

1 Abstract

2 Introduction

3 Pigs eat everything. But they eat some things more than others. And when those “some things”
4 are agricultural crops, it can result in extensive economic damage. [estimate]. However, crops
5 are one of many forage resources competing for a feral swine’s attention on a landscape. As
6 extreme generalists, feral swine consume anything from sapling trees, fungus, and masting seeds
7 to small invertebrates and livestock. The availability of these different forage resources varies
8 in both space and time, such that pigs will switch their foraging behavior availability of these
9 various potential resources

10 Methods

11 Data

12 To analyze how forage availability affected pig movement patterns, we used GPS collar data
13 collected on 500 pigs in the United States of America and Canada (Fig. 1). Of the X pigs, X
14 were male and Y were female. These data are from X different studies and were collected from
15 May, 2004 to November, 2017. The median collaring time per pig across all studies was X (95%
16 quantiles) (see supplementary material for additional details). The average time between GPS
17 fixes varied by study, with the median fix time being X across all Y studies [95% CI].

18 For each pig used in the study, we cleaned the movement trajectory using the following
19 criteria. First, we excluded all 2D GPS fixes from the analysis (Bjorneraas *et al.* 2010). Second,
20 we eliminated the first and last 25 fixes for each pig to account for capture effects [citation].
21 Third, we eliminated all pig movements in which pigs moved faster than 50 km per hour over
22 any distance or moved greater than 15 km per hour over a distance of 8 km or more. These
23 movements are unlikely given previously observed patterns of feral swine movement (Mayer &
24 Brisbin 2009). Finally, we also removed large temporal gaps between fix times as these can lead
25 to a large amount of uncertainty regarding a pig’s location in between two temporally distant
26 fixes. To do this, we split a single pig movement trajectory into multiple sequences when the time
27 between two adjacent fixes was greater than *c.* 130 minutes. We then discarded all sequences
28 that had less than 150 fixes at least 130 minutes apart. The remaining sequences were considered
29 independent “runs” for a given pig. We chose to only include sequences of 150 fixes as shorter
30 runs would not always span the length of a day, precluding analysis of diel movement patterns,

31 and longer runs excluded too many of the pigs from the analysis. After this cleaning, our analysis
32 contained X pigs and a total of X fix times (Fig. 1).

33 Covariates

34 The goal of our analysis was to understand how the availability of anthropogenic and natural
35 forage on a landscape affected the movement and resource selection of feral swine. To address
36 these goals, we identified covariates related to natural forage and anthropogenic forage resources
37 that were comparable across the different studies used in this analysis. For the availability of
38 natural forage resources, we considered two proxies: plant productivity as measured by Normal-
39 ized Difference Vegetation Index (NDVI) and density of masting trees (Table 1). We included
40 masting tree density as masting events are an important component of pig diets and can have
41 significant implications on population-level pig growth rates (e.g. Bieber & Ruf 2005). NDVI
42 is a general correlate of plant productivity in an area (Pettorelli *et al.* 2005) and, given that
43 pigs are extreme generalists and plant material often makes up a large percentage of their diet
44 (Mayer & Brisbin 2009), we use NDVI to reflect the availability of herbaceous forage.

45 For anthropogenic forage resources, we used agricultural data available from the National
46 Agricultural Statistics Service (NASS). These data are obtained yearly at a 30m by 30m scale
47 across the contiguous US. Each pixel specifies the dominant type of crop grown in that pixel.
48 While the NASS data enumerates 105 types of crop, we chose not to distinguish between crop
49 types in our initial analyses. Instead, we used a generic grouping of “crop” vs. “no-crop”, which
50 allowed us to more easily explore how the effect of agricultural forage on pig movement varied
51 across space and time [though maybe specific croptypes would be better]. We did consider
52 specific crop types in our post-hoc analyses, in which we grouped the 105 crop types into 11
53 groups: cereals, oilseed, tobacco, beverage and spice, leguminous, grasses, sugar, root and tuber,
54 fruit and nuts, vegetables and melons, and other crops [according the criteria outlined in...].
55 WHAT ABOUT CANADA?]

