 2- or 3-way interactions become insignificant, eliminate them also per Rule 1.2 and Rule 2.1.

3.5 Spatial correlation

It is important to model spatial correlation because if correlation exists, but it is not accounted for, then the model variance and all *p*-values are underestimated. One drawback to using a random effect is that most software packages which implement a random effect assume stationarity– that the covariance function is the same throughout the spatial domain. In complex and inconsistent terrain such as that in ENP, this assumption may not be valid. Another drawback is that using a random effect method accounts for the effects of location within the park in the random part of the model as opposed to the deterministic part of the model. If there is a cause-effect relationship between spatial coordinates and the response variable, as there probably is for alligator nest-building, then it is better to account for spatial correlation in the deterministic part of the model. This also allows for modeling the interaction between spatial coordinates and other predictor variables.

3.6 Implementation in R

Analyses were performed in R (R Core Team 2019) using the mgcv package (Wood 2004, 2011, and 2017). To take advantage of multicore machines that allow parallel processing, we used Microsoft R Open (https://mran.microsoft.com/open) which includes multi-threaded math kernel libraries from Intel. A setting of *setMKLthreads(7)* was used.

In the mgcv package, the settings *family* = *binomial*, *scale* = 0 and *method* = *"ML"* was used. We used the *te()* smooth argument for main effects and the *ti()* smooth for interactions. Mgcv uses *k* for specification of the number of columns and Section 4 details how values of *k* were chosen with exploration of each variable. Natural cubic spline (*bs* = *'cr'*) smoothing is the default with *te()* and *ti()* functions because it is appropriate when variables are not on the same scale, is easily interpreted, and is appropriate where other smooth options are not (Wood 2017). Additional details of the selection of a GAM model and its arguments are provided in Section 4.

4 MODEL-BUILDING RESULTS

4.1 Alligator holes

Alligator hole information could be expressed as one of three variables (Table 1): *dist_AH, hole*, and *hole_count*. The metric *dist_AH* provided the effect of an alligator hole on multiple grid cells even if they did not have alligator holes. Although the number and location of alligator holes may change from year to year, the alligator hole dataset does not have time-varying information. We determined which metric (*dist_AH, hole*, or *hole_count*) best explained the effect of alligator holes on nesting.

The value of *k* needed to be selected for *dist_AH*, and a decision had to be made as to whether to use default knots for the calculation of the cubic spline basis functions, or whether to specify custom knots. This allowed for more flexibility of the functions where distance was close to zero and in near proximity to alligator holes. A spline curve is a piecewise polynomial curve, meaning it joins two or more polynomial curves. The locations of the joins are known as "knots".

In the full dataset, the proportion of cell-years that have a nest is 0.0285. The graph of nest proportion vs. *dist_AH* in Figure 3 shows that for distances near zero, nest proportion starts higher than the full-data value of 0.0285, then drops as distance from an alligator hole increases. At close range, an alligator hole provides a place to mate and better feeding opportunities during the dry season. Alligators find proximity to alligator holes suitable for nesting in both wet and dry years.

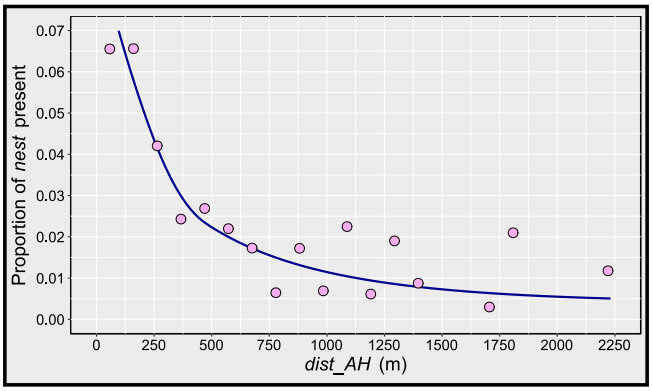


Figure 3. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 vs *dist_AH*.

Here, knots are values of the predictor variable— in this case, *dist_AH*— that are chosen to divide the predictor variable values into bins. To allow greater flexibility closer to distances of zero, we used more knots closer to distance = 0. The knots locations were 6, 200, 450, 500, and 3500 m and stayed fixed even when other variables were added to the model. Figure 3 shows the predicted probability curve based on specified fixed location of knots.

Since the total number of observations for *hole_count* = 1, 2, 3, 4, 5, 6, 7, 8, respectively, are 7992, 3264, 1128, 360, 216, 72, 120, 24 in a grid-cell, we decided to lump *hole_count* >= 4 to get a larger number of observations. Table 2 uses the new categorical variable *hole_count2* in place of discrete *hole_count*. The definition of *hole_count2* is that if *hole_count* >= 4, then *hole_count2* = '4+', otherwise *hole_count2* = *hole_count*.

Table 2 gives the value of *hole_count2* in the columns, and the value of nest (= 0 or 1) in the rows (grand total [55656] = grid cells in SRF domain [2319] * no. of years [24]). Now the goal was to investigate whether the probability that *nest* = 1 is different for the different values of *hole_count2* from '0', '1', '2', '3', '4+'. (Since *hole_count2* is a new categorical variable, the values of *hole_count2* are put in quotation marks).

Table 2. Bivariate frequency table comparing *nest* and *hole_count2*.

nest	Item	*hole_count2*					
		'0'	'1'	'2'	'3'	'4+'	Total
0	Frequency	41723	7617	3036	990	706	54072
	Percent	74.97	13.69	5.45	1.78	1.27	97.15
	Row %	77.16	14.09	5.61	1.83	1.31	100
	Col %	98.22	95.31	93.01	87.77	89.14	NA
1	Frequency	757	375	228	138	86	1584
	Percent	1.36	0.67	0.41	0.25	0.15	2.85
	Row %	47.79	23.67	14.39	8.71	5.43	100
	Col %	1.78	4.69	6.99	12.23	10.86	NA
Total	Frequency	42480	7992	3264	1128	792	55656
	Percent	76.33	14.36	5.86	2.03	1.42	100

We estimated marginal means (aka least-squares means) for levels of hole counts in a linear model [*gam(nest ~ 1 + as.factor(hole_count2), scale=0, select=TRUE, family=binomial(link="logit"), method = "ML", data=gator.trn)*] and computed contrasts among them to determine whether the observable differences in proportion of alligator holes that have nests across values of *hole_count2* are statistically significant, using the R package emmeans (Lenth 2019). The test adjusts the *p*-values because multiple tests are being performed simultaneously to preserve the effective significance level of $\alpha = 0.05$. The results presented in Table 3 show that the proportion of nests for *hole_count2* = '2', '3', and '4+' are not significantly different from one another.

Table 3. Multiple comparisons of effects of values of *hole_count2* on probability of a nest.

Comparison of *hole_count2* Least Squares Means ($\alpha = 0.05$)			
Contrast	Estimate	*p*-value	Different?
'0' – '1'	-0.927	<.0001	Yes
'0' – '2'	-1.522	<.0001	Yes
'0' – '3'	-1.865	<.0001	yes
'0' – '4+'	-1.936	<.0001	Yes
'1' – '2'	-0.594	<.0001	Yes
'1' – '3'	-0.937	<.0001	Yes
'1' – '4+'	-1.009	<.0001	Yes
'2' – '3'	-0.343	0.206	No
'2' – '4+'	-0.415	0.156	No
'3' – '4+'	-0.072	0.997	No

Based on these results, a new variable was created called *hole_count3*, defined as follows. If *hole_count2* = '2', '3', or '4+', then *hole_count3* = '2+'; otherwise, *hole_count3* = *hole_count2*. This leaves possible values for *hole_count3* as '0', '1', and '2+'. Next, a comparison was made of three single-variable models of predicting probability of *nest* = 1, as shown in Table 4. The capitalized SVM name in this report (e.g., *HOLE*) indicates only that variable is present in the model.

Table 4. Performance statistics for models containing *hole* and *hole_count3*.

Model	AUC		Rank	*edf*
	Training	Validation		
HOLE	0.64	0.65	2	2.0
HOLE_COUNT3	0.65	0.65	3	3.0
DIST_AH	0.70	0.70	5	4.3

Each model was fit to the training-set, and the AUC was calculated based on predictions to both the training-set and validation-set data. The *HOLE* model had training and validation AUCs of 0.64 and 0.65, the *HOLE_COUNT3* model had training and validation AUCs of 0.65 and 0.65, and the *DIST_AH* model had training and validation AUCs of 0.70 and 0.70. Based on these results, we chose to use *dist_AH* in models moving forward. The code below was used to fit the model with *dist_AH*—

```
DM <- gam(nest ~ 1 + te(dist_AH),
          scale = 0, select = TRUE, knots = list(dist_AH=c(6,
              200, 450, 500, 3500)),
          family = binomial(link = "logit"), method =
              "ML", control =  ctrl, data = gator.trn)
```

where "*gator.trn*" is the alligator training-set data.

4.2 Space and distance variables

The distance variables were considered at the same time as the spatial coordinates because distances from canals and roads may be highly correlated with, and thus confounded with, spatial coordinates. We wanted to tackle the challenge of disentangling the effects early in the process.

The distance variables (Table 1) include *dist_AH*, discussed in previous Section 4.1; *dist_canals*, which is the perpendicular distance between the center of the grid cell and the closest canal; and *dist_ENPrds*, which is the perpendicular distance between the grid cell and the closest road (only heavily traveled roads were considered).

Unlike *dist_AH*, which had a nest proportion of about 0.07

when close to alligator holes (Figure 3), *dist_canals* and *dist_ENPrds* did not show as high proportion at close distance to canals and roads (Figure 4). For this reason, we decided to use default knots for calculation of the cubic spline basis functions for *dist_canals* and *dist_ENPrds*. Figure 4 shows the fitted predicted probability curve based on default knots and appears to capture the trend well.

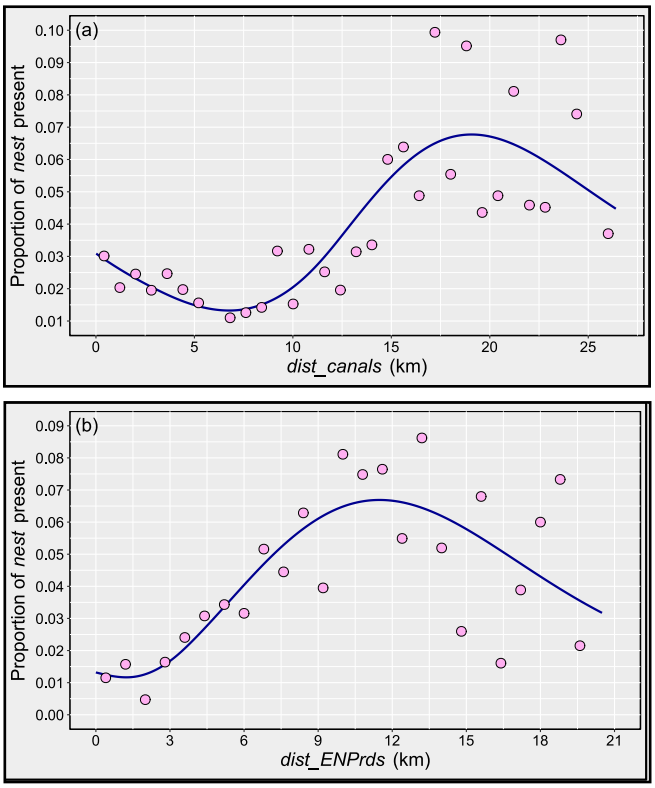


Figure 4a and 4b. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 vs (a) *dist_canals* and (b) *dist_ENPrds*.

The final model chosen at the end of this step in the process had the following specification in mgcv—

```
DiM_S <- gam(nest ~ 1 + te(dist_AH) + te(dist_canals) +
            te(dist_ENPrds) + ti(dist_AH,dist_canals)
            + ti(dist_AH, dist_ENPrds) + ti(dist_canals,
            dist_ENPrds) + te(xCentroid) + te(yCentroid)
            + ti(xCentroid,yCentroid),

      knots = list(dist_AH=c(6, 200, 450, 500, 3500)),
          scale=0, select=TRUE,

      family=binomial(link="logit"), method = "ML",
          control=ctrl, data=gator.trn)
```

The performance statistics of this model are given in Table 5. All the *p*-values were very small, except for *xCentroid*, showing that, thus far, all the terms included are needed.

Table 5. Performance statistics for model with alligator hole, spatial coordinates, and distance variables.

Model	AUC		Rank	*edf*
	Training	Validation		
DiM_S	0.83	0.83	85	45

4.3 Hydrological variables

The SVMs were used to determine which of the hydrological variables (Table 1) should be included in the model. The first step in choosing among them was to use the settings—*method* = '*ML*' and *select* = *TRUE*— to fit SVMs and compare the results. Table 6 shows the fit statistics for each of these single-variable models. In terms of the validation AUC and the fit statistics, the *DEPTH_WY* model had the highest performance, followed by *HYDROPERIOD*.

The *depth_bp*, *depth_cm*, and *depth_nb* variables were created to replace *depth_wy* (Figure 2) to determine impacts of water depth on the different stages of the reproduction cycle. The *d3* model contained these three replacement variables and the *d3i* model contained these three plus the three pairwise interactions among them (Table 6). The *d3* model outperforms the *DEPTH_WY* model in terms of the outside-the training-set data fit statistic validation AUC.

Table 6. Performance statistics for SVMs fit to the hydrological variables and for two multi-variable models containing depth for the breeding potential, courtship and mating, and nest building periods, without and with interactions respectively.

Model	AUC		Rank	*edf*	Deviance explained
	Training	Validation			
DEPTH_WY	0.72	0.71	5	3.8	0.076
HYDROPERIOD	0.70	0.72	5	4.3	0.072
DRYDAYS_MAX	0.69	0.71	5	4.4	0.066
DEPTH_CM	0.70	0.70	5	4.2	0.062
DEPTH_BP	0.69	0.70	5	3.9	0.063
DEPTH_NB	0.69	0.67	5	3.8	0.051
DEPTH_MAX	0.67	0.68	5	3.8	0.054
DEPTH_MIN	0.66	0.68	5	4.0	0.041
**d3*	0.72	0.72	13	8.3	0.073
**d3i*	0.73	0.73	61	15.4	0.081

*multi-variable models containing depth for the breeding potential, courtship and mating, and nest building periods, without (*d3*) and with (*d3i*) interactions.

Note that in the *DEPTH_WY* model, the rank is the full number of columns, and this would normally be the model degrees of freedom, but due to the penalty parameter, the effective degrees of freedom (*edf*) is 3.8. Similarly, for the *d3* model, the rank is 13, but the *edf* is 8.3. The *d3i* model outperforms both the *DEPTH_WY* and the *d3* models in terms of all fit statistics. All the terms in *d3* and *d3i*

model were statistically significant ($\alpha = 0.001$), except one interaction term- *ti(depth_bp,depth_nb)*.

4.3.1 Progression of models

Adding the terms from the *d3i* model (Table 6) to the terms in the *DiM_S* model (Table 5) resulted in model *d3i_DiMS* (Table 7). The number of columns in the model matrix increased by 188, from 85 to 273, though the *edf* increased by a small fraction of that (29), which shows that not all the columns are needed. The training AUC increased from 0.83 to 0.85, and the validation AUC increased from 0.83 to 0.84.

Table 7. Progression of models through adding hydrological variables.

Model	AUC		Rank	*edf*
	Training	Validation		
DiM_S	0.83	0.83	85	44.8
d3i_DiMS	0.85	0.84	273	74.0
d3i_DiMS_e	0.85	0.84	193	69.5

Since not all the terms in the *d3i_DiMS* model were statistically significant, all interaction terms that were not significant at the $\alpha = 0.001$ level were removed from the model following rules in Section 3.4. All main effect terms were kept in the model to allow interactions between terms currently in the model with the new main effect terms that would be added when meteorological and habitat variables were included. The resulting model is the *d3i_DiMS_e* shown in Table 7. The reduction in the number of columns has resulted in rank being much closer to the *edf* value, while the training and validation AUC remained the same.

4.4 Meteorological variables

The meteorological variables consist of three rain variables and three temperature variables (Table 1). The rain variables vary over both time and space, but the temperature variables vary over time only, not over space. That means there are at most 24 distinct values of each temperature variable—one for each year—and it was important to limit the degrees of freedom for temperature so that it did not become a surrogate for an effect of year, which would result in overfitting.

4.4.1 Rain

Figure 5 and Table 8 show the results of the SVM runs for the rain variables. For *rain_nb*, the fitted curve appears to be overfitting the training-set data, but this is not the case for *rain_bp and rain_cm*, for which there is some underfit. The rank shown in Table 8 is the full-model rank, which includes one column for the intercept and four columns for each continuous variable, since the SVMs were run using the default value of $k = 5$ in the *te()* smoother.

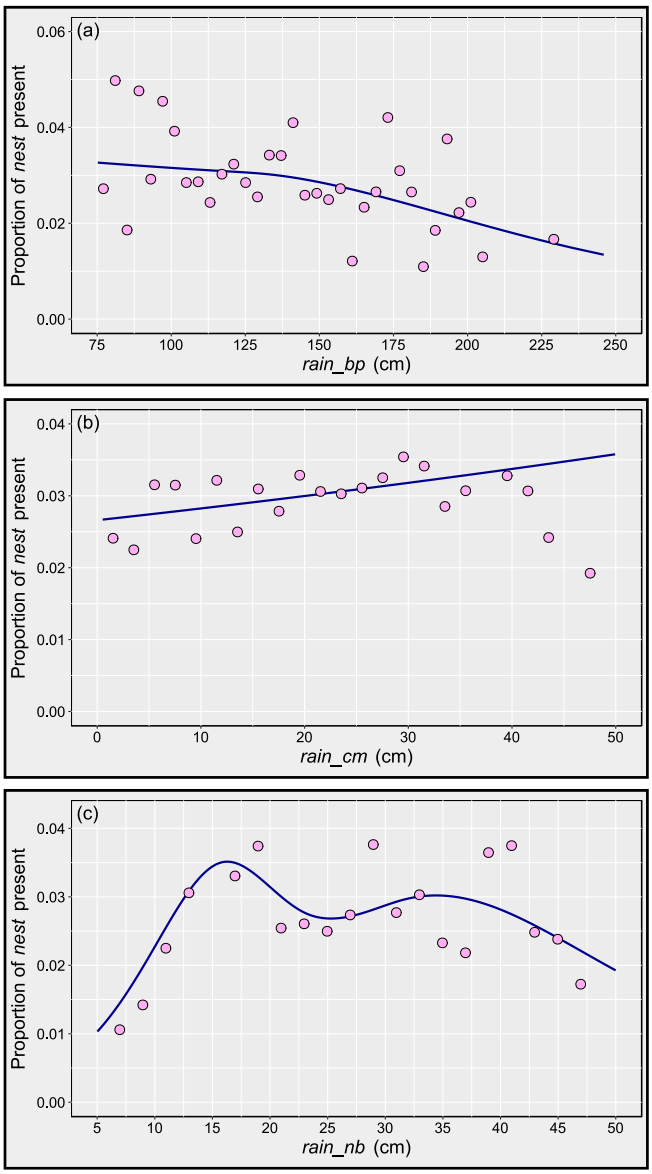


Figure 5a, 5b, and 5c. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 as a function of (a) *rain_bp*, (b) *rain_cm*, and (c) *rain_nb*.

