Ecological Controls on the Campanian Distribution of Hesperornis (Aves: Hesperornithiformes) in the Western Interior Seaway

Blake Chapman

Fort Hays State University, brchapman187@gmail.com

Follow this and additional works at: https://scholars.fhsu.edu/theses

Recommended Citation


This Thesis is brought to you for free and open access by the Graduate School at FHSU Scholars Repository. It has been accepted for inclusion in Master's Theses by an authorized administrator of FHSU Scholars Repository.
ECOLOGICAL CONTROLS ON THE CAMPANIAN DISTRIBUTION OF
HESPERORNIS (AVES: HESPERORNITHIFORMES) IN THE
WESTERN INTERIOR SEAWAY

being

A Thesis Presented to the Graduate Faculty
of Fort Hays State University in
Partial Fulfillment of the Requirements for
the Degree of Master of Science

by

Blake Chapman
B.S., The University of Texas at Austin

Date 11/24/2021

Approved
Major Professor

Approved
Dean of the Graduate School
This thesis for
the Master of Science Degree

By
Blake R. Chapman

has been approved

Dr. Laura Wilson, Committee Chair

Dr. Robert Channell, Committee Member

Dr. Richard Lisichenko, Committee Member
ABSTRACT

The epicontinental Western Interior Seaway (WIS) of Late Cretaceous North America provided a unique marine habitat for cephalopods, fish, marine reptiles, and the foot-propelled diving seabird *Hesperornis*. While several predator-prey relationships among *Hesperornis* or other hesperornithiforms and other WIS animals have been hypothesized based on gut contents, bite marks, and coprolites/colonites, ecological relationships have not been quantitatively tested. Paleontological species distribution modeling (SDM) studies have focused on extinct non-marine taxa and marine invertebrates, with only two marine vertebrate studies of extant taxa. Here, two SDM methods were used to test the influence of vertebrate faunas, sedimentary rock type, paleogeography, and outcrop area on *Hesperornis* occurrences: generalized linear models (GLMs) and occupancy models (an extension of standard GLMs). Results of both model types indicate a decreased probability of *Hesperornis* presence at elasmosaur-occupied sites and an increased probability of *Hesperornis* presence with higher paleolatitudes. The negative association between elasmosaurs and *Hesperornis* in all highest-ranked models, coupled with evidence of similar diets, may indicate food resource competition, but the geographic distributions of both taxa may instead reflect different habitat preferences. The positive impact of paleolatitude may reflect seasonal migrations resulting in large breeding congregations at higher paleolatitudes (based on probable immature specimens), preference for specific environmental conditions or feeding areas (e.g., along oceanographic fronts or upwelling zones), or migration mortality. While Spearman’s correlation test did not identify a western, middle, or eastern WIS preference for *Hesperornis*, the inclusion of this regional covariate with paleolatitude in the highest-
ranked occupancy models may indicate an interaction effect between these two covariates that favored higher *Hesperornis* abundance in South Dakota and Manitoba along probable mixing zones or fronts. Taxa hypothesized by previous authors to have had ecological interactions with hesperornithiforms were not supported by the models as major factors in the *Hesperornis* distribution. These include *Tylosaurus* and polycotylid plesiosaurs, hypothesized feeders on *Hesperornis*, *Enchodus*, a teleost consumed by the smaller hesperornithiform *Fuminollis*, and *Squalicorax*, an opportunistic shark that consumed hesperornithiforms. This study represents the first paleogeographic model for *Hesperornis* occurrences and the first GLM and occupancy modeling applications to extinct marine vertebrates. These models examine previous hypotheses and provide novel interpretations for hesperornithiform ecology.
ACKNOWLEDGMENTS

I thank my thesis committee, Laura Wilson (advisor), Robert Channell, and Richard Lisichenko, for overseeing and guiding my thesis project. I wish to thank Medhavi Ambardar (Assistant Professor of Biological Sciences) of Fort Hays State University for sponsoring my oral presentation at the 60th annual Society of Integrative and Comparative Biology Conference (SICB) in Austin, Texas (2020). I would like to thank Dr. Dale A. Winkler (Director) and Dr. Michael J. Polcyn (Research Associate and Director of the Digital Earth Sciences Laboratory) of the Shuler Museum of Paleontology at Southern Methodist University (SMU), Dr. Darrin C. Pagnac (Curator of Vertebrate Paleontology) of the Museum of Geology at the South Dakota School of Mines and Technology (SDSMT), and James C. Sagebiel (Collections Manager) and the staff of the Vertebrate Paleontology Laboratory (VPL), part of the Jackson School Museum of Earth History at The University of Texas at Austin, for allowing access to collections, providing specimen lists and copies of literature sources (the latter from the staff of SMU), and answering questions regarding localities and associated specimens. I thank Chase Shelburne (interim Paleontology Collections Manager) of the Sternberg Museum of Natural History for allowing access to the collections. I thank Sally Y. Shelton (Associate Professor of Practice and Associate Chair of Heritage and Museum Sciences) of the Museum of Texas Tech University (contacted when she was the Associate Director of the Museum of Geology at SDSMT), Vanessa Rhue (Collections Manager, Vertebrate Paleontology) of the Yale Peabody Museum of Natural History, Maureen Walsh (Collections Manager) of Dinosaur Institute at the Natural History Museum of Los Angeles County, and Megan Sims (Collection Manager of Vertebrate Paleontology) of
the KU Biodiversity Institute & Natural History Museum for providing locality
information for *Hesperornis* specimens. I also thank Amanda Millhouse (Deputy
Collections Manager, Vertebrate Paleontology) of the National Museum of Natural
History of the Smithsonian Institution, Dr. Emily L. Bamforth, Assistant Curator of
Palaeontology of the *T. rex* Discovery Centre at the Royal Saskatchewan Museum and
Adjunct Professor in the Department of Geological Sciences at the University of
Saskatchewan, and Jacob Van Veldhuizen (Collection Manager of Vertebrate
Paleontology) of the University of Colorado Museum of Natural History for providing
information on Campanian vertebrate occurrences and assemblages. I thank Dr. Kenshu
Shimada (Professor of Paleobiology) of the Departments of Biological Sciences and
Environmental Science and Studies at the College of Science and Health at DePaul
University for discussion of a shark-bitten hesperornithiform bone. I wish to thank the
Forsyth Library of Fort Hays State University for providing access to numerous research
articles. I thank Gary Kelner and Alexander Lyles for their help in designing distribution
maps and troubleshooting ArcMap and Jonathan Camelo for helping with data
conversion for GPlates. I thank Mary Carpenter and Ed Welsh for their helpful comments
on my thesis defense presentation. I also thank the graduate students of the Department of
Geosciences who supported me through the writing and data collection process and my
parents, Jenny and Russell Chapman, and maternal grandparents, Carol and Richard
Young, for their constant love and support. I thank Cassie Chapman, my late beloved cat,
for her comfort and companionship when I visited home. Rest in peace, dear Cassie.

Funding for travel and lodging in Texas to visit the SMU and VPL collections and
to present my research at 60th annual SICB conference was provided by the Graduate
Scholarly Experience grant from the Fort Hays State University Graduate School and the Department of Geosciences, Fort Hays State University.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF EQUATIONS</td>
<td>x</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Institutional Abbreviations</td>
<td>10</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>12</td>
</tr>
<tr>
<td>Data Collection and Processing</td>
<td>12</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>23</td>
</tr>
<tr>
<td>Generalized Linear Modeling</td>
<td>24</td>
</tr>
<tr>
<td>Occupancy Modeling</td>
<td>28</td>
</tr>
<tr>
<td>Non-SDM Tests</td>
<td>31</td>
</tr>
<tr>
<td>RESULTS</td>
<td>33</td>
</tr>
<tr>
<td>Results of Generalized Linear Model Analyses</td>
<td>33</td>
</tr>
<tr>
<td>Results of Occupancy Model Analyses</td>
<td>38</td>
</tr>
<tr>
<td>Model Comparisons</td>
<td>39</td>
</tr>
<tr>
<td>Results of Non-SDM Tests</td>
<td>40</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>47</td>
</tr>
<tr>
<td>Covariate Groupings from One Model Type</td>
<td>55</td>
</tr>
</tbody>
</table>
Other Covariates of Note ........................................................................................................60
Model Comparison and Discussion ..............................................................................63
CONCLUSIONS..................................................................................................................70
Future Directions and Applications for Paleontological Analyses ......................71
REFERENCES ....................................................................................................................74
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothesized interactions between hesperornithiforms and other WIS taxa</td>
</tr>
<tr>
<td>2</td>
<td>Taxonomic covariates used for modeling</td>
</tr>
<tr>
<td>3</td>
<td>Non-taxonomic covariates used for modeling</td>
</tr>
<tr>
<td>4</td>
<td>Full dataset GLM results for $\Delta AIC \leq 2$</td>
</tr>
<tr>
<td>5</td>
<td>Reduced dataset GLM results for $\Delta AIC \leq 2$</td>
</tr>
<tr>
<td>6</td>
<td>Occupancy model results for $\Delta AIC \leq 2$</td>
</tr>
<tr>
<td>7</td>
<td>Covariate groupings shared among the highest-ranked models ($\Delta AIC \leq 2$)</td>
</tr>
<tr>
<td>8</td>
<td>Spearman’s rank correlation results for the number of <em>Hesperornis</em> specimens vs. non-binary covariates</td>
</tr>
<tr>
<td>9</td>
<td>Advantages and disadvantages of model types used</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Campanian distribution of <em>Hesperornis</em> within the WIS and High Arctic</td>
</tr>
<tr>
<td>2</td>
<td>Map of Campanian fossil-bearing localities included in the full dataset GLMs</td>
</tr>
<tr>
<td>3</td>
<td>Map of Campanian fossil-bearing localities included in the reduced dataset GLMs and occupancy models</td>
</tr>
<tr>
<td>4</td>
<td>Association of outcrop area (km$^2$) and the number of <em>Hesperornis</em> specimens using Spearman’s rank correlation</td>
</tr>
<tr>
<td>5</td>
<td>Association of latitude or paleolatitude (°N) and the number of <em>Hesperornis</em> specimens using Spearman’s rank correlation</td>
</tr>
<tr>
<td>6</td>
<td>Comparison of the number of <em>Hesperornis</em> specimens across the WIS using a Kruskal-Wallis analysis</td>
</tr>
<tr>
<td>7</td>
<td>Association of position in the WIS and the number of <em>Hesperornis</em> specimens using Spearman’s rank correlation</td>
</tr>
<tr>
<td>8</td>
<td>Distribution of Campanian <em>Hesperornis</em> and elasmosaurs in the WIS</td>
</tr>
</tbody>
</table>
## LIST OF EQUATIONS

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Calculation of AIC values</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>Calculation of ΔAIC values</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>Calculation of the explained deviance/variation</td>
<td>27</td>
</tr>
</tbody>
</table>
INTRODUCTION

During the Campanian Age of the Late Cretaceous, the epicontinental Western Interior Seaway (WIS) divided North America with water depths of 60 m or more (Gill and Cobban, 1966; Hanczaryk and Gallagher, 2007; Slattery et al., 2018). In addition to fish, marine reptiles, cephalopods and other invertebrates, the WIS was also home to hesperornithiforms (Hesperornithiformes), flightless foot-propelled diving seabirds most comparable to cormorants (Phalacrocoracidae) and diving ducks (Aythya and Mergus) in hindlimb morphology (Bell et al., 2019). *Hesperornis* is the most well-known hesperornithiform with the largest number of preserved skeletal remains and the widest North American geographical distribution, ranging from Arkansas to the Canadian Arctic (Figure 1) (Wilson et al., 2011; Wilson et al., 2016; Aotsuka and Sato, 2016; Bell and Chiappe, 2016). Most *Hesperornis* specimens are isolated elements from the hind limbs, pelvis, and vertebrae, but both partial and nearly complete skeletons are known (e.g., Carpenter, 2006; Aotsuka and Sato, 2016; Bell and Chiappe, 2016). WIS species ranged in size from *H. macdonaldi*, ~0.8 m long, to *H. gracilis*, *H. lumgairi*, and *H. regalis*, ≤ 1.9-2.0 m (length estimates based on Martin and Lim, 2002; Wilson et al., 2011; Aotsuka and Sato, 2016). The genus was largely marine to transitional-marine, but non-marine occurrences are known from the Campanian of Montana, Wyoming, and Alberta (Shufeldt, 1915; Fox, 1974; Case, 1987; DeMar and Breithaupt, 2008; Unpublished specimens: one UCMP specimen and nine YPM specimens [Supplement S1]) and the Maastrichtian of Montana (Tanaka et al., 2018; Supplement S1). Despite the abundance of *Hesperornis* and inferences made about their ecology, possible ecological relationships have not been quantified or predicted using distribution models. The purpose of this study...
is to determine the sedimentological, paleogeographical, and faunal community factors that most influence the WIS distribution of *Hesperornis* with emphasis on potential predator-prey and competitive relationships. A secondary objective is to test generalized linear modeling and occupancy modeling on extinct marine vertebrates and to apply novel techniques in marine vertebrate paleontology.

Figure 1. Campanian distribution of *Hesperornis* within the WIS and High Arctic. *Hesperornis*-bearing sites are reconstructed to the early middle Campanian (80 Ma) and depicted using an 80 Ma paleogeographic map (modified from Scotese, 2016). Symbols representing fossil localities are not to scale. Data were obtained from the Paleobiology Database, museum collections records, and published literature (see Supplement S2).
Almost all hesperornithiform dietary inferences are based on tooth and jaw morphology and mode of swimming (Elzanowski, 1983; Dumont et al., 2016; Wilson et al., 2016). While considered generalist feeders or pursuit divers of fish or squid (Elzanowski, 1983; Dumont et al., 2016; Wilson et al., 2016; Everhart, 2017), the only probable direct evidence of hesperornithiform diet are coprolites or a colonite associated with the *Fumicollis hoffmani* holotype from Kansas (UNSM 20030, formerly *Baptornis advenus*) (Martin and Tate, 1976; Bell and Chiappe, 2015). These trace fossils contain an *Enchodus* jaw and other fish bones, but the coprolite/colonite association with the skeleton remains tentative (Table 1) (Martin and Tate, 1976; Wilson et al., 2016; Everhart, 2017). While the fossil record has not yet revealed much direct evidence of the hesperornithiform diet, some fossils from South Dakota demonstrate interactions with predators and scavengers (Table 1). Based on known specimens, three taxa are identified as hesperornithiform predators or scavengers: polycotylid plesiosaurs, *Tylosaurus proriger*, and *Squalicorax* cf. *falcatus*. 
Table 1. Hypothesized interspecific interactions between hesperornithiforms and other WIS taxa. The column “Ecological Role” refers to the hesperornithiform role in the interspecific interaction.

