Generalizing animal movement predictions across landscapes: a scalable framework grounded in empirical telemetry data

Chalkowski K[1]*, Sack A[2]*, Wittemyer G[13], Miller RS[4], Boughton RK[3], Snow NP[1], Beasley JC[5], Ditchkoff SS[6], Evans TS[7], Gray SM[8], Lewis JS[9], Parsons MA[10], VerCauteren K[1], Webb SL[11], Young JK[12] and Pepin KM[1]
**These authors contributed equally*

Author affiliations:

1. National Wildlife Research Center, USDA-APHIS, Fort Collins, CO 80521, USA
2. The Carter Center Atlanta, GA 30307, USA
3. University of Florida, Institute of Food and Agriculture Sciences, Range Cattle Research & Education Center, 3401 Experiment Station, Ona, FL 33865
4. USDA/APHIS/ Veterinary Services, Center for Epidemiology and Animal Health, 2150 Centre Avenue, Fort Collins, Colorado 80526, USA
5. University of Georgia, Savannah River Ecology Laboratory, Warnell School of Forestry and Natural Resources, Aiken, SC 29803
6. College of Forestry, Wildlife and Environment, Auburn University, Auburn, AL 36849, USA
7. Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Mississippi State, Mississippi 39762, USA
8. U.S. Geological Survey, Michigan Cooperative Fish and Wildlife Research Unit, Michigan State University, East Lansing, MI 48824, USA
9. Arizona State University, College of Integrative Sciences and Arts, Mesa, AZ 85212, USA Orcid ID: 0000-0002-3071-5272
10. Department of Wildland Resources and Ecology Center, Utah State University, Logan, UT 83422, USA ORCID iD: 0000-0002-4229-8277
11. Texas A&M Natural Resources Institute, and Department of Rangeland, Wildlife and Fisheries Management, Texas A&M University, College Station, TX 77843, USA. ORCID: 0000-0001-6034-5164
12. Department of Wildland Resources and Ecology Center, Utah State University, Logan, UT 84322, USA ORCID iD: 0000-0003-4522-0157
13. Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523 OrchidID: 0000-0003-1640-5355

Corresponding author:
Kayleigh Chalkowski, kchalkowski@gmail.com

Keywords
predictive ecology; wild pig; *Sus scrofa*; landscape ecology; scale; modeling framework

Competing interests
The authors declare they have no competing interests

Funding
The research was supported by NIFA-AFRI Agricultural Biosecurity program (GRANT13428587) and the United States Department of Agriculture's National Feral Swine Damage Management Program.

Acknowledgements
The authors thank multiple collaborators who shared wild pig geolocation data including Susan M. Cooper, Tyler Campbell, Lindsay Holmstrom, Steve Hartley, and Michael D. White. We also acknowledge the contributions of Jonathan Potts, Luca Börger, Bronson K. Strickland, and Garrett M. Street whose publicly available wild pig geolocation data were used in this work. The authors also thank Josh Hewitt for feedback on the modeling approach. The research was supported by NIFA-AFRI Agricultural Biosecurity program (GRANT13428587) and the United States Department of Agriculture's National Feral Swine Damage Management Program. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official US Government determination or policy. Mention of commercial products or companies does not represent an endorsement by the US government

Abstract

Background: Accurately predicting animal movement across broad ecological contexts is vital for effective wildlife management, yet most models are developed for specific locations and fail to generalize across diverse landscapes. New approaches are needed to scale individual-level movement processes in a way that accounts for ecological variation.

Methods: We developed a modeling framework to predict movement distance and habitat selection across large geographic areas using GPS telemetry data from 564 wild pigs collected in 25 studies across the United States. Movement distances were modeled using gamma distribution parameters, and habitat selection was modeled using integrated step selection analysis. We used gradient-boosted regression models with nested cross-validation to optimize predictions and reduce overfitting. To account for spatial non-stationarity in habitat preference, individuals were grouped by habitat availability.

Results: Our framework captured broad-scale spatial and temporal variation in movement behavior. Mean daily movement distances ranged from 152 to 1,404 meters, with the greatest seasonal variation observed in northern populations. Habitat selection models revealed consistent preference for wetlands and avoidance of croplands, with variation in other land types depending on the local ecological context. Leave-one-out sensitivity analyses indicated greater model reliability in subsets with more individuals, sites, and land types.

Conclusions: This framework offers a scalable approach to predicting animal movement across heterogeneous landscapes. It enables estimation of movement parameters in regions lacking local telemetry data, supporting more informed decision-making for invasive species management, disease control, and conservation planning.

Introduction

Decisions about animal health and ecosystem services are often made across broader extents than data collection may be feasibly collected, requiring methods for predicting outcomes in novel contexts (i.e., generalizability). Example applications include ecosystems where an invasive species or novel disease might spread upon arrival (Crowl et al, 2008), how climate change or other anthropogenic influences might change species distributions (Swab et al, 2015), or for mitigation of ecological challenges such as agricultural damage (Boyce et a. 2020) and species conservation (Fraser et al. 2018). Animal movement is a process intertwined with numerous ecological patterns (Thompson et al. 2021, Aiello et al. 2023, Manlove et al. 2022), but much of what we know about animal movement decisions come from analyses inferring local ecological processes from data at unique locations that are difficult to extrapolate to novel contexts. Determining drivers of animal movement that offer mechanistic insight to population or disease processes and is a core objective in wildlife research (Nathan et al. 2008, Manlove et al. 2022). However, animal movement depends on local ecological conditions (Nathan et al. 2008), meaning local understanding must be generated to infer spatial processes. Exploring approaches and developing methods for predicting animal movement across divergent ecological contexts would be of great value, given inference could then be applied across ecosystems.