Table 8. SVM results for rain variables.

Model	Graph appearance	AUC		Rank	*edf*	*p*-value
		Training	Validation			
RAIN_BP	Some effect	0.52	0.53	5	2.8	3.04E-02
RAIN_CM	Some effect: Underfit	0.52	0.49	5	2.0	9.65E-02
RAIN_NB	Overfit	0.55	0.57	5	4.7	8.73E-04

For both *rain_bp* and *rain_cm*, the *edf*s that result after the smoothing/penalty parameter was fit was much lower than 5, and for *rain_nb*, it was close to 5, but this resulted in overfit (Figure 5). Thus, for all the rain terms, $k = 3$ was used to allow two columns for each variable, which provided more flexibility than a linear function, but avoided overfitting. Although the SVM results for *rain_bp* and *rain_cm* were not statistically significant, they were included in the model in case they became significant by way of an interaction with another variable, such as one of the water depth variables.

In the progression of models below, interactions among the rain variables were considered, and external interactions between each of the rain variables and each of the other main effect variables, except for the spatial coordinates, were considered.

4.4.2 Temperature

The temperature variables do not vary spatially. Each has a maximum of 24 distinct values, one for each year. When interaction terms are created with other variables that vary over both space and time, the interaction terms themselves will also vary over space and time, and will therefore have many more distinct values. Temperature-with-temperature interactions were avoided since there were too few distinct values to support them. In the progression of models below, the models contained external interactions between each of the temperature terms and each other main effect variable, but no internal interactions among the temperature variables.

Table 9 and Figure 6 show the results of fitting SVMs to the temperature variables. For these models, the default value of $k = 5$ was used. The *edf* after the smoothing/penalty parameter was fit remains close to 5 for both *TEMP_BP* and *TEMP_NB*, and the training and validation AUC are both higher than the *TEMP_CM*, but the graphs show that these results are due to overfitting. The variable *temp_cm* did not show a strong effect as the fitted model was a near horizontal line with a small slope, both AUC values were 0.52, and the *p*-value was not significant. Insignificant terms can become significant, however, when interactions are considered. Results that tell about the interaction of hotness and dryness could be useful. The terms that tell about dryness are the rain terms, the water depth terms, and the proximity to canal terms. To keep from overfitting, $k = 3$ was used for all three of the temperature terms.

Table 9. SVM results for temperature variables.

Model	Graph appearance	AUC		Rank	*edf*	*p*-value
		Training	Validation			
TEMP_BP	Overfit	0.57	0.55	5	4.9	3.82E-12
TEMP_CM	Some-effect	0.52	0.52	5	2.1	1.63E-01
TEMP_NB	Overfit	0.57	0.54	5	4.7	3.01E-05

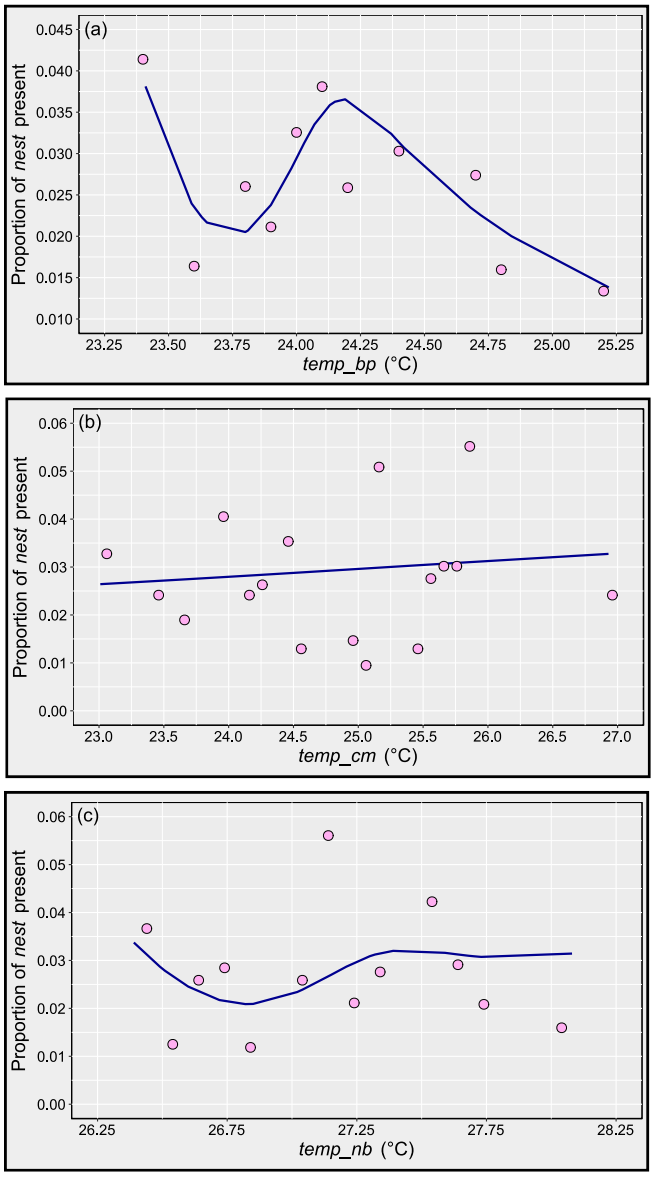


Figure 6a, 6b, and 6c. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 as a function of (a) *temp_bp*, (b) *temp_cm,* and (c) *temp_nb*.

4.4.3 Progression of models

The main effects of each of the rain and temperature variables, the internal interactions among the rain variables, and the external interactions— rain-with-temperature, rain-with-distance, rain-with water-depth, temperature-with-distance, temperature-with-water-depth— were added to the terms that remained in model *d3i_DiMS_e* in Table 7. The model that resulted was model *d3iDiMS_r3it3_i* in Table 10. The rank of the model matrix increased by 348 columns, from 193 to 541, while the *edf* increased by only 39, showing that not all 348 columns were needed. The training AUC increased from 0.85 to 0.88, while the validation AUC increased from 0.84 to 0.85.

Table 10. Progression of models from adding hydrological variables to adding meteorological variables

Model	AUC		Rank	*edf*
	Training	Validation		
d3i_DiMS_e	0.85	0.84	193	69.5
d3iDiMS_r3it3_i	0.88	0.85	541	108.8
d3iDiMS_r3it3_i_e	0.87	0.85	193	79.0

All interaction terms with *p*-value greater than $\alpha = 0.001$ were removed following rules in Section 3.4 resulting in model *d3iDiMS_r3it3_i_e* in Table 10. The number of columns in the model matrix for this model was 193, with *edf* = 79. The training AUC was 0.87, and the validation AUC was 0.85. The terms remaining in this model are shown in Table 11.

4.5 Habitat variables

In addition to the alligator hole variables already discussed in Section 4.1, there were other habitat variables (Table 1) motivated by habitat categories used in the APSI model (Shinde et al. 2014). For all these variables, the values vary over space but not over time in the data, although they most likely do vary slowly over time.

4.5.1 Canal variables

There were three variables to choose from to represent the effect of canals on the probability of nest building: *canal*, *canal_pcent*, and *dist_canals* (already considered and evaluated in Section 4.2). Table 12 shows that only 696 cells have a value of *canal* = 1. For cells with *canal* = 1, 8.48% have *nest* = 1, but for those with *canal* = 0, only 2.77% have *nest* = 1. Using *dist_canals* makes the distance effectively 0 for those 696 cells, allowing them to have a higher probability of *nest* = 1. A continuous function of *dist_canals* allows nearby cells to also have a higher probability of *nest* = 1, and that probability can decrease with distance. The nest-building probability was a function of a continuous variable for canal using either *dist_canals* or *canal_pcent*. However, Figure 7 shows that as a continuous variable, *canal_pcent* does not provide much more information than did *canal*; for 54,960 of the 55,656 observations, the value is 0.

Figure 7 shows that when dividing *canal_pcent* into bins to calculate the proportion of observations for which *nest* = 1, only three bins meet the criterion that there must be at least 50 observations in a bin. Figure 4(a), in contrast, shows that considering the effect of the distance from a canal allowed the model to predict different probabilities of *nest* = 1 for all different values of distance, with a gentle curve differentiating the probabilities for different distances. We decided *dist_canals* would be the best variable to use in the model because it allowed a canal to have an effect on a grid cell even if that grid cell did not contain a canal, where the other two variables did not (see Section 4.2 for a description of how *dist_canals* was modeled).

Table 11. Terms in model *d3iDiMS_r3it3_i_e*.

Term	Estimate_*edf**	Std.Error_*rdf***	*p*-value
(Intercept)	-4.9	0	3.22E-127
te(dist_AH)	3.0	4	6.60E-30
te(dist_canals)	3.6	4	1.12E-26
te(dist_ENPrds)	2.7	4	5.49E-05
ti(dist_AH,dist_canals)	7.1	16	9.22E-09
ti(dist_AH,dist_ENPrds)	6.9	16	1.09E-05
ti(dist_canals,dist_ENPrds)	6.0	16	4.78E-08
te(xCentroid)	0.0	4	9.20E-01
te(yCentroid)	0.1	4	2.79E-01
ti(xCentroid,yCentroid)	12.6	16	5.73E-31
te(depth_bp)	2.5	4	6.61E-08
te(depth_cm)	0.9	4	8.36E-05
te(depth_nb)	0.0	4	1.00E+00
ti(depth_bp,depth_cm)	4.6	16	7.52E-09
ti(depth_bp,dist_AH)	2.5	14	1.63E-07
te(rain_bp)	0.0	2	9.14E-01
te(rain_cm)	0.0	2	6.55E-01
te(rain_nb)	0.0	2	7.17E-01
te(temp_bp)	1.6	2	1.26E-04
te(temp_cm)	0.0	2	8.09E-01
te(temp_nb)	0.7	2	5.43E-02
ti(rain_bp,temp_bp)	2.3	4	1.24E-06
ti(rain_cm,temp_cm)	3.8	4	3.09E-10
ti(rain_cm,temp_nb)	1.0	4	2.43E-09
ti(depth_bp,temp_cm)	2.7	8	4.55E-10
ti(depth_nb,temp_cm)	2.9	8	7.23E-06
ti(depth_cm,temp_nb)	3.2	8	1.32E-04
ti(dist_ENPrds,temp_bp)	3.9	8	2.53E-07
ti(dist_canals,temp_cm)	3.5	8	6.01E-06

* The regression coefficient estimates if the term is represented by a single column in the model matrix; the estimated degrees of freedom (*edf*) if the term is a smooth function of a continuous variable, represented by multiple columns in the model matrix.
** The standard error of the regression coefficient estimates if the term is represented by a single column in the model matrix; the reference degrees of freedom (*rdf*) if the term is a smooth function of a continuous variable.
** Value of the standard error of the parameter estimates if the term is a categorical variable, and it contains the reference degrees of freedom (*rdf*).

Table 12. Bivariate frequency table comparing the binary variables *nest* and *canal*.

nest	Item	*canal*		
		0	1	Total
0	Frequency	53435	637	54072
	Percent	96.01	1.14	97.15
	Row %	98.82	1.18	100
	Col %	97.23	91.52	NA
1	Frequency	1525	59	1584
	Percent	2.74	0.11	2.85
	Row %	96.28	3.72	100
	Col %	2.77	8.48	NA
Total	Frequency	54960	696	55656
	Percent	98.75	1.25	100

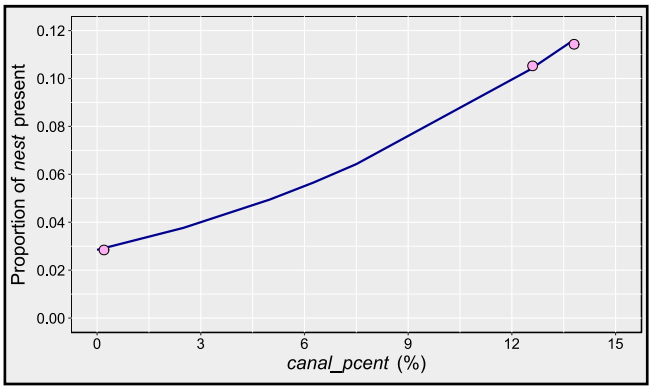


Figure 7. Empirical probability (plum dots) of *nest* = 1 as a function of *canal_pcent*.

4.5.2 Edge variables

Two variables explained the differences in the probability of a nest being built as a function of whether the grid cell was on the edge between marsh and upland: the binary variable *edge* and the continuous variable *edge_pcent* (Table 1). It is usually better to use a continuous variable to reduce information loss rather than just a binary separation of values for probability *nest* = 1. Figure 8 shows that there were enough continuous values of *edge_pcent* to allow its use as a continuous predictor variable without large gaps between values such as those seen in Figure 7 for *canal_pcent*.

Figure 9(a) shows the SVM-predicted probability of *nest* = 1 as a function of *edge_pcent*. A default value of $k = 5$ was used in the SVM in this case. Figure 9(b) shows the same for *edge_pcent*, except that in this SVM, $k = 20$. The results in Table 13 show that when $k = 5$ or 20, the only difference between training or validation AUCs is that the p-value for $k = 5$ is slightly lower. Since the *edf* for the $k = 5$ SVM was equal to 2.0, we chose $k = 3$ for *edge_pcent* to provide more flexibility than a linear function.

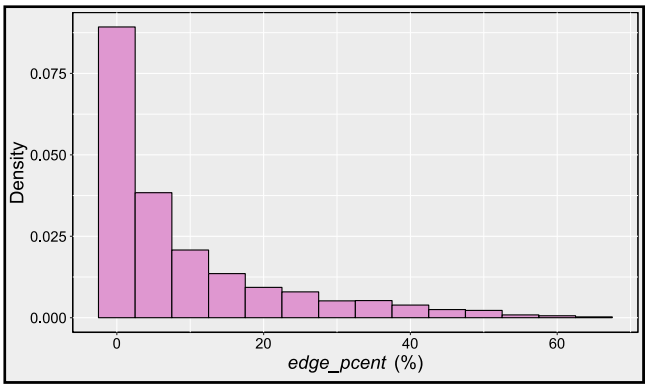


Figure 8. Histogram of *edge_pcent*.

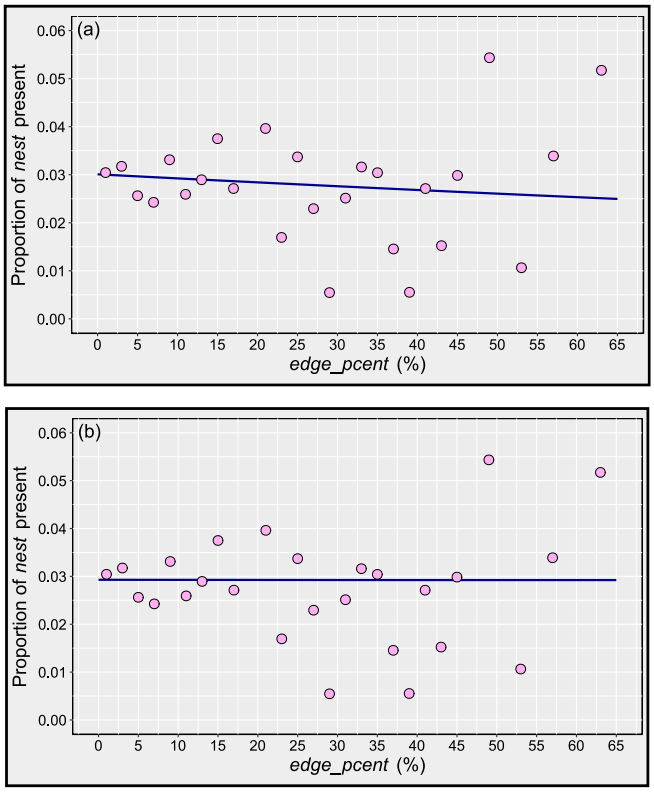


Figure 9a and 9b. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 as a function of *edge_pcent* when (a) $k = 5$ and (b) $k = 20$ (blue curves).

4.5.3 Excluded variables

Excluded variables do not refer to variables that are not included in the model. It refers to the binary variable *excluded* and the continuous variable *excluded_pcent* (Table 1) with sub-cells classified as excluded habitat (e.g., salt marsh). It is preferable to use a continuous variable, if possible, but the numbers of observations were low for *excluded_pcent* over 40%. Additionally, 96% of the observations in the data had *excluded_pcent* = 0 (Table 14). Therefore, the binary variable *excluded* was chosen for the model.

Table 13. SVM results for *edge_pcent*.

Model	*k*	Graph appearance	AUC		Rank	*edf*	*p*-value
			Training	Validation			
EDGE_PCENT	5	Some-effect	0.51	0.53	5	2.0	0.304
EDGE_PCENT	20	No-effect	0.51	0.53	20	1.0	0.328

Table 14. Bivariate frequency table comparing *nest* and *excluded*.

nest	Item	excluded		
		0	1	Total
0	Frequency	52099	1973	54072
	Percent	93.61	3.54	97.15
	Row %	96.35	3.65	100
	Col %	97.08	99.05	NA
1	Frequency	1565	19	1584
	Percent	2.81	0.03	2.85
	Row %	98.80	1.20	100
	Col %	2.92	0.95	NA
Total	Frequency	53664	1992	55656
	Percent	96.42	3.58	100

Of the 53,664 observations with no excluded habitat (i.e., *excluded* = 0; Table 14), the empirical probability of *nest* = 1 was 0.0292, which was slightly higher than the full-data value of 0.0285. Of the 1,992 observations that have *excluded* = 1, the empirical probability of *nest* = 1 is 0.0095. Since these two probabilities were different, *excluded* was chosen as a useful variable to be included in the model.

4.5.4 Marsh variables

The marsh variables consisted of the binary variable *marsh* and the continuous variable *marsh_pcent* (Table 1). The histogram of *marsh_pcent* has no major gaps, and has the full range of values from 0 to 100 covered by enough observations that it is reasonable to estimate a value of the empirical probability of *nest* = 1 for each (Figure 10).