<table>
<thead>
<tr>
<th>Ecological Role</th>
<th>Age</th>
<th>Hesperornithiform Taxon</th>
<th>Other Taxa</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predator</td>
<td>Late Coniacian to early Campanian</td>
<td><em>Fumicollis hoffmani</em> (UNSM 20030; holotype; formerly <em>Baptornis advenus</em>)</td>
<td><em>Enchodus</em> sp. and indeterminate fish (preserved in tentatively associated coprolites / colonites)</td>
<td>Martin and Tate (1976); Bell and Chiappe (2015)</td>
</tr>
<tr>
<td>Prey</td>
<td>Middle Campanian</td>
<td><em>Hesperornis</em> sp. (YPM VPPU 17208.D)</td>
<td>Polycotylidae indet. (inferred from bite marks)</td>
<td>Martin et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Early Campanian</td>
<td><em>Hesperornis</em> sp. (SDSM 10439; Note: All five taxa have the same catalog number.)</td>
<td><em>Tylosaurus proriger</em></td>
<td>Martin and Bjork (1987); Martin et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Middle to late Turonian</td>
<td><em>Hesperornithiformes</em> indet. (SMM P2000.12.10)</td>
<td><em>Squalicorax</em> cf. <em>falcatus</em> (may have scavenged the hesperornithiform; inferred from bite marks)</td>
<td>Hanks and Shimada (2002); Shimada and Hanks (2020)</td>
</tr>
</tbody>
</table>
Potential predator-prey relationships across geographic space are topics addressable by species distribution models (SDMs), a term used interchangeably with ecological niche models (ENMs). These models are a common form of distribution or niche modeling techniques applied in modern ecological and paleontological studies. SDMs rely on modeling algorithms to predict regional habitat suitability (i.e., the fit of environmental and ecosystem conditions), estimate the probability of presence, or predict and describe the factors that most influence the distribution of the target organism(s) (Elith and Leathwick, 2009; Elith et al., 2011; Robinson et al., 2011; Peterson et al., 2012; Elith and Franklin, 2013; Myers et al., 2015). Each model includes one or more environmental or biological variable(s) that could affect the pattern of species occurrence or abundance such as temperature, water depth, predation/competitive taxa, etc. These variables are termed covariates (meaning independent variables that may covary) or predictors.

Researchers have used SDMs for modern ecosystems and paleoenvironments to study niche stability, interspecific interactions, dispersal, conservation, invasive species proliferation, community dynamics, extinction event/biodiversity crisis survivorship, speciation and extinction rates, and climate change effects (e.g., Maguire and Stigall, 2008; Menke et al., 2009; Malizia and Stigall, 2011; Robinson et al., 2011; Svenning et al., 2011; Rovelli and Myers, 2016; Waterson et al., 2016; Carrier, 2018; MacKenzie et al., 2018). Non-marine SDM applications outnumber marine SDMs, with most modern analyses of marine ecosystems examining conservation planning, method evaluation, and theoretical ecology for fish and mammals compared to relatively few marine invertebrate studies (see Dambach and Rödder, 2011; Robinson et al., 2011; Melo-Merino et al.,
2020, for reviews of modern marine research topics). Paleontological marine SDMs, while less abundant than non-marine applications, focus on invertebrates (e.g., Stigall Rode and Lieberman, 2005; Dudei and Stigall, 2010; Malizia and Stigall, 2011; Brame and Stigall, 2014; Pimiento et al., 2016; Rovelli and Myers, 2016; 2018; Carrier, 2018; Carotenuto et al., 2020). This leaves a gap in the application of SDMs to marine vertebrate paleoecosystems.

SDM species occurrence information includes four data types: presence-only, presence/background, presence/absence, and occupancy/detection data. The type of species occurrence information will determine which modeling algorithms should be used and the prediction generated using those algorithms. Presence-only models (sensu stricto Pearson, 2010; Guillera-Arroita et al., 2015), such as those created using BIOCLIM, use surveyed sites where the species was detected or found (species presences) to make inferences about the species distribution without environmental conditions from surrounding sites (Nix, 1986; Busby, 1991; Booth et al., 2014; Guillera-Arroita et al., 2015). Presence/background modeling algorithms (e.g., MaxEnt and GARP) focus on known species presences and incorporate these and background data to estimate relative habitat suitability (Stockwell and Peters, 1999; Phillips et al., 2004; 2006; Elith and Leathwick, 2009; Malizia and Stigall, 2011; Guillera-Arroita et al., 2015; Myers et al., 2015). Background data are randomly-chosen sites within the study area that characterize species distributions using the environmental conditions throughout the study area. These background points may exclude or include known species presences depending on the preference of the researcher and may be described as pseudo-absences (e.g., Elith and Leathwick, 2009; Pearson, 2010; Barbet-Massin et al., 2012).
In contrast to presence-only and presence/background methods, presence/absence models, such as generalized linear models (GLMs), include absence data recorded in field surveys (known absences) and environmental or biological conditions from all surveyed sites (e.g., Elith et al., 2006; Guillera-Arroita et al., 2015). Presence/absence models are advantageous because these models predict the probability of occupancy (i.e., probability of presence) rather than the relative habitat suitability predicted by presence/background models (e.g., Elith et al., 2011; Guillera-Arroita et al., 2015; Gelfand and Shirota, 2019). In other words, occupancy estimates the probability of presence for the target species rather than ranking sites by suitability. However, the estimation of occupancy or presence is biased due to imperfect detection; absences (or non-observations) may be either true absences or a failure to detect the target species (e.g., Elith and Franklin, 2013; Guillera-Arroita et al., 2015; MacKenzie et al., 2018). Imperfect detection is due to differences in an individual’s surveying ability, differential outcrop preservation, or uneven collection practices.

Correction of biased occupancy requires a conceptual extension of GLMs known as occupancy modeling (alternatively, occupancy/detection or site-occupancy modeling) that uses occupancy/detection data. Both occupancy models and other presence/absence models estimate occupancy from presence/absence data, but occupancy models also estimate the probability of detection (p): the probability that the target species was observed if present at a site (i.e., the probability of observation) (MacKenzie et al., 2002; Elith and Franklin, 2013; Guillera-Arroita et al., 2015; Koshkina et al., 2017; MacKenzie et al., 2018). The detection probability for a site denotes the combined probability that a taxon was observed and the probability that it was truly present (i.e., was it living there
and was it falsely detected?) (e.g., Guillera-Arroita et al., 2015; MacKenzie et al., 2018).

For paleontological analyses, the detection probability can be considered both “preservation and discovery” (Liow, 2013:194). The detection probability differs from the occupancy probability with the latter denoting the probability of presence. The detection probability is estimated using the detection history, the collection of non-detections \( (p = 0) \) and detections \( (p = 1) \) recorded during repeated site sampling.

Estimation of detection probabilities is most crucial when the target taxon is not observed (i.e., observed absence). Here, the detection probability allows a researcher to discriminate between true absences, where a taxon was not observed and does not inhabit a site, and false-negatives, where a taxon was not observed but was truly present (e.g., Lahoz-Monfort et al., 2014; Guillera-Arroita et al., 2015; MacKenzie et al., 2018).

The simultaneous estimation of detection and occupancy in occupancy modeling facilitates estimation of true occupancy \( (\psi) \), the probability of presence when imperfect detection is modeled. For clarity, the biased occupancy probability from most presence/absence models (including GLMs) is hereafter termed the probability of occurrence to distinguish it from the probability of true occupancy estimated by occupancy models. Despite biased occupancy estimates, GLMs and other presence/absence models have one advantage over occupancy models: GLMs can incorporate sites that are only sampled once and sites with multiple samples, leading to larger sample sizes than occupancy models which only use repeatedly sample sites.

Only six studies have used SDM to examine extinct nektonic marine species and two studies examined the past distribution of extant nektonic marine taxa (Bigg et al., 2008; Provan et al., 2009). Of these eight studies, only one investigated an extinct marine
vertebrate (Pimiento et al., 2016). Rovelli and Myers (2016; 2018), Carrier (2018), Carrier and Myers (2018), and Carotenuto et al. (2020) incorporated ammonites and nautiloids which lived as nektonic to nektobenthic organisms. Pimiento et al. (2016) used fossil tooth occurrences of *Otodus (= Carcharocles) megalodon (sensu Shimada et al., 2017)* to examine possible extinction-causing factors. Bigg et al. (2008) and Provan et al. (2009) studied the distributions of the extant Atlantic cod (*Gadus morhua*) and the North Atlantic copepod (*Calanus finmarchicus*), respectively, during the Last Glacial Maximum. Paleontological SDM literature includes two occupancy model and two GLM applications. Liow (2013) used occupancy models to examine occupancy shifts for Late Ordovician brachiopods from the Cincinnatian Arch. Like most SDM studies regarding extinct marine organisms, that study did not include any nektonic organisms or vertebrates. Lawing et al. (2021) applied occupancy models to study the influence of detection variability on relative abundance of late Quaternary pollen taxa. The two GLM studies focused on non-marine taxa. Melchionna et al. (2018) used GLMs to investigate the extinction-causing triggers for Neanderthals (*Homo neanderthalensis*) during the Pleistocene. Eduardo et al. (2018) used GLMs to model South American mammal and caiman distributions during the late Pleistocene. Additional paleontological SDM applications are reviewed by Svenning et al. (2011) and Varela et al. (2011). It should be noted that Myers and Lieberman (2011) and Myers et al. (2013) investigated competition amongst predatory and durophagous marine vertebrates, and the relationship between mollusc geographic range size and survivorship, respectively, in the WIS. On the surface these studies may appear to represent SDMs (though the authors do not make this claim), but these studies cannot be considered SDMs because minimum bounding polygons were
used to generate paleogeographic ranges from occurrences of included marine taxa rather than to relate the occurrences to paleoenvironmental or biological data.

This study represents the first GLM and occupancy modeling applications to characterize extinct marine vertebrate ecosystems and the first SDM analysis of *Hesperornis* occurrences. The methods used here incorporate absence data to enable estimation of occupancy rather than relative habitat suitability (Guillera-Arroita et al., 2015; MacKenzie et al., 2018). GLMs test larger samples of sites than occupancy models by including locality information from single-sample sites. Occupancy modeling, through collection of site detection histories, permits the identification of factors that may influence detectability of fossil taxa, such as differential collection and sampling methods or preservational biases (e.g., Liow, 2013; MacKenzie et al., 2018). The separation of these factors from covariates that influence occupancy should improve occupancy probability estimates and reduce false-absence frequency (e.g., Tyre et al., 2003; Dorazio, 2014; Lahoz-Monfort et al., 2014; Guillera-Arroita et al., 2015; MacKenzie et al., 2018). Comparison between GLMs and occupancy models identifies covariates or covariate groupings predicted by multiple model types to most influence *Hesperornis* occurrences. Each model type, while overlapping for 22 localities, contributes different information: single-sample sites from GLMs and detection histories from occupancy models.

**Institutional abbreviations:** AMNH: American Museum of Natural History, New York City, New York; CFDC: Canadian Fossil Discovery Center, Morden, Manitoba; CMN: Canadian Museum of Nature, Ottawa, Ontario; FHSM: Sternberg Museum of Natural History, Fort Hays State University, Hays, Kansas; SDSM: Museum of Geology, South
Dakota School of Mines and Technology, Rapid City, South Dakota; SMU SMP: Shuler Museum of Paleontology, Southern Methodist University, Dallas, Texas; TMM: Jackson School Museum of Earth History (formerly the Texas Memorial Museum), The University of Texas at Austin, Austin, Texas; UCM: University of Colorado Museum of Natural History, University of Colorado Boulder, Boulder, Colorado; UCMP: University of California Museum of Paleontology, Berkeley, California; UNSM: University of Nebraska State Museum, Lincoln, Nebraska; YPM VPPU: Yale Peabody Museum, Vertebrate Paleontology, Princeton University Collection, New Haven, Connecticut.
MATERIALS AND METHODS

Data Collection and Processing

Faunal occurrence and lithologic data were collected from the Paleobiology Database (PBDB), Integrated Digitized Biocollections database (iDigBio), museum collections records, and published literature using the search terms “Hesperornis,” “Vertebrata,” “Campanian,” and a geopolitical region (i.e., state, province, or territory). These data represent Campanian fossil vertebrate-bearing localities across the WIS. Faunal occurrences from Texas were also obtained during collections visits to SMU SMP and the TMM. Collections visits to SDSM and FHSM provided additional faunal occurrences for South Dakota and Kansas, respectively. Site lithology and geochronologic age were determined using published literature.

Liow (2013:195) defined faunal sites as “a unique combination of a geographic location, a depositional sequence, and a facies for the single-time-interval occupancy model.” This study follows a similar approach with lithology (mudrock, sandstone, or carbonate) used rather than facies. Faunal sites were grouped into a single locality if they fell within an approximately 10-km radius and were from the same formation or member. The 10-km buffer combines isolated faunal occurrences to reduce preservation and collection biases. While subjective, the 10-km buffer accounts in-part for the high mobility of WIS taxa that were likely a component of regional faunal assemblages. Localities were only recorded if they contained at least two taxa of genus or higher rank. Sites that lacked chronostratigraphic clarity (i.e., the source rock unit was unclear or spanned several geologic ages), had faunas generalized across many variable lithologies and depositional environments, or had an unknown georeferenced position were not
included. These constraints resulted in a total of 61 fossil localities of which 26 bear *Hesperornis* material (Fig. 2). Some of the models used 22/61 Campanian fossil localities of which 11 bear *Hesperornis* material (Fig. 3). These models are described in the Data Analysis section below.

PBDB and iDigBio localities were checked against the referenced literature or museum collections to ensure accuracy. These localities were georeferenced using coordinates provided by the referenced article or through approximation in Google Earth Pro 7.3 (Google, LLC, 2020) using either a locality map or a site description. Localities listed in Public Land Survey System (PLSS) notation were placed in the center of the smallest known PLSS unit (unless the article noted otherwise) using Google Earth Pro. Localities recorded following the Dominion Land Survey (DLS) and the Alberta Township Survey System variant, the western Canada equivalents of the PLSS, were estimated using township and range grid shapefiles for Alberta and Saskatchewan in ArcMap 10.7.1 (ESRI, Inc., 2019). Geographic positions of sites follow decimal degree notation. It is noted in Supplement S2 if the online collections data aggregates did not properly place the specimens in geographic space.