Previous movement model frameworks have considered the alteration of behavioral states across different scales within track patterns of individual animals (Torney et al. 2021, Leos-Barajas et al. 2017, Buderman et al. 2018, Benhamou et al. 2014), with a few studies modeling animal movement across broad scales and ecological contexts (Wilber et al. 2019, Kay et al. 2017). Building on these outputs, new approaches focusing on optimizing movement predictions are needed to fully leverage insights gained from such efforts. The statistical methods employed

in these contexts (e.g., generalized additive models), though, may be inadequate for modeling movement across large extents where numerous multi-scale environmental covariates are needed to optimize predictions and inferences, due to convergence issues (Kay et al. 2017). Thus, for the goal of cross-system prediction and inference, new methodological frameworks are needed to develop broad-scale movement models.

Scaling up animal movement processes across ecological contexts poses some common challenges. First, prediction across greater environmental heterogeneity than exists in the observed data is risky given ecological contexts change across space (i.e., Tobler's first law of geography; Werkowska et al. 2017). Both the strength and direction of species' functional responses can differ across ecological contexts (Pease et al. 2022)—a phenomenon known as spatial non-stationarity (Rousseau and Betts 2022,Sultaire et al. 2021), where species exhibit divergent functional responses to their environment across their range. Spatial non-stationarity has been observed in animal movement models in changing environments, frequently in the context of altered resource availability (Kaszta et al. 2021, Fushing et al. 2006, Gilbert et al. 2016). Thus, the potential for changes in movement responses across different ecological contexts should be accounted for when generalizing animal movement models (Kaszta et al. 2021). The second challenge to scaling-up animal movement processes is, the relationship between environmental variables and animal movement can differ across spatial and temporal scales of analysis. For example, explanatory variables that influence response variables at smaller scales may not transfer to coarser, larger scales, and vice versa (i.e., "ecological fallacy"; Wu 2004). The use of coarse scales can help identify instances of non-stationarity and other emergent relationships (Werkowska et al. 2017), allowing for greater generality, but this

technique reduces local extremes (Paton and Mathiopoulos 2016) and results in loss of information (Fritsch, Lischke, and Meyer 2020).

Animal movement studies have unique challenges when predicting across vast ecological contexts. One challenge stems from the use of geolocation data, via collaring of animals, on which movement models are dependent. While telemetry technologies have become more efficient over the last few decades (Thums et al. 2018), the process of geotagging animals can still be costly and time-consuming (Kie et al. 2010), and many movement studies therefore only track a limited number of individuals, which can make population-level inference infeasible (Buderman et al. 2016). Another difficulty arises from the volume of geolocation data where a large number of geolocations per individual can result in computationally intensive models as more individuals are included in a dataset (Byrne et al. 2014). In addition to improving generalizability across broad scales, aggregation of movement parameters can ameliorate computational challenges by reducing the resolution of the data, but this can result in underestimation of extreme movements (Byrne et al. 2014). This underestimation can be a problematic loss of information since spatial variation can inform us about ecological processes of interest (Shaw 2020).

While there have been few studies aimed at modeling animal movement parameters across spatial scales (Potts and Borger 2022, Kay et al 2017, Allen et al. 2016), more attention has been given to extrapolation across ecological contexts for species distribution models. These advances may offer solutions to improve generalizability of animal movement patterns across different geographic areas as well. Aggregation during different parts of the modeling process to optimize out-of-sample prediction or inference during pre-model data processing, in-model scaling, or post-model scaling (Fritsch 2020) can be a useful approach for predicting across

spatial extent. For example, one method that accounted for spatial non-stationarity modeled species distribution data to regional locations on grids by grouping data by availability of habitat in nearby cells, which resulted in better out-of-sample predictions than models that did not group data by differences in habitat availability (Paton and Mathiopoulos 2016). Recently, cross-validation (CV) methods that fit spatially separated data (spatially-blocked CV) have further improved upon these techniques (Valavi et al. 2023) and their ability to predict across spatially diverging ecological contexts.

We present a methodological framework (Figure 1) to improve generalizability of individual-level movement processes including habitat selection and movement distance. In our framework, we present a step-selection model on movement preference (RSF) that combines data from multiple study sites, and a predictive model of movement distance across spatial extent where local data is lacking. We challenged our methodological framework with global positioning system (GPS) data from wild pigs (*Sus scrofa*) across the United States of America (U.S.)(Figure 2). Wild pigs are an ideal system to evaluate methodological approaches for scaling up movement processes because of their wide-ranging distribution (Lewis et al. 2017, Lewis et al. 2019, Snow et al. 2017) and generalist habitat use across a variety of ecosystems (McClure et al. 2015, Wilber et al. 2019). Furthermore, as a pest species, environmental predictors of their density and distribution are well-studied on a number of spatial and temporal scales (McClure et al. 2018, Lewis et al. 2017), which provides direction for selection of environmental predictors. The application of our movement scaling framework provides a national-scale map of wild pig movement metrics including both mean and dispersion parameters which can be leveraged for simulation modeling. Further, this application serves to evaluate the functionality of our framework when challenged with nuanced, real-world data

114 while also providing ecological estimates for national-scale management decisions about wild

115 pig populations and damage, and foreign animal disease preparedness.