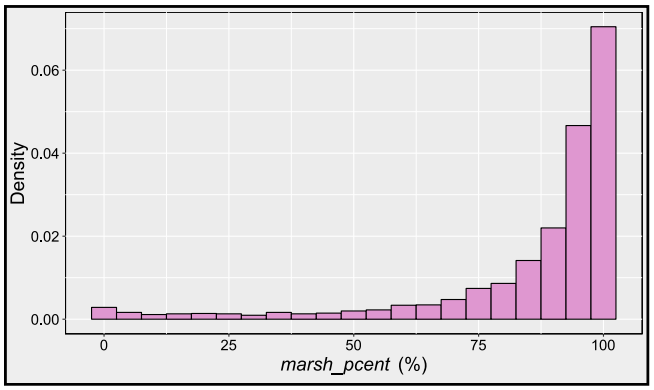


Figure 10. Histogram of *marsh_pcent*.

Figures 11(a, b) show the SVM fits to *marsh_pcent* for $k = 5$ and $k = 20$, respectively, and Table 15 gives fit statistics for these SVMs. Although both the training and validation AUCs increase when $k = 20$, which results in *edf* = 8, Figure 11(b) shows that the $k = 20$ SVM overfits the data. Since the *edf* for the $k = 5$ SVM was 4.7 with slight overfitting of data, $k = 4$ was chosen for *marsh_pcent* moving forward.

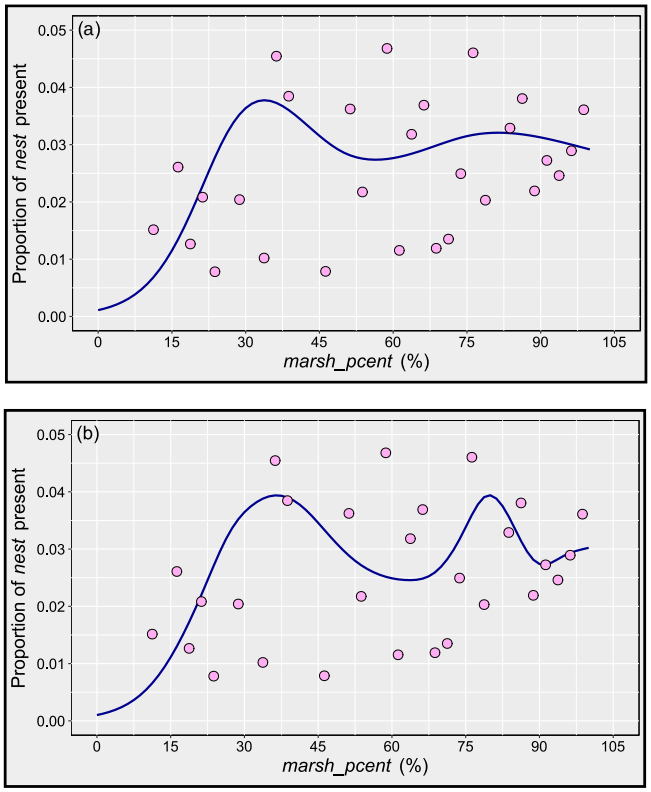


Figure 11a and 11b. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 as a function of *marsh_pcent* when (a) $k = 5$ and (b) $k = 20$ (blue curves).

4.5.5 Upland variables

As with the other habitat variables, there were two variables that could be used to describe the effect of upland on nest building probability. The first was the binary variable *upland* and second was the continuous variable *upland_pcent* (Table 1).

Since it is best to use a continuous variable when possible, and since there were no major gaps in the values taken by *upland_pcent*, *upland_pcent* was chosen for the model. Table

Table 15. SVM results for *marsh_pcent*.

Model	*k*	Graph appearance	AUC		Rank	*edf*	*p*-value
			Training	Validation			
MARSH_PCENT	5	Some-overfit	0.53	0.50	5	4.7	1.35E-03
MARSH_PCENT	20	Overfit	0.55	0.52	20	8.0	1.54E-03

Table 16. SVM results for *upland_pcent*.

Model	*k*	Graph appearance	AUC		Rank	*edf*	*p*-value
			Training	Validation			
UPLAND_PCENT	5	Gentle curve	0.50	0.52	5	2.8	3.80E-02
UPLAND_PCENT	20	Gentle curve	0.50	0.52	5	3.5	6.98E-02

16 gives the results from the SVM with *upland_pcent* for *k* = 5 and *k* = 20. The *edf* from the *k* = 5 SVM was 2.8 and *k* = 20 SVM was 3.5. The predicted curves in Figure 12 show that there is only a slight difference from k = 5 to k = 20. Thus, *k* = 5 was chosen for *upland_pcent* to provide more flexibility in capturing the SVM trend.

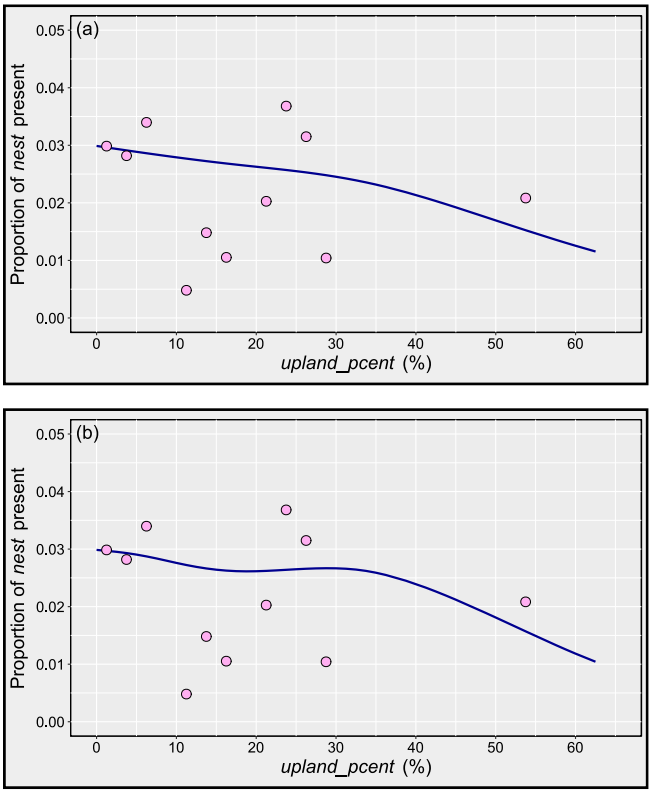


Figure 12a and 12b. Empirical probability (plum dots) and SVM-predicted probability of *nest* = 1 as a function of *upland_pcent* when (a) *k* = 5 and (b) *k* = 20 (blue curves).

4.5.6 Progression of models

The main effect smooth terms for *edge_pcent*, *marsh_pcent*, and *upland_pcent*, along with the categorical term for *excluded* (excluded habitat) were added to model *d3iDiMS_r3it3_i_e* from Table 10. Interactions between each of *edge_pcent*, *marsh_pcent*, and *upland_pcent* and each of the continuous variables already in the model were also added. Fit statistics from the resulting model, called *d3DMSr3t3_hi_i*, are given in Table 17. Although 323 columns were added to the model matrix, the training AUC went up by only 0.007, and the validation AUC by 0.003.

Table 17. Progression of models from adding meteorological variables to adding habitat variables.

Model	AUC		Rank	*edf*
	Training	Validation		
d3iDiMS_r3it3_i_e	0.868	0.847	193	79.0
d3DMSr3t3_hi_i	0.875	0.850	516	103.4
d3DMSr3t3_hi_i_e	0.871	0.850	210	85.2

In model *d3DMSr3t3_hi_i* the effect of *excluded*, with a *p*-value of 0.461, was not statistically significant. Of the main effects of *edge_pcent*, *marsh_pcent*, and *upland_pcent*, only *marsh_pcent* was statistically significant. Of these 36 interactions involving habitat variables, only two were statistically significant: *ti(edge_pcent,depth_cm)* and *ti(upland_pcent,dist_canals)*.

The low impact of adding habitat variables in *d3iDiMS_r3it3_i_e* showed that the variables already in the model were sufficient without adding habitat variables. All interaction terms that were not statistically significant were removed from the model at the same time. The remaining interaction terms were assessed for significance and removed if they were not significant. This process was repeated until all interaction terms in the model were significant, following rules in Section 3.4. The model that resulted is model *d3DMSr3t3_hi_i_e* in Table 17 which has 17 more columns than did model *d3iDiMS_r3it3_i_e*. This back-elimination process resulted in a slightly lower training AUC. The terms remaining in this model are displayed in Table 18.

Table 18. Terms in model *d3DMSr3t3_hi_i_e*.

Term	Estimate_*edf*	Std.Error_*rdf*	*p*-value
(Intercept)	-4.9	0	3.27E-125
te(dist_AH)	2.9	4	1.88E-26
te(dist_canals)	3.6	4	8.24E-27
te(dist_ENPrds)	2.5	4	8.32E-04
ti(dist_AH,dist_canals)	7.3	16	6.54E-09
ti(dist_AH,dist_ENPrds)	7.1	16	1.07E-05
ti(dist_canals,dist_ENPrds)	5.8	16	7.39E-08
te(xCentroid)	0.0	4	8.78E-01
te(yCentroid)	0.0	4	5.17E-01
ti(xCentroid,yCentroid)	12.4	16	2.19E-32
te(depth_bp)	2.4	4	5.16E-09
te(depth_cm)	0.9	4	1.14E-06
te(depth_nb)	0.0	4	1.00E+00
ti(depth_bp,depth_cm)	4.2	16	4.42E-08
ti(depth_bp,dist_AH)	2.5	14	3.28E-07
te(rain_bp)	0.0	2	1.00E+00
te(rain_cm)	0.5	2	1.35E-01
te(rain_nb)	0.0	2	7.19E-01
te(temp_bp)	1.4	2	1.51E-03
te(temp_cm)	0.0	2	4.58E-01
te(temp_nb)	0.2	2	2.18E-01
ti(rain_bp,temp_bp)	2.0	4	7.35E-07
ti(rain_cm,temp_cm)	3.6	4	3.32E-13
ti(rain_cm,temp_nb)	1.8	4	9.85E-13
ti(depth_bp,temp_cm)	3.5	8	9.12E-12
ti(depth_nb,temp_cm)	3.5	8	9.10E-08
ti(depth_cm,temp_nb)	3.5	8	3.49E-08
ti(dist_ENPrds,temp_bp)	4.1	8	1.01E-07
ti(dist_canals,temp_cm)	3.6	8	1.84E-06
te(edge_pcent)	0.0	2	8.46E-01
te(marsh_pcent)	1.4	3	2.74E-05
te(upland_pcent)	0.6	4	1.28E-01
ti(edge_pcent,depth_cm)	3.0	8	2.00E-05

4.6 Space interactions

All the models in the progression from *DiM_S* model (Table 5) onward contained a main effect for *xCentroid*, *yCentroid*, and an interaction between the two. As each category of variables was considered and added to the model, no interactions of spatial coordinates with main effects of the other variables were included. Because spatial coordinates could function as a surrogate for any other variable, the other variables were considered first to allow them to explain as much variability in the probability that *nest* = 1 as possible.

It is possible, however, that the effect of some variables is different at different locations due to unobserved variables not included in the model, and this phenomenon would be captured using interactions between spatial coordinates and other variables. In this final stage of model development, 2-way interactions between each *xCentroid* and *yCentroid* and the other continuous main effect terms are added to the model, as is a 3-way interaction among *xCentroid*, *yCentroid*, and each of the continuous main effect terms. Results from this model, called *d3DMSr3t3hi_Si*, are given in Table 19. Adding these terms resulted in increasing the number of columns in the model matrix by over 1000, and the *edf* of 152 showed that not all were necessary.

Table 19. Progression of models from adding habitat variables to adding interactions with spatial coordinates.

Model	AUC		Rank	*edf*
	Training	Validation		
d3DMSr3t3_hi_i_e	0.87	0.85	210	85.2
d3DMSr3t3hi_Si	0.90	0.86	1281	152.4
d3DMSr3t3hi_Si_e	0.88	0.86	479	114.0

Backward elimination of insignificant terms was performed using the rules enumerated in Section 3.4. The model left after application of these rules was called *d3DMSr3t3hi_Si_e* (Table 19). The terms included in this model are displayed in Table 20. Though the training and validation AUC for model *d3DMSr3t3hi_Si_e* are only 0.01 more than that in model *d3DMSr3t3_hi_i_e*, Table 20 showed that adding the interactions with spatial coordinates resulted in removing some of the terms in *d3DMSr3t3_hi_i_e*, and the result was a model that gave credit to a different set of variables.

5 MODEL PERFORMANCE ASSESSMENT

This section assesses the performance of the final model. Figure 13 shows the ROC for each of the training-set and validation-set when the model was fit to the training-set. A model that contained only an intercept term, but no predictor variables, would have a ROC curve with two points, one at (0, 0), and one at (1, 1); the curve would be a straight line connecting these two points, and the AUC would be 0.5. When comparing two models, the one for which the AUC is greater, has greater sensitivity and specificity for a wider range of cut-off values (Hastie et al. 2001), and since our objective was to characterize the probability surface rather than to classify grid cells as having a nest or not having a nest, this was the metric we chose to assess model performance.

Table 19 showed that when the final model was fit to the training-set data and used to predict the validation-set data, the area under the ROC— that is, the validation AUC— was equal to 0.86. When the model fit to the training-set data was

Table 20. Terms in model *d3DMSr3t3hi_Si_e*, the final model.

Term	Estimate_*edf*	Std. Error_*rdf*	*p*-value
(Intercept)	-7.0	0	5.58E-155
te(dist_AH)	1.0	4	8.02E-14
ti(dist_AH,dist_canals)	4.6	16	2.45E-07
ti(dist_AH,dist_ENPrds)	7.9	16	1.99E-14
te(xCentroid)	2.12	4	4.46E-09
te(depth_cm)	1.0	4	5.96E-11
ti(depth_bp,depth_cm)	3.6	16	1.77E-04
te(temp_bp)	1.7	2	4.04E-05
ti(rain_bp,temp_bp)	1.7	4	1.63E-04
ti(rain_cm,temp_cm)	3.8	4	6.56E-12
ti(rain_cm,temp_nb)	1.0	4	1.07E-09
ti(depth_bp,temp_cm)	4.5	8	2.83E-12
ti(depth_nb,temp_cm)	1.6	8	7.94E-04
ti(depth_cm,temp_nb)	3.2	8	8.71E-06
ti(dist_ENPrds,temp_bp)	4.0	8	3.23E-06
ti(dist_AH,xCentroid,yCentroid)	14.8	64	2.51E-21
ti(dist_canals,xCentroid)	2.2	16	1.07E-06
ti(dist_canals,yCentroid)	1.9	16	5.21E-05
ti(dist_canals,xCentroid,yCentroid)	14.1	64	1.20E-34
ti(dist_ENPrds,xCentroid)	5.2	16	4.77E-13
ti(dist_ENPrds,xCentroid,yCentroid)	6.2	64	3.67E-07
ti(edge_pcent,yCentroid)	2.7	8	5.40E-06
ti(marsh_pcent,yCentroid)	5.7	12	4.57E-07
ti(marsh_pcent,xCentroid,yCentroid)	10.6	48	6.54E-09
ti(upland_pcent,xCentroid,yCentroid)	8.0	64	6.08E-06

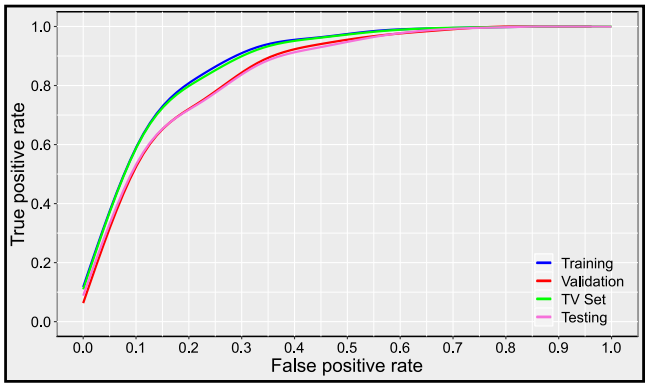


Figure 13. ROC curves for training-set and validation-set when the model was fit to the training-set and ROC curves for the TV-set and test-set when the model was fit to the TV-set.

used to predict the training-set data, the AUC was higher and equaled 0.88. Since the final model was fit to the training-set data, it does a better job predicting the training-set data than it does predicting the validation-set data. The difference in area between the two curves (Figure 13) is a measure of the degree to which the model overfit the training-set data. In addition to capturing a true signal, it also treated some of the noise in the training-set data as a signal, and it showed why we needed to use the AUC calculated on the validation-set data, not the training-set data, to make model-building decisions. Because both the training-set and validation-set data were used to build the model, use of the AUC based on either the training-set or validation-set to assess final model performance would produce overly optimistic results (Hastie et al. 2001, p. 200).

Therefore, to assess the performance of the final model, the training-set and validation-set are combined and referred to as the "TV-set". Final model *d3DMSr3t3hi_Si_e* is then re-fit to the TV-set. This refitting resulted in values of parameters estimated from the TV-set data being slightly different from what they were when the final model was fit to the training-set data, but the model terms used and values of *k* were the same. This re-fitted model was then used to predict the probability that *nest* = 1 for the test-set data, which was set aside and not used during model construction. The AUCs turned out to be the same for the fitting and verification data in the two cases (Table 21). The rank was same no matter which dataset was used to fit the model, because the rank is the number of columns in the model matrix. The *edf* increased when fitting the model using the TV-set. The TV-set most likely provides more information about signal and noise than did the training-set, since it was the training-set augmented by 50% additional observations. More effective columns were required to capture this additional information.

Table 21. Comparing final model AUCs for different training and validation datasets.

Case	AUC		AUC		Rank	*edf*
	Fitting data	Value	Verification data	Value		
Final Model	Training-set	0.88	Validation-set	0.86	479	114
†Final Model re-fitted	*TV-set	0.88	Test-set	0.86	479	148
*TV set = Training-set + Validation-set; † No terms were added or deleted during re-fitting model						

Figure 13 compares the ROC curves for the predictions to the different fitting and verification datasets. Given that the AUC values were the same for the training-set and TV-set, it was not surprising to see how close the green curve for the TV-set was to the blue curve for the training-set. Furthermore, just as the validation-set (red) and test-set (plum) have similar curves, the difference in area between the training-set and validation-set curves was similar to the difference in area between the TV-set and test-set curves. While the comparisons displayed in Table 21 and Figure 13 are interesting, the point of this exercise was to obtain the AUC for the test-set data for the purposes of quantifying the

performance of the final model. None of these results were used to modify any of the decisions made when building the model.