If an updated taxonomic identification was found for sampled sites, the specimen identification was changed to follow the most recent article. Two examples are *Trinacromerum bonneri* and *Hainosaurus pembinensis* from the CFDC Manitoba Escarpment sites; these taxa were re-identified as *Dolichorhynchops* and *Tylosaurus* species, respectively (O’Keefe, 2008; Bullard and Caldwell, 2010). Reidentification of included specimens outside of literature revisions was beyond the scope of this project. Traditional terrestrial groups were ignored from sampled localities for analysis (although
noted as present in Supplement S2). Excluded vertebrate taxa include all non-avian dinosaurs, non-mosasaurian squamates, Lissamphibia, and Mammalia. These taxa were either not present across most dataset localities or would not have lived within the same community. Non-tetrapod sarcopterygians were excluded because they are present at only three sites (low sample size): *Megalocoelacanthus dobiei* of Kansas (AMNH FF 20267 and the unpublished FHSM VP-18758) and *Coelacanthus granulatus* from Utah (UCM 95431) (Schwimmer et al., 1994; Dutel et al., 2012; Harrell and Ehret, 2019; Jacob Van Veldhuizen, pers. comm., 2020). Their limited geographical distribution within 1.5° of latitude also merits exclusion. Because this research is focused on vertebrate communities, invertebrate and plant taxa were also excluded.
Figure 2. Map of Campanian fossil-bearing localities included in the full dataset GLMs. Orange triangles represent Campanian *Hesperornis* occurrences; blue circles represent Campanian fossil localities (inclusive of the *Hesperornis*-bearing sites). Symbols used to represent fossil localities are not to scale. Data were obtained from the Paleobiology Database, museum collections records, and published literature (see Supplements S2-S3).
Figure 3. Map of Campanian fossil-bearing localities included in the reduced dataset GLMs and occupancy models. Orange triangles represent Campanian *Hesperornis* occurrences; blue circles represent Campanian fossil localities (inclusive of the *Hesperornis*-bearing sites). Symbols used to represent fossil localities are not to scale. Data were obtained from the Paleobiology Database, museum collections records, and published literature (see Supplements S2 and S4).

Thirty covariates were used for modeling the distribution of *Hesperornis*, the target species (Tables 2 and 3). Of these 30 variables, 24 covariates represent contemporaneous taxonomic occurrences and include both genera and larger clades.
Of these covariates, *Tylosaurus*, Polycotylidae, *Squalicorax*, and *Enchodus* represent known or inferred ecological interactions with *Hesperornis* and other hesperornithiforms (Table 1). While interactions between *Enchodus* and *Fumicollis hoffmani* (UNSM 20030) or *Squalicorax* and an indeterminant hesperornithiform (SMM P2000.12.10) do not include *Hesperornis* (Martin and Tate, 1976; Bell et al., 2015; Shimada and Hanks, 2020), the similarly-sized *Hesperornis* might have had a similar ecological relationship with these organisms. Eight other genera were included because they occurred at six or more localities and were present in three or more faunal communities with *Hesperornis* (Table 2). Genera were examined rather than species to increase sample size and avoid taxonomic disputes (see Wilson et al., 2011; Aotsuka and Sato, 2016; Bell and Chiappe, 2016 for taxonomic discussions concerning *Hesperornis* spp.). Genera identified with a ‘?’ or the comparative abbreviation ‘cf.’ were included to increase sample sizes and follow an occupancy modeling assumption that there are no incorrect identifications (i.e., false-positives) (MacKenzie et al., 2018). While false-positives can be modeled (e.g., Royle and Link, 2006; Chambert et al. 2015), it is beyond the scope of the current project. Eight larger taxonomic groups (‘order’ or ‘class’ rank in Linnean classification) and five smaller clades (‘family,’ ‘parafamily,’ or ‘sub-family’ rank, including Polycotylidae) were also examined (Table 2). The detection probabilities for each of the non-*Hesperornis* taxonomic occurrences are currently unknown as they were not the target species of this study.

Only a handful of paleo-SDMs incorporate biotic factors into their models. Most of the SDM/ENM papers by A. L. Stigall and co-authors examined Ordovician and Devonian benthic invertebrates with models incorporating biofacies (Stigall Rode and
Lieberman, 2005; Dudei and Stigall, 2010; Malizia and Stigall, 2011; Walls and Stigall, 2011; Brame and Stigall, 2014). These covariates include paleocommunities that represent invertebrate communities found in varying levels of water turbulence, different substrate types (i.e., deep/muddy to shallow/firm), anoxic conditions, continental settings, or dominance by one species (Stigall Rode and Lieberman, 2005; Dudei and Stigall, 2010; Malizia and Stigall, 2011; Walls and Stigall, 2011; Brame and Stigall, 2014). The influence of biofacies on the target species, however, was not interpreted despite being scored as a categorical or continuous variable. Maguire and Stigall (2009) use crocodilian presence/absence data as a proxy for paleotemperature in the Miocene-Pliocene Great Plains of North America. Crocodilian occurrences are used as further evidence for paleotemperature and vegetative shifts across the region that influenced species distributions within Equinae. Carotenuto et al. (2020) is the only paleo-SDM study to use occurrences of contemporaneous taxa as direct predictors (rather than a proxy) for the paleogeographic distribution of their target species. Both Maguire and Stigall (2009) and Carotenuto et al. (2020) interpolate occurrences of coeval species within their study area.

Three covariates describe general sedimentary lithologies: mudrock, sandstone, and carbonate (Table 3). The mudrock covariate contains shales, mudstones, and siltstones (i.e., fine-grained siliciclastics) while the carbonate covariate describes limestones, including chalks. Marls and calcareous mudstones/shales were split between the mudrock and carbonate covariates. The 24 taxa and three sedimentary lithologies were coded as binary values with 0 and 1 representing absence and presence, respectively.
Three spatial covariates were examined (Table 3): paleolatitude (°N) at 80 Ma, relative position in the WIS (WIS position), and outcrop area. Paleolatitude (a continuous variable) required conversion from modern latitude in GPlates 2.1.0 (Müller et al., 2018) using the Scotese (2016) paleogeographic reconstructions. For the GLMs, paleolatitudes were standardized between 0 and 1 by dividing the paleolatitudes by 90°N; 90°N represents the maximum latitude for the Northern Hemisphere. A different standardization method was used for the occupancy models because division of paleolatitude by 90°N tended to create model errors. Z-score transformation (e.g., Schumacker and Tomek, 2013) was used instead. This transformation subtracts the sample mean before dividing by the standard deviation of the sample. Resultant values follow the same statistical distribution shape (Warren et al., 2011) and maintain interval-value differences proportional to the original data (pers. obs.).

WIS position (a categorical variable) divides the WIS into thirds: west, middle, and east. This covariate was used instead of paleolongitude because the qualitative nature accounts for WIS shoreline curvature across paleolongitude lines. The WIS position covariate also roughly reflects bathymetry with the deepest water in the middle WIS (e.g., Longman et al., 1998; Lowery et al., 2018). Fossil localities were sorted into west, middle, and east categories by comparing modern positions to the Blakey (2014) and Scotese (2016) paleogeographic reconstructions. The Blakey (2014) reconstructions include the early-late Campanian WIS whereas the Scotese (2016) reconstructions only include the middle and late Campanian WIS. The rough groupings of sites into west, middle, and east should closely resemble quantified distances from either the western or eastern paleoshorelines. Quantified distances are less subjective when dividing sites into
regions, but this quantitative covariate has not yet been tested. Paleolatitude and WIS position were used to examine potential habitat preferences of *Hesperornis* linked to geographic location in the WIS. These preferred regions or paleolatitudes may reflect sea surface temperature, bathymetry, or ecology.

Outcrop area (km$^2$) was approximated by clipping mapped geologic unit polygons to United States county boundaries and Canadian census division boundaries using ArcMap (sources included in Supplement S5). While not necessarily correlated with exposure area, outcrop area is a proxy for the relative amount of available rock and has been applied by multiple authors to address detection or sampling biases (e.g., Crampton et al., 2003; Uhen and Pyenson, 2007; Wall et al., 2009; Dunhill, 2011; 2012; Dunhill et al., 2012; Irmis et al., 2013). This covariate is used to assess if preservational bias influences *Hesperornis* occurrences and detection. The z-score transformation was used for outcrop area to reduce the magnitude difference between binary numbers (i.e., presences and absences) and tens to thousands of square kilometers.
Table 2. Taxonomic covariates used for modeling. Larger taxonomic groups denote clades that are designated as ‘order’ or ‘class’ rank in traditional Linnaean classification whereas taxonomic subgroups correspond to ‘family,’ ‘parafamily,’ or ‘sub-family’ rank.

<table>
<thead>
<tr>
<th>Larger Taxonomic Group</th>
<th>Taxonomic Subgroup</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chondrichthyes</td>
<td></td>
<td>Squalicorax *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cretalamna</td>
</tr>
<tr>
<td>Actinopterygii</td>
<td></td>
<td>Xiphactinus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enchodus *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cimolichthys</td>
</tr>
<tr>
<td>Testudines</td>
<td></td>
<td>Toxochelys</td>
</tr>
<tr>
<td>Non-<em>Hesperornis</em> Aves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crocodyliformes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pterosaurs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plesiosauria</td>
<td>Elasmosauridae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Polycotylidae *</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dolichorhynchops</td>
<td></td>
</tr>
<tr>
<td>Mosasauridae / Mosasauria</td>
<td>Russellosaurina</td>
<td>Tylosaurus *</td>
</tr>
<tr>
<td></td>
<td>Plioplatecarpinae</td>
<td>Platecarpus</td>
</tr>
<tr>
<td></td>
<td>Plioplatecarpus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mosasaurinae</td>
<td>Clidastes †</td>
</tr>
</tbody>
</table>

* Indicates a taxon hypothesized to have paleoecological interactions with *Hesperornis* or another hesperornithiform (see Table 1).

† Specimens included as *Clidastes* should be interpreted as non-monophyletic and likely a paraphyletic basal mosasaurine grade until the genus is better resolved, a problem that has been known for over 20 years (e.g., Bell, 1997; Christiansen and Bonde, 2002; Dutchak and Caldwell, 2009; Palci et al., 2013; Lively, 2018; Madzia and Cau, 2020).
Table 3. Non-taxonomic covariates used for modeling.

<table>
<thead>
<tr>
<th>Sedimentary Lithology</th>
<th>Covariate Name</th>
<th>Spatial Covariate</th>
<th>Transformation for Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mudrock</td>
<td>Paleolatitude (°N) *</td>
<td></td>
<td>Divided by 90°N (GLMs)</td>
</tr>
<tr>
<td></td>
<td>- Standardized at 80 Ma -</td>
<td></td>
<td>Z-score transformed (occupancy models)</td>
</tr>
<tr>
<td>Sandstone</td>
<td>WIS position</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Carbonate</td>
<td>Outcrop area (km²)</td>
<td></td>
<td>Z-score transformed</td>
</tr>
</tbody>
</table>

* Paleolatitude is hypothesized to influence the distribution of *Hesperornis* by Nicholls and Russell (1990), Feduccia (1999), and Rees and Lindgren (2005).

Maps depicting *Hesperornis*-bearing localities and other sampled sites were created in ArcMap 10.7.1 and ArcGIS Pro 2.8.0 (ESRI, 2021). Site data and maps of North America were formatted for the North American Datum 1983 and projected to North America Lambert Conformal Conic. For Figure 1, modern geographic coordinates of Campanian *Hesperornis* localities were converted to a paleogeographic location (both in decimal degrees) using the Scotese (2016) paleogeographic reconstructions and GPlates to create a locality shapefile. This geographic conversion was used to account for tectonic movement of the North American Plate that included translation and rotation of localities. All localities were reconstructed for the early middle Campanian (80 Ma) which is more site representative of the dataset (early-late Campanian) than the late Campanian (75 Ma) reconstruction (Scotese, 2016) (Fig. 1). Figures 2 and 3 were created using political boundary shapefiles derived from ArcGIS Online (ESRI, Inc., 2020). The modern coordinates for Campanian fossil localities and *Hesperornis*-bearing sites were used.
The supplemental materials contain data on fossil assemblages and localities (Supplements S2-S4). Sources of faunal assemblages and map data are also provided (Supplement S5).

Data Analysis

All results except for occupancy models were produced in the software and programming language R 4.0.4 (R Core Team, 2021). R codes and formatted data used for analyses are in Supplements S3, S4, and S6. Prior to running the GLMs and occupancy models, the 24 taxonomic covariates (Table 2) were compared to each other, not *Hesperornis*, to assess independence using Jaccard’s similarity index (Boyle et al., 1990; Real and Vargas, 1996). This was accomplished using the R packages ‘vegan’ and ‘ade4’ (Chessel et al., 2004; Dray and Dufour, 2007; Dray et al., 2007; Oksanen et al., 2019). Though typically applied to investigate community similarity between two sites using presence and absence data, Jaccard’s index was used to compare pairwise occurrences of the 24 taxonomic covariates and their degree of similarity. This index was chosen over other similarity indices because it does not depend on shared site absences or the total sample size (Boyle et al., 1990; Real and Vargas, 1996; Robert Channell, pers. comm., 2020). Similarity values ≥ 80% among taxa or communities are generally considered noteworthy (Robert Channell, pers. comm., 2020) and this threshold is followed here. Pairings between a taxonomic covariate and itself or with its more inclusive taxonomic group (e.g., *Xiphactinus* and Actinopterygii) lacked independence and were ignored. Predicted values range between 0 and 1 with 0 representing a pairwise relationship where the occurrences are most dissimilar and 1 representing full similarity. High similarity
values indicate two taxa whose co-occurrences likely reflect an ecological relationship or both taxa shared environmental preferences. If two taxa had similar occurrences (similarity values $\geq 80\%$), this would indicate nonindependence between the two covariates and only one taxon of the pair could provide strong insight into the distribution of *Hesperornis*. The occurrences of the second taxon would provide no additional information and this taxonomic covariate would be removed before to running the SDMs.

*Generalized Linear Modeling*

Generalized linear models (GLMs) are modified linear regression models that can incorporate data that violate normal distribution assumptions common to ecological data (Guisan et al., 2002; Agresti, 2013). These models have two components: maximum likelihood estimation and a link function (Guisan et al., 2002; Elith et al., 2006). Likelihood is the probability, multiplied by a constant, that a hypothesis explains the observed data and maximum likelihood estimation is the method that identifies the combination of covariate estimates that produce the largest value from a likelihood function (Hastie et al., 2009; James et al., 2013; Etz, 2018; MacKenzie et al., 2018). A hypothesis with a higher likelihood has more support given the data than a hypothesis with a smaller likelihood (Etz, 2018). The second GLM component, the link function, expresses covariate estimates as a linear equation to explain the response variable (Guisan et al., 2002; Hastie et al., 2009; MacKenzie et al., 2018). In this study, *Hesperornis* occurrences are the response variable.
Two sets of GLMs were generated in R for this study. One GLM set was created using a modified R script originally written by Dr. Robert Channell of Fort Hays State University. This set incorporated the full dataset (61 localities) (Fig. 2). The second set used a reduced dataset of 22 localities and is comparable to the occupancy models (see below) (Fig. 3). Unfortunately, outcrop area tended to create fitting errors in both GLM sets, likely caused by a difference in magnitude between outcrop area and binary covariates. Removal of this covariate reduced the number of models afflicted by errors and resulted in only 29 covariates used for most GLM analyses. Outcrop area, however, did not produce errors in the occupancy models.