116

117

Methods

We evaluate two workflows for scaling-up movement models including for movement distance and preference (habitat step-selection; HSS). We evaluated two movement metrics: daily step length and displacement from geolocation data. Daily step length is the distance between geolocations at successive daily intervals, and daily displacement is the net distance between a starting point (averaged for the first seven days of each season) and median daily location. For movement distance prediction, we aggregate step-level data for each individual to predict gamma distribution parameters for watersheds across the U.S. wild pig distribution, then generate step-level predictions from these individual-level distributions. We evaluate the success of scaling-up distance prediction in terms of ability to accurately predict heterogeneity in individual and step-level movement distances. For habitat step-selection (HSS), we group data by habitat availability during data pre-processing to determine broad-scale HSS coefficients for land classes used by wild pigs in the U.S. We evaluate the success of scaling-up HSS in terms of ability to successfully capture non-stationarity in movement preference.

To test these frameworks, we used geolocation data from individual wild pigs fit with GPS telemetry collars in studies across the continental U.S. from 10 states, 25 studies, and 564 individual wild pigs (Figure 2; Kay et al. 2017, Wilber et al. 2019, Potts et al. 2022; Data S1). Geolocation processing methods are described in Supplementary Methods 1. All analyses were conducted in R version 4.2.0 (R Core team 2022).

Scaling-up movement distance models

Pre-processing of data: Our methods for estimation of daily step length and displacements are described in Supplementary Methods 2. To characterize local heterogeneity of the movement of

individuals, we fit a gamma distribution to the distribution of step lengths and displacements of daily geolocations for each wild pig and determined both the mean and variance (dispersion) of these individual distributions as responses (Supplementary Methods 3). Lastly, environmental predictors joined to wild pig geolocation data were spatially and temporally aggregated on a per-individual basis (Supplementary Methods 4, Data S2-S4) to match the scale of the response data. To model movement distance response variables (step length mean, step length dispersion, displacement mean, displacement dispersion) across a potentially large number of complex interactions, we used gradient-boosted regression because this approach does not assume linear relationships, and is ideal for prediction (Elith et al. 2008). All boosted regression models were conducted using the *dismo* library (Hijmans et al. 2022) with a Poisson loss function for step length and displacement means, and a least squares loss function for step length and displacement dispersions.

Optimizing predictions during cross-validation: We compared two aggregation methods in our nested n x k cross-validation (CV) process during hyperparameter optimization to reduce over-fitting. Spatial grouping of validation sets during cross-validation has been proposed as a method to improve predictions to out-of-sample spatial extents (Roberts et al. 2017). Thus, we compared both random and spatially-blocked n x k-fold CV methods for optimizing hyperparameters (number of trees, learning rate, tree complexity, and bag fraction) (Supplementary Methods 5). We considered the best aggregation method for each model to be the one with the lowest root mean square error (RMSE) value on the outer fold, and used its corresponding optimal hyperparameters for the predictions (Supplementary Methods 6). To evaluate whether our nested CV was successful at reducing overfitting, we compared RMSE values between inner and outer

folds, because a symptom of overfitting is RMSE values that are greater on the testing than the training set (Freeman et al. 2015). During CV, we also conducted feature selection with the full set of predictors, relative influence (removing variables that had <1% influence on responses), lasso regression, and an intercept-only model (Supplementary Methods 7). After hyperparameter optimization and feature selection, we used the best hyperparameter and parameter sets (those which minimized RMSE) to predict movement distances to watersheds across the U.S. wild pig distribution (Supplementary Methods 8).

Maintaining heterogeneity of predictions: To maintain extreme values in our predictions, we took a two-step approach. First, we predicted individual mean and dispersion responses to each watershed across the U.S. wild pig distribution using a stochastic gradient boosted (SGB) ensemble (Malinin et al. 2020a), which is an ensemble of gradient boosted decision tree (GBDT) ensembles (Malinin et al. 2020a, Friedman, 2002) with optimal hyperparameters and selected features determined via RMSE during the hyperparameter optimization and model selection process. Since ensembles can reduce uncertainty compared to singular models (Dietterich 2000), we used the mean of each SGB ensemble watershed prediction distribution as the predicted response (Supplementary Methods 9). The variance of the resulting SGB distributions of predictions was also used to estimate the epistemic uncertainty of our models (Zhou et al. 2022, Malinin et al. 2020a), reflected by the disagreement of the ensemble model predictions (Malinin et al. 2020a, Malinin et al. 2020b, Yang and Li 2023).

Second, to obtain predictions that contain extreme values of movement distance distributions, we generated step-level predictions with a gamma distribution using the predicted mean and dispersion for each watershed. Through this process, we ended up with two sets of

both observed and predicted values. That is, 1) the observed individual wild individual-level mean movement distances with the predicted wild pig-level means and 2) the step-level movement distance distributions. Lastly, we compared the predicted values for similarity to their respective observed value distributions using a Kolmogorov-Smirnov (KS) test (Press and Teukolsky 1988).

Example 2: Scaling-up HSS models: Given context-dependence of habitat selection (Mason and Fortin 2017), we chose to group individuals by habitat availability, as defined by availability of our land class predictor variable. This resulted in 12 groups based on the availability of four key land cover types: forest, cropland, wetland, and urban. The four land cover types have all been supported by previous research to be important to wild pig movement or habitat selection (Wilber et al. 2019, Brown 1985, Campbell and Long 2009, Jones et al. 2018, Engeman et al. 2007, Grey et al, 2020, Vercauteren et al. 2020). The number of individuals and study sites were calculated for each group, as was the distribution of the days across the year.