Table 22 gives the information in the ROC curves (Figure 13) in tabular form for the final model fitted to the TV-set and used to predict to the test-set. Sensitivity is the true positive rate expressed as a percent. Specificity is the true negative rate expressed as a percent, and accuracy is the percent of all predictions that were correct. The probability cutoff is *z*. To understand the information in this table, find the row where $z = 0.50$. If a probability cutoff of 0.5 is chosen, that means that any cell-year in the test-set data for which the predicted probability that *nest* = 1 was 0.5 or higher will be classified as having a nest. When these predicted classifications are compared to the actual values of nest for the observations in the test-set data, the statistics in the table can be calculated. In the table, at 0.5, sensitivity is equal to 0, which means that of all the observations in the test-set data that truly had *nest* = 1, 0% were classified as having *nest* = 1; specificity = 100, which means that of all the observations in the test-set data that truly had *nest* = 0, 100% were classified as having *nest* = 0; and accuracy = 97, which means that of all the observations in the test-set data, 97% were correctly classified.

Clearly, if classification was the goal, a probability cutoff of 0.5 would not be useful for finding nests. Recall that in the full dataset, 2.85% of the cell-years had nests, or close to 3%. Consider the probability cutoff of $z = 0.03$. In this row, the sensitivity is 84%, which means that of all the observations in the test-set data that had nests, 84% were classified as having a nest. The specificity is 74%, which means that of all the observations in the test-set data that did not have a nest, 74% were classified as not having a nest. The accuracy is also 74%, which means that of all the observations in the test-set data, 74% were classified correctly. Because the proportion of cell-years that have nests was very low, a low probability cutoff was needed to find nests, and specificity and accuracy are very close to each other for all probability cutoffs.

Use of this nest-building model is not to be based on classifying each cell-year as having a nest or not having a nest, however. For water management applications, it is sufficient to look at the probability surface for different scenarios, as Section 6 demonstrates.

Table 22. Sensitivity, specificity, and accuracy of final model (fit to the TV-set and predicting to the test-set) for probability cut-offs from 0.01 to 0.99.

Sensitivity	Specificity	Accuracy	*z*
94	56	57	0.01
90	67	67	0.02
84	74	74	0.03
80	79	79	0.04
76	83	82	0.05
48	92	91	0.10
25	97	95	0.15
13	98	96	0.20
6	99	97	0.25
2	100	97	0.30
1	100	97	0.35
0	100	97	0.40
0	100	97	0.45
0	100	97	0.50
0	100	97	0.55
0	100	97	0.60
0	100	97	0.65
0	100	97	0.70
0	100	97	0.75
0	100	97	0.80
0	100	97	0.85
0	100	97	0.90
0	100	97	0.95
0	100	97	0.99

6 APPLICATION

The purpose of this section is to use the model to examine differences in the probability that a nest will be built under different hydrological profiles because the Everglades is undergoing active hydrologic restoration. Hydrologic period-based (*BP* and *CM*; Figure 2) wet, dry, and typical years were selected from Figure 14 (based on *depth_bp*) and Figure 15 (based on *depth_cm*) to investigate predicted probabilities and prediction interval width of nesting. The variable *depth_bp* provides a yearly representation of wetness (Figure 2). This variable influences alligator body condition, since body condition depends on food availability and mobility in the marsh, though other variables in the model also influence nesting. The variable *depth_cm* shows how suitable conditions are (Figure 2) during the breeding season for courtship and mating leading to subsequent nesting.

For the determination of wet, dry and typical years, quantiles (Q) of *depth_bp* and *depth_cm* for the period of record (POR) were compared with quartiles of individual years to find a dry year ($max\{abs[Q1(t) - Q1(POR)]\}$; $Q1[t] < Q1[POR]$, $t = year$), a wet year ($max[Q3(t) - Q3(POR)]$; $Q3[t] > Q3[POR]$), and a typical year ($min\{abs[Q2(POR) - Q2(t)]\}$).

6.1 Nesting under wet, dry, and typical years

Based on *depth_bp* (Figure 14), a wet year was 1996; a dry year was 2012; and typical years were 1993 and 2003. Similarly, based on *depth_cm* (Figure 15), a wet year was 1995, a dry year was 2011, and a typical year was 2015.

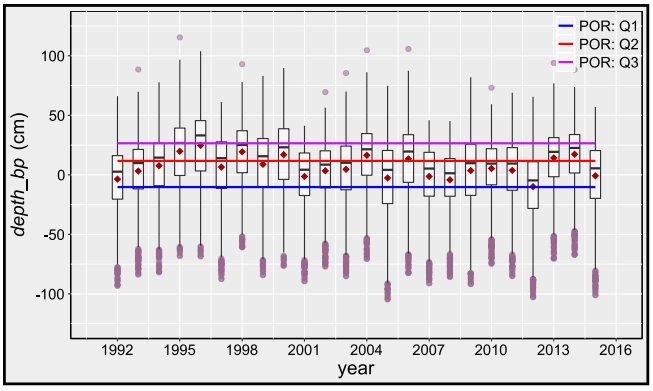


Figure 14. Boxplots of *depth_bp* (cm) vs. year

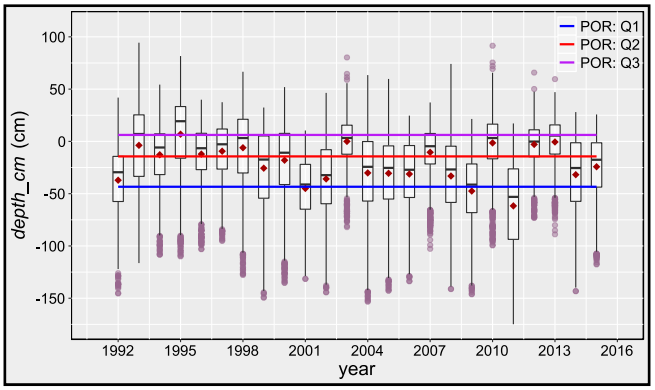


Figure 15. Boxplots of *depth_cm* (cm) vs. year.

6.1.1 Maps of depth_bp and depth_cm

Figure 16(a-d) shows maps of *depth_bp* for wet-1996, dry-2012, and typical-1993 and 2003 years based on the *depth_bp* criterion (Figure 14). Figure 17(a-c) show maps of *depth_cm* for wet-1995, dry-2011, and typical-2015 years based on the *depth_cm* criterion (Figure 15). The bracketed numbers in the legend indicate the percent of grid cells for that range in all the maps. Darker blue and purple colors indicate wetter conditions, while peach and dark red indicate drier conditions. The map for 1996 (Figure 16[a]) is relatively darker blue and purple, with drier upland cells on the eastern border of ENP in the Rocky Glades (RG) and in the Long Pine Key (LPK) areas on either side of the Main Park Road (Ingraham Highway).

On the 1993 and 2003 maps, which represent typical years, the drier regions in RG expanded farther west into the freshwater marl prairie on the eastern border of upper Shark Slough (USS) and lower Shark Slough (LSS) basins, also expanding in the Taylor Slough (TS) and Panhandle (PH) basins (see Figures 16[c, d]). Some of the darker blue areas in the East Slough (ES) on the border between ENP and BCNP (Big Cypress National Preserve) are now a lighter shade of blue.

On the 2012 dry year map (Figure 16[b]), about 54% of cells had water levels below ground (peach and dark red areas), and they expanded even farther from the drier areas (~36%) in the 1993 and 2003 map into the East Slough. Year 2012 was the driest year in the 24-year span of data used (based on *depth_bp*; Figure 16[b]).

The map for 1995 (Figure 17[a]) shows the wettest conditions during the courtship and mating period but there are still drier conditions (~35%) in the RG, LPK, TS and PH basins. Year 2011 had the driest courtship and mating period (Figure 17[b]) where nearly all the SRF area (95%) was dry (<0 cm water depth).

In the typical year of wetness (2015; Figure 17[c]) for courtship and mating, only the upper (USS) and lower Shark Slough (LSS), some areas adjacent to the L-67 EXT canal in northeast Shark Slough (NESS), southern East Slough, central Taylor Slough, and southern Panhandle appear wet (~23%).

6.1.2 Predicted probabilities (*nest* = 1) and 95% prediction interval widths

Prediction interval here means there was a 95% probability that *nest* = 1 fell within the prediction interval, i.e., a 2.5% probability that *nest* = 1 fell below the prediction interval, and a 2.5% probability that *nest* = 1 fell above the prediction interval. The scales are the same on all maps (Figures 18, 19, 20 and 21) and they range from 0 to 70% (the highest is 66.2% during the POR) for the predicted probability and 0 to 100% for the prediction interval. The darker blue and purple in the probability maps correspond to the highest probabilities and the peach and dark red correspond to the lowest. Darker blue, purple, and green in the 95% prediction interval maps indicate wider prediction intervals, and thus greater uncertainty in the predicted probabilities, while peach, magenta, and dark red indicate very narrow prediction interval widths and thus more certainty in the predicted probabilities.

Tables 23 and 24 summarize predicted probabilities and 95% prediction interval widths, respectively, for hydrological basins corresponding to the years reported in Figures 18, 19, 20 and 21.

Though 1996 had the wettest breeding potential period, and 2012 had the driest, the spatial distribution of the probabilities for these two years look visually similar to some extent and also similar to 1993, a typical year (Figures 18[a, c] and 19[a]). Year 2003 was also a typical year, but does not appear to have the same spatial distribution as 1993 (Figure 19[a, c]).

In the four years in which we considered breeding potential period (Table 23), in 1993 (typical year), LSS had the highest

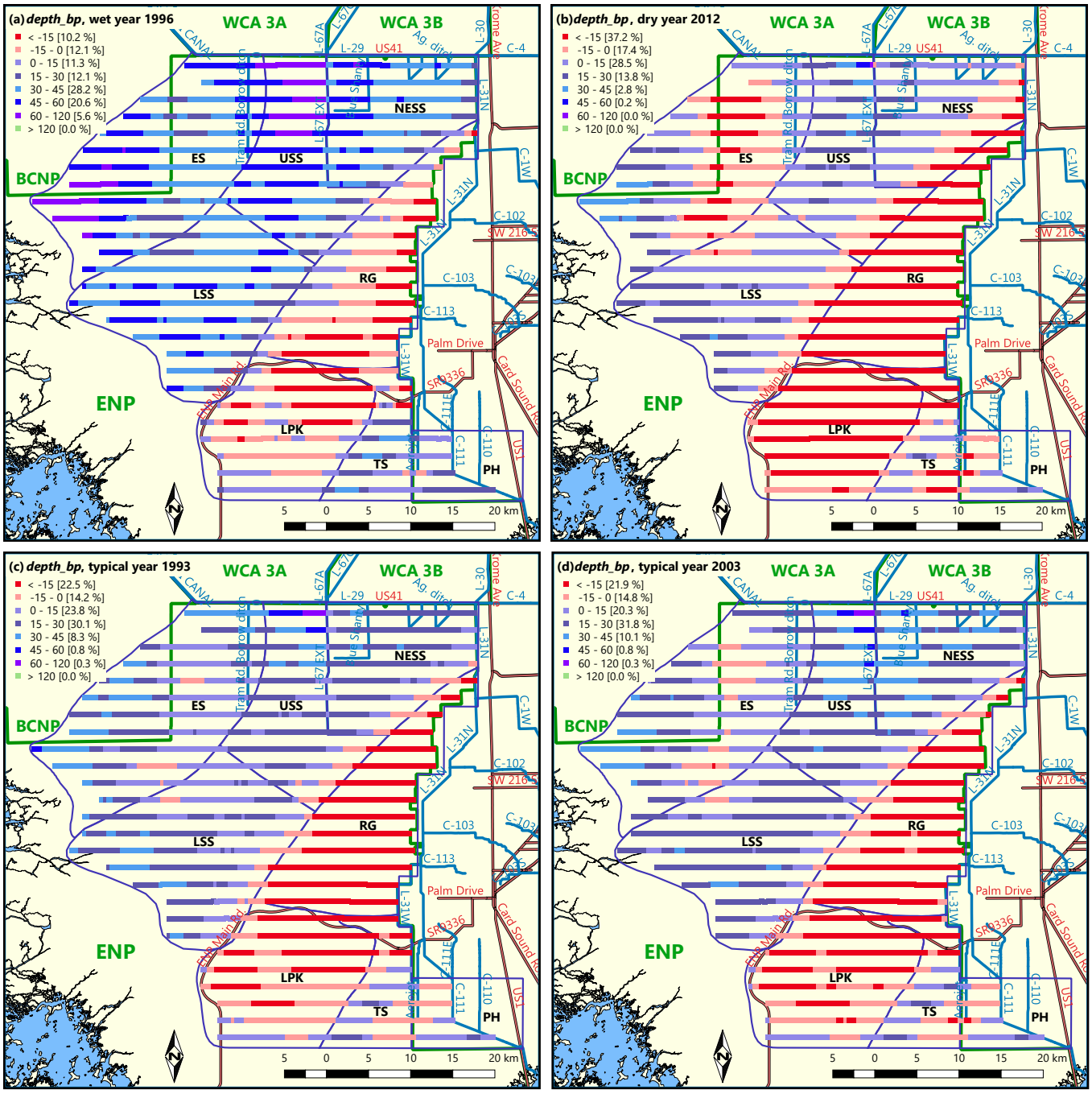


Figure 16a, 16b, 16c, and 16d. Map of *depth_bp* (cm) values in a wet year (1996), a dry year (2012), and typical years (1993 and 2003) along SRF transects. Numbers in brackets in the legend show percent of grid cells in that category.

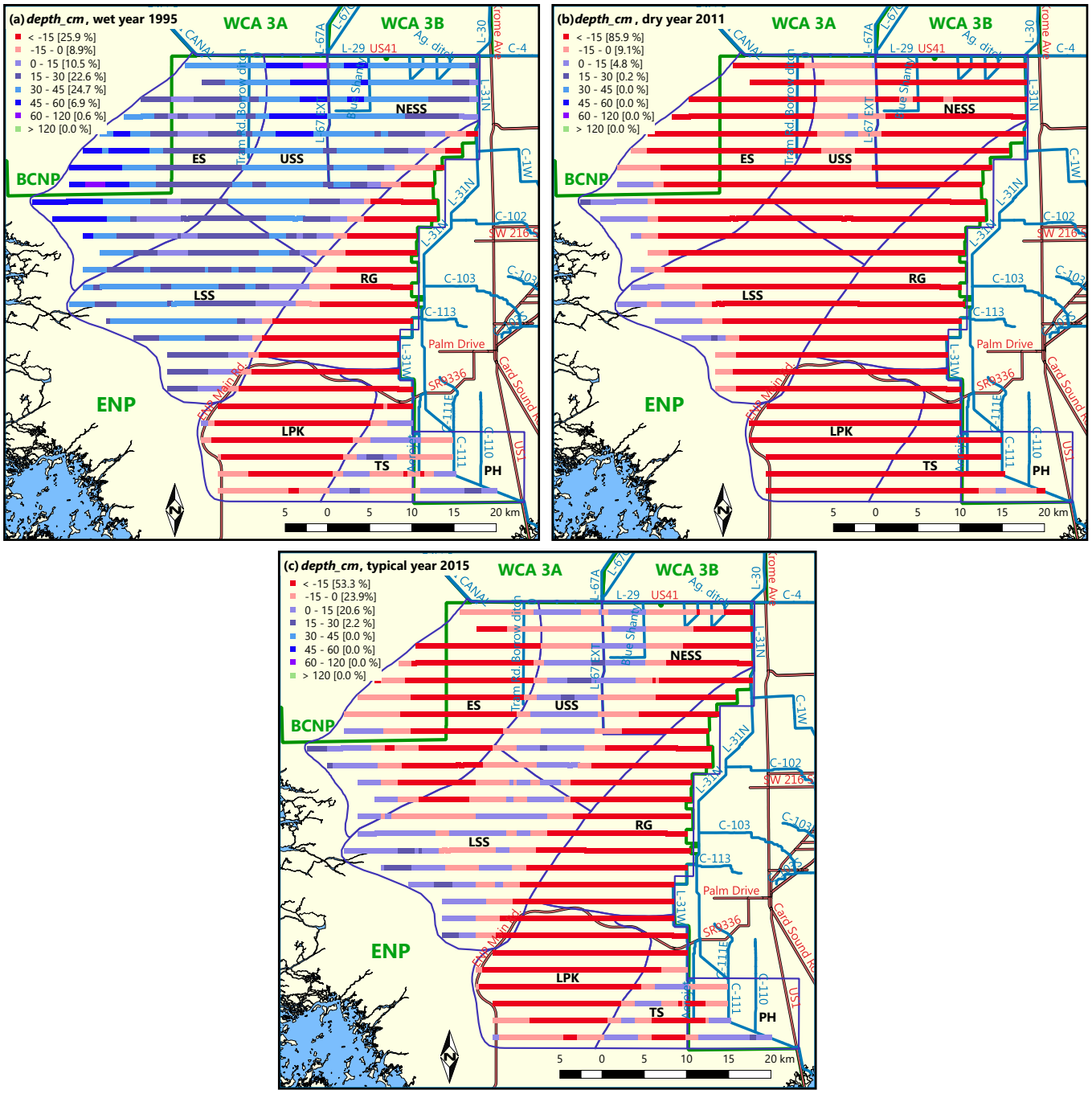


Figure 17a, 17b, and 17c. Map of *depth_cm* (cm) values in a wet year (1995), a dry year (2011), and a typical year (2015) along SRF transects. Numbers in brackets in the legend show percent of grid cells in that category.

Table 23. Predicted mean probabilities (reported in percent) in hydrological basins for mapped years; Figures 16 and 17.

Case	Year	ES	USS	LSS	NESS	RG	TS	LPK	PH
All data	POR§	3.3	5.7	6.7	1.4	0.1	1.9	0.2	0.3
a. *depth_bp*; wet	1996	3.8	7.5	7.6	1.6	0.1	1.5	0.2	0.2
b. *depth_bp*; dry	2012	3.9	6.3	8.6	1.5	0.3	3.2	0.4	0.5
c. *depth_bp*; typical	1993	4.5	9.2	9.4	2.7	0.1	2.4	0.2	0.2
d. *depth_bp*; typical	2003	1.7	3.3	3.1	1.3	0.2	3.1	0.5	0.4
a. *depth_cm*; wet	1995	2.4	7.0	6.7	1.3	0.2	0.9	0.3	0.1
b. *depth_cm*; dry	2011	2.8	4.0	5.5	1.0	0.0	1.1	0.1	0.3
c. *depth_cm*; typical	2015	4.5	10.4	9.2	2.4	0.4	3.1	0.7	0.3
Basins- ES: East Slough, USS: Upper Shark Slough, NESS: Northeast Shark Slough, LSS: Lower Shark Slough, RG: Rocky Glades, TS: Taylor Slough, LPK: Long Pine Key, and PH: Panhandle; § Period of Record									

Table 24. 95% mean prediction interval widths (reported in percent) in hydrological basins for mapped years; Figures 16 and 17.