The R script requires the packages ‘ResourceSelection,’ ‘boot,’ and ‘jtools’ (Davison and Hinkley, 1997; Lele et al., 2019; Canty and Ripley, 2020; Long, 2020). After loading the full dataset and specifying the model covariates, the variables were scaled with a z-score standardization (Schumacker and Tomek, 2013) and incorporated into the model dataframe (i.e., the data table or array). The dataframe was then split into *Hesperornis* presences and absences. Jackknife resampling (e.g., Schumacker and Tomek, 2013) was completed for 1000 randomizations to subsample 26 of the 35 absences from the dataset. This subsampling created a balanced GLM (i.e., equal presences and absences) to reduce statistical error that could be generated from unequal presences and absences. Once subsampling was complete, the absences were recombined with the presences to form a single balanced GLM dataframe. The GLM was then built using a logistic link (logit) function that transforms covariate values into probabilities to follow a linear model (e.g., Burnham and Anderson, 2002; Agresti, 2013; James et al., 2013; MacKenzie et al., 2018). This GLM was fit to the dataframe. Overall GLM
significance was assessed using the R function `anova()` to compare the GLMs with covariates to the null model that lacked covariates.

GLM performance was assessed using Akaike’s Information theoretic Criterion (AIC) to describe the relative information loss when the model is retained (Akaike, 1973; 1974; MacKenzie et al., 2018). AIC can be calculated using the following equation:

\[
AIC = -2 \ln(L(\hat{\theta}|x)) + 2\delta \tag{Eq. 1}
\]

Where \(\ln(L(\hat{\theta}|x))\) represents the likelihood function and \(\delta\) indicates the number of modeled variables (Akaike, 1973; 1974; MacKenzie et al., 2018). The first term, , describes the unexplained model variation and the second, , is a penalty term that prevents model improvement every time a new covariate is added. The second term encourages parsimony in model selection. Models with a lower AIC value retain the most information and are “estimated to be ‘closest’ to the unknown reality that generated the data” (Burnham and Anderson, 2002:62; MacKenzie et al., 2018). The differences in AIC values (\(\Delta AIC\) or \(\Delta_i\)) are examined to rank models by the amount of information lost with lower AIC values being better (Burnham and Anderson, 2002; Burnham et al., 2011; MacKenzie et al., 2018). \(\Delta AIC\) is determined by the following equation:

\[
\Delta AIC_k = AIC_k - AIC_{\text{min}} \tag{Eq. 2}
\]

where \(k\) indicates the model rank and \(\text{min}\) identifies the model with the lowest AIC value (Burnham and Anderson, 2002; Burnham et al., 2011; MacKenzie et al., 2018). Models with small \(\Delta AIC\) compared to the model with the smallest AIC are most supported as explainative of the data. The relative proportion of model support is based on an evidence ratio, that is, the probability or likelihood ratio between the highest-ranked model and any
other model (Burnham et al., 2011). The evidence ratio enables ΔAIC values to be described by relative likelihood values representing the amount of model support. ΔAIC ≤ 2 is a common threshold for the most supported models (Burnham and Anderson, 2002; Burnham et al., 2011; MacKenzie et al., 2018). With ΔAIC = 2, a model is approximately 2.7 times less likely than the highest-ranked model (Burnham et al., 2011, Table 1). In this study, the highest-ranked models are those with ΔAIC ≤ 2, following Burnham and Anderson (2002), and are used for ecological interpretations.

GLM data were assumed to be derived from a binomial distribution because occurrence has two states: present (1) and absent (0). GLM analysis in R returns p-values and estimates for each covariate coefficient and the intercept. Each coefficient has a sign (positive or negative) that indicates the effect (increase or decrease, respectively) the covariate exerts on Hesperornis occupancy. The intercept identifies the occurrence probability when the model covariates equal 0 or are not present at a site. R also calculates an AIC value and null and residual deviances for each model. Null and residual deviances were used to calculate the explained deviance or the explained data variation in Hesperornis occurrences. Explained deviance is a ‘pseudo-R²’ value derived from the following equation (modified from McFadden, 1974:Eq. 30; Dobson and Barnett, 2008:137):

\[
\text{Explained deviance} = \frac{\text{Null deviance} - \text{Residual deviance}}{\text{Null deviance}} = 1 - \frac{\text{Residual deviance}}{\text{Null deviance}} \quad (\text{Eq. 3})
\]

The second GLM set includes a reduced dataset of 22 Campanian fossil-bearing localities with 11 preserving Hesperornis (Fig. 3). These are the same localities that contain two or more sites in the occupancy models (see below) for a more direct
comparison. The modified R script was not necessary for the reduced dataset GLMs because the count of presences and absences is equal. Instead, the standard GLM function in R, glm() was used. GLMs were considered a good fit for the data if the models were significantly different from the null model (i.e., a model with no covariates).

**Occupancy Modeling**

Single-season occupancy models were created and run in PRESENCE v. 2.13.16 (Hines, 2006), an occupancy modeling program that estimates the occupancy and detection probabilities simultaneously. These models, which use maximum likelihood estimation and link functions (e.g., MacKenzie et al., 2018), can be constructed for one taxon (single-species model) or multiple taxa (multi-species model) and for one or more seasons. A season is one surveying period of sampled sites. For this study, only single-season models were used representing the entire Campanian; subdividing the Campanian into early, middle, and late would split the sample size of the current dataset and likely reduce result reliability.

Occupancy modeling necessitates multiple visits to a site for taxonomic sampling to build the detection history. Consequently, the dataset used in PRESENCE was restricted to 22 localities that combined two or more sites within a 10-km buffer to estimate occupancy and detection probabilities (Supplement S4). Of these 22 localities, *Hesperornis* was present in half (Fig. 3). *Hesperornis* occurrence data were entered into the Data Input Form in PRESENCE. The 30 covariates (Tables 2 and 3) were imported into the Site Covars tab of the Data Input Form. All PRESENCE analyses were run as
simple single-season models using 1000 bootstraps, 31 surveys (the maximum number of sites combined into one sampled locality), and one observation method (fossil occurrence data collection from previously identified specimens) within the season. No sampling occasion covariates were used because, with the lack of additional information, the localities were assumed to have been collected and studied in a similar manner.

Initially, models were run with a single covariate for either the occupancy or detection probability of *Hesperornis*, while the other probability was held constant. This established a covariate baseline. The 24 taxonomic covariates, paleolatitude, and WIS position were used for the occupancy probability, while outcrop area was used only in the detection probability. Outcrop area is controlled by erosional and depositional rates and this covariate would not have impacted *Hesperornis* occurrences in life. The lithological covariates were used for either probability but not both simultaneously. Each lithology reflects depositional environments that may be preferred by *Hesperornis* or more conducive to preservation of *Hesperornis* fossils. All single-covariate models were compared with the null occupancy model, \( \psi(.),p(.) \), where \( \psi \) and \( p \) are held constant for all sites, to evaluate if covariates affect predicted values of \( \psi \) and \( p \). In the null occupancy model, \( \psi \) equals the proportion of known *Hesperornis* presences over the total sites (0.500 in this study) and \( p = 1 \) (perfect detection). PRESENCE automatically calculates the covariate coefficients and the probabilities of occupancy and detection for each site and model. As with the GLMs, the covariate coefficient sign denotes the effect on the *Hesperornis* distribution. Occupancy model performance was assessed by AIC. Model and covariate statistical significance were not assessed in the occupancy models because \( p \)-values were not estimated, as likelihood estimation is the focus of these models.
The covariates from GLMs and occupancy models that performed better than the null model (i.e., returned a lower AIC value) were run again in various combinations until new models failed to fall below the AIC value of the highest-ranked model. This method was used to determine which covariate combinations (hypotheses) most explain the data because automatic testing of covariate combinations via automated step-wise regression methods are limited. These functions only evaluate a few models, introduce parameter estimation bias, tend to include covariates that are irrelevant to the target species, suggest overconfidence in the highest-ranked model, and do not present models that are ranked below the highest-ranked model (Babyak et al., 2004; Whittingham et al., 2006; Burnham et al., 2011). Covariates with AIC values below the null reduced the performance of models even when the low AIC covariates were paired with higher AIC covariates above the null model.

Occupancy model fit was assessed using Pearson’s chi-square test to compare the observed $\chi^2$ value from the raw dataset to the average $\chi^2$ generated by conducting a parametric bootstrap of the dataset (Burnham and Anderson, 2002; MacKenzie et al., 2018). If the observed $\chi^2$ value is significantly larger than the bootstrapped $\chi^2$ value, then the model is considered as a poor fit for the data. Prior to assessing each model, however, the ‘global’/most general model, the model that incorporates the most covariates, must be assessed for model fit using Pearson’s chi-square test and parametric bootstrapping. Fit of this model to the data is considered indicative that more parsimonious models can also fit the data (Burnham and Anderson, 2002; MacKenzie et al., 2018).
Non-SDM Tests

Separate tests were conducted in R to determine if outcrop area (km$^2$), modern latitude (°N), paleolatitude (°N), and WIS position significantly influence the number of *Hesperornis* specimens (p-value < 0.0125 using the Bonferroni Correction) (e.g., Armstrong, 2014). These tests use *Hesperornis* specimen counts rather than occurrence data to provide additional insight into preservational biases against *Hesperornis* and/or *Hesperornis* habitat preferences. Spearman’s rank correlation was used to evaluate raw outcrop area, modern latitude, paleolatitude, and region of the WIS because the occurrence data did not meet the assumptions required by linear regression or Pearson’s product moment correlation (Schumacker and Tomek, 2013; Heumann and Shalabh, 2016). Specifically, the data failed to meet normality and homoscedasticity (equal variance) assumptions. R returned the Spearman correlation coefficient ($\rho$) and a p-value. $\rho$ ranges from -1 to +1 with positive and negative values indicating positive or negative linear rank associations, respectively (e.g., Artusi et al., 2002). Spearman’s rank correlation would indicate the strength and direction (positive or negative) of the association between latitudinal/paleolatitudinal position, WIS region, or outcrop area and the number of *Hesperornis* specimens. Plots of the data may also demonstrate heavily sampled latitudes, paleolatitudes preferred by *Hesperornis*, or regions of the WIS and preservational biases toward certain outcrop area sizes.

A Kruskal-Wallis test was used to examine potential geographic bias of fossil localities using three samples (west, middle, and east). An analysis of variance (ANOVA) could not be used because the *Hesperornis* occurrence data failed to meet the normality assumption (Dalgaard, 2008). The Kruskal-Wallis test evaluates the differences in sample
population medians (Dalgaard, 2008). In this study, a significant difference ($p < 0.0125$) in the number of *Hesperornis* specimens between WIS regions could indicate a preservational bias or habitat preference toward one or more regions.
RESULTS

Jaccard’s index of similarity results find that the occurrences of non-*Hesperornis* avians or Crocodyliformes with *Plioplatecarpus* are the most dissimilar (0%). Plesiosauria and *Enchodus* are the most similar at 60% of sites. The relatively low similarity scores (< 80%) returned in the Jaccard’s index analysis in general supports independence among the 24 taxonomic covariates. Full results of the Jaccard’s index analysis, highest-ranked GLMs, highest-ranked occupancy models, and non-SDM tests for preservational bias or habitat preference are presented in Supplements S7-S12.

Results of Generalized Linear Model Analyses

Eighty-three GLMs were generated with the full dataset. Eight of these models were excluded because they returned errors when fit to the data. Of the remaining 75 models, all but five have AIC values that are smaller than the null model (AIC = 85.23; ΔAIC = 32.73). Only the two highest-ranked models (ΔAIC ≤ 2) are presented in Table 4 with the rest included in Supplement S9. All two highest-ranked models are statistically significant (p-value < 0.001) and contain the covariates Elasmosauridae, *Toxochelys*, and paleolatitude. The highest-ranked GLM (AIC = 52.503) contains four covariates: Elasmosauridae, Actinopterygii, *Toxochelys*, and paleolatitude. The model has a total explained deviance/variation of 62.2%. The second highest-ranked model (ΔAIC = 0.740) is similarly constructed with the exception of Actinopterygii. Nearly all covariates in the two highest-ranked models were significant (p-value < 0.05) with only Actinopterygii recovered as non-significant in Model 2 (p-value = 0.121). Coefficient
estimates for all two highest-ranked models were positive except for Elasmosauridae (Supplements S7 and S10).

The reduced GLM dataset was used to create 115 GLMs. Of these, 30 had to be excluded due to fitting or convergence errors. Of the 85 included models, 30 models return higher AIC values than the null model (AIC = 32.498; ΔAIC = 17.250). The 15 highest-ranked reduced dataset GLMs (ΔAIC ≤ 2) are listed in Table 5. The results of the remaining models are included in Supplements S7 and S11. The reduced dataset GLM with the lowest AIC includes the covariates of Actinopterygii, non-Hesperornis Aves, Elasmosauridae, and paleolatitude (AIC = 15.248; ΔAIC = 0.000; p-value < 0.001; explained deviance = 0.828). While the overall model is significant, no covariates in the model show significance, though Elasmosauridae (p-value = 0.088) and paleolatitude (p-value = 0.078) were marginally significant. Of the 12 covariates present in the 15 highest-ranked models, only paleolatitude (p-values < 0.048) and Elasmosauridae (p-value = 0.039) have statistically significant p-values for 3/15 and 1/15 highest-ranked models, respectively (Supplements S7 and S10). No other covariates show statistical significance in the highest-ranked models. Elasmosauridae, Plioplatecarpus, carbonate (except for one model), and mudrock have a negative impact on Hesperornis occurrences based on the covariate coefficient signs. The other covariates are positive. All 15 models with ΔAIC ≤ 2 are statistically significant (p-value < 0.001). The 15 highest-ranked reduced dataset GLMs are characterized by three repeating groupings of covariates that include some model overlap. The most common grouping appeared seven times: Elasmosauridae + Squalicorax + paleolatitude. Elasmosauridae + non-Hesperornis Aves + paleolatitude...
appeared five times while Elasmosauridae + Testudines + paleolatitude appeared three times.