56 While our primary goal of this analysis was to understand the effect of forage availability
57 on pig movement, pig movement and home range size is also driven by a number of other vari-
58 ables, including cover, water availability, temperature, pressure, human development, mammal
59 diversity, sex, and age (McClure *et al.* 2015; Garza *et al.* 2017; Kay *et al.* 2017) [MORE]. Of
60 these additional predictors of pig movement, cover and water are critically important as both
61 are necessary for pigs to thermoregulate (Choquenot *et al.* 1996) and cover, in particular, is
62 necessary for protection from predators [citation] (citations in Mayer2009). We used tree canopy

density data [check wording] from the National Landcover Database (NLCD) as a covariate for cover (Table 1). These data are available at a 30m by 30m scale for the contiguous United States. For a water, we used the National Wetland Inventory (NWI) Database which identifies X types of water bodies larger than X m² in the contiguous United States. For this study, we only considered the water sources that were permanent or semi-permanent as defined by the NWI and computed the distance to the nearest water source at a 30m by 30m resolution (see Table 1 and Sup Mat).

We also included mean monthly temperature, total monthly precipitation, and distance to developed land/roads as additional covariates (Table 1). Note that while the inclusion of these covariates is import for capturing pig movement, our questions were focused the foraging resources, such that we considered non-foraging covariates "blocking" covariates and did not exhaustively explore their potential relationships and interactions beyond what had already been shown in the literature.

Finally, we also considered a number of study-level variables to allow us to explore how and why the effect of natural forage resources and anthropogenic forage resources varied across different studies. To this end, we explored study-level covariates such as drought severity, ecoregion, local pig density, and MORE (see Table 1).

Continuous-time, discrete-space movement model

To understand role of forage resources on pig movement, we used the modeling framework of Hanks *et al.* (2015) and Wilson *et al.* (2018). Generally, this framework leverages autocorrelated animal movement data and gridded raster covariates to make inference about the resource utilization of an animal (Hanks *et al.* 2015; Buderman *et al.* 2018; Wilson *et al.* 2018).

Specifically, this approach can be broken into two distinct steps. Given a trajectory of GPS fixes (not necessarily with equal fix times), the first step of this approach estimates animal movement as a function of continuous time at some particular temporal grain (e.g. 15 minutes). To do this, we used a phenomenological functional movement model (FMM) (Buderman *et al.* 2016; Hooten *et al.* 2017), which is a non-mechanistic, continuous-time movement model that can capture an animal's movement patterns at some desired-level of detail (Buderman *et al.* 2016). The phenomenological FMM can be represented as a series of basis functions, which allow for large flexibility in animal movement patterns. In particular, we used a B-spline basis expansion to model the longitude and latitude of an animal as a function of time (see Supplementary Material for additional detail).

95 In the second step of the analysis requires translating our continuous movement model into
 96 discrete, rasterized space. To do this we used our fitted FMM to predict a pig's location at 15
 97 minute intervals. We chose 15 minutes as this provided a reasonable trade-off between compu-
 98 tational time and computing the exact amount of time spent in each cell, which is equivalent to
 99 letting the time between fixes go to zero [NOTE: Would this be hard to do? Double Check inter-
 100 pretation of 15 minute intervals...]. We then converted this predicted pig path into discrete 30m
 101 by 30m grid cells such that we were left with a continuous time movement path in discrete space
 102 (Fig. X). [We chose 30m by 30m] To account for the uncertainty in our movement model, we
 103 generated 20 continuous-time discrete space movement paths and performed all of the analyses
 104 described below on each of the movement paths (Hanks *et al.* 2015; Buderman *et al.* 2018).

105 Given these continuous-time, discrete-space trajectories, we then explored how agricultural
 106 and natural forage resources on a landscape affected pig movement using a continuous-time
 107 Markov Chain (CTMC) approach Hanks *et al.* (2015). The CTMC approach considers continuous-
 108 time animal movement (i.e. the FMM model described above) through a discrete landscape.
 109 From this point-of-view, animal movement can be considered as a series of rates of moving from
 110 cell i to an adjacent cell j , λ_{ij} . As for any continuous-time Markov Chain, the process can be
 111 decomposed into the waiting time before a state change occurs (i.e. the time an animal spends
 112 in a cell) and the new state once a change occurs (i.e. the new cell to which the animal has
 113 moved) (Allen 2003). With this interpretation, one can then model the rate of moving between
 114 cell i and j λ_{ij} as a function of the environmental covariates in cell i and j . Hanks *et al.* (2015)
 115 showed this inference can be re-expressed as a latent-variable, Poisson Generalized Linear Model,
 116 where the response variable for adjacent cell j is one if a pig moved to that cell from cell i , and
 117 0 otherwise. Specifically, let z_{ij} be the zero/one latent variable, then