Case	Year	ES	USS	LSS	NESS	RG	TS	LPK	PH
All data	POR§	40.9	36.0	35.0	54.2	68.2	63.3	73.7	88.9
a. *depth_bp*; wet	1996	39.6	35.2	33.2	53.1	64.4	61.7	71.5	88.8
b. *depth_bp*; dry	2012	38.7	33.4	33.2	53.5	66.9	62.3	72.8	88.9
c. *depth_bp*; typical	1993	38.8	35.9	33.7	53.6	65.5	61.9	71.9	88.8
d. *depth_bp*; typical	2003	40.7	36.6	36.9	53.3	65.7	62.1	71.9	88.8
a. *depth_cm*; wet	1995	46.9	40.9	38.2	55.0	68.2	65.2	74.5	89.3
b. *depth_cm*; dry	2011	43.5	36.9	38.8	56.3	86.3	66.2	79.5	88.9
c. *depth_cm*; typical	2015	40.2	34.8	33.3	54.3	66.8	62.6	72.9	88.8
Basins- ES: East Slough, USS: Upper Shark Slough, NESS: Northeast Shark Slough, LSS: Lower Shark Slough, RG: Rocky Glades, TS: Taylor Slough, LPK: Long Pine Key, and PH: Panhandle; § Period of Record									

mean (9.4%). Other basins (ES, USS and NESS) also had higher mean probabilities in 1993. Among the basins (ES, USS, LSS) showing higher mean nesting probabilities, ES showed the least variability (1.4 std dev; Appendix: II), but LSS in general had a higher mean probability (Table 23 and Appendix: II) and a tighter prediction interval width (Table 24 and Appendix: II) during most years. NESS, RG, TS, LPK and PH basins had lowest probability and maximum uncertainty (Tables 23 and 24, respectively).

Examination of Figures 20(a, c) and 21(a) for the years selected on the basis of courtship and mating water depths (*depth_cm*) show higher probabilities in Shark Slough in the typical year of 2015 followed by the wet year of 1995 and dry year of 2011. The means in Table 23 provide a numeric confirmation of visual observations and shows that USS had the highest mean probability (10.4 in 2015) followed by LSS (9.4 in 1993) and then ES (4.5 in 1993 and 2015). Table 24 shows that LSS had a tighter prediction interval width (33.2 in 1996 and 2012) followed by USS (33.4 in 2012) and ES (38.7 in 2012).

An examination of *depth_bp* for wet, dry, and typical years (Figure 16[a-d]) along with predicted nesting probabilities for the same years (Figures 18 and 19) shows higher probability of nesting in areas that are wet most of the time. Central Shark Slough and central Taylor Slough (Figure 1) are more likely to be wet most of the year. It is interesting to note that nesting probabilities (Figures 18-21) were higher between the lower end of the L-67 EXT canal and the Tram Road borrow ditch. The presence of alligator holes (Figure 1) in this area and proximity to canals had a positive influence as alligators congregate here when there is less water in the marsh.

Because the overall proportion of cell-years in the data that had a nest was so small— 0.0285, or 2.85%— the narrowest, and thus most useful, prediction interval widths correspond to the highest predicted probabilities. Put another way, the cell-years for which there is the most certainty in predictions tend to also be the ones that have higher probabilities of having a nest. That said, the area where there are high predicted probabilities and also narrow prediction intervals is the lower part of Shark Slough (LSS) followed by upper Shark Slough (USS) (Tables 23 and 24, and Figures 18-21).

Figures 16(c, d) show hydrologically typical years (based on *depth_bp*) 1993 and 2003 and the predicted probabilities for the same are displayed in Figure 19(a, c). Here, 1993 shows higher probabilities in Shark Slough and 2003 has a minimal response. Table 23 shows USS having 9.2% and 3.3% and LSS having 9.4% and 3.1% mean probabilities in 1993 and 2003, respectively. Thus, representative years may not be representative (based on a specific criterion as in our case- dry, wet, and typical years) and one should look at the

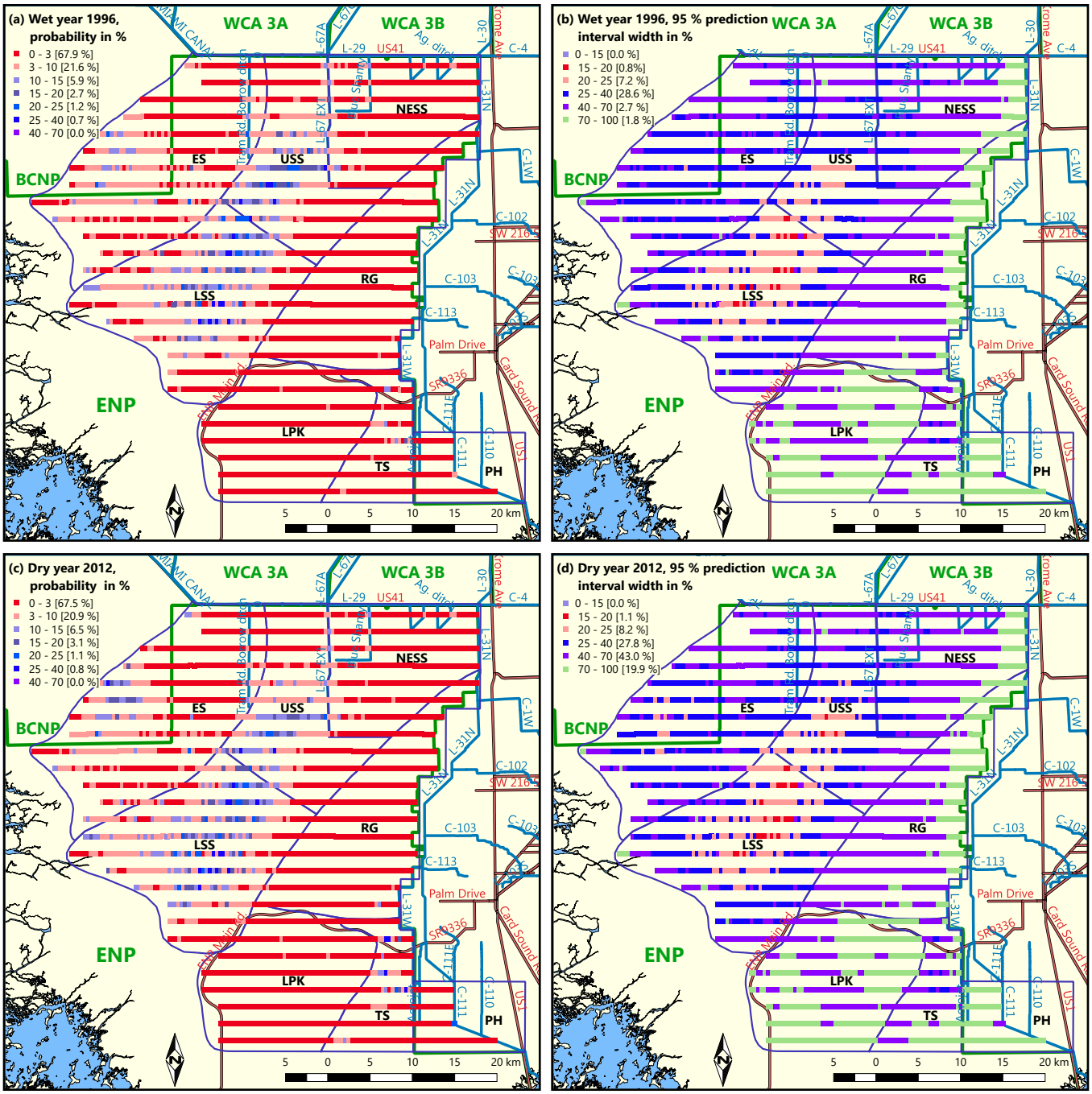


Figure 18a, 18b, 18c, and 18d. Map of predicted probability (*nest* = 1) and prediction interval widths corresponding to Figure 16(a, b) conditions. Numbers in brackets in the legend show percent of grid cells in that category.

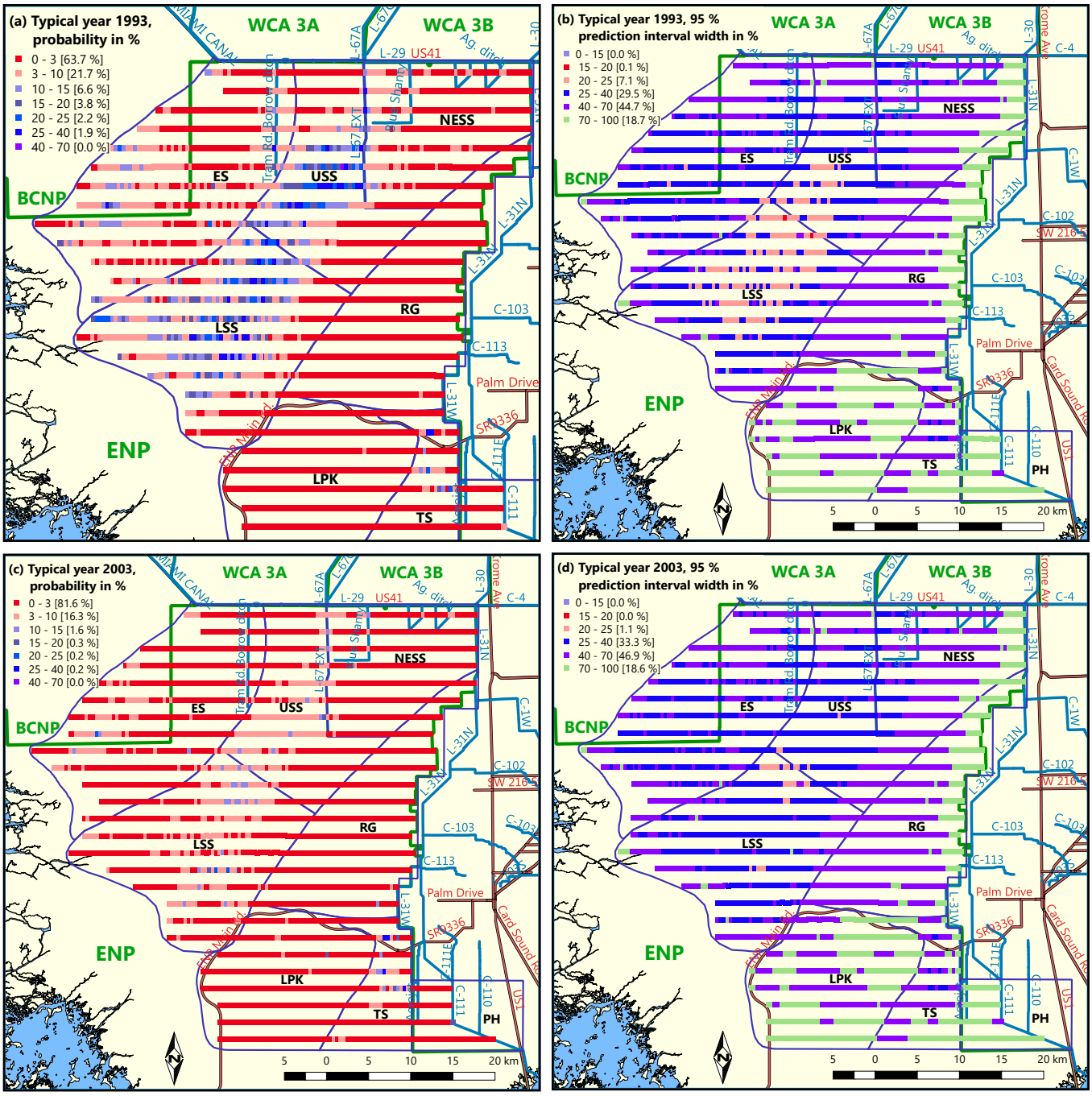


Figure 19a, 19b, 19c, and 19d. Map of predicted probability (*nest* = 1) and prediction interval widths corresponding to Figure 16(c, d) conditions. Numbers in brackets in the legend show percent of grid cells in that category.

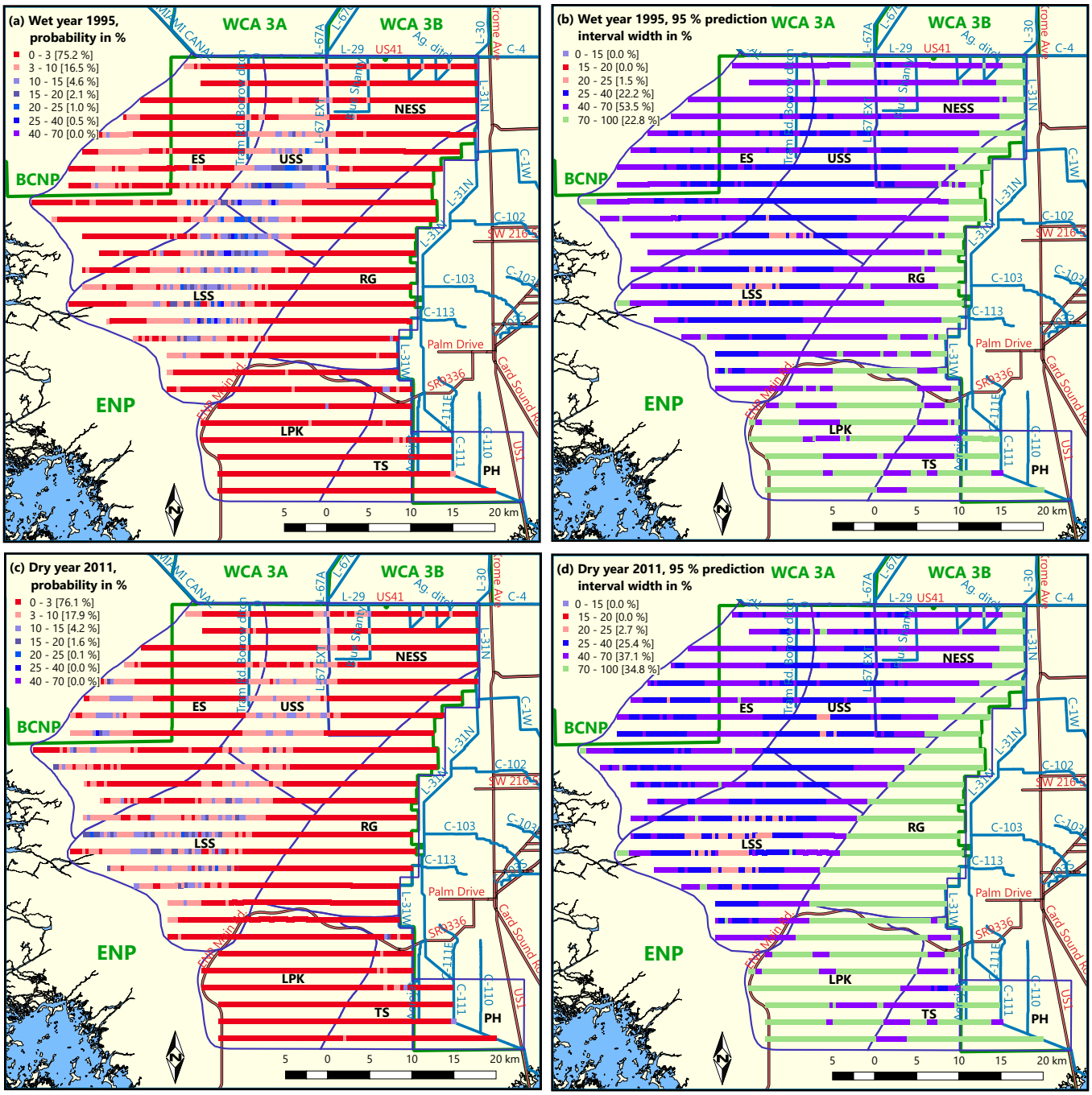


Figure 20a, 20b, 20c, and 20d. Map of predicted probability (*nest* = 1) and prediction interval widths corresponding to Figure 17(a, b) conditions. Numbers in brackets in the legend show percent of grid cells in that category.

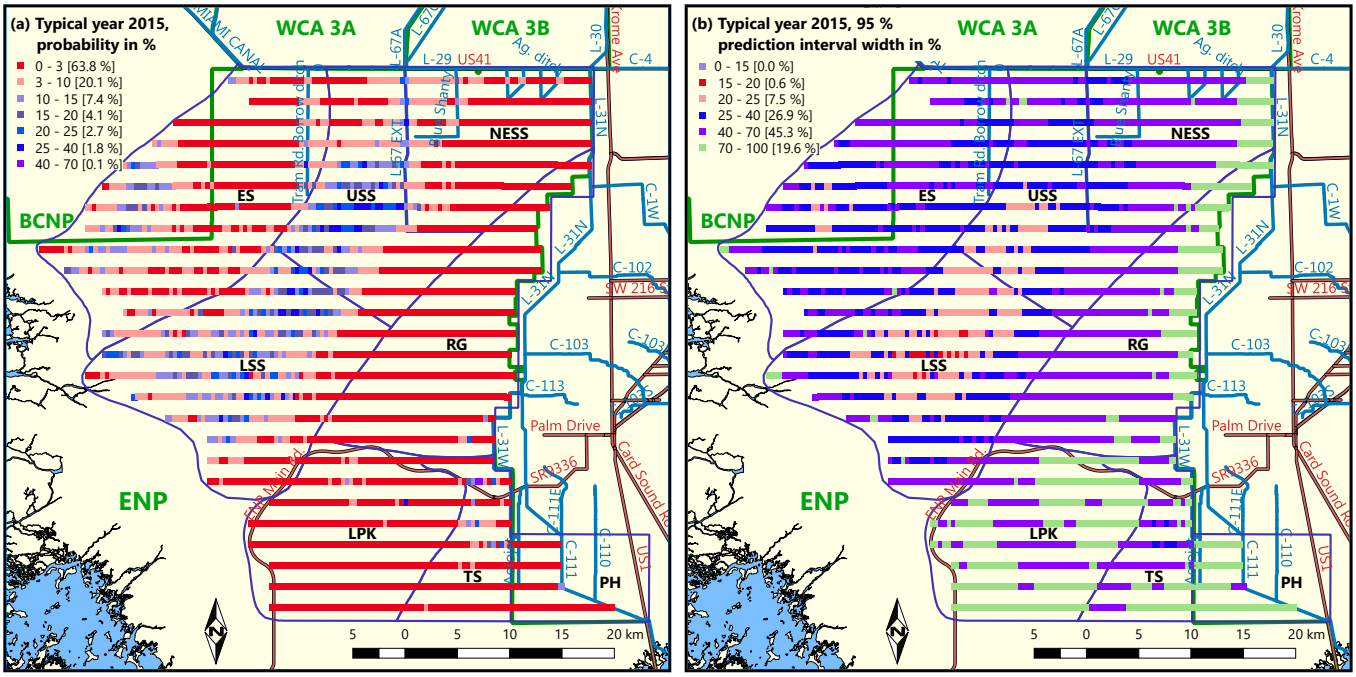


Figure 21a and 21b. Map of predicted probability (*nest* = 1) and prediction interval widths corresponding to Figure 17(c) conditions. Numbers in brackets in the legend show percent of grid cells in that category.

full range of outputs/results because other factors also play a role. In this case, it appears to be the influence of spatial scale used to select the representative years. The whole spatial SRF model domain was used (Figures 14 and 15) in selecting representative years. The local habitat conditions seem to play a role in expected probability of nesting. Local basin conditions appear to have a major influence on water depths in basins where a major difference in predicted probabilities can be observed during typical years based on *depth_bp* (Table 23; Figure 19[a, c]). Mean basin water depths (Table 25) during breeding cycle periods (Figure 2) show quite a difference between 1993 and 2003. Results of predicted probabilities for all years are provided in Appendix: II.