The reduced dataset GLMs seem to be the least reliable of the three model types. This model type did not produce a general covariate grouping for the highest-ranked models like the other two model types. Instead, nine covariate groupings were identified that differed by one covariate. In addition, most covariates in the highest-ranked reduced dataset GLMs were not statistically significant. This contrasts with the highest-ranked full dataset GLMs that have nearly all significant covariates and is likely related to the small sample size of the reduced dataset GLMs. Only Elasmosauridae and paleolatitude were occasionally significant in the reduced dataset GLMs for 1/15 and 3/15 models, respectively. Lastly, the reduced dataset GLMs also returned more fitting errors (30 total) resulting in more excluded models than the other two model types. This count of excluded models is nearly four times higher than the full dataset GLMs (8 excluded) and one and one-half as many exclusions as the occupancy models (21 excluded) (Supplements S7, S9, S11, and S12). The combination of a smaller sample size (22 localities) and the inability to estimate detection probabilities likely reduced the reliability of reduced dataset GLM results and their ability to determine covariate significance. The reduced dataset GLMs will not be discussed in-depth further.
Table 4. Full dataset GLM results for $\Delta$AIC ≤ 2. Explained deviances were calculated using Eq. 3. ‘Std’ is an abbreviation for the use standardization techniques described in the Materials and Methods. Significant covariates and p-values are highlighted in red. The p-value column represents the significance of the overall model. Null and residual deviances for these models are presented in Supplements S7 and S9.

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Model</th>
<th>AIC (ΔAIC)</th>
<th>Explained Deviance</th>
<th>Overall Significance (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Elasmosauridae, Actinopterygii, Toxochelys, and StdPaleolatitude</td>
<td>52.503 (0.000)</td>
<td>0.622</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>2</td>
<td>Elasmosauridae, Toxochelys, and StdPaleolatitude</td>
<td>53.243 (0.740)</td>
<td>0.601</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td><strong>Threshold for Highest-Ranked Models</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null</td>
<td>None</td>
<td>85.230 (32.727)</td>
<td>0.000</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
Table 5. Reduced dataset GLM results for ΔAIC ≤ 2. Significant covariates and p-values (< 0.05) are highlighted in red. Other values are presented in Supplements S7 and S11.

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Model</th>
<th>AIC (ΔAIC)</th>
<th>Explained Deviance</th>
<th>Overall Significance (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Elasmosauridae, Actinopterygii, Non-Hesperornis Aves, and StdPaleolatitude</td>
<td>15.248 (0.000)</td>
<td>0.828</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>2</td>
<td>Elasmosauridae, Non-Hesperornis Aves, and StdPaleolatitude</td>
<td>15.419 (0.171)</td>
<td>0.757</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>3</td>
<td>Elasmosauridae, Actinopterygii, Squalicorax, and StdPaleolatitude</td>
<td>15.436 (0.188)</td>
<td>0.822</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>4</td>
<td>Elasmosauridae, Testudines, Non-Hesperornis Aves, and StdPaleolatitude</td>
<td>15.611 (0.363)</td>
<td>0.816</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>5</td>
<td>Elasmosauridae, Squalicorax, and StdPaleolatitude</td>
<td>15.765 (0.517)</td>
<td>0.745</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>6</td>
<td>Elasmosauridae, Chondrichthyes, Non-Hesperornis Aves, and StdPaleolatitude</td>
<td>15.900 (0.652)</td>
<td>0.807</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>7</td>
<td>Elasmosauridae, Testudines, Squalicorax, and StdPaleolatitude</td>
<td>15.940 (0.692)</td>
<td>0.805</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>8</td>
<td>Elasmosauridae, Crocodyliformes, and StdPaleolatitude</td>
<td>16.201 (0.953)</td>
<td>0.731</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>9</td>
<td>Elasmosauridae, Squalicorax, Enchodus, and StdPaleolatitude</td>
<td>16.213 (0.965)</td>
<td>0.796</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>10</td>
<td>Elasmosauridae, Non-Hesperornis Aves, Pterosauria, and StdPaleolatitude</td>
<td>16.419 (1.171)</td>
<td>0.790</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>11</td>
<td>Elasmosauridae, Mosasaurinae, Squalicorax, and StdPaleolatitude</td>
<td>16.439 (1.191)</td>
<td>0.789</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>12</td>
<td>Elasmosauridae, Testudines, and StdPaleolatitude</td>
<td>16.685 (1.437)</td>
<td>0.715</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>13</td>
<td>Elasmosauridae, Mosasaurinae, Squalicorax, and StdPaleolatitude</td>
<td>16.826 (1.578)</td>
<td>0.711</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>14</td>
<td>Pterosauria, Elasmosauridae, Squalicorax, and StdPaleolatitude</td>
<td>16.984 (1.736)</td>
<td>0.771</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>15</td>
<td>Elasmosauridae, Pliopletecarpinae, Squalicorax, and StdPaleolatitude</td>
<td>17.187 (1.939)</td>
<td>0.764</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td><strong>Threshold for Highest-Ranked Models</strong></td>
<td><strong>(2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Null</strong></td>
<td>None</td>
<td>32.498 (17.250)</td>
<td>0.000</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Results of the Occupancy Model Analyses

One hundred and eighteen models were generated in PRESENCE (Supplements S7 and S12). Twenty-one of these models were excluded due to potential convergence errors or negative values in the variance-covariance matrix included in the PRESENCE output. Only five of the remaining 97 models have ΔAIC ≤ 2 (Table 6) and 12 models had larger AIC values than the null model (Supplement S12). The highest-ranked model (AIC = 148.625) includes Elasmosauridae, paleolatitude, and position in the WIS as covariates that influenced ψ while carbonate and outcrop area influence the detection probability (p). The remaining four models differ from the highest-ranked model by the addition of an extra taxon, *Squalicorax*, *Tylosaurus*, or *Platecarpus*, or the subtraction of outcrop area as a detection covariate. Covariate coefficient signs (+ or -) are constant across all four highest-ranked models with positive signs returned for paleolatitude, position in the WIS, *Squalicorax*, and *Tylosaurus* (Supplement S10). Negative signs were returned for Elasmosauridae, carbonate, and outcrop area. The other 90 occupancy models are presented in Supplement S12.
Table 6. Occupancy model results for \( \Delta AIC \leq 2 \). Symbols: \( \psi \): probability of occupancy; \( p \): probability of detection; (.) : period notation, which indicates that the probability was held constant. Additional model information is included in Supplements S7 and S12.

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Model</th>
<th>AIC (( \Delta AIC ))</th>
<th>Model Likelihood</th>
<th>-2*Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \psi(\text{Elasmosauridae, StdPaleolatitude, WIS Position}), p(\text{Carbonate, StdOutcrop Area}) )</td>
<td>148.625 (0.000)</td>
<td>1.000</td>
<td>138.625</td>
</tr>
<tr>
<td>2</td>
<td>( \psi(\text{Elasmosauridae, Squalicorax, StdPaleolatitude, WIS Position}), p(\text{Carbonate, StdOutcrop Area}) )</td>
<td>148.694 (0.069)</td>
<td>0.970</td>
<td>136.694</td>
</tr>
<tr>
<td>3</td>
<td>( \psi(\text{Elasmosauridae, Tylosaurus, StdPaleolatitude, WIS Position}), p(\text{Carbonate, StdOutcrop Area}) )</td>
<td>149.165 (0.540)</td>
<td>0.763</td>
<td>137.165</td>
</tr>
<tr>
<td>4</td>
<td>( \psi(\text{Elasmosauridae, StdPaleolatitude, WIS position}), p(\text{Carbonate}) )</td>
<td>149.995 (1.370)</td>
<td>0.1369</td>
<td>141.995</td>
</tr>
<tr>
<td>5</td>
<td>( \psi(\text{Elasmosauridae, Platecarpus, StdPaleolatitude, WIS Position}), p(\text{Carbonate, StdOutcrop Area}) )</td>
<td>150.299 (1.674)</td>
<td>0.434</td>
<td>138.299</td>
</tr>
</tbody>
</table>

Threshold for Highest-Ranked Models (2)

Null | \( \psi(\cdot), p(\cdot) \) | 186.287 (37.662) | 0.000 | 182.287 |

**Model Comparisons**

A comparison between the full dataset GLMs and occupancy models is presented in Table 7 using highest-ranked models (\( \Delta AIC \leq 2 \)). The full dataset GLMs have an overall covariate grouping of Elasmosauridae + *Toxochelys* + paleolatitude. The covariate grouping of Elasmosauridae + carbonate + paleolatitude + WIS position + outcrop area defines most of the occupancy models. The highest-ranked full dataset GLMs and occupancy models share one model grouping: Elasmosauridae + paleolatitude.
Table 7. Covariate groupings shared among the highest-ranked models (ΔAIC ≤ 2).

<table>
<thead>
<tr>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Dataset GLMs (2 models)</strong></td>
</tr>
<tr>
<td>Elasmosauridae</td>
</tr>
<tr>
<td><em>Toxochelys</em></td>
</tr>
<tr>
<td>Paleolatitude</td>
</tr>
<tr>
<td><strong>Occupancy Models (5 models)</strong></td>
</tr>
<tr>
<td>Elasmosauridae</td>
</tr>
<tr>
<td>Paleolatitude</td>
</tr>
<tr>
<td>Position in the WIS</td>
</tr>
<tr>
<td>Carbonate *</td>
</tr>
<tr>
<td>Outcrop area *</td>
</tr>
<tr>
<td>Note: These covariates are in all five models except for outcrop area (4/5 models).</td>
</tr>
<tr>
<td><strong>All Model Types (7 models)</strong></td>
</tr>
<tr>
<td>Elasmosauridae</td>
</tr>
<tr>
<td>Paleolatitude</td>
</tr>
</tbody>
</table>

* Covariate was modeled as a factor of the detection probability.

Results of Non-SDM Tests

The Spearman’s rank correlation test for outcrop area finds no significant linear association between the number of *Hesperornis* specimens and outcrop area (ρ = 0.005, S = 2909.5, p-value = 0.980) (Table 8; Fig. 4). There are three outliers that are qualitatively located well outside of the main cluster of plotted points (the red circles in Fig. 4A). These sites correspond to the Pembina Member of the Manitoba Escarpment where 99 CFDC specimens of *Hesperornis* were recorded, the expansive Smoking Hills Formation deposits along the Anderson River, Northwest Territories, and the Pierre Shale deposits of east-central Saskatchewan. Removal of these outliers also results in a non-significant correlation (ρ = -0.135, S = 2297.2, p-value = 0.539).

A comparison between the number of *Hesperornis* specimens and latitude using Spearman’s rank correlation test finds a non-significant linear relationship (ρ = 0.147, S =
2493.8, p-value = 0.472) (Table 8; Fig. 5A). Four outliers are present and represent the CFDC Pembina Member-bearing sites of the Manitoba Escarpment (49.3°N) and three Arctic localities (> 66.5°N). If the four outlying points are removed, the association becomes stronger but remains non-significant (\(\rho = 0.208, S = 1402.8, p\)-value = 0.353) and the data remain scattered (Fig. 5B). Spearman’s correlation tests for paleolatitude and the number of *Hesperornis* specimens returned similar non-significant correlations with outliers (\(\rho = 0.043, S = 2799.8, p\)-value = 0.836) and without outliers (\(\rho = 0.077, S = 1634, p\)-value = 0.732) (Table 8; Figs. 5C-D). The lack of non-significant linear correlation between paleolatitude and the number of *Hesperornis* specimens is at odds with the results of the SDMs which predict paleolatitude as a major influence on *Hesperornis* occurrences. Figures 5C and 5D, however, indicate a possible non-linear association between *Hesperornis* specimen counts and paleolatitude with most *Hesperornis* specimens concentrated at mid-latitudes (~49-58°N). Aggregations at these paleolatitudes may help explain the ecological role of latitude on *Hesperornis* as detected by the SDMs (see Discussion).

The Kruskal-Wallis rank sum test for position in the WIS finds no significant difference in median *Hesperornis* specimen counts among regions (\(\chi^2 = 1.751, df = 2, p\)-value = 0.417) (Fig. 6). Spearman’s rank correlation test did not find a significant linear correlation between position in the WIS and the number of *Hesperornis* specimens (\(\rho = 0.094, S = 2650.5, p\)-value = 0.648) (Table 8; Fig. 7). Plots of the data show an outlier present in the sample of points from the eastern region of the WIS that represents the CFDC Pembina Member-bearing sites of the Manitoba Escarpment (Figs. 6A and 7A). Removal of this outlier does not create significance using the Kruskal-Wallis test (\(\chi^2 =")
2.285, df = 2, p-value = 0.319) nor covariance with position in the WIS using Spearman’s correlation (ρ = 0.019, S = 2551.6, p-value = 0.930) (Table 8; Fig. 6B and 7B).

Table 8. Spearman’s rank correlation results for the number of *Hesperornis* specimens vs. non-binary covariates. The outliers are considered as the red circles in Figs. 4-7 and were qualitatively assessed. The letter S is a test distribution statistic. ρ is the Spearman correlation coefficient. Full results of these non-SDM tests are presented in Supplement S7.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Outlier(s) Present?</th>
<th>S</th>
<th>ρ</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcrop Area</td>
<td>Yes</td>
<td>2909.5</td>
<td>0.005</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2297.2</td>
<td>-0.135</td>
<td>0.539</td>
</tr>
<tr>
<td>Latitude</td>
<td>Yes</td>
<td>2493.8</td>
<td>0.147</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1402.8</td>
<td>0.208</td>
<td>0.353</td>
</tr>
<tr>
<td>Paleolatitude</td>
<td>Yes</td>
<td>2799.8</td>
<td>0.043</td>
<td>0.836</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1634.0</td>
<td>0.077</td>
<td>0.732</td>
</tr>
<tr>
<td>Position in the WIS</td>
<td>Yes</td>
<td>2650.5</td>
<td>0.094</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2551.6</td>
<td>0.0186</td>
<td>0.930</td>
</tr>
</tbody>
</table>
Figure 4. Association of outcrop area (km$^2$) and the number of *Hesperornis* specimens using Spearman’s rank correlation. All locality data are included in A with the red circles indicating substantial outliers that were removed for B. Outcrop area data sources are listed in Supplement S5.
Figure 5. Association of latitude or paleolatitude (°N) and the number of *Hesperornis* specimens using Spearman’s rank correlation. Modern latitudes are depicted in A and B. Paleolatitudes at 80 Ma are used in C and D. All locality data are included in A and C with the red circles indicating substantial outliers that were removed for B and D.
Figure 6. Comparison of the number of *Hesperornis* specimens across the WIS using a Kruskal-Wallis analysis. The western, central, and eastern regions of the WIS are coded as 0, 1, and 2, respectively. The outlier (Pembina Member localities in southern Manitoba) is denoted as a red circle in A. No difference in the number of *Hesperornis* specimens was identified when the outlier is retained in A ($\chi^2 = 1.751$, df = 2, p-value = 0.417) or removed in B ($\chi^2 = 2.285$, df = 2, p-value = 0.319).
Figure 7. Association of position in the WIS and the number of *Hesperornis* specimens using Spearman’s rank correlation. The western, central, and eastern regions of the WIS are coded as 0, 1, and 2, respectively. All locality data are included in A with the red circle indicating an outlier (Pembina Member localities in southern Manitoba) that was removed for B.
DISCUSSION

SDM paleoecological interpretations are limited to relationships between *Hesperornis* (the target species) and other taxa or environmental variables. In this study, *Hesperornis* is interpreted as having strong paleoecological relationships when covariates are supported by both the highest-ranked full dataset GLMs and occupancy models. Covariate groupings for one model type (Table 7) and ecological factors previously hypothesized to influence hesperornithiforms (Tables 1-3) are also discussed.