$$z_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad (1)$$

$$\log \lambda_{ij} = \log \tau_{ij} + \beta \mathbf{X} \quad (2)$$

118 where τ_{ij} is the waiting time before moving from cell i to cell j , \mathbf{X} is a vector of landscape
 119 covariates, and β is the effect of these covariates on movement.

120 Considering \mathbf{X} , we explored two classes of covariates: location-based drivers and directional
 121 drivers of movement (Hanks *et al.* 2015). Location-based drivers are a result of the cell that an
 122 animal is currently in and affect how long an animal remains in the current cell. For example,

123 if masting tree density was a negative location-based driver of pig movement then a pig in a cell
 124 with high masting tree density would tend to remain in that cell longer than a cell with lower
 125 masting tree density. Directional drivers of movement determine the direction that a pig might
 126 move once it leaves the cell it is currently occupying. For example, if masting tree density was a
 127 positive directional driver of pig movement then, upon leaving the currently occupied cell, a pig
 128 tends to move in the direction of increasing masting tree density, relative to its current position
 129 [TODO: THIS IS A LOCAL dirIENT EFFECT...SHOULD WE ALSO EXPLORE GLOBAL
 130 EFFECTS AT SOME POINT? IN OTHER WORDS, does the presence of masting trees some
 131 where in the vicinity affect how a pig uses crop?]. Table 1 shows which of the covariates described
 132 in the previous section we considered as location-based drivers, directional drivers, or both.

133 **Model specification and fitting**

134 We analyzed three variations of the CTMC model with the goal of understanding the role of
 135 foraging resources on the movement of feral swine across populations. Each model is described
 136 below

137 *Model 1: Main effects model*

138 The first, and simplest, model that we analyzed only considered time-invariant main effects
 139 of anthropogenic and natural foraging resources as well as three other non-foraging covariates
 140 known to affect pig movement: canopy cover, sex, and distance to water. The model is defined
 141 as

$$\log(\lambda_{ij}) = \log(\tau_{ij}) + \beta_0 + \beta_1 \text{directional persistence} + \beta_2 \text{male} \quad (3)$$

$$+ \beta_3 \text{canopy cover}_{\text{loc}} + \beta_4 \text{canopy cover}_{\text{dir}} \quad (4)$$

$$+ \beta_5 \text{distance to water}_{\text{loc}} + \beta_6 \text{distance to water}_{\text{dir}} \quad (5)$$

$$+ \beta_7 \text{distance to development}_{\text{loc}} + \beta_8 \text{distance to development}_{\text{dir}} \quad (6)$$

$$+ \beta_9 \text{distance to crops}_{\text{loc}} + \beta_{10} \text{distance to crops}_{\text{dir}} \quad (7)$$

$$+ \beta_{11} \text{masting tree density}_{\text{loc}} + \beta_{12} \text{masting tree density}_{\text{dir}} \quad (8)$$

$$+ \beta_{13} \text{NDVI}_{\text{loc}} + \beta_{14} \text{NDVI}_{\text{dir}} \quad (9)$$

142 where “directional persistence” describes the propensity of pigs to continue to move in the
 143 direction they moved in the previous time step. the subscript “loc” refers to a location-based

144 covariate (i.e. how that covariate effects the rate an animal is moving) and the subscript “dir”
145 refers to a directional covariate (i.e. how that covariate effect the direction in which an animal
146 moves). We fit this model in two steps. First, we fit the model using LASSO regularization using
147 the `glmnet` package (Friedman *et al.* 2010). LASSO simultaneously performs variable selection
148 and shrinks [non-predictive or highly correlated] covariates to zero James *et al.* (2013). To select
149 the best-model in terms of predictive performance, we performed 5-fold cross-validation [NOTE:
150 that I probably need to do cross-validation for case-control studies] and chose the model with
151 the maximum penalization parameter for which the Poisson deviance was within one standard
152 error of the minimum out-of-sample deviance (provided in `glmnet`). This approach increases
153 regularization, with only a small cost to deviance [and allowed us to more easily perform our
154 second set .