For our HSS model, we used integrated step selection analysis (iSSA), which uses a conditional logistic regression model (Avgar et al. 2016). The iSSA model estimates movement and resource selection parameters and allows inference about habitat selection within a mechanistic movement model. The data was resampled to every six hours and daily steps (24 hours with 3 hour window) were calculated for four times a day. Individuals without a minimum of ten consecutive days were removed from the model. For the iSSA model, ten random steps were calculated using the *amt* package (Signer et al. 2019) drawn from the full distribution of steps and turn angles across all years and study sites. Each individual was run in a separate conditional regression model to examine habitat selection using deviation coding to compare the

mean for each land cover type to the overall mean of land types. Only land cover types that each group of wild pigs visited at least once was included in each model. The average was then taken of all estimates that were significant for each habitat availability subset. All available land cover types from the National Land Cover Database (NLCD) were included, with the four levels of developed land combined into a single category of developed land. Our model was tested for its sensitivity of inclusion of step length as an interaction term, considering significant interactions and AIC.

To understand data-driven uncertainty in our top model, we also conducted a leave-one-out sensitivity analysis. In our leave-one-out analysis within each habitat availability subset, one study was removed at a time, the model was rerun, and the mean square error was calculated for each group and the full model. The standard error was calculated for the estimates for each land type and each availability subset for every subset with more than one study. To further characterize uncertainty in our findings, we compared the correlation of these standard error results based on the number of studies, individuals, and total observations per availability subset using Pearson's correlation.

Results

Scaling-up movement distance models

We captured broad-scale spatial and temporal patterns across the U.S. for step length (Figure 3) and displacement means (Figure 4). Across seasons, mean daily step length predictions ranged between 152.41-1404.21 m, and between 325.73-3048.05 m for mean daily displacements. Step length and displacement variance (dispersion parameter; Supplementary Results 1, 2) varied from 0.39-3.05 and 0-0.99, respectively. The largest seasonal changes

233 predicted for step length and displacement mean and were in the northern ranges of wild pigs.
234 This is reflected in daylight duration mean and variance being among the top most influential
235 variables for all responses (Figure 5; Supplementary Results 3-6).

236 All top models used either all variables, or all but those with the lowest 10% relative
237 influence, and all top models used the random cross-validation compared to the spatially-
238 segregated method (Table 2). Our nested cross-validation resulted in lower RMSE for outer folds
239 compared to inner folds (Supplementary Results 7), and step-level distributions. Observed
240 individual-level responses were captured by our predictions (Supplementary Results 8), with
241 study-level heterogeneity RMSE (Supplementary Results 9) and fit of predicted to observed
242 values (Supplementary Results 10-13). The fit of the predicted step-level step length and
243 displacement distributions was closer to observed step level values than the same comparison
244 between individual-level responses, where Kolmogorov-Smirnov test statistics were lower for all
245 step-level daily distance pairs compared to mean daily distance pairs (Table 3) and predicted
246 values were closer to observed values (Figure 3, 4). Model uncertainty evaluations for the
247 movement distance models varied regionally for each response, and ranged from 2.05e-05 to
248 8.50e-05 for step length mean, 1.29e-05 to 6.27e-05 for displacement mean, 6.6e-05 to 7.40e-04
249 for step length variance, and 2.57e-05 to 9.08e-05 for displacement variance (Supplementary
250 Results 14).

251

252 *Scaling-up HSS models*

253 We captured differences in habitat selection based on land cover type availability (Figure 6).
254 There were 12 habitat availability subsets (Supplementary Results 15). The largest subset
255 contained 219 individuals and comprised of all included studies, and comprised of all land class

types (Supplementary Results 15). In general, emergent herbaceous wetland and barren land was preferred over other land cover types, and pasture and cropland were avoided (Figure 6, Supplementary Results 16, 17). Selection for other land cover types was mixed with grasslands being generally selected against, and deciduous forests selected for. However, not all groups had data from all year (Supplementary Results 18). Step length was considered as an interaction term and only seven of the interactions had an effect greater than 1.1 or less than 0.9 (Supplementary Results 19). Outside of the group made of an individual pig with access only to open water and emergent herbaceous wetlands, these effects were all towards urban (developed) or barren land cover types. Some of these long, regular movements were investigated for individuals to verify they were not errors and were confirmed as landfills or a location with food waste.

Our leave-one-out sensitivity analysis allowed us to identify which availability subsets had less variation in the estimates, and further characterize this through smaller standard errors between estimates and changes in mean square error (Supplementary Results 20-22). Subset 1 and 2 were excluded from this analysis, as they contained a single study. We saw less variation in the availability subsets generally in the larger subsets that also had access to a greater number of land types. Removal of studies from subset 12 had minimal effect, as the standard errors of the estimates ranged from 0.02 to 0.006. However, the highest three standard errors of the estimates were for shrubs in subset 7 (2 studies, 8 individuals, 6 land types) and crops and evergreen forests in subset 11 (6 studies, 13 individuals, 10 land types). The number of individuals, studies, total GPS points and land types used in the availability subset was significantly correlated with the standard errors of the estimates from the leave-one-out analysis ($p<0.001$ for all four) (Supplementary Results 23).

Discussion

Estimation of local animal movement processes has already been used to improve management of invasive species (Thompson et al. 2021), mitigate agricultural damage (Boyce et al. 2020), reduce threats to species of conservation concern (Fraser et al. 2018), and control disease outbreaks (Aiello et al. 2023, Manlove et al. 2022, Wilbur et al. 2022, Pepin et al. 2022). However, previous statistical movement models either dealt with a limited number of ecosystems, or were not developed to extrapolate models across divergent ecological contexts. We presented here an approach for scaling up high-resolution animal movement for national-scale inference. We predicted two features of movement: movement distance and habitat selection. We optimized predictions, reduced overfitting, and represented heterogeneity in our broad-scale movement distance models. We also accounted for spatial non-stationarity in broad-scale habitat -selection models through grouping by habitat availability. In this context, our framework can be leveraged to parameterize mechanistic models in areas where limited movement studies have been conducted due to difficult terrain or lack of resources, throughout large ranges across which continuous telemetry studies would be infeasible, or in regions where animals are absent but may occupy in the future due to climate change, anthropogenic land-use change, or direct relocation efforts.