The current study finds paleolatitude (a continuous covariate) in all seven highest-ranked models (Tables 4, 6, and 7). Positive covariate coefficients indicate that, as paleolatitude increases, the *Hesperornis* occurrence/occupancy probability also increases. This result has several possible paleoecological interpretations. Feduccia (1999) and Rees and Lindgren (2005) hypothesized that hesperornithiforms migrated along the WIS, perhaps breeding in the High Arctic. Northward migration of *Hesperornis* for breeding is supported by a sub-adult (Wilson et al., 2014; 2016) and several small, possibly immature specimens from the High Arctic (Russell, 1967; Bryant, 1983; Nicholls and Russell, 1990; Feduccia, 1999; Hills et al., 1999; Rees and Lindgren, 2005; Wilson et al., 2016). However, multiple small and potentially immature *Hesperornis* specimens have been found in the Manitoba Escarpment of southern Manitoba (Nicholls and Russell, 1990; Aotsuka and Sato, 2016), with potentially more osteologically immature specimens from South Dakota and Manitoba depending on the validity of *Hesperornis macdonalldi* and *H. mengeli* (Martin and Lim, 2002; Aotsuka and Sato, 2016; Bell and Chiappe, 2016). One interpretation is that the presence of potentially immature individuals in Manitoba and South Dakota may not support migration to breed exclusively in the High Arctic, but the
original breeding hypothesis (Feduccia, 1999; Rees and Lindgren, 2005) could be modified to also include southern Manitoba as a potential breeding area (Aotsuka and Sato, 2016). Unfortunately, bone histology is inconclusive regarding whether *Hesperornis* overwintered in or migrated to the High Arctic (Wilson and Chin, 2014).

Alternatively, the immature *Hesperornis* individuals in southern Manitoba may represent mortality of skeletally immature individuals during migration. Modern fledgling guillemots (*Uria aalge*) have been observed swimming ~200-280 km on average across the Baltic Sea, a journey that begins when they are around 25% of adult body size (Hedgren and Linnman, 1979; Olsson et al., 1999). Perhaps young *Hesperornis* also began their southern migration at an early ontogenetic stage and the high abundance of *Hesperornis* specimens between 48.6-58°N paleolatitude reflects mortality during migration due to geographically restricted oceanographic processes such as shifts in current upwelling strength (reflective of biological productivity), storms, starvation, and exhaustion (e.g., Matović et al., 2017; Jones et al., 2018; Tavares et al., 2020). Mortality in Magellanic penguins (*Spheniscus magellanicus*), for example, is best explained by wave height and atmospheric pressure, both related to storm intensity, and reduction in primary productivity (Tavares et al., 2020). Perhaps some of these same stressors impacted migratory *Hesperornis* resulting in the higher abundance of specimens between southwestern South Dakota and Saskatchewan. This is supported by the hypothesized presence of oceanic fronts and mixing areas in the Central Rocky Mountain/Black Hills region during the late Cenomanian to early Turonian (e.g., Slingerland et al., 1996; Fisher and Hay, 1999; Elderbak and Leckie, 2016; Lockshin et al., 2017) and the Coniacian to early Campanian (Longman et al., 1998). These pre-Campanian fronts and seawater
mixing areas also extend into North Dakota, Manitoba, and Saskatchewan (Slingerland et al., 1996; Elderbak and Leckie, 2016; Lockshin et al., 2017). With multiple late Cenomanian to early Campanian studies reporting similar oceanographic mixing hypotheses, it is likely that the same fronts and mixing regions remained throughout the Campanian. Continued mixing of water masses in the Campanian is further supported by the presence of an upwelling zone or oceanographic front in the Denver Basin of Colorado during the early-middle Campanian (Timm and Sonnenberg, 2018).

Another explanation for the positive influence of paleolatitude on *Hesperornis* occurrences is preference of certain paleolatitudes, as indicated by denser *Hesperornis* concentrations, despite a broad paleogeographical range. While Spearman’s correlation test did not find a significant linear association between the number of *Hesperornis* specimens and paleolatitude (Table 8), Fig. 4 qualitatively demonstrates that *Hesperornis* specimens are most numerous between 48.6-58°N paleolatitude (43-53°N modern latitude) or southwestern South Dakota to central Saskatchewan. This region includes 222/237 specimens in the dataset (93.7%) (Fig. 1; Supplements S2, S6, and S7). Latitudinal differences in sea surface temperatures could drive increased *Hesperornis* occurrence or occupancy probabilities at these paleolatitudes, with an average sea surface temperature range of 13.7-21.6°C (late middle-late Campanian, southern Canada) to 16-24.2°C (late Campanian non-seep deposits, South Dakota) (He et al., 2005; Dennis et al., 2013; Landman et al., 2018). This temperature range may have been ideal for *Hesperornis* occurrences and may also reflect a region of increased death during migration. Alternatively, the highest abundance can be explained by the oceanic fronts running through Wyoming/South Dakota and southern Canada (Slingerland et al., 1996;
Fisher and Hay, 1999; Elderbak and Leckie, 2016; Lockshin et al., 2017). As regions of high productivity, the mixing zone along the fronts could serve as important feeding grounds for *Hesperornis* similar to how modern Magellanic penguins feed on prey concentrations in these zones (Boersma et al., 2009). The possibility that the influence of paleolatitude on *Hesperornis* occurrences may reflect variation in sampling effort across WIS deposits cannot be discounted at this time. Further study with the current or similar datasets should include covariates that quantify collection bias, perhaps by comparing the total specimens collected over exposure area or the number of active dig years, and covariates for geographic biases such as proximity to cities and towns.

Results from this study found elasmosaurid plesiosaurs present in all highest-ranked occupancy models and full dataset GLMs (Tables 4, 6, and 7). Negative covariate coefficients indicate that the presence of elasmosaurs reduced the probability of occurrence/occupancy for *Hesperornis* and/or that the models support high biogeographical disparity between Elasmosauridae and *Hesperornis*. This implies a possible predator-prey relationship, competitive interactions, or differences in habitat preference. It is unlikely that elasmosaurs exerted a predative pressure on adult *Hesperornis*, despite a large body size, due to their small heads and slender teeth (e.g., Everhart, 2017; Kear et al., 2017). The restricted width between elasmosaur quadrates (15 cm in the Cenomanian and largest WIS elasmosaur *Thalassomedon haningtoni*) and inflexible cranial articulations (Cicimurri and Everhart, 2001; Araújo and Poley, 2013) would not enable the swallowing of large adult *Hesperornis*. It is possible that the negative effect elasmosaurids exert on the *Hesperornis* distribution is a competitive pressure. Wilson et al. (2016) hypothesized that hesperornithiforms were generalists who
consumed fish, squid, and soft-bodied crustaceans. Similar diets are reported or inferred for most elasmosaurids based on gut contents (fish, cephalopods, crustaceans, other benthic invertebrates, and gastroliths), tooth morphology and wear, and cranial musculature models (Massare, 1987; Cicimurri and Everhart, 2001; McHenry et al., 2005; Araújo and Polcyn, 2013; Everhart, 2017; Kear et al., 2017). One exception is the Campanian-Maastrichtian Aristonectinae (*sensu* Otero, 2016) whose members are inferred filter-feeders (e.g., O’Keefe et al., 2017) and limited to the Southern Hemisphere (O’Gorman; 2019); therefore, aristonectine elasmosaurs would not have likely fed on or competed with hesperornithiforms, which were exclusive to the Northern Hemisphere (Bell and Chiappe, 2016; Tanaka et al., 2018).

Despite ecological similarities and some geographic overlap between Campanian *Hesperornis* and elasmosaurs, it is difficult to provide conclusive evidence for interspecific competition in the fossil record. Competing taxa must share similar feeding ecologies, be contemporaries, live in the same environments, and overlap geographically (Benton, 1996a,b; Ando & Fordyce, 2014; Scott and Anderson, 2021). Both the elasmosaur and *Hesperornis* distributions fit these conditions during the Campanian, but key information is lacking. Supporting evidence for ancient competition is often demonstrated across long time frames (i.e., multiple epochs or ages) and with evidence for character displacement (morphological shifts) or competitive replacement/displacement through negative diversity, abundance, or geographic correlations, or isotopic and microwear analyses on enamel (Feranec et al., 2009; Myers and Lieberman, 2011; Tyler and Leighton, 2011; Ando and Fordyce, 2014; Smith and DeSantis, 2020; Scott and Anderson, 2021).
While a competitive relationship between *Hesperornis* and elasmosaurs is not conclusive, it is possible these taxa shared a complex ecological relationship. Delphinids frequently herd fish into schools that become more accessible for seabirds and some whales are known to steal fish herded by diving seabirds (e.g., Vaughn et al., 2008; Anderwald et al., 2011; Thiebault et al., 2016; Veit and Harrison, 2017). Increases to individual success from group foraging between cetaceans and seabirds likely outweigh competitive pressures (Thiebault et al., 2016; Veit and Harrison, 2017) and this may have been true for elasmosaurs and *Hesperornis*. Alternatively, if *Hesperornis* was capable of diet-switching (*sensu* Wilson et al., 2016) like Adélie penguins (*Pygoscelis adeliae*) during competition with cetaceans (Ainley et al., 2006), a high diversity of prey species in a region may enable *Hesperornis* to reduce competition with elasmosaurs.

The absence of additional evidence for competition or another ecological relationship does not exclude differences in habitat preferences or other abiotic factors that may drive biogeographical disparity patterns between elasmosaurs and *Hesperornis*. Elasmosaurs and *Hesperornis* are most abundant in the cooler waters of mid- to high-paleolatitude localities (i.e., the Northern Interior Faunal Subprovince) (Fig. 8; Nicholls and Russell, 1990). These shared latitudinal distributions, and associated seawater temperature conditions, are unlikely to explain the negative correlation between these taxa. Small-bodied elasmosaurs (both osteologically immature and small-bodied adult individuals) and *Hesperornis* are also known from non-marine and lagoonal settings during the Campanian (e.g., Shufeldt, 1915; Fox, 1974; Sato et al., 2005; DeMar and Breithaupt, 2008; Street, 2020; Campbell et al., 2021) (Supplements S1, S2, and S13) with large-bodied adult and some smaller adult elasmosaurs found further offshore (Sato
and Wu, 2006; Campbell et al, 2021). Most occurrences of Elasmosauridae and *Hesperornis*, however, are in marine settings (Carpenter, 1999; Kubo et al., 2012; Wilson et al., 2016; Campbell et al., 2021) (Supplement S2) with Campanian *Hesperornis* abundance highest along the likely locations of oceanographic fronts or mixing zones (see paleolatitude discussion). Congregations of modern seabirds tend to form along mixing and upwelling regions where prey are abundant (e.g., Shealer, 2002; Ballance et al., 2006; Boersma et al., 2009; Fauchald, 2009; Wilson et al., 2019) and the *Hesperornis* occurrences may reflect this trend. Elasmosaurs, by comparison, have a broader distribution across the Campanian WIS (e.g., Carpenter, 1999; Kubo et al., 2012) (Fig. 8). This is not surprising given that elasmosaurs are thought to travel great distances within the WIS to consume gastroliths because those associated with elasmosaur skeletons differ in composition from the surrounding rock matrix or have shapes associated with nearshore to fluvial environments while the elasmosaur skeletons were found in marine deposits (Cicimurri and Everhart, 2001; Kubo et al., 2012). It is possible that elasmosaurs regularly traveled between different regions of the WIS to feed, but this is currently unknown. Perhaps the negative correlation between *Hesperornis* and elasmosaurs is related to habitat distributional patterns with elasmosaurs lacking preferred regions outside of mid- to high-paleolatitudes and *Hesperornis* congregating around oceanic fronts.
Figure 8. Distribution of Campanian *Hesperornis* and elasmosaurs in the WIS.

Campanian sites are reconstructed to the early middle Campanian (80 Ma) and depicted using an 80 Ma paleogeographic map (modified from Scotese, 2016). Symbols representing fossil localities are not to scale. Data were obtained from the Paleobiology Database, museum collections records, and published literature (Supplements S2 and S13). NiD identifies elasmosaur specimens that were not included in the current dataset because they were isolated occurrences without a vertebrate community (Supplement S13).
Covariate Groupings from One Model Type

It is possible that the full dataset GLM or occupancy model covariate groupings could reflect reality rather than covariate groupings shared by both model types (Table 7). In the full dataset GLM grouping, all covariates (Elasmosauridae + Toxochelys + paleolatitude) have a positive influence on Hesperornis occurrences except for Elasmosauridae. Elasmosauridae and paleolatitude likely have the same relationship to Hesperornis as discussed previously, but it is unclear how Toxochelys relates to Hesperornis, if at all. Current dietary interpretations for Toxochelys describe a consumer of fish, hard-shelled invertebrates, and soft-bodied organisms, supporting a generalist diet like Hesperornis and elasmosaurs. This is based on fish-bearing coprolites associated with UCM 70738 and snout emargination on other skulls (Carpenter, 2006; Carrino, 2007). An overlap in diet could imply competition, but the positive impact of Toxochelys on Hesperornis occurrences from the full dataset GLMs negates this conclusion. Alternatively, Toxochelys could be preferred to Hesperornis as prey of larger predatory species such as Tylosaurus and Squalicorax as Toxochelys has been found as stomach contents and with bite marks (Druckenmiller et al., 1993; Schwimmer et al., 1997; Martin and Bjork, 1987; Bell and Barnes, 2007; Everhart et al., 2013). A preference for Toxochelys as food would increase the probability of Hesperornis occurrence with reduction of predation pressure on the latter. However, despite partial geographic overlap in Manitoba, South Dakota, Kansas, and Arkansas (Gentry and Ebersole, 2018) (Fig. 1), there is no clear biogeographic pattern or biological explanation for why the full dataset GLMs support higher Hesperornis occurrence probabilities when Toxochelys was present. Any effects of Toxochelys-elasmosaur interactions on Hesperornis occurrences
are equally unclear. Although untested, perhaps these three generalist taxa co-occurred because all exhibited dietary flexibility through diet-switching.