155 While advantageous for model selection, the aforementioned LASSO approach does not pro-
156 vide easy access to parameter uncertainty or incorporation of random effects. We then refit the
157 model model X, excluding the parameters removed by the LASSO, in a Bayesian framework. For
158 each β_i coefficient in the model, we included a random effect of individual. Specifically, given k
159 individual pigs in a population, we allowed $\beta_{i,k} \sim N(\mu_{\beta_i}, \sigma^2 \phi_i)$, which indicates that individual-
160 level effects for parameter i are drawn from a population-level distribution (Buderman *et al.*
161 2018). When $i > 1$, $\log(\phi_i) \sim N(0, 0.04)$, which weights the variance population-level variance
162 for each i . When $i = 1$, $\phi_1 = 1$. For the population-level means, we specified the hyper-prior
163 $\mu_{\beta_i} \sim N(0, 3)$. In this model, the random effects of individuals for each covariate account for
164 variation in resource use between individuals that were sampled at different times, in addition
165 to inherent between-individual variability.

166 Given the size of the “data” for each study (e.g. most studies had > 1 million predicted
167 positions from the FMM model), we used automatic differentiation variational inference (ADVI)
168 to approximate the posterior distributions of this Bayesian model (Kucukelbir *et al.* 2015).
169 [Description of ADVI?]. Combining both ADVI and the preliminary LASSO regularization,
170 allowed us to fit these models in a reasonable amount of time (e.g. X hours).

171 *Model 2: Time-varying movement model*

172 The second model we considered allowed pig movement and resource use to vary with time.
173 We allowed overall movement speed to vary on both a daily and monthly time-scale and also
174 allowed selection for forage resource, cover, and water to vary on a monthly time scale. The
175 model is given by

$$\log(\lambda_{ij}(h, m)) = \log(\tau_{ij}) + \beta_0(h) + \beta_0(m) \quad (10)$$

$$+ \beta_1 \text{directional persistence} + \beta_2 \text{male} \quad (11)$$

$$+ \beta_3(m) \text{canopy cover}_{\text{loc}} + \beta_4(m) \text{canopy cover}_{\text{dir}} \quad (12)$$

$$+ \beta_5(m) \text{distance to water}_{\text{loc}} + \beta_6(m) \text{distance to water}_{\text{dir}} \quad (13)$$

$$+ \beta_7 \text{distance to development}_{\text{loc}} + \beta_8 \text{distance to development}_{\text{dir}} \quad (14)$$

$$+ \beta_9(m) \text{distance to crops}_{\text{loc}} + \beta_{10}(m) \text{distance to crops}_{\text{dir}} \quad (15)$$

$$+ \beta_{11}(m) \text{masting tree density}_{\text{loc}} + \beta_{12}(m) \text{masting tree density}_{\text{dir}} \quad (16)$$

$$+ \beta_{13}(m) \text{NDVI}_{\text{loc}} + \beta_{14}(m) \text{NDVI}_{\text{dir}} \quad (17)$$

176 where h indicates the hour of the day and m indicates the month of the year. $\beta_0(h)$ describes
 177 how overall movement rate changes as a function of the hour of the day. $\beta_i(m)$ describes how the
 178 effect of a particular covariate varies with the month of the year. We modeled $\beta_0(h)$ using cyclic
 179 splines to ensure movement rate at the start of a 24 hour period was the same as movement
 180 rate at the end of the 24 hour period. We modeled $\beta_i(m)$ using cyclic-cubic regression splines
 181 when the data for pigs in a populations spanned 12 months (Wood 2006). This ensured that the
 182 effect of a particular covariate was the same at the end December and the beginning of January.
 183 When the data for a particular population did not span 12 months [NOTE: rarely, if ever, does
 184 a single pig span twelve months! Should we be fitting each pig separately?], we modeled $\beta_i(m)$
 185 using a non-cyclic cubic spline [check terminology].