We found the nested CV resulted in RMSE values were lower for the outer fold for all models, suggesting we did not overfit models to sampled data. We also found that spatially blocking CV was not the best technique for all response variables to optimize predictions, where random groupings for CV resulted in lower RMSE for all models (Valavi et al. 2019). Many of the movement distance models had low model uncertainty for predicting out-of-sample movement path characteristic average values across a range of habitat availability. While our

302 predictive uncertainty estimate provided some guidelines for understanding the variability in our
303 predictions, these models were computationally intensive. Alternative frameworks using virtual
304 gradient-boosted ensemble models would achieve similar outcomes, while reducing
305 computational resource needs (Malinin et al. 2019) and generating multiple kinds of uncertainty
306 estimates (NGBoost), including separate estimations of epistemic and aleatoric uncertainty (due
307 to noise in the data). Future work to improve scaling up movement may also consider
308 incorporating uncertainty estimates in tandem with loss functions during model optimization to
309 improve accuracy of predictions (Sarawgi et al. 2021). Such benefits have been seen in neural
310 network predictions (Sarawgi et al. 2021), though to our knowledge this has not been attempted
311 with supervised machine learning methods.

312 Our HSS results indicated differences in habitat selection across different ecological
313 contexts, suggesting spatial non-stationarity exists across ecoregions within the range
314 distribution of wild pigs. By grouping according to habitat availability, we were able to indicate
315 general and divergent selection across differing contexts by grouping individuals by available
316 land cover types. For example, while wetlands were consistently positively selected and pasture
317 and croplands negatively selected, preference for other land types were more varied. This nuance
318 would have been lost if a single group would have been used in a single selection model.
319 Overall, our method of grouping iSSA models by habitat availability provides a simple
320 framework for leveraging movement data across many ecological contexts.

321 There were some limitations in our HSS modeling. We had limited data in some
322 geographic areas, especially in the northern regions of the continental United States, where wild
323 pig populations are sparse. Also, due to the variance in GPS-collar fix rates, we had to resample
324 all datasets to the lowest frequency rate, which was every six hours. This meant the level of

detail was equal over each study, but some information was lost. This is a consistent and known issue with upscaling data at the pre-model stage (Fritsch, Lischke, and Meyer 2020), and is a common limitation when modeling geolocation data. Lastly, the variability in effect size was evident for habitat availability groups containing fewer samples. Potentially, as these land cover combinations occurred more rarely in our data set, they are likely less relevant on a national scale.

Our step-selection model did not incorporate seasonal differences. Seasonal differences may be especially relevant for crop use, as seasonality affects the value of some crops that can serve as a source of protective cover as well as food (Wilber et al. 2019, Boyce et al. 2020). This may explain why crops were frequently selected against in the HSS models when generalizing groups across the entire year and across spatial extent, but this might change if seasonality or different behavioral states were incorporated (e.g., Wilber et al. 2019). Furthermore, not all groups included a full year of data. Including more parameters for seasonality beyond year quarters in the context of different behavioral states may have also improved our movement distance model predictions. For example, we did not incorporate local information about seasonal breeding patterns or habitat availability changes at finer scales, both of which can be important in predicting wild pig movement patterns (Schlichting et al. 2020, Clontz et al. 2021, Chinn et al. 2023). Further work to incorporate multiple scales representing periodic seasonal changes in behavioral states may expand the utility of this framework, as incorporation of multiple scales has been shown to improve movement model predictions at local scales as well (Torney et al. 2021, Leos-Barajas et al. 2017, Buderman et al. 2018, Benhamou et al. 2014).

There are cases when management strategies and understanding of movement would improve from a cohesive view of a larger geographic area. For example, both the U.S. and

348 Australia have wild pig management plans at the national level (United States Department of
349 Agriculture 2020, Australian Pork Limited 2021), where understanding national-level differences
350 in movement behaviors could inform appropriate allocation of resources and adaptation of
351 control strategies according to ecological context. Also, many countries have foreign animal
352 disease preparedness plans that involve risk from wildlife and require an understanding of animal
353 movement across regional extents to plan response. While the application of our framework
354 focused on a single species, it could readily be applied to better predict and understand species
355 interactions for conservation planning, disease spillover and spillback mitigation, and invasive
356 species management.
357

References

Aiello, C. M., Galloway, N. L., Prentice, P. R., Darby, N. W., Hughson, D., & Epps, C. W. (2023). Movement models and simulation reveal highway impacts and mitigation opportunities for a metapopulation-distributed species. *Landscape Ecology*, *38*(4), 1085–1103. https://doi.org/10.1007/s10980-023-01600-6

Australian Pork Limited. (2021). *National Feral Pig Action Plan: 2021-2031*. Australian Pork Limited, Kingston ACT 2600. https://feralpigs.com.au/wp-content/uploads/2023/02/National-Feral-Pig-Action-Plan-February-2023.pdf

Benhamou, S. (2014). Of scales and stationarity in animal movements. *Ecology Letters*, *17*(3), 261–272. https://doi.org/10.1111/ele.12225

Boyce, C. M., VerCauteren, K. C., & Beasley, J. C. (2020). Timing and extent of crop damage by wild pigs (*Sus scrofa* Linnaeus) to corn and peanut fields. *Crop Protection*, *133*, 105131. https://doi.org/10.1016/j.cropro.2020.105131

Brown, L. N. (1985). Elimination of a Small Feral Swine Population in an Urbanizing Section of Central Florida. *Florida Scientist*, *48*(2), 120–123.