Table 25. Mean water depths (cm) during typical years breeding cycle periods in hydrological basins (Figure 1).

Breeding cycle period	Year	Basin		
		USS	LSS	ES
BP	1993	21.4	15.7	19.1
	2003	26.6	17.8	13.4
CM	1993	26.3	9.6	17.8
	2003	18.6	10.0	0.7
NB	1993	33.6	27.1	28.4
	2003	40.0	30.4	22.2

In general, either based on *BP* or *CM* periods, considering the higher nesting probability regions (USS, LSS, ES), typical years had higher probability followed by wet years and lastly by dry years. The only exception was typical year 2003, which had the lowest probability (Table 23). It appears that in typical years (1993 and 2003), local conditions do affect breeding potential, courtship and mating, and subsequent probability of nesting.

6.2 Nesting under Combined Operational Plan - Alternative-Q

The model was applied to assess the change in nesting pattern with simulated water management operational changes influencing the hydrological regime in ENP. Figure 22 shows simulated hydrological conditions of the Combined Operational Plan (COP) – Alternative Q for Everglades restoration conducted by the SFWMD with the Regional Simulation Model (SFWMD 2005; https://www.sfwmd.gov/science-data/rsm-model). The Regional Simulation Model simulated operations of the South Florida Water Management System using climate data from 1965 to 2005, assuming operational changes proposed in Alternative-Q of COP (COP-Alt Q) are implemented (USACE 2020). Specific years shown in Figure 22 are for comparison with maps in Figures 16(a, c, d) and 17(a) which show water depths based on interpolated observed data from EDEN (used in model development). COP-Alt Q is one possible water management possibility out of several alternative hydrological scenarios evaluated for Everglades restoration and is used here only for the purpose of an example application of this alligator model.

With restoration of the Everglades, a significant

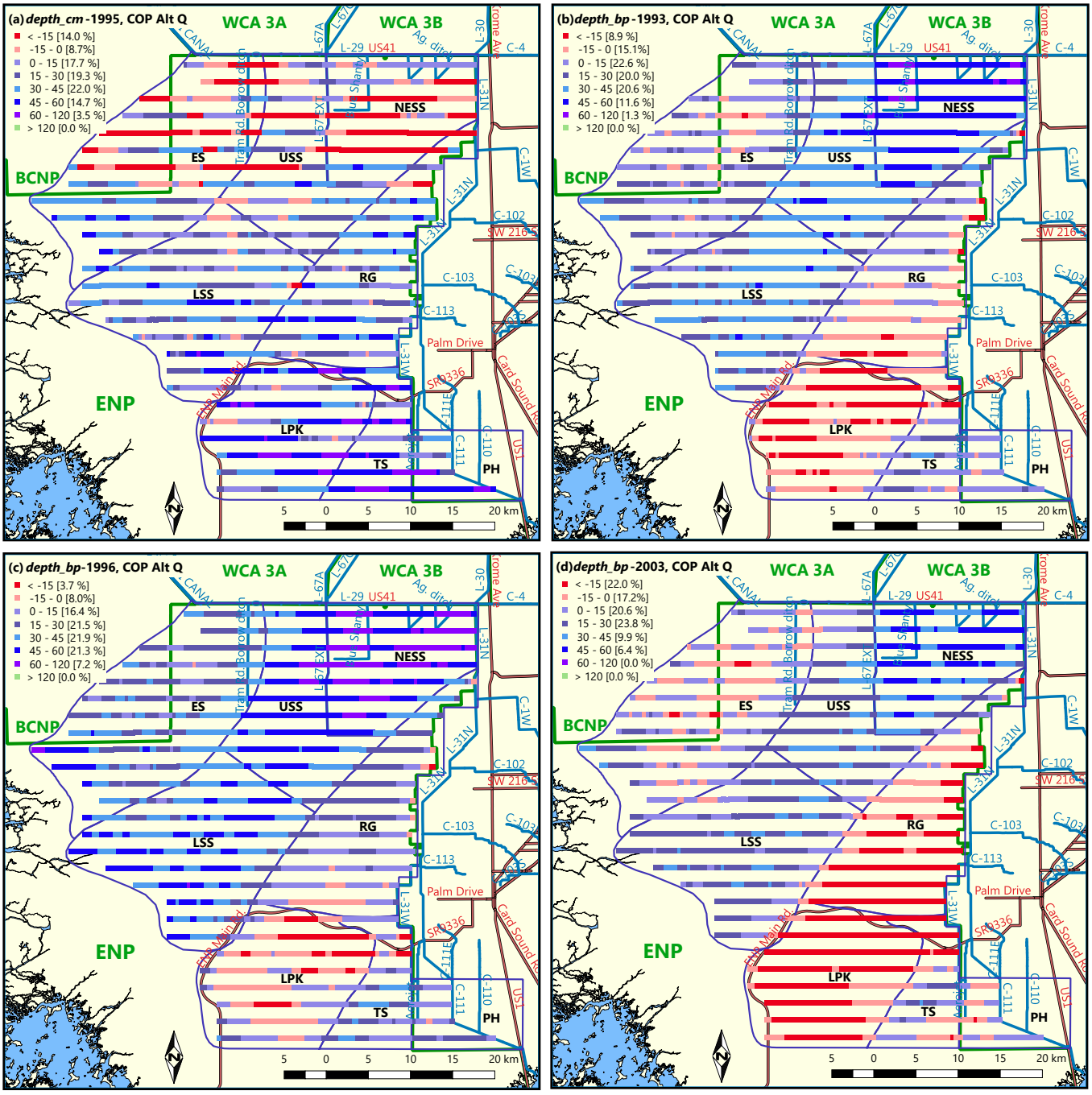


Figure 22a, 22b, 22c, and 22d. Map of water depth values (cm) along SRF transects in 1995 (*depth_cm*) and 1993, 1996, and 2003 (*depth_bp)* obtained from Alternative-Q of COP simulation for comparison with maps in Figures 16 and 17.

redistribution of water is expected in ENP. Comparing *CM* period in 1995 for COP-Alt Q in Figure 22(a) with Figure 17(a) shows that NESS and ES will be drier. However, RG, LPK, and TS appear to become substantially wetter (Figure 22[a]). Comparing *BP* period in 1996 and 1993 for COP-Alt Q in Figure 22(c, b) with Figure 16(a, c), respectively, shows NESS, RG, TS, PH, USS and LSS to be relatively wetter. However, a comparison of year 2003 in these figures shows only NESS, USS and LSS to become relatively wetter and ES relatively drier under COP-Alt Q. To examine more closely the differences in water depths spatially (in basins) and temporally (in *BP*, *CM*, and *NB* time periods; Figure 2) under COP-Alt Q and EDEN, a quantitative assessment is provided in Figures 23-25 as differences were not as discernible using maps. COP-Alt Q had consistently lower water depths in ES, higher in NESS and RG, and some years higher in other basins during *BP* period (Figure 23). Figures 24 and 25 show temporal distribution of water depths in COP-ALT Q comparison with EDEN during *CM* and *NB* periods in different basins. Water depths were consistently high under COP-ALT Q in eastern and southern peripheral basins of RG, TS, LPK, and PH and consistently lower or equal in ES, USS, LSS, and NESS during *CM* and *NB* periods.

Alligator habitat may alter hydrologically in all basins under simulated hydrological conditions for COP-Alt Q, which may affect alligator nesting in these basins. A quantitative assessment of change in area under different probability ranges of nesting is presented (Table 26) for EDEN and COP-Alt Q hydrological regimes. The influence varies with year. For example, in years 1995 and 1996, COP-Alt Q showed less grid-cells in '0 – 3'% probability, whereas years 1993 and 2003 showed more grid cells in '0 – 3'% probability. This lowering in percent probability during 1993 and 2003 was offset by increased percent in higher probability categories in COP-Alt Q. In general, a very small percent probability increase was observed in the very high category, >40% probability range, during years shown in Table 26 under COP-Alt Q where the EDEN hydrologic conditions showed near 0% probability. This change in probability of nesting spatially shows the influence of redistribution of water hydrologically under simulated COP-Alt Q conditions.

Table 27 shows mean probability in percent for two hydrologic scenarios in different years organized by basins. Application of the model to COP-Alt Q altered hydrologic scenario predicts that probability of nesting may change in these basins during all the years reported here. Rocky Glades, LPK and PH basins have very low observed nesting as also predicted by the model under EDEN hydrology used for model development (Table 27) in all years. Under COP-Alt Q expected hydrology (1992-2005), the model predicted that in general nesting increased in TS, decreased by small amount in ES, USS, LSS, and NESS, and slightly increased in RG, LPK, and PH basins (Table 27). Water distribution in COP-Alt Q spatially (in basins) and temporally (*BP*, *CM*, and *NB* time periods; Figures 23-25) influenced predicted probabilities.

7 DISCUSSION

This report has presented the development of a model of the probability of an alligator building a nest in a cell-year as a function of non-time-varying habitat variables, and of water depth, rain, and temperature from each of three time periods— the breeding potential (*BP*), courtship and mating (*CM*), and nest building (*NB*) time periods, defined in Figure 2. All the variables considered are believed by American alligator experts to impact some aspect of the alligator breeding cycle, whether through contributions to the overall wellness of the alligator, to the logistics of mating, or to the suitability of a location for a nest. The final model contains the following variables either as a main effect, an interaction, or both: *dist_AH*, *dist_canals*, *dist_ENPrds*, *depth_bp*, *depth_cm*, *depth_nb*, *rain_bp*, *rain_cm*, *temp_bp*, *temp_cm*, *temp_nb*, *marsh_pcent*, *edge_pcent*, *upland_pcent*, *xCentroid*, and *yCentroid*.

Though it is tempting to do so, it is important to think carefully before attributing cause-effect interpretations to the results of the variable elimination process during the model building process. Different modelers may choose different training, validation, and test datasets, or use a different decision-making process, and therefore could get a model with equivalent performance that has different variables in it. For example, *rain_nb* was not present in the final model despite *rain_nb* being an important driver for alligator nesting at specific locations.

In the APSI model (Shinde et al. 2014), grid cells containing canals were classified as "excluded" from alligator habitat. Canals do not provide suitable habitat for juvenile alligators and are typically inhabited by adult alligators. Canals may also act as reproductive sinks. Chopp (2003) noted in areas adjacent to some canals, nests may experience rapid and extreme changes in water depths during incubation resulting in reduced nest success and increased hatchling mortality. Our analysis of the SRF data (Section 4.5.1) and the presence of *dist_canals* in the present final model showed that the proximity (Figure 4[a]) to a canal is predictive (in combination with *xCentroid* and *yCentroid* – as interaction effect; Table 20) of the probability of a nest being built. The variables *dist_canal* and *dist_ENPrds* has sizable equal correlation with *xCentroid* (-0.73) and between themselves (-0.71)— and all three of these variables remained in the final model.

Similarly, *dist_AH* (distance to alligator holes; Figure 3) and *dist_ENPrds* (distance to roads; Figure 4[b]) were also found to be predictive (Table 20). Both canals and roads represent an anthropogenic influence as they are not part of alligator

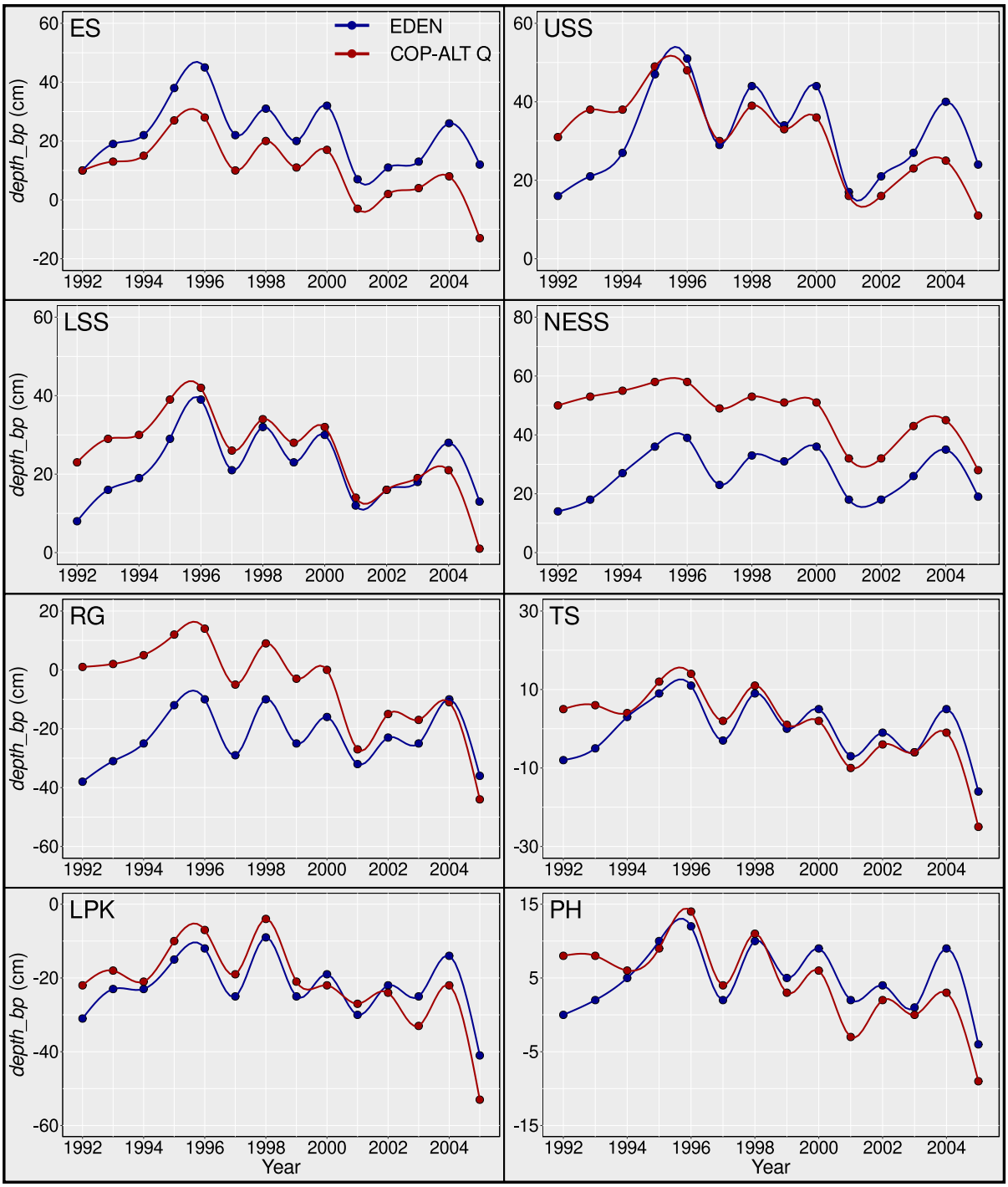


Figure 23. Mean water depths during *BP* period in basins.

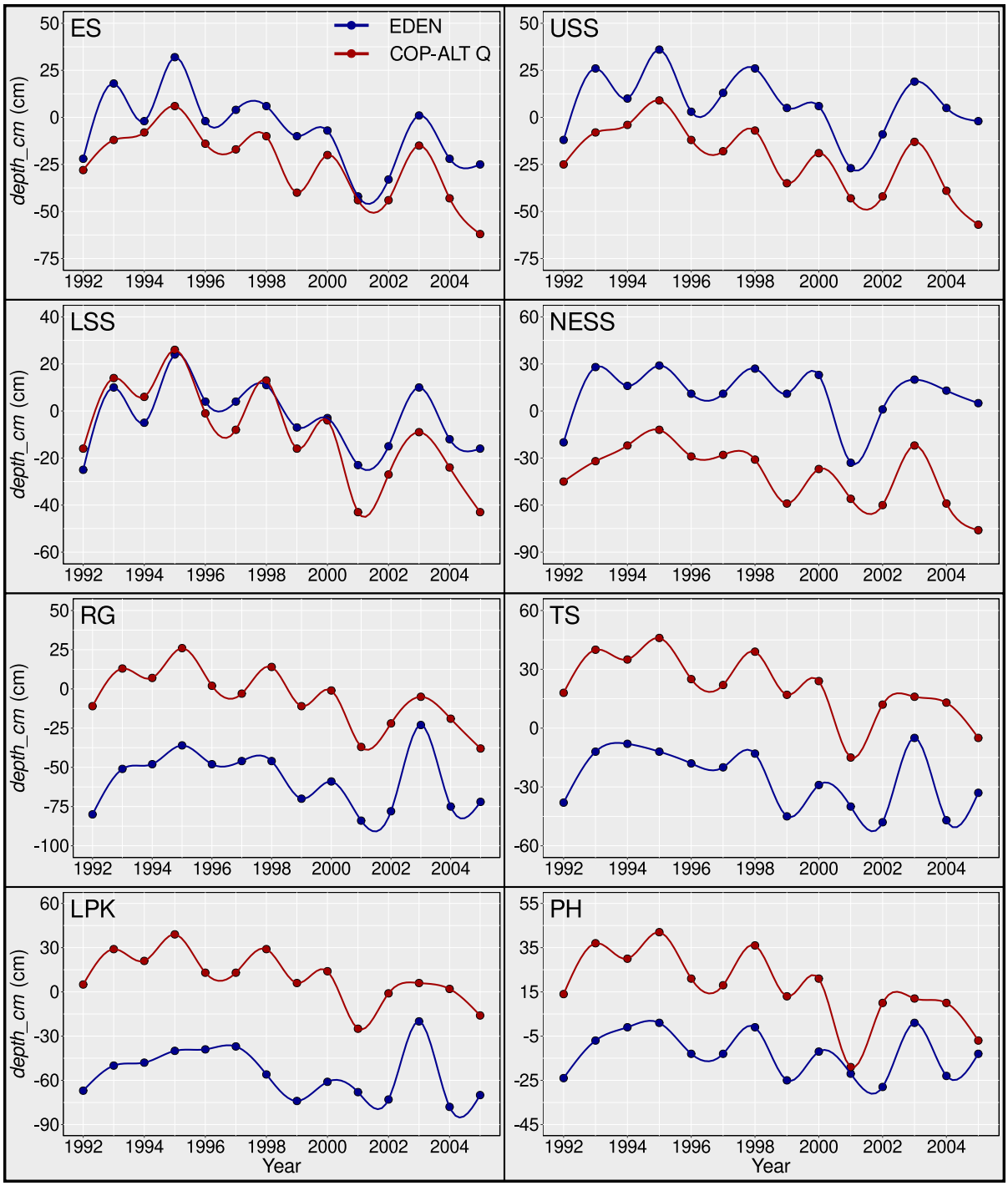


Figure 24. Mean water depths during *CM* period in basins.