186 Similar to model 1, we performed LASSO regularization followed by an approximate Bayesian
 187 analysis using ADVI (see Supplementary Material).

188 *Model 3: Environmentally-driven movement model*

189 The final model we considered still allowed for overall movement speed to vary both daily
 190 and monthly, but now sought to describe the seasonal changes in selection for anthropogenic
 191 and natural forage in terms of monthly changes in precipitation and temperature. Specifically,
 192 we updated the basis function form of model 3 to include interactions between temperature,
 193 precipitation and forage resources.

$$\begin{aligned}
\log(\lambda_{ij}(h, m)) = & \log(\tau_{ij}) + \beta_0(h) + \beta'_1 \text{season}_m \\
& + \beta_2 \text{directional persistence} + \beta_3 \text{male} \\
& + \text{canopy cover}_{\text{loc}}(\beta_4 + \beta'_4 \text{season}_m) \\
& + \text{canopy cover}_{\text{dir}}(\beta_5 + \beta'_5 \text{season}_m) \\
& + \text{distance to water}_{\text{loc}}(\beta_6 + \beta'_6 \text{season}_m) \\
& + \text{distance to water}_{\text{dir}}(\beta_7 + \beta'_7 \text{season}_m) \\
& + \text{distance to crops}_{\text{loc}}(\beta_8 + \beta'_8 \text{season}_m) \\
& + \text{distance to crops}_{\text{dir}}(\beta_9 + \beta'_9 \text{season}_m) \\
& + \text{masting tree density}_{\text{loc}}(\beta_{10} + \beta_{10'} \text{season}_m) \\
& + \text{masting tree density}_{\text{dir}}(\beta_{11} + \beta_{11'} \text{season}_m) \\
& + \text{NDVI}_{\text{loc}}(\beta_{12} + \beta_{12'} \text{season}_m) \\
& + \text{NDVI}_{\text{dir}}(\beta_{13} + \beta_{13'} \text{season}_m)
\end{aligned}$$

194 where $\beta'_i \text{season}_m$ is the monthly effect of temperature and precipitation and is specifically given
195 by $\beta_{i,1} \text{temperature}_m + \beta_{i,2} \text{precipitation}_m + \beta_{i,3} \text{temperature}_m \text{precipitation}_m$. We used the same
196 approach as model 1 when fitting the LASSO model. When incorporating random effects of
197 individuals, we assumed that the effect of season was the same across individuals and that the
198 random effect of individual was manifested in the β_i coefficients (i.e. not the β'_i coefficients).
199 Otherwise, this random effects model was implemented as in model 1.