Buderman, F. E., Hooten, M. B., Ivan, J. S., & Shenk, T. M. (2018). Large-scale movement behavior in a reintroduced predator population. *Ecography*, *41*(1), 126–139. https://doi.org/10.1111/ecog.03030

Byrne, A. W., Quinn, J. L., O'Keeffe, J. J., Green, S., Paddy Sleeman, D., Wayne Martin, S., & Davenport, J. (2014). Large-scale movements in European badgers: Has the tail of the movement kernel been underestimated? *Journal of Animal Ecology*, *83*(4), 991–1001. https://doi.org/10.1111/1365-2656.12197

Campbell, T. A., & Long, D. B. (2009). Feral swine damage and damage management in forested ecosystems. *Forest Ecology and Management*, *257*(12), 2319–2326. https://doi.org/10.1016/j.foreco.2009.03.036

Chinn, S.M., J. Hepinstall-Cymerman, and J.C. Beasley. 2023. Reproduction drives changes in space use and habitat selection in a highly adaptable invasive mammal. Journal of Mammalogy 104:479-495

Clontz, L. M., Pepin, K. M., VerCauteren, K. C., & Beasley, J. C. (2021). Behavioral state resource selection in invasive wild pigs in the Southeastern United States. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-86363-3

Commission for Environmental Cooperation (CEC). (2020). *2015 Land Cover of North America at 30 Meters. North American Land Change Monitoring System. Ed. 2.0, Raster digital data [30-m].* Canada Centre for Remote Sensing (CCRS), U.S. Geological Survey (USGS), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Comisión Nacional Forestal (CONAFOR), Instituto Nacional de Estadística y Geografía (INEGI).

Data.gov. (2021). *TIGER/Line Shapefile, 2019, nation, U.S., Primary Roads National Shapefile* [dataset]. Unknown. https://catalog.data.gov/dataset/tiger-line-shapefile-2019-nation-u-s-primary-roads-national-shapefile

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In G. Goos, J. Hartmanis, & J. Van Leeuwen (Eds.), *Multiple Classifier Systems* (Vol. 1857, pp. 1–15). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45014-9_1

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

Engeman, R. M., Stevens, A., Allen, J., Dunlap, J., Daniel, M., Teague, D., & Constantin, B. (2007). Feral swine management for conservation of an imperiled wetland habitat: Florida's vanishing seepage slopes. *Biological Conservation*, *134*(3), 440–446. https://doi.org/10.1016/j.biocon.2006.08.033

Franckowiak, G. A., Torres-Poché, Z., & Poché, R. M. (2018). Activity Patterns by Feral Hogs in the Texas Panhandle. *The American Midland Naturalist*, *180*(2), 233–245. https://doi.org/10.1674/0003-0031-180.2.233

Fraser, K. C., Davies, K. T. A., Davy, C. M., Ford, A. T., Flockhart, D. T. T., & Martins, E. G. (2018). Tracking the Conservation Promise of Movement Ecology. *Frontiers in Ecology and Evolution*, *6*. https://www.frontiersin.org/articles/10.3389/fevo.2018.00150

Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance. *Canadian Journal of Forest Research*, *46*(3), 323–339. https://doi.org/10.1139/cjfr-2014-0562

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. *Methods in Ecology and Evolution*, *11*(11), 1368–1378. https://doi.org/10.1111/2041-210X.13466

Fushing, H., Hwang, C.-R., Lee, H.-C., Lan, Y.-C., & Horng, S.-B. (2006). Testing and mapping non-stationarity in animal behavioral processes: A case study on an individual female bean weevil. *Journal of Theoretical Biology*, *238*(4), 805–816. https://doi.org/10.1016/j.jtbi.2005.06.031

Gibert, J. P., Chelini, M.-C., Rosenthal, M. F., & DeLong, J. P. (2016). Crossing regimes of temperature dependence in animal movement. *Global Change Biology*, *22*(5), 1722–1736. https://doi.org/10.1111/gcb.13245

Hijmans, R. J., Philips, S., & Elith, J. (2022). *dismo: Species Distribution Modeling* (1.3-9) [R package version 1.3-9]. https://CRAN.R-project.org/package=dismo

Jones, K. C., Gorman, T. A., Rincon, B. K., Allen, J., Haas, C. A., & Engeman, R. M. (2018). Feral swine *Sus scrofa*: A new threat to the remaining breeding wetlands of the Vulnerable reticulated flatwoods salamander *Ambystoma bishopi*. *Oryx*, *52*(4), 669–676. https://doi.org/10.1017/S0030605316001253

Kaszta, Ż., Cushman, S. A., & Slotow, R. (2021). Temporal Non-stationarity of Path-Selection Movement Models and Connectivity: An Example of African Elephants in Kruger National Park. *Frontiers in Ecology and Evolution*, *9*. https://www.frontiersin.org/articles/10.3389/fevo.2021.553263