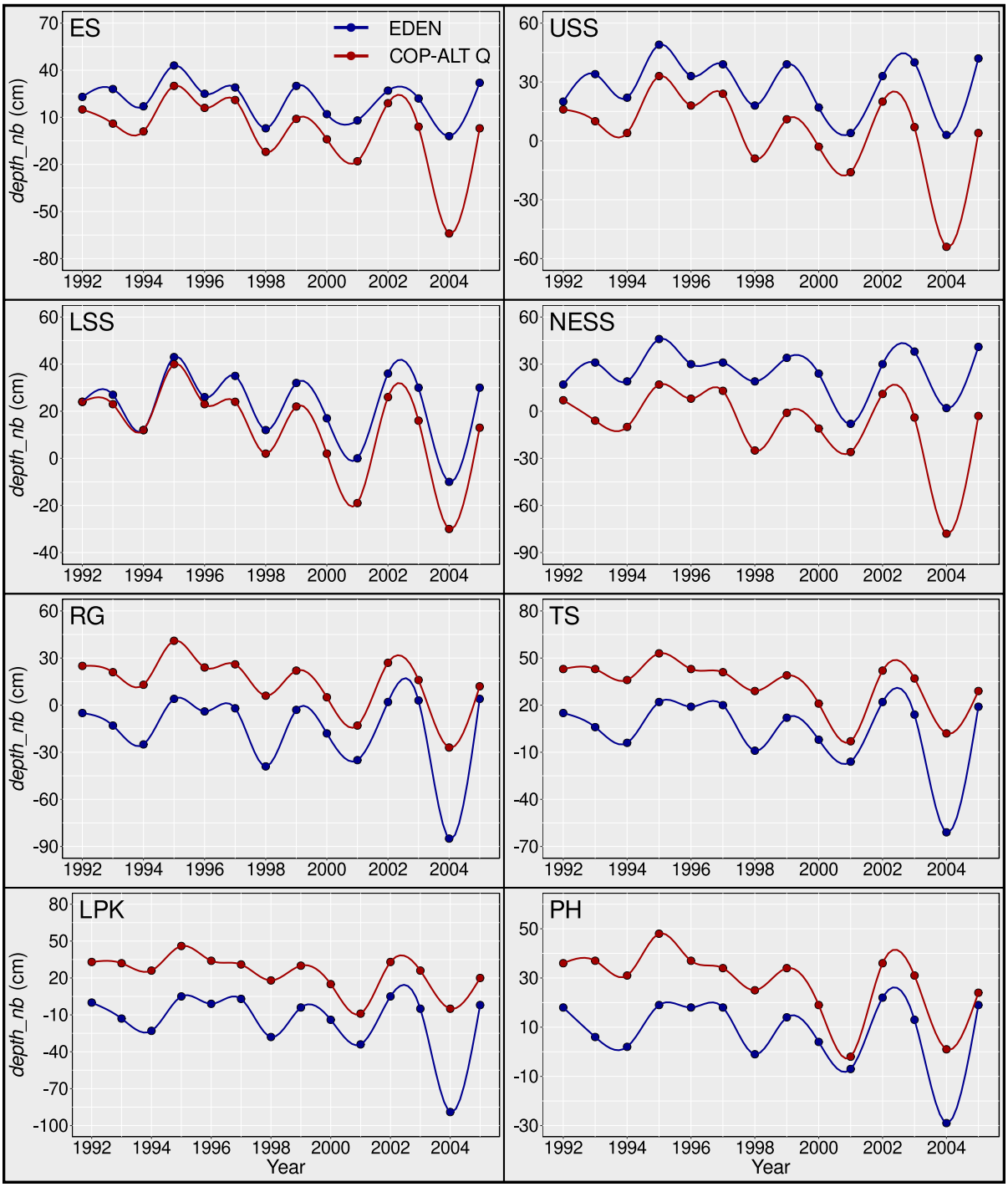


Figure 25. Mean water depths during *NB* period in basins.

Table 26. Comparison of grid-cell percent under different nest probability (reported in percent) ranges for two hydrological scenarios- EDEN (used for model development) and COP-Alt Q (expected hydrology).

Nest probability in %	1995		1993		1996		2003	
	EDEN	COP-Alt Q	EDEN	COP-Alt Q	EDEN	COP-Alt Q	EDEN	COP-Alt Q
0 – 3	75.25	66.49	63.69	66.24	67.87	66.49	81.59	83.70
3 – 10	16.52	19.02	21.73	23.93	21.65	22.21	16.26	14.23
10 – 15	4.61	5.95	6.64	4.96	5.86	5.48	1.55	1.25
15 – 20	2.11	3.75	3.84	2.41	2.67	3.10	0.26	0.30
20 – 25	1.03	2.07	2.16	1.25	1.21	1.68	0.17	0.09
25 – 40	0.47	2.16	1.90	0.69	0.69	0.86	0.17	0.26
> 40	0.00	0.56	0.04	0.52	0.04	0.13	0.00	0.18

Table 27. Comparison of mean nest probability (reported in percent) for two hydrological scenarios- EDEN (used for model development) and COP- Alt Q (expected hydrology) for different years in hydrologic basins.

Basin	Hydrology	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
ES	EDEN	2.6	4.5	3.9	2.4	3.8	2.6	2.9	1.4	2.1	2.0	2.8	1.7	2.6	3.5
	COP-Alt Q	3.7	4.3	4.1	2.4	3.0	1.6	2.1	1.1	2.0	1.9	2.0	1.4	2.3	3.0
USS	EDEN	4.8	9.2	7.2	7.0	7.5	5.5	6.9	2.4	3.3	3.4	6.0	3.3	3.5	3.0
	COP-Alt Q	2.4	3.2	8.3	10.9	8.0	4.1	4.0	1.9	2.8	2.5	3.7	2.5	2.6	2.1
LSS	EDEN	4.6	9.4	8.3	6.7	7.6	5.4	6.5	3.4	4.5	4.0	6.5	3.1	5.3	8.5
	COP-Alt Q	4.2	7.4	11.3	9.6	8.3	5.2	7.2	3.2	4.1	2.8	5.5	2.6	4.9	7.6
NESS	EDEN	0.9	2.7	1.7	1.3	1.6	1.1	1.6	0.4	0.7	0.7	1.6	1.3	0.8	1.0
	COP-Alt Q	0.3	0.6	2.0	4.6	2.1	1.0	0.7	0.3	0.6	0.4	1.2	1.2	0.5	0.4
RG	EDEN	0.1	0.1	0.2	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.2	0.1
	COP-Alt Q	1.0	1.0	0.6	0.4	0.3	0.3	0.4	0.1	0.4	0.2	0.3	0.3	0.4	0.1
TS	EDEN	3.0	2.4	2.2	0.9	1.5	1.0	1.6	0.5	1.4	1.3	1.3	3.1	1.7	1.6
	COP-Alt Q	9.4	7.7	5.1	2.2	3.3	3.3	6.3	1.7	3.8	2.3	3.8	5.4	4.6	2.7
LPK	EDEN	0.2	0.2	0.2	0.3	0.2	0.1	0.0	0.1	0.1	0.1	0.1	0.5	0.2	0.2
	COP-Alt Q	2.5	2.9	1.5	0.5	0.8	0.6	1.3	0.3	0.8	0.4	0.8	1.2	1.1	0.5
PH	EDEN	0.4	0.2	0.4	0.1	0.2	0.1	0.2	0.0	0.2	0.2	0.2	0.4	0.2	0.3
	COP-Alt Q	0.7	0.5	0.5	0.1	0.3	0.2	0.3	0.1	0.3	0.1	0.3	0.4	0.3	0.2

natural habitat. It is expected that roads would be an interference to alligator natural habitat and nesting would be influenced by presence of roads in wilderness. The same is exhibited in Figure 4(b) and shows an increasing proportion of nesting as distance increases from the roads.

The probability of nesting was not strongly influenced by cells containing more than two alligator holes (Section 4.1). A female alligator's mean annual home range is 36 ha (Morea, 1999), or 1.8 grid cells in the current model. In contrast, a male alligator's mean annual home range is 122 ha (Morea, 1999) and would cover 6.1 grid cells. The model predicted that having a large number of alligator holes in an alligator's home range would not necessarily have a positive influence on nesting. Instead of alligator hole count, we used distance to an alligator hole as the metric. This metric provides the effect of an alligator hole on multiple grid cells even if they do not have alligator holes.

The final model had sixteen predictors, which are either a main effect, an interaction, or both (Table 20). In the past, this could cause concern that the model was overly complex and not parsimonious. However, newer modeling realities, such as large datasets and widely accessible advanced computing resources, limit the validity of those criticisms for the present model. There are adequate observations in this work to support not only a complex model but also model selection and assessment via data splitting. Ecological systems are usually complex, and there are many variables that describe them. For a model to describe an ecological system adequately, it sometimes must also be complex and contain many variables and their interactions.

Habitat suitability index (HSI) and GAM models are different and serve different purposes for resource managers. Earlier models (see Shinde et al. 2014, Rice et al. 2004, Palmer et al. 2004 and Newsom et al. 1987) predict habitat suitability indices (HSI; 0-1) and were developed relying heavily on

expert judgment and with limited data. The indices have no associated assessment of uncertainty with expected nesting or alligator production. This GAM model, on the contrary, is based on extensive data analysis with minimal expert judgment. It builds on the results from previous HSI models and adds new variables describing the structure of the landscape and weather. The present GAM model provides the probability of nesting at a specific location and quantifies the associated uncertainty.

While we can use examples of wet, dry and average (typical) hydrological years to evaluate the model across a range of likely hydrological conditions, we cannot assume that those single years are representative of other wet, dry, or average years. The selected years are being influenced in part by the previous year (temporal extent), there are seasonal differences among the annual categories of wet/dry/average (temporal scale), and there are real differences among basins within the Everglades region (spatial scale).

Prior to water management practices that reduced water flows to Shark River Slough, alligators were more abundant on the edges of the sloughs (Craighead 1968), but now they are most abundant in the central sloughs (Kushlan 1990, Morea 1999), because existing management practices resulted in a very short hydroperiod on the edges of the sloughs (Mazzotti and Brandt 1994, Mazzotti et al. 2009). If water management changes to make the hydrological conditions more like what they were before human intervention, those changes are expected to result in a re-distribution of alligators. To the extent that a space for time substitution (Johnson and Miyanishi 2008, Banet and Trexler 2013) can be considered a reasonable approach for forecasting, we believe this model can be a valid exploration of impacts on alligator nesting (such as shown in Section 6.2) since the 24-year time span in this model includes a wide range of wet to dry conditions. Simulated COP-Alt Q conditions altered the hydrology of the basins (Figure 22) and showed a relative increased wetness in some basins (Figures 23-25). When applying the GAM model to COP-Alt Q hydrological conditions, only water depths changed, and all other input variables retained their original values. Application of the GAM model to COP-Alt Q operations (1992-2005) that redistributed water (spatially and temporally; Figures 23-25) indicates that nesting increases in peripheral regions of RG, TS, LPK, and PH (Table 27) due to apparent shifts from very shallow water to less shallow water depths and nesting decreases in the central regions of ES, USS, LSS, and NESS (Table 27) seemingly due to shifts from optimal to sub-optimal water depths. This highlights the importance of using a GAM to provide water managers information on how the water redistribution plan might be tweaked to get desired results.

7.1 Model limitations

This GAM model may not be applicable to other regions of the Everglades (such as WCAs, LNWR, and BCNP) as it was developed using SRF nesting data and local spatial information not representative of other regions. However, the methods used to build this model could be used to develop models for other Everglades regions or to generalize across regions (see Section 3).

7.1.1 Limitations of habitat data

Habitat data were obtained from land classification and alligator hole surveys, which were costly in terms of time, labor, and funding, and thus cannot be repeated at short intervals/frequency. Change may not be detected or expected year-to-year but is occurring over longer periods, especially in areas where restoration is proving to be effective. Marsh communities can change with changing hydrological regimes in 3 to 5 years (Nott et al. 1998; Armentano et al. 2006; Sah et al. 2014). Integrating the alligator nesting model with a vegetation succession model (e.g., see Pearlstine et al. 2011) and remote sensing could help overcome the challenges of large-scale field surveys.

7.1.2 Over-estimating model performance

We are underestimating the prediction errors because the observations in the test set are correlated with the observations in the training and validation datasets. The leave-out-blocks method (Roberts, et al. 2017), is an alternative approach to consider that may give a more accurate assessment of what the AUC curve really looks like when using the model for prediction under various hydrological scenarios. Using the current model to predict years 2015 and onwards will also give a better assessment of the AUC curve.

7.2 Future work

1. Account for potentially imperfect nest detection in the SRF survey.

    1.1 Design and implement a study to quantify the probability of nest detection in the SRF survey. Test the null hypothesis that the nest detection probability is equal to 1. Ugarte (2006) estimated 53.3% detection in a small study on SRF data. Systematic Reconnaissance Flight observers check roughly 80% of detected nests in follow-up nest visits each year and record if a nest is

from a previous year (Mark Parry, ENP, personal communication). A larger study is required to assess uncertainty remaining with likely detectability.

1.2 If the probability of nest detection in the SRF survey is significantly different from 1, modify the statistical model of nest-building to account for the nest-detection probability.

2. Create a unified statistical model of alligator production

2.1 Use data from follow-up nest visits to develop a statistical model of the number of alligator hatchlings from a nest in a given grid cell in a given year, conditioned on a nest having been built in that grid cell that year, as a function of variables believed to affect nest success.

2.2 Use probability theory to link the statistical model of alligator hatchlings to the statistical model of nest-building to create a unified statistical model of alligator production.

Use more distance variables: Sections 4.1, 4.2, and 4.5.1 explain why distance variables- *dist_AH*, *dist_ENPrds*, and *dist_canals* were used as predictor variables over their binary or continuous alternatives. It may have been more effective to use *dist_edge, dist_marsh, dist_excluded, dist_upland* as well. It would at least have been worthwhile to have considered those variables including smaller roads and levees, though it might have been better to use *dist_edge*, to capture the effects of marsh and upland without using *dist_marsh* and *dist_upland*.

8 REFERENCES

Armentano, T.V., J.P. Sah, M.S. Ross, D.T. Jones, H.C. Cooley, and C.S. Smith. 2006. Rapid responses of vegetation to hydrological changes in Taylor Slough, Everglades National Park, Florida, USA. Hydrobiologia 569: 293-309.

Barr, B. 1997. Food Habits of the American Alligator, *Alligator mississippiensis*, in the Southern Everglades. Unpublished Ph.D. thesis, University of Miami, Miami, Florida. 243 pp.

Banet, A. and J. Trexler. 2013. Space-for-Time Substitution Works in Everglades Ecological Forecasting Models. PLoS ONE 8

Brandt, L.A. 2018. Baseline data on alligator nesting in A.R.M. Loxahatchee National Wildlife Refuge to inform future monitoring. Journal of Fish and Wildlife Management 10(1):266–276; e1944-687X. https://doi.org/10.3996/092017-JFWM-078

Bugbee, C.D. 2008. Emergence dynamics of American alligators (*Alligator mississippiensis*) in Arthur R. Marshall Loxahatchee national wildlife refuge: life history and application to statewide alligator surveys. Unpublished M.S. Thesis, University of Florida, Gainesville, FL.

Burtner, B.F., P.C. Frederick. 2017. Attraction of nesting wading birds to alligators (*Alligator mississippiensis)*. Testing the 'nest protector' hypothesis. Wetlands 37 (4): 697-704.

Chabreck, R.H. 1965. The movement of alligators in Louisiana. Proc. Southeast. Assoc. Game Fish Comm. 19:102-110.

Craighead, F.C. 1968. The role of the alligator in shaping plant communities and maintaining wildlife in the Southern Everglades. Florida Naturalist. 41:2-7, 69-74, 94.

Chopp, M.D. 2003. Everglades alligator (*Alligator mississippiensis*) production and natural history in interior and canal habitats at Arthur R. Marshall Loxahatchee national wildlife refuge. Unpublished M.S. Thesis, University of Florida, Gainesville, FL.

Dalrymple, G.H. 1996a. Growth of American Alligators in the Shark Valley Region of Everglades National Park. Copeia, Vol. 1996, No. 1, pp. 212-216.

Dalrymple GH. 1996b. The effect of prolonged high water levels in 1995 on the American alligator in the Shark Valley area of Everglades National Park. Pages 125-136 in Armentano T.V., editor. Proceedings of the Conference: Ecological Assessment of the 1994-1995 High Water Conditions in the Southern Everglades. Held at Florida International University, Miami, Florida. August 22-23, 1996.

Dalrymple, G.H. 2001. American Alligator Nesting and Reproductive Success in Everglades National Park. An analysis of the systematic reconnaissance flight data (SRF) from 1985 -1998. Statistical Analysis of Environmental Factors Influencing Spatial and Temporal Patterns of American Alligator Nesting in the Southern Everglades. Final Report, University of Miami-Everglades National Park Cooperative Agreement # CA528-03-9013.

Deitz, D.C. and D.R. Jackson. 1979. Use of American alligator nests by nesting turtles. Journal of Herpetology 13:510–512.

Fleming, D.M. 1990. American alligator distribution and abundance in relation to landscape pattern and temporal characteristics of the Everglades. National Park Service, Everglades National Park, South Florida Research Center Unpublished Report, Homestead, Florida.

Fleming, D.M. 1991. Annual Report- Wildlife ecology studies. Everglades National Park South Florida Research Center, Homestead, FL.

Fujisaki, I., F.J. Mazzotti, K.M. Hart, K.G. Rice, D. Ogurcak, M. Rochford, B.M. Jeffery, L.A. Brandt, and M.S. Cherkiss. 2012. Use of alligator hole abundance and

occupancy rate as indicators for restoration of a human-altered wetland. Ecological Indicators 23:627–633.

Frederick, P., D.E. Gawlik, J.C. Ogden, M.I. Cook, and M. Lusk. 2009. The white ibis and wood stork as indicators for restoration of the everglades ecosystem. Ecological Indicators 9: 83-95.

Graham J.A. 2004. Establishing a method to assess detectability of American alligator nests in the Arthur R. Marshall Loxahatchee National Wildlife Refuge. Report to U.S. Fish and Wildlife Service Arthur R. Marshall Loxahatchee National Wildlife Refuge, Boynton Beach, Florida.