The occupancy model covariate grouping (Table 7) is divided into two parts with potential influence on the *Hesperornis* distribution. Elasmosaurs, paleolatitude, and WIS position were modeled as important to the probability of *Hesperornis* occupancy while the detection probability was influenced by carbonate rocks and outcrop area. Of the five covariates, only paleolatitude and WIS position have a positive association with *Hesperornis*. For occupancy probabilities, elasmosaurs, paleolatitude, and the elasmosaur-paleolatitude interaction likely had the same effects as discussed above. Abundance of *Hesperornis* did not significantly vary west to east across the WIS (Figs. 6-7; Table 8), likely reflecting no preference by *Hesperornis* and the lack of preservational bias among these regions. The WIS position covariate was also not incorporated into any highest-ranked full dataset GLMs (Table 4; Supplements S7 and S9), further supporting this conclusion. The five highest-ranked occupancy models, however, detected a positive association of WIS position and *Hesperornis* occurrences. This association is likely not caused by differences in *Hesperornis* occurrence counts between regions because all are nearly equivalent between the west (3), middle (4), and east (4). Rather, the occupancy models may be detecting the interaction between WIS position and other occupancy probability covariates in the model. The Elasmosauridae-WIS position interaction is not likely to affect the *Hesperornis* distribution because elasmosaurs do not appear to have a strong preference for any west to east region. It is more likely that the interaction between WIS position and paleolatitude influenced *Hesperornis* occurrence probabilities. The interaction may manifest as the concentration
of *Hesperornis* occurrences in southwestern South Dakota (middle WIS) and Manitoba (eastern WIS), along the mixing region of southern and northern water masses, within 48.6-58°N paleolatitude where the *Hesperornis* specimen count is most numerous (222/237 specimens in the dataset) (Figs. 5-7).

The occupancy models predict that carbonate decreases the probability of *Hesperornis* detection, which likely reflects decreasing eustacy during the middle to late Campanian (e.g., Haq et al., 1987; Müller et al., 2008; Blakey, 2014). During the early Campanian, chalk and other calcareous deposits stretched from the Cretaceous Gulf Coast to the Dakotas and potentially to the Manitoba Escarpment (represented by the upper Boyne Member of the Carlile Formation) along the middle to eastern WIS, but after the early Campanian, ocean-derived carbonates disappear from the WIS as sea levels decrease (Shurr and Reiskind, 1984; Roberts and Kirschbaum, 1995; Bamburak et al., 2013; Shaw et al., 2017). This is reflected in the current full dataset where only two of the early Campanian carbonate-bearng sites are in South Dakota and Kansas while the other four early Campanian carbonate-bearing sites are located along the Cretaceous Gulf Coast in modern Texas. The 11 middle-late Campanian carbonate-bearing sites also fall along the Cretaceous Gulf Coast in Texas and Arkansas. *Hesperornis* is present at 4/17 carbonate-bearing sites in the full dataset with one occurrence each in Kansas and South Dakota (early Campanian) and two occurrences in Arkansas (middle-late Campanian).

In the occupancy model data, only 4/22 total sites bear carbonate lithologies, and *Hesperornis* is present at only one of these sites: the early Campanian Niobrara deposits in Kansas. The *Hesperornis*- and carbonate-bearing areas in Arkansas were not included in the occupancy models because the number of localities in Hempstead and Clark
Counties are unknown. Regardless, if these two combined sites were included in the occupancy models, the number of *Hesperornis*- and carbonate-bearing sites (3/22) would remain much smaller than *Hesperornis*- and mudrock-bearing sites (8/22). Therefore, it remains likely that eustatic shifts drive the apparent decrease in *Hesperornis* detection when carbonate is present.

Outcrop area does not have a linear relationship with *Hesperornis* specimen counts (Table 8; Fig. 4), but outcrop area is present in 4/5 highest-ranked occupancy models. In these models, outcrop area has a negative association with *Hesperornis* occurrences implying that larger outcrop areas decrease the probability of *Hesperornis* occupancy. This result is reasonable when compared to Fig. 4, which shows that most *Hesperornis* specimens came from < 2000 km² outcrop areas. Perhaps larger outcrop areas are not as heavily or evenly sampled and *Hesperornis* specimens are less likely to be observed. Larger outcrop areas may also not be completely sampled or were unevenly sampled because of their broader geographic size compared to small rock outcrops. It is unlikely that differences in outcrop area among lithologies influenced *Hesperornis* occurrences. *Hesperornis* specimens have been found primarily in mudrock deposits (Supplements S1 and S2) which have higher rates of sedimentation and may appear less fossiliferous than carbonaceous deposits. Sedimentation rates for the Sharon Springs Formation (shale) and the Campanian portion of the Niobrara Formation (marl to chalk) have been estimated at 10-104 m/myr (1-10.4 cm/kyr) and 9-26 m/myr (0.9-2.6 cm/kyr), respectively (Carpenter, 2006; Locklair and Sageman, 2008). Unfortunately, outcrop area trends in the occupancy models cannot be verified by the full dataset GLMs at this time due to model fitting issues that usually occur when outcrop area is an included covariate.
Perhaps with better resolution on preserved outcrop in North America, a possible influence on *Hesperornis* occurrences could also be identified from the full dataset GLMs.

Outcrop area is likely among the most important determinants of detection probability because this property provides the total amount of space available for sampling. This information helps characterize biases including collection effort, survey completeness for an outcrop, and variable exposure (e.g., Crampton et al., 2003; Peters and Heim, 2010). Unfortunately, outcrop area is often difficult to estimate at the level of rock exposure for a locality. Both Liow (2013) and this study encountered issues with precise measures of outcrop area because these data are typically not published online and may not have been measured. The lack of measurement is particularly challenging for older localities where determining outcrop size was not the purpose of the excavation (e.g., sites collected by O. C. Marsh and E. D. Cope) or the site may no longer be accessible (e.g., demolished road cuts or the now-submerged Bailey Island of the Missouri River in South Dakota) (Welles and Bump, 1949; pers. obs.). While outcrop area estimation from geologic maps is a useful proxy, these estimates do not typically account for recent geological cover (e.g., topsoil, alluvium, or glacial till) and human development (Uhen and Pyenson, 2007; Dunhill, 2011; 2012). Part of this problem can be circumvented by obtaining estimates using remote sensing and GIS to map the true rock exposure area rather than mapped geologic unit areas (Dunhill, 2011; 2012).

Alternatively, for localities whose dimensions are not provided in published literature, museum/institution records can be consulted to see how individual sites were sampled, if recorded. Both options, however, are quite time-intensive and would likely require
numerous research teams/projects at smaller scales before true outcrop exposure area analysis at the continental scale. For example, inquiry regarding the size of Campanian deposits for this study would require information on 61 sites (including 21 localities of 2-31 smaller sites each) and a search through the records of at least 26 institutions. If information on outcrop exposure area is lacking in collections records, use of remote sensing and GIS data (sensu Dunhill, 2011; 2012), or personal measurement of outcrops would be necessary if the outcrops remain accessible. Georeferenced coordinate uncertainty could constrain outcrop area from geologic maps to a more finite area around a known fossil locality. This restriction may better reflect true rock exposure area, but further tests are needed.

Other Covariates of Note

Polycotylids are another group of plesiosaurs in the WIS, but these taxa possess a shorter neck and longer skull unlike the long neck and short skull length characterizing elasmosaurs. Martin et al. (2016) hypothesized that a polycotylid plesiosaur attempted to eat a juvenile *Hesperornis* based on tooth marks discovered on an adult *Hesperornis* tibiotarsus (YPM VPPU 017208.D) in South Dakota. Since these bite marks show evidence of healing, they likely reflect an act of predation and not scavenging.

Campanian polycotylids could reach 6-7 m long with heads around 1 m long (Schumacher and Martin, 2016; Everhart, 2017). While perhaps not large enough to feed on adult *Hesperornis regalis*, *H. lumgairi*, and *H. gracilis* (≤ 1.9-2.0 m in length), polycotylids may have been able to eat juveniles or smaller *Hesperornis* species such as *H. macdonaldi* (~0.8 m long) (size estimates based on Martin and Lim, 2002; Wilson et
al., 2011; Aotsuka and Sato, 2016). In this study, however, polycotylid plesiosaurs did not appear as a covariate in any of the seven highest-ranked models among all model types (Table 7) and occurrences outside of the highest-ranked models had variable positive or negative covariate coefficients. Together, these results indicate that polycotylids likely had minimal impact on Hesperornis occurrences. Consequently, YPM VPPU 017208.D may represent a rare attack on Hesperornis rather than Hesperornis being the preferred prey of polycotylids. Additionally, stomach contents of polycotylids provide evidence of feeding on teleosts and ammonites, and their tooth morphology suggests a generalist diet such as fish and squid (Massare, 1987; Sato and Tanabe, 1998; Cicimurri and Everhart, 2001; Janzic, 2006).

Tylosaurus, a large mosasaur capable of reaching 13-14 m in length (Everhart, 2017), was hypothesized to be an opportunistic predator by Martin and Bjork (1987) based on stomach contents of SDSM 10439 that include Hesperornis sp. (Table 1). The wide variety of vertebrate taxa found as stomach contents from this specimen and other Tylosaurus individuals supports their hypothesis (e.g., Everhart, 2004a; Bell and Barnes, 2007). As a covariate in the distribution of Hesperornis, Tylosaurus was found in only one of the highest-ranked occupancy models (Table 6; Supplements S7 and S12) and likely reflects noise in the models. This indicates that Tylosaurus is unlikely to have exerted a strong influence, predative or otherwise, on Hesperornis despite occasionally feeding on the latter. Model results from this study do not refute the hypothesis that Tylosaurus was opportunistic (e.g., Martin and Bjork, 1984; Everhart, 2004a; Bell and Barnes, 2007) rather than preying on specific taxa.
Martin and Tate (1976) hypothesized that *Fumicollis hoffmani* (*sensu* Bell and Chiappe, 2015) consumed the teleost *Enchodus* and other fish based on coprolites/colonites associated with UNSM 20030 from Kansas (Table 1) (Martin and Tate, 1976; Bell and Chiappe, 2015). *Enchodus* could have been small enough to be consumed by *Hesperornis*. The largest *Enchodus* species (*E. petrosus*) reached 1.5 m long, but smaller individuals and other *Enchodus* species could be prey for a large hesperornithiform (Martin and Tate, 1976; Bell and Chiappe, 2015; Everhart, 2017). The largest WIS species of *Hesperornis* (*≤* 1.9-2.0 m long) were bigger than *F. hoffmani* (~1.1-1.3 m long) (based on measurements in Wilson et al., 2011; Bell and Chiappe, 2015; Aotsuka and Sato, 2016); therefore, these *Hesperornis* species could also consume smaller *Enchodus* individuals. However, *Enchodus* was present in none of highest-ranked models (Tables 4, 6, and 7), which indicates that *Enchodus* was not a major food source, or it may have been seasonally or regionally consumed by *Hesperornis*. This result is not surprising given the ubiquitous nature of *Enchodus* in the WIS (e.g., Goody, 1976; Martin et al., 1998; Cavin et al., 2012; Supplement S2) and the generalist diet and potentially migratory behavior of *Hesperornis* (Feduccia, 1999; Rees and Lindgren, 2005; Wilson et al., 2016). Modern seabirds often search vast distances of water for prey patches, with these patches behaving dynamically in geographic location and areal extents (e.g., Shealer, 2002; Vlietstra 2005; Boersma et al., 2009; Fauchald, 2009). These fluctuations can lead to migration between feeding areas and/or prey-switching. Seasonal or interannual environmental shifts and the distance of prey aggregations from seabird breeding colonies also influence the prey species targeted (e.g., Ainley et al., 1996; Vlietstra, 2005; Wells et al., 2017).
Hanks and Shimada (2002) and Shimada and Hanks (2020) hypothesized that hesperornithiforms were prey of the generalist shark *Squalicorax* based on *Squalicorax* cf. *falcatus* bite marks on a partial hesperornithiform tibiotarsus (SMM P2001.12.10) (Table 1). *Squalicorax* species could reach lengths of 2.0-4.8 m and lived throughout the WIS with the most Campanian occurrences limited to southern Canada and lower latitudes (e.g., Nicholls and Russell, 1990; Schwimmer et al., 1997; Shimada and Cicimurri, 2006) except for one occurrence on Devon Island, Nunavut in the High Arctic (Chin et al., 2008). *Squalicorax* had a tooth morphology resembling tiger sharks (*Galeocerdo cuvier*) who feed indiscriminately and are known to eat seabirds and carrion (e.g., Lowe et al., 1996; Compagno, 2001; Simpfendorfer et al., 2001; Dicken et al., 2017). *Squalicorax* species are probable scavengers or opportunists with evidence of feeding on traditional marine reptile taxa, dinosaurs, fish, pterosaurs, starfish, and hesperornithiforms (e.g., Druckenmiller et al., 1993; Schwimmer et al., 1997; Neumann, 2000; Everhart, 2004b; Shimada and Cicimurri, 2005; Ehret and Harrell, 2018; Shimada and Hanks, 2020). Despite feeding on hesperornithiforms, *Squalicorax* is unlikely to affect the *Hesperornis* distribution because, like the *Tylosaurus* and *Platecarpus* covariates, *Squalicorax* only appears in one of the highest-ranked occupancy models and is not a shared covariate in most of these models (Table 6).

**Model Comparison and Discussion**

Each SDM type has advantages and disadvantages. For GLMs and occupancy models, these differences are largely sampling and the detection probability estimation (Table 9). Unlike occupancy models, GLMs do not estimate the detection probability
(i.e., the probability of observing the target species) and do not provide unbiased occupancy probability estimates (i.e., the probability that the target species is present) for a site unless the species is always detected at all sites where it is present (Pearson, 2010; Guillera-Arroita et al., 2015). Unless this condition is met, occupancy probabilities for GLMs (exclusive of occupancy models) cannot be separated from detection probabilities (e.g., Guillera-Arroita et al. 2015; MacKenzie et al., 2018). In this way, occupancy models are more advantageous than GLMs because these models simultaneously model occupancy and detection probabilities.

On the other hand, GLMs are advantageous because these models can include information from single-sample and multi-sample sites for a combined higher site count. Detection probability estimation in occupancy modeling requires site detection histories built using multiple samples from each site. Most fossil sites have not been sampled or collected for determining the detection history which would necessitate repeated collection expeditions or sampling of multiple beds within a single time interval at one locality (Liow, 2013). Rather, paleontological localities are often only sampled once (e.g., Meyer, 1974; Kass, 1999; Sato, 2003; Bell et al., 2014), while other localities and regions are repeatedly sampled (e.g., Bardack, 1968; Lillegraven and McKenna, 1986; Case, 1987; Brinkman, 1990; Brinkman et al., 2005; DeMar and Breithaupt, 2008; Aotsuka and Sato, 2016). Detailed information on collection practices is typically unavailable in the published literature with fossil assemblages often combined into broad geological units (e.g., Case, 1987; Hoganson et al., 1996; Eaton et al., 1999; Peng et al., 2001). In this study, the larger GLM site availability is demonstrated with 61 localities,
but only 22 of these localities (36.1%) were repeatedly sampled regions appropriate for occupancy modeling.