Kay, S. L., Fischer, J. W., Monaghan, A. J., Beasley, J. C., Boughton, R., Campbell, T. A., Cooper, S. M., Ditchkoff, S. S., Hartley, S. B., Kilgo, J. C., Wisely, S. M., Wyckoff, A. C., VerCauteren, K. C., & Pepin, K. M. (2017). Quantifying drivers of wild pig movement across multiple spatial and temporal scales. *Movement Ecology*, *5*(1), 14. https://doi.org/10.1186/s40462-017-0105-1

Kie, J. G., Matthiopoulos, J., Fieberg, J., Powell, R. A., Cagnacci, F., Mitchell, M. S., Gaillard, J.-M., & Moorcroft, P. R. (2010). The home-range concept: Are traditional estimators still relevant with modern telemetry technology? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2221–2231. https://doi.org/10.1098/rstb.2010.0093

King, A. W. (1991). Translating models across scales in the landscape. In *Quantitative methods in landscape ecology: The analysis and interpretation of landscape heterogeneity* (Vol. 82, pp. 470–517). Springer.

Lainder, A. D., & Wolfinger, R. D. (2022). Forecasting with gradient boosted trees: Augmentation, tuning, and cross-validation strategies: Winning solution to the M5 Uncertainty competition. *International Journal of Forecasting*, *38*(4), 1426–1433. https://doi.org/10.1016/j.ijforecast.2021.12.003

Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., van Beest, F. M., Nabe-Nielsen, J., & Morales, J. M. (2017). Multi-scale Modeling of Animal Movement and General Behavior Data Using Hidden Markov Models with Hierarchical Structures. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(3), 232–248. https://doi.org/10.1007/s13253-017-0282-9

Lessmann, S., Sung, M.-C., Johnson, J. E. V., & Ma, T. (2012). A new methodology for generating and combining statistical forecasting models to enhance competitive event prediction. *European Journal of Operational Research*, *218*(1), 163–174. https://doi.org/10.1016/j.ejor.2011.10.032

Lewis, J. S., Corn, J. L., Mayer, J. J., Jordan, T. R., Farnsworth, M. L., Burdett, C. L., VerCauteren, K. C., Sweeney, S. J., & Miller, R. S. (2019). Historical, current, and potential population size estimates of invasive wild pigs (*Sus scrofa*) in the United States. *Biological Invasions*, *21*(7), 2373–2384. https://doi.org/10.1007/s10530-019-01983-1

Lewis, J. S., Farnsworth, M. L., Burdett, C. L., Theobald, D. M., Gray, M., & Miller, R. S. (2017). Biotic and abiotic factors predicting the global distribution and population density of an invasive large mammal. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep44152

Malinin, A. (2019). *Uncertainty Estimation in Deep Learning with application to Spoken Language Assessment*. https://www.repository.cam.ac.uk/handle/1810/298857

Malinin, A., Mlodozeniec, B., & Gales, M. (2020). *ENSEMBLE DISTRIBUTION DISTILLATION*.

Malinin, A., Ustimenko, A., & Prokhorenkova, L. (2021). *UNCERTAINTY IN GRADIENT BOOSTING VIA ENSEMBLES*.

Manlove, K., Wilber, M., White, L., Bastille-Rousseau, G., Yang, A., Gilbertson, M. L. J., Craft, M. E., Cross, P. C., Wittemyer, G., & Pepin, K. M. (2022). Defining an epidemiological landscape that connects movement ecology to pathogen transmission and pace-of-life. *Ecology Letters*, *25*(8), 1760–1782. https://doi.org/10.1111/ele.14032

Mason, T. H. E., & Fortin, D. (2017). Functional responses in animal movement explain spatial heterogeneity in animal–habitat relationships. *Journal of Animal Ecology*, *86*(4), 960–971. https://doi.org/10.1111/1365-2656.12682

McClure, M. L., Burdett, C. L., Farnsworth, M. L., Lutman, M. W., Theobald, D. M., Riggs, P. D., Grear, D. A., & Miller, R. S. (2015). Modeling and Mapping the Probability of Occurrence of Invasive Wild Pigs across the Contiguous United States. *PLOS ONE*, *10*(8), e0133771. https://doi.org/10.1371/journal.pone.0133771

Mesgaran, M. B., Cousens, R. D., & Webber, B. L. (2014). Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, *20*(10), 1147–1159. https://doi.org/10.1111/ddi.12209

Miller, R. S., Bevins, S. N., Cook, G., Free, R., Pepin, K. M., Gidlewski, T., & Brown, V. R. (2022). Adaptive risk-based targeted surveillance for foreign animal diseases at the wildlife-livestock interface. *Transboundary and Emerging Diseases*, *69*(5), e2329–e2340. https://doi.org/10.1111/tbed.14576

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*. https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021

National Drought Mitigation Center, U.S. Department of Agriculture, & National Oceanic and Atmospheric Administration. (2022). *U.S. Drought Monitor*.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. The Review of Economics and Statistics, MIT Press, vol. 94(1), pages 20-36, February.

Paton, R. S., & Matthiopoulos, J. (2016). Defining the scale of habitat availability for models of habitat selection. *Ecology*, *97*(5), 1113–1122. https://doi.org/10.1890/14-2241.1

Pease, B. S., Pacifici, K., & Kays, R. (2022). Exploring spatial nonstationarity for four mammal species reveals regional variation in environmental relationships. *Ecosphere*, *13*(8), e4166. https://doi.org/10.1002/ecs2.4166

Pepin, K. M., Davis, A. J., Beasley, J., Boughton, R., Campbell, T., Cooper, S. M., Gaston, W., Hartley, S., Kilgo, J. C., Wisely, S. M., Wyckoff, C., & VerCauteren, K. C. (2016). Contact heterogeneities in feral swine: Implications for disease management and future research. *Ecosphere*, *7*(3), e01230. https://doi.org/10.1002/ecs2.1230

Potts, J. R., Börger, L., Strickland, B. K., & Street, G. M. (2022). Assessing the predictive power of step selection functions: How social and environmental interactions affect animal space use. *Methods in Ecology and Evolution*, *13*(8), 1805–1818. https://doi.org/10.1111/2041-210X.13904

Press, William H., and Saul A. Teukolsky. "Kolmogorov-Smirnov test for two-dimensional data." Comput. Phys 2.74 (1988): 10-1063.