Goodwin, T.M. and W.R. Marion. 1978. Aspects of the Nesting Ecology of American Alligators (*Alligator mississippiensis*) in North- Central Florida. Herpetologica, Vol. 34, No. 1, pp. 43-47.

Hall, P.M. and A.J. Meier. 1993. Reproduction and behavior of western mud snakes (*Farancia abacura reinwardtii)* in American alligator nests. Copeia 1993(1):219–222.

Hastie, T.J., R. Tibshirani, and J. Friedman. 2001. Elements of Statistical Learning. Springer-Verlag, New York, NY.

Howarter, S.R. 1999. Thermoregulation of the American alligator in the Everglades. M.S. Thesis. University of Florida, Gainesville, Florida. 73 pp.

Johnson, E.A. and K. Miyanishi. 2008. Testing the assumptions of chronosequences in succession. Ecology Letters 11:419-431.

Joanen, T. and L. McNease. 1970. A telemetric study of nesting female alligators on Rockefeller Refuge, Louisiana Proc. Southeast Assoc. Game Fish Comm. 24:175-193.

Joanen, T. and L. McNease. 1972. A telemetric study of adult male alligators on Rockefeller Refuge, Louisiana. Proc. Southeast Assoc. Game Fish Comm. 26:252-275.

Kushlan, J.A. 1974. Observations on the role of the American alligator (*Alligator mississippiensis*) in the southern Florida wetlands. Copeia 1974:993–996.

Kushlan, J.A. 1990. Wetlands and wildlife, the Everglades perspective in freshwater wetlands and wildlife. In R.R. Sharitz and J.W. Gibbons, editors. CONF-8603101, DOE Symposium Series No. 61. U.S. Department of Energy, Office of Scientific and Technical Information, Oak Ridge, Tennessee.

Kushlan, J.A. and M.S. Kushlan. 1980. Everglades alligator nests: Nesting sites for marsh reptiles. Copeia 1980:930–932.

Kushlan, J.A. and T. Jacobsen. 1990. Environmental variability and the reproductive success of Everglades alligators. Journal of Herpetology 24(2):176–184.

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3. https://CRAN.R-project.org/package=emmeans.

Mazzotti, F.J. and L.A. Brandt. 1994. Ecology of the American alligator in a seasonally fluctuating environment. pp. 485-505 In S. Davis and J. Ogden, (eds.), Everglades: The Ecosystem and its Restoration. St. Lucie Press, Delray Beach, Florida.

Mazzotti, F.J., G.R. Best, L.A. Brandt, M.S. Cherkiss, B.M. Jeffery, and K.G. Rice. 2009. Alligators and crocodiles as indicators for restoration of Everglades ecosystems. Ecological Indicators, 9s, s137–s149.

Mazzotti, F.J., L.A. Brandt, P. Moler, and M.S. Cherkiss. 2007. The American crocodile (*Crocodylus acutus*) in Florida: Recommendations for endangered species recovery and ecosystem restoration. Journal of Herpetology 41(1):122–132.

McNease, L. and T. Joanen. 1974. A telemetric study of immature alligators on Rockefeller Refuge, Louisiana. Proc. Southeastern Assoc. Game and Fish Commissioners Conf. 28: 482-500.

Morea, C.R. 1999. Home range, movement, and habitat use of the American alligator in the everglades. Unpublished M.S. Thesis, University of Florida, Gainesville, FL.

Newsom, J.D., T. Joanen, and R. Howard. 1987. Habitat suitability index models: American alligator. U.S. Fish and Wildlife Service Biological Report 82(10.136). 14 pp.

Nott, M.P., O.L. Bass Jr., D.M. Fleming, S.E. Killeffer, N. Fraley, L. Manne, J.L. Curnutt, T.M. Brooks, R. Powell, and S.L. Pimm, 1998. Water levels, rapid vegetational changes, and the endangered Cape Sable Seaside Sparrow. Animal Conservation 1: 23–32.

Ogden, J.C. and S.M. Davis. 1999. The use of conceptual ecological landscape models as planning tools for the south Florida ecosystems restoration programs. Unpublished report. South Florida Water Management District. West Palm Beach, Florida.

Ogden, J.C., S.M. Davis, and L.A. Brandt. 2003. Science strategy for a regional ecosystem monitoring and assessment program: The Florida Everglades example. Pages 135–163 in: D. Busch and J. Trexler, editors. Monitoring ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives. Island Press, Washington, DC.

Palmer, M.R., L. Gross, and K.G. Rice. 2004. ATLSS American Alligator Production Index Model- Basic Model Description. The Institute for Environmental Modeling, University of Tennessee, Knoxville, Tennessee. Accessed September 26, 2018, at http://atlss.org/cerp_runs/mod_info/od_gator.html

Parry, M.W. and O.S. Bass. 2009. Annual Report- Alligator Systematic Reconnaissance Flights (SRF): Summary of American alligator nesting effort and success in Everglades National Park, South Florida Natural Resources Center, Homestead, FL.

Pearlstine, L., S. Friedman, and M. Supernaw. 2011. Everglades Landscape Vegetation Succession Model (ELVeS) Ecological and Design Document: Freshwater Marsh & Prairie Component version 1.1. South Florida Natural Resources Center, Everglades National Park, National Park Service, Homestead, Florida. 128 pp.

Percival, H.F., K.G. Rice, S.R. Howarter. 2000. American alligator distribution, thermoregulation, and biotic potential relative to hydroperiod in the Everglades. Fla. Coop. Fish and Wildl. Res. Unit, USGS Tech. Rep. 155 pp.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RECOVER. 2014. Greater Everglades Performance Measure Wetland Trophic Relationships – American Alligator Abundance, Body Condition, Hole Occupancy, and Production Suitability Index. (Last Date Revised: June 5, 2014), http://141.232.10.32/pm/recover/perf_ge.aspx, accessed Oct 26, 2018.

Rice, K.G. and F.J. Mazzotti. 2007. American alligator distribution, size, and hole occupancy and American crocodile juvenile growth and survival. MAP RECOVER Annual Report. Unpublished report submitted to U.S. Army Corps of Engineers, Jacksonville, Florida.

Rice, K.G., F.J. Mazzotti, L.A. Brandt, and K.C. Tarboton. 2004. Alligator habitat suitability index. Pages 93–110 in K.C. Tarboton, M.M. Irizarry-Ortiz, D.P. Loucks, S.M. Davis, and J.T. Obeysekera, editors. Habitat suitability indices for evaluating water management alternatives. Unpublished report. South Florida Water Management District, West Palm Beach, Florida.

Rice, K.G, H.F. Percival, and A.R. Woodward. 2000. Estimating sighting proportions of American alligator nests during helicopter survey. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 54:314-321.

Roberts, D.R., V. Bahn, S. Ciuti, M.S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J.J. Lahoz-Monfort, B. Schroder, W. Thuiller, D.I. Warton, B.A. Wintle, F. Hartig, and C.F. Dormann. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40: 913-929.

Rootes, W.L. and R.H. Chabreck. 1993. Reproductive Status and Movement of Adult Female Alligators. Journal of Herpetology, Vol. 27, No. 2, pp. 121-126.

Sah, J.P., M.S. Ross, S. Saha, P. Minchin, and J. Sadle. 2014. Trajectories of vegetation response to water management in Taylor Slough, Everglades National Park, Florida. Wetlands 34 (Suppl 1): S65-S79

SFWMD. 2005. Theory Manual, Regional Simulation Model (RSM). South Florida Water Management District, Office of Modeling, West Palm Beach, FL 33406, May 16, 2005.

Shinde, D., L. Pearlstine, L.A. Brandt, F.J. Mazzotti, M.W. Parry, B. Jeffrey, and A. LoGalbo. 2014. Alligator Production Suitability Index Model (GATOR-PSIM v. 2.0): Ecological and Design Documentation. South Florida Natural Resources Center, Everglades National Park, Homestead, Florida. Ecological Model Report. SFNRC Technical Series 2014:1.

Taylor, D. 1984. Management implications of an adult female alligator telemetry study. Proc. Ann. Conf. Southeast. Assoc. Fish and Wildl. Agencies 38:222-227.

Ugarte, C.A. 2006. Long term (1985-2005) spatial and temporal patterns of alligator nesting in Everglades National Park, Florida, USA. School of Natural Resources and the Environment. University of Florida, Gainesville, FL.

USACE. 1992. Modified Water Deliveries to Everglades National Park, Florida. General design memorandum and Environmental Impact Statement. June 1992. U.S. Army Corps of Engineers, Jacksonville, Florida.

USACE. 2020. Combined Operational Plan: Final Environmental Impact Statement. U.S. Army Corps of Engineers, Jacksonville, Florida.

Wood, S.N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association. 99:673-686.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36.

Wood, S.N. 2017. Package 'mgcv'. Documentation of R package. Date/Publication: 2017-09-19 00:27:56 UTC.

APPENDICES

I Estimation of spatial hydrological and habitat information

This applies to the hydrological and habitat variables. The data are based on 400 m (horizontal width: east-west) x 500 m (vertical width: north-south) grid cells on SRF transects (Figure 1) running east to west. The east and west cell boundaries were aligned with EDEN grid cells. If the edges of the transect on the east and west end covered >70% of the EDEN grid cell width, they were included in analysis, otherwise they were discarded. In some transects, the SRF grid cell overlapped 3 EDEN cells in the north to south direction (Figure 26). The following information describes how we obtained data for predictor variables at these grid cells on SRF transects.

- Hydrological data was obtained from EDEN which has a 400 m x 400 m resolution.
- SRF transects for nest sightings are 500 m wide.

For hydrological metrics, we took the weighted average by area overlapped of EDEN grid cells within the SRF grid cell as the hydro metric value for that SRF grid cell (Figure 26).

For habitat metrics, since the habitat cells are at 50 m resolution, we collected all the 50 m cells (total 80) within each SRF grid cell increment and counted the presence cells to get percent presence (100 x number of presence cells/80).

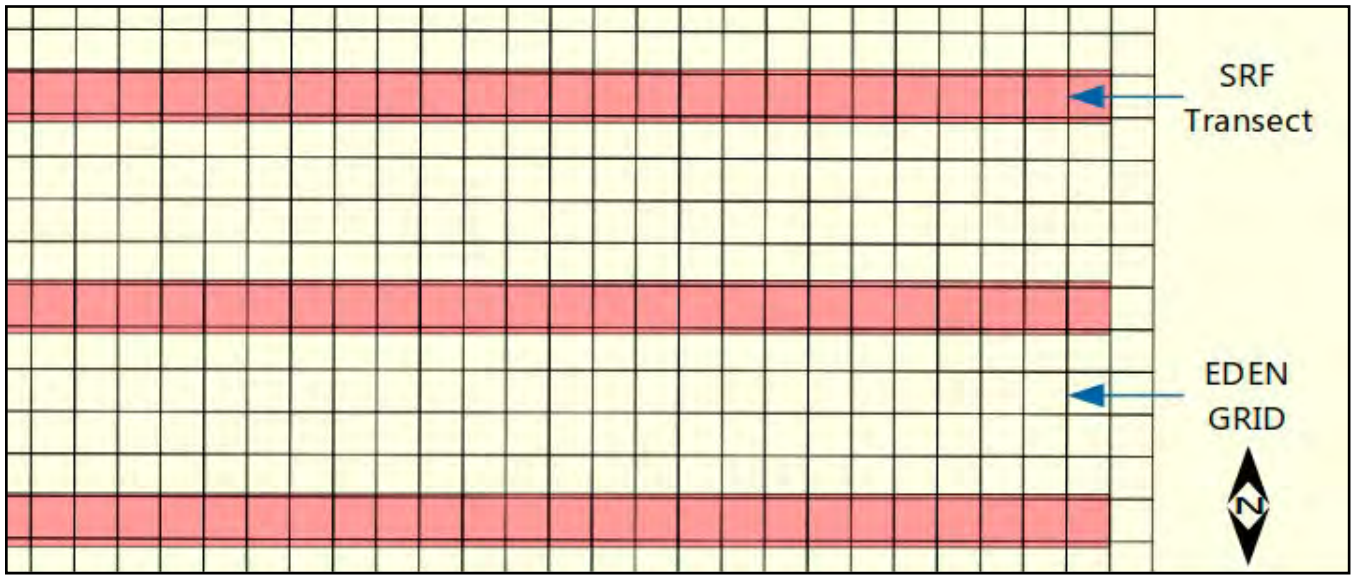


Figure 26. Illustration of EDEN hydrological grid overlaid on SRF transect for estimation of hydrological and habitat predictor data on spatial scale

II Tables of predicted probability (nest = 1) and 95% prediction interval width (EDEN hydrological conditions)

Table 28. Predicted maximum and mean probabilities (reported in percent) in hydrological basins (Figure 1).

Year	ES		USS		LSS		NESS		RG		TS		LPK		PH	
	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean
POR[§]	66.2	3.3	48.7	5.7	50.0	6.7	53.9	1.4	22.4	0.1	49.7	1.9	27.6	0.2	21.3	0.3
SD	10.1	1.4	9.4	2.6	8.7	2.2	10.2	0.9	4.7	0.1	11.2	0.7	6.1	0.1	5.0	0.1
1992	27.4	2.6	23.7	4.8	23.8	4.6	14.4	0.9	8.0	0.1	47.1	3.0	3.8	0.2	16.5	0.4
1993	34.0	4.5	40.6	9.2	38.3	9.4	36.9	2.7	7.3	0.1	36.5	2.4	3.7	0.2	8.8	0.2
1994	28.6	3.9	31.9	7.2	40.3	8.3	14.1	1.7	9.2	0.2	33.6	2.2	6.1	0.2	15.4	0.4
1995	21.9	2.4	29.8	7.0	33.4	6.7	10.9	1.3	10.2	0.2	14.6	0.9	13.3	0.3	5.3	0.1
1996	24.7	3.8	31.7	7.5	42.1	7.6	15.1	1.6	4.9	0.1	17.4	1.5	9.9	0.2	8.2	0.2
1997	22.9	2.6	26.3	5.5	31.0	5.4	9.3	1.1	5.5	0.1	13.1	1.0	3.5	0.1	3.1	0.1
1998	27.7	2.9	29.9	6.9	32.2	6.5	12.3	1.6	1.2	0.0	18.0	1.6	0.5	0.0	8.4	0.2
1999	13.9	1.4	13.6	2.4	19.7	3.4	4.2	0.4	3.4	0.1	7.9	0.5	1.9	0.1	2.0	0.0
2000	17.0	2.1	16.9	3.3	31.7	4.5	6.8	0.7	5.8	0.1	20.3	1.4	2.0	0.1	9.5	0.2
2001	16.1	2.0	15.0	3.4	19.2	4.0	10.8	0.7	4.1	0.1	23.4	1.3	2.2	0.1	8.6	0.2
2002	28.4	2.8	26.2	6.0	38.5	6.5	16.6	1.6	4.6	0.1	16.2	1.3	2.8	0.1	8.4	0.2
2003	20.8	1.7	15.0	3.3	21.9	3.1	13.5	1.3	11.7	0.2	37.4	3.1	17.0	0.5	17.5	0.4
2004	18.3	2.6	16.6	3.5	25.6	5.3	11.9	0.8	8.5	0.2	31.1	1.7	5.2	0.2	8.8	0.2
2005	22.0	3.5	17.5	3.0	39.0	8.5	19.4	1.0	5.5	0.1	29.4	1.6	3.3	0.2	12.4	0.3
2006	19.6	2.4	15.6	3.3	26.3	5.0	10.6	0.7	3.5	0.1	18.9	1.4	1.2	0.1	17.0	0.4
2007	29.6	4.0	21.6	4.4	36.4	7.6	14.3	1.0	7.4	0.2	32.3	2.5	3.6	0.3	17.6	0.4
2008	23.9	3.4	29.6	7.2	41.1	8.2	19.4	1.6	5.1	0.1	27.2	1.9	2.4	0.2	14.2	0.3
2009	29.2	3.4	16.6	3.6	37.7	7.5	16.0	0.9	7.7	0.2	36.5	1.7	6.9	0.3	8.5	0.2
2010	66.2	8.8	48.7	12.9	50.0	11.9	53.9	5.0	11.5	0.3	34.2	2.1	5.2	0.3	16.2	0.4
2011	27.9	2.8	17.4	4.0	24.9	5.5	15.6	1.0	1.2	0.0	15.2	1.1	1.3	0.1	12.7	0.3
2012	26.6	3.9	29.8	6.3	46.7	8.6	13.9	1.5	14.5	0.3	39.7	3.2	7.2	0.4	21.3	0.5
2013	23.4	3.7	18.5	4.3	25.5	5.7	17.1	1.2	11.3	0.2	36.7	2.4	9.5	0.4	14.7	0.4
2014	28.2	3.8	33.7	7.5	45.5	8.1	15.1	1.5	2.2	0.0	24.2	2.0	1.3	0.1	16.9	0.4
2015	35.9	4.5	37.4	10.4	35.2	9.2	26.3	2.4	22.4	0.4	49.7	3.1	27.6	0.7	12.0	0.3

ES: East Slough, USS: Upper Shark Slough, NESS: Northeast Shark Slough, LSS: Lower Shark Slough, RG: Rocky Glades, TS: Taylor Slough, LPK: Long Pine Key, and PH: Panhandle; [§] Period of Record

Table 29. Predicted mean 95% prediction interval width (reported in percent) in hydrological basins (Figure 1).

Year	ES	USS	LSS	NESS	RG	TS	LPK	PH
POR[§]	41	36	35	54	68	63	74	89
1992	42	38	37	55	73	64	75	89
1993	39	36	34	54	65	62	72	89
1994	38	33	33	53	65	62	72	89
1995	47	41	38	55	68	65	75	89
1996	40	35	33	53	64	62	72	89
1997	41	35	35	54	67	63	73	89
1998	43	37	37	55	70	65	77	89
1999	48	41	41	58	69	66	76	89
2000	39	35	34	55	65	62	72	89
2001	41	35	35	54	68	63	73	89
2002	41	34	33	53	68	63	74	89
2003	41	37	37	53	66	62	72	89
2004	39	35	33	53	67	63	74	89
2005	41	38	35	55	68	65	75	89
2006	40	36	34	54	67	62	74	89
2007	38	33	33	53	67	63	73	89
2008	39	35	33	54	68	62	73	89
2009	42	37	35	54	71	64	74	89
2010	40	36	35	54	66	63	73	89
2011	43	37	39	56	86	66	80	89
2012	39	33	33	54	67	62	73	89
2013	38	35	35	53	65	62	72	89
2014	41	36	34	54	70	64	75	89
2015	40	35	33	54	67	63	73	89

ES: East Slough, USS: Upper Shark Slough, NESS: Northeast Shark Slough, LSS: Lower Shark Slough, RG: Rocky Glades, TS: Taylor Slough, LPK: Long Pine Key, and PH: Panhandle; [§] Period of Record

These pages intentionally left blank

This page intentionally left blank