Table 9. Advantages and disadvantages of model types used.

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Dataset GLMs</strong></td>
<td>• Require less data than occupancy models (i.e., do not need site detection histories)</td>
<td>• Cannot estimate probability of detection → biased occupancy estimates</td>
</tr>
<tr>
<td>(And Other non-Occupancy Model Presence/Absence Methods)</td>
<td>• Do not require multiple samples per site → more sites available for sampling</td>
<td></td>
</tr>
<tr>
<td><strong>Occupancy Models</strong></td>
<td>• Account for detection probability → unbiased occupancy estimates</td>
<td>• Require multiple samples per site → fewer sites available for sampling</td>
</tr>
<tr>
<td>(An Extension of GLMs)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ultimately, including both GLMs and occupancy models in a single study offers comparison among techniques and incorporate different data types. Agreement in individual covariates or covariate groupings across model types offers the most support for covariate/covariate grouping relationships with the target species. The agreement shows that, despite differing in total site or survey count, the models can inform on overall ecological patterns. Covariates present in multiple high-ranking models (ΔAIC ≤ 2) of one model type (e.g., Table 7) indicate that those covariates may be important for estimating occurrence/occupancy or detection probabilities for the target species at a given locality. Covariates that are irregularly included in models (e.g., Actinopterygii and *Clidastes* in the full dataset GLMs or *Platecarpus, Tylosaurus*, and *Squalicorax* in the occupancy models) (Tables 4 and 6; Supplements S7, S9, and S12) are unlikely to be
important to the target species because they are only explaining some of the variation in that model whereas the covariates consistently in the highest-ranked models explain more variation.

The SDMs generated in this study were successfully run for the Campanian WIS ecosystem. For the highest-ranked full dataset GLMs, this was demonstrated when these models were found as significant compared to the null model (Table 4; Supplements S7 and S9). For the occupancy models, the global model was assessed for fit and found to have a non-significant $\chi^2$ statistic compared to the bootstrapped $\chi^2$ value ($p\text{-value} = 0.611$) (Supplements S7 and S12). This indicates that the model and more parsimonious models should describe the data well. For the current study, the global occupancy model was represented by the following occupancy covariates: the larger taxonomic groups (Table 2), StdPaleolatitude, and WIS position. The detection covariates for this model were the three sedimentary lithologies and StdOutcrop area. The highest-ranked occupancy models also returned observed $\chi^2$ values less than the bootstrapped $\chi^2$ values (Supplements S7 and S12) indicating that these models appropriately fit the data.

Overall, the SDMs were useful for elucidating ecological patterns in the WIS. They determined that paleolatitude strongly influenced Hesperornis occurrences and enabled quantitative testing of taxa hypothesized to have ecological interactions with hesperornithiforms based on qualitative fossil evidence alone, namely Tylosaurus, Squalicorax, polycotylids, and Enchodus. One downside is that they can identify covariates as important to the target species that, upon further examination, might be in-contrast with some paleobiogeographical, sedimentological, and ecological data and not as relevant to the species of interest (e.g., elasmosaurs, Toxochelys, and carbonate for
Hesperornis occurrences). Future analyses using fewer and more targeted covariates will likely prune out unimportant covariates to result in models more accurate to the ecosystem. There also remains some unexplained variation in the data according to full dataset GLM results. These models have an average explained variation or deviance of 0.535 (range = 0.427-0.622) (Table 4; Supplements S7 and S9). This indicates that 37.8-57.3% of the variation is not included in current full dataset GLMs. Additional variation in Hesperornis occurrence data may be explained by testing other occupancy and detection covariates not yet considered (see Future Directions).

There are several potential taphonomic issues with the current SDMs that must be addressed. The models evaluate the ecological factors that may have influenced Hesperornis occurrences across the Campanian, a geologic age with a duration of 11.5 million years (Cohen et al., 2013; updated). Compared to other geologic ages of the Late Cretaceous, the Campanian lasted nearly twice as long as the second longest geologic age (Cenomanian) and over four times longer than the shortest geologic age (Santonian) (Cohen et al., 2013, updated), which could create stronger time-averaging problems. The issue is partially addressed by using some combined sites that were from the same rock unit and time of deposition (i.e., early, middle, and late Campanian). These Campanian subdivisions are approximately 3.2, 4.3, and 4.2 million years in duration, respectively (Ogg and Hinnov, 2012). Despite being a better fit to other Late Cretaceous geologic age durations, Campanian subdivisions cannot be used for evaluating Hesperornis occurrences because these divisions are limited by site sample size (Supplement S3) and are likely not as reliable as SDMs using the entire Campanian.
Preservational bias may have played a role in creating high abundance of *Hesperornis* between 43-53°N modern latitude because Campanian-aged rocks are largely absent north of ~53°N modern latitude along the northern WIS either by low outcrop exposure or preservation (Stott et al., 1993; Fenton et al., 2013; Prior et al., 2013). Additionally, Campanian-aged deposits along the eastern shoreline of WIS have eroded away (e.g., Slattery et al., 2015), which leads to information loss for *Hesperornis* occurrences. Fortunately for the models, these erosional issues affect all taxa and not just *Hesperornis*, so inferences regarding ecological relationships between *Hesperornis* and other taxa should be unaffected in the models.

Potential transport of *Hesperornis* bones from the site of death by currents has not been modeled in this study. The dominance of hindlimb, pelvic, or vertebral elements from *Hesperornis* in the Campanian WIS (Carpenter, 2006; Aotsuka and Sato, 2016; Bell and Chiappe, 2016) has been interpreted by Carpenter (2006) as evidence of post-mortem disarticulation while floating. This potential for transport could be addressed in future studies by including a covariate for skeletal completeness as a proxy for the transport time or disarticulation through scavenging or predation.

The occupancy models in this study inform on taphonomic bias using the detection history data to better characterize overall paleocommunities in an area. These models explicitly account for differential effects of lithology and outcrop area on the detectability of *Hesperornis*. Only carbonate was considered as important to *Hesperornis* detectability in the occupancy models among lithologies, but this may be an artifact of regression during the middle and late Campanian. The full dataset GLMs do not explicitly account for taphonomic issues because outcrop area, a covariate that only
affects detection probabilities, could not be modeled and the effect of lithology cannot be segregated between occupancy and detection probabilities. Inclusion of additional covariates that exclusively describe taphonomic bias (e.g., skeletal completeness and erosional or weathering proxies) could avoid the problem faced by lithological covariates in the GLMs.
CONCLUSIONS

Paleontological SDMs have focused on non-marine taxa and marine invertebrates, with only two studies examining the past distribution of marine vertebrates. No SDM studies have examined extinct marine vertebrates from the Late Cretaceous, leaving a gap in paleontological SDM applications and research. The current study uses GLMs and occupancy models to determine the ecological covariates that most explain the distribution of *Hesperornis*, an extinct flightless seabird for which several paleoecological interpretations have been proposed, but none have been quantitatively evaluated.

This study has two objectives: 1) Identify the ecological controls on the distribution of *Hesperornis* and 2) Test the applicability of SDMs in the WIS. The strongest ecological controls are elasmosaurs and paleolatitude. Elasmosaurs may negatively influence the *Hesperornis* distribution through competition or the model results for elasmosaurs may represent different habitat preferences from *Hesperornis*. Paleolatitude may reflect *Hesperornis* migration behavior, migration mortality, or true environmental preference. The negative association between carbonate lithologies and *Hesperornis* likely reflects changes in eustacy during the middle to late Campanian rather than an influence on the probability of *Hesperornis* detectability. No support is provided by the highest-ranked models (ΔAIC ≤ 2) for two taxa previously hypothesized to prey upon *Hesperornis*, namely *Tylosaurus* and Polycotylidae (Martin and Bjork, 1987; Martin et al., 2016), for *Enchodus* as preferred prey of *Hesperornis*, or for *Squalicorax* as a predator or scavenger of *Hesperornis*. *Hesperornis* did not prefer a particular west to east region of the WIS, but where the genus occurred may be related to an interaction
between WIS position and paleolatitude. This may be reflected in higher abundance of *Hesperornis* along hypothesized mixing zones and oceanic fronts within the WIS (see Discussion). Lastly, higher detectability of *Hesperornis* specimens may be limited to smaller outcrop area sizes, but more testing is needed.

Species distribution modeling is useful to model WIS taxa by providing quantitative insight into ecological patterns with more information obtained from comparing SDM techniques (e.g., GLMs and occupancy models). Maximum likelihood estimation using ΔAIC enables comparison of model support to determine the combination of covariates that is most explained by the current *Hesperornis* dataset. Modeling software for SDMs (e.g., R and PRESENCE) makes it easy to simultaneously compare numerous variables from environmental and biological elements of the WIS such as endemic taxa, lithologies, and paleogeographical information. The software also provides detailed model outputs to describe models and diagnose problems. GLMs and occupancy models are both useful to model extinct taxa because they incorporate different data (i.e., more sites in GLMs and the detection history in occupancy models). Use of GLMs with datasets meant to match occupancy models is not recommended because these reduced dataset GLMs have examine fewer sites, return higher error rates, and do not converge on an overall covariate grouping.

**Future Directions and Applications for Paleontological Analyses**

The potential for customizing occupancy models and GLMs leaves many avenues for expanding current analyses. There remain many additional occupancy/occurrence
covariates to consider for WIS studies: distance from paleoshoreline, inferred water depth, proximity to river mouths, models of WIS circulation or sea surface temperatures, lithological sub-categories and features, invertebrate occurrences, and size classes of included taxa. Other covariates can be used to account for more sampling/collection and detection biases: proximity to areas of high human development, bedding plane area, researcher/collection bias for sampling effort, research effort as gauged by publication or author counts per stratigraphic/chronologic unit, land use, and taphonomy, including skeletal completeness (e.g., Uhen and Pyenson, 2007; Dunhill, 2011; Dunhill et al., 2012; Benton et al., 2013; Vilhena and Smith, 2013). Proximity to areas of higher human development, which can be characterized by location of outcrops or human population size, may act as a proxy for ease of access to a site (e.g., Uhen and Pyenson, 2007; Dunhill et al., 2012). Bedding plane size can reflect preservational or sampling biases in the presence of bedding surface heterogeneity (Liow, 2013; Marenco and Hagadorn, 2019). It is currently unknown if the size of the site buffer impacts model results in this study. Future studies should experiment with different site buffer dimensions using real and simulated data to evaluate if the choice of modeling buffer size creates or reduces significant sampling bias.

The methods used in this study could easily be applied in new regions of the world and to different time periods. The current *Hesperornis*-focused dataset could be transformed to emphasize a different taxon (e.g., Mosasauridae) with *Hesperornis* and Hesperornithiformes becoming covariates. Further studies using the current dataset, with more covariate characterization of WIS paleoenvironmental conditions, could examine
species co-occurrence and predict community occupancy rather than occupancy for a single taxon (e.g., Zipkin et al., 2010; Pollock et al., 2014).
REFERENCES


Aotsuka, K., and T. Sato. 2016. Hesperornithiformes (Aves: Ornithurae) from the Upper


Everhart, M. J. 2004a. Plesiosaurs as the food of mosasaurs; new data on the stomach contents of Tylosaurus proriger (Squamata; Mosasauridae) from the Niobrara Formation of Western Kansas. The Mosasaur 7:41-46.


Paleontology Honoring the 90th Birthday of Alexander Whitmore, Smithsonian Institution Press, Washington, D.C. Smithsonian Contributions to Paleobiology 27.


dispersal potential has maintained long-term population stability in the North
Atlantic copepod Calanus finmarchicus. Proceedings of the Royal Society B:

R Core Team. 2021. R 4.0.4: A language and environment for statistical computing. R
Foundation for Statistical Computing, Vienna, Austria. Available at
https://www.r-project.org/.


Rees, J., and J. Lindgren. 2005. Aquatic birds from the Upper Cretaceous (lower
Campanian) of Sweden and the biology and distribution of hesperornithiforms.

of the Western Interior of middle North America – Coal distribution and sediment

and A. J. Richardson. 2011. Pushing the limits in marine species distribution
modelling: lessons from the land present challenges and opportunities. Global

Rovelli, R., and C. Myers. 2016. Using PaleoENM to determine the effects of available
ecological niche size on extinction susceptibility across the K/Pg extinction in the
Mississippi Embayment [sic]. Geological Society of America Abstracts with
Programs 48(7):287-2.


Schumacher, B. A., and J. E. Martin. 2016. Polycotylus latipinnis Cope (Plesiosauria, Polycotylidae), a nearly complete skeleton from the Niobrara Formation (Early


Shaw, D. J., Nicolas, M. P. B., and Chow, N. 2017: Stratigraphy and geochemistry of the Cretaceous Boyne Member, Carlile Formation, in the Manitoba Potash Corporation core at 3-29-20-29W1, southwestern Manitoba (parts of NTS 65K1); pp. 173–182 in Report of Activities 2017. Manitoba Growth, Enterprise and


I hereby grant Fort Hays State University an irrevocable, non-exclusive, perpetual license to include my thesis ("the Thesis") in FHSU Scholars Repository, FHSU's institutional repository ("the Repository").

I hold the copyright to this document and agree to permit this document to be posted in the Repository, and made available to the public in any format in perpetuity.

I warrant that the posting of the Thesis does not infringe any copyright, nor violate any proprietary rights, nor contains any libelous matter, nor invade the privacy of any person or third party, nor otherwise violate FHSU Scholars Repository policies.

I agree that Fort Hays State University may translate the Thesis to any medium or format for the purpose of preservation and access. In addition, I agree that Fort Hays State University may keep more than one copy of the Thesis for purposes of security, back-up, and preservation.

I agree that authorized readers of the Thesis have the right to use the Thesis for non-commercial, academic purposes, as defined by the "fair use" doctrine of U.S. copyright law, so long as all attributions and copyright statements are retained.

To the fullest extent permitted by law, both during and after the term of this Agreement, I agree to indemnify, defend, and hold harmless Fort Hays State University and its directors, officers, faculty, employees, affiliates, and agents, past or present, against all losses, claims, demands, actions, causes of action, suits, liabilities, damages, expenses, fees and costs (including but not limited to reasonable attorney's fees) arising out of or relating to any actual or alleged misrepresentation or breach of any warranty contained in this Agreement, or any infringement of the Thesis on any third party's patent, trademark, copyright or trade secret.

I understand that once deposited in the Repository, the Thesis may not be removed.

Thesis: Ecological Controls on the Distribution of Hesperornis (Aves: Hesperornithiformes) in the Western Interior Seaway

Author: Blake Chapman

Signature: Blake Chapman

Date: 11-29-2021