200 **Model comparison and population-level analyses**

201 For LASSO models...just compare CV deviance with standard error.

202 For ADVI models...check out LOO techniques. These might work.

203 Results

204 Effects of non-foraging covariates on movement

205 Effects of foraging covariates on movement

206 Main effects models

207 Time-varying effects

208 Seasonal effects model

209 Population-level analysis

210 Discussion

211 Temporal variation in selection for foraging resources...

212 With regards to management,

213 References

- 214 Allen, L.J. (2003) *An Introduction to Stochastic Processes with Applications to Biology*. Pearson
215 Education, Inc, Upper Saddle River, New Jersey.
- 216 Bieber, C. & Ruf, T. (2005) Population dynamics in wild boar *Sus scrofa*: Ecology, elasticity of
217 growth rate and implications for the management of pulsed resource consumers. *Journal of*
218 *Applied Ecology*, **42**, 1203–1213.
- 219 Bjørneraas, K., Moorter, B.V., Rolandsen, C.M. & Herfindal, I. (2010) Screening Global Posi-
220 tioning System Location Data for Errors Using Animal Movement Characteristics. *Journal of*
221 *Wildlife Management*, **74**, 1361–1366.
- 222 Buderman, F.E., Hooten, M.B., Alldredge, M.W., Hanks, E.M. & Ivan, J.S. (2018) Predatory
223 behavior is primary predictor of movement in wildland-urban cougars. *bioRxiv*.
- 224 Buderman, F.E., Hooten, M.B., Ivan, J.S. & Shenk, T.M. (2016) A functional model for char-
225 acterizing long-distance movement behaviour. *Methods in Ecology and Evolution*, **7**, 264–273.
- 226 Choquenot, D., McIlroy, J. & Koen, T. (1996) Managing vertebrate pests: feral pigs. *Beureau*
227 *of Resource Sciences*, pp. 1–171.
- 228 Friedman, J., Hastie, T. & Tibshirani, R. (2010) Regularization Paths for Generalized Linear
229 Models via Coordinate Descent. *Journal of Statistical Software*, **33**, 1–22.
- 230 Garza, S.J., Tabak, M.A., Miller, R.S., Farnsworth, M.L. & Burdett, C.L. (2017) Abiotic and
231 biotic influences on home-range size of wild pigs (*Sus scrofa*). *Journal of Mammalogy*, pp.
232 1–11.
- 233 Hanks, E.M., Hooten, M.B. & Alldredge, M.W. (2015) Continuous-time discrete-space models
234 for animal movement. *Annals of Applied Statistics*, **9**, 145–165.
- 235 Hooten, M.B., Johnson, D.S., McClintock, B.T. & Morales, J.M. (2017) *Animal Movement:*
236 *Statistical Models for Telemetry data*. CRC Press, New York, USA.

237 James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *Introduction to Statistical Learning*
238 *with Applications in R*. Springer, New York, USA.

239 Kay, S., Fischer, J., Monaghan, A., Beasley, J., Boughton, R., Campbell, T., Cooper, S.,
240 Ditchkoff, S., Hartley, S., Kilgo, J., Wisely, S., Wyckoff, A., VerCauteren, K. & Pepin, K.
241 (2017) Quantifying drivers of wild pig movement across multiple spatial and temporal scales.
242 *Movement Ecology*, **5**, 1–15.

243 Kucukelbir, A., Gelman, A. & Blei, D.M. (2015) Automatic Variational Inference in Stan. *Nips*,
244 **1**, 1–22.

245 Mayer, J.J. & Brisbin, I.J.J. (2009) Wild Pigs: Biology, Damage, Control Techniques and Man-
246 agement. Technical report, Savannah River National Laboratory, Aiken, South Carolina.

247 McClure, M.L., Burdett, C.L., Farnsworth, M.L., Lutman, M.W., Theobald, D.M., Riggs, P.D.,
248 Grear, D.A. & Miller, R.S. (2015) Modeling and mapping the probability of occurrence of
249 invasive wild pigs across the contiguous United States. *PLoS ONE*, **10**, 1–17.

250 Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J. & Stenseth, N.C. (2005) Us-
251 ing the satellite-derived NDVI to assess ecological responses to environmental change. *Trends*
252 *in Ecology and Evolution*, **20**, 503–510.

253 Wilson, K., Hanks, E. & Johnson, D. (2018) Estimating animal utilization densities using
254 continuous-time Markov chain models. *Methods in Ecology and Evolution*, **In press**.

255 Wood, S.N. (2006) *Generalized additive models : An introduction with R*. CRC Press, Boca
256 Raton.

Table 1: Description of covariates used in analysis

Covariate	Description	Data Source	Spatial Resolution	Temporal Resolution	Location or direction?
Normalized Difference Vegetation Index (NDVI)	A proxy for plant productivity and natural forage availability.	MODIS and NASS	250 m \times 250 m	Monthly	Both
Density of hard-masting trees		Tabak et al.	1 km \times 1 km	Time-invariant	Both
Distance to crops	The distance to the nearest crop field. A measure of anthropogenic forage.	NASS	30 m \times 30 m	Yearly	Both
Distance to water	The distance to the nearest permanent or semi-permanent water source	NWI	30 m \times 30 m	Time-invariant	Both
Canopy density [CHECK]	A proxy for habitat cover	NLCD	30 m \times 30 m	Time-invariant	Both
Distance to developed-land	A measure of human presence	NLCD	30 m \times 30 m	Time-invariant	Both
Temperature	Mean monthly temperature	NOAA	50 km \times 50 km	Monthly	Location
Precipitation	Total monthly precipitation	NOAA	50 km \times 50 km	Monthly	Location