R Core Team. (2022). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rousseau, J. S., & Betts, M. G. (2022). Factors influencing transferability in species distribution models. *Ecography*, *2022*(7), e06060. https://doi.org/10.1111/ecog.06060

Sarawgi, U., Khincha, R., Zulfikar, W., Ghosh, S., & Maes, P. (2021). *Uncertainty-Aware Boosted Ensembling in Multi-Modal Settings* (arXiv:2104.10715). arXiv. http://arxiv.org/abs/2104.10715

Schlichting, P. E., Beasley, J. C., Boughton, R. K., Davis, A. J., Pepin, K. M., Glow, M. P., Snow, N. P., Miller, R. S., VerCauteren, K. C., & Lewis, J. S. (2020). A Rapid Population Assessment Method for Wild Pigs Using Baited Cameras at 3 Study Sites. *Wildlife Society Bulletin*, *44*(2), 372–382. https://doi.org/10.1002/wsb.1075

Shaw, A. K. (2020). Causes and consequences of individual variation in animal movement. *Movement Ecology*, *8*(1), 12. https://doi.org/10.1186/s40462-020-0197-x

Shi, H., & Yin, G. (2018). Boosting conditional logit model. *Journal of Choice Modelling*, *26*, 48–63. https://doi.org/10.1016/j.jocm.2017.07.002

Signer, J., Fieberg, J. and Avgar, T., 2019. Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. Ecology and evolution, 9(2), pp.880-890.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, *39*(5), 1–13. https://doi.org/10.18637/jss.v039.i05

Streng, K., de Best, P. A., Timen, A., Koopmans, M. P. G., van der Poel, W. H. M., & Sikkema, R. S. (2023). Rapid response screening for emerging zoonotic pathogens, barriers and opportunities: A study for enhanced preparedness of the Netherlands. *One Health*, *16*, 100507. https://doi.org/10.1016/j.onehlt.2023.100507

Thompson, B. K., Olden, J. D., & Converse, S. J. (2021). Mechanistic invasive species management models and their application in conservation. *Conservation Science and Practice*, *3*(11), e533. https://doi.org/10.1111/csp2.533

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S.-C., & Wilson, B. E. (2022). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1. *ORNL DAAC*. https://doi.org/10.3334/ORNLDAAC/2129

Thums, M., Fernández-Gracia, J., Sequeira, A., Eguíluz, V., Duarte, C., & Meekan, M. (2018). How Big Data Fast Tracked Human Mobility Research and the Lessons for Animal Movement Ecology. *Frontiers in Marine Science*, *5*. https://doi.org/10.3389/fmars.2018.00021

Torney, C. J., Morales, J. M., & Husmeier, D. (2021). A hierarchical machine learning framework for the analysis of large scale animal movement data. *Movement Ecology*, *9*(1), 6. https://doi.org/10.1186/s40462-021-00242-0

United States Department of Agriculture. (2020). *APHIS National Feral Swine Damage Management Program*. https://www.aphis.usda.gov/aphis/ourfocus/wildlifedamage/operational-activities/feral-swine/feral-swine-program

USDA Forest Service. (2019). *NLCD 2016 Tree Canopy Cover (CONUS)*. Salt Lake City, UT.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2023). Flexible species distribution modelling methods perform well on spatially separated testing data. *Global Ecology and Biogeography*, *32*(3), 369–383. https://doi.org/10.1111/geb.13639

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, *92*(1), e01486. https://doi.org/10.1002/ecm.1486

"Wild Pig Spatial Ecology and Behavior" in VerCauteren, K.C., J.C. Beasley, S.S. Ditchkoff, J.J. Mayer, G.J. Roloff, and B.K. Strickland (Eds.). Invasive Wild Pigs in North America: Ecology,

Impacts, and Management. Boca Raton, FL: CRC. Press; 2020.

Werkowska, W., Márquez, A. L., Real, R., & Acevedo, P. (2017). A practical overview of transferability in species distribution modeling. *Environmental Reviews*, *25*(1), 127–133. https://doi.org/10.1139/er-2016-0045

Wilber, M. Q., Chinn, S. M., Beasley, J. C., Boughton, R. K., Brook, R. K., Ditchkoff, S. S., Fischer, J. W., Hartley, S. B., Holmstrom, L. K., & Kilgo, J. C. (2019). Predicting functional responses in agro-ecosystems from animal movement data to improve management of invasive pests. *Ecological Applications*, *30*(1), e02015.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*(2), 125–138. https://doi.org/10.1023/B:LAND.0000021711.40074.ae

Yang, A., Schlichting, P., Wight, B., Anderson, W. M., Chinn, S. M., Wilber, M. Q., Miller, R. S., Beasley, J. C., Boughton, R. K., VerCauteren, K. C., Wittemyer, G., & Pepin, K. M. (2021). Effects of social structure and management on risk of disease establishment in wild pigs. *Journal of Animal Ecology*, *90*(4), 820–833. https://doi.org/10.1111/1365-2656.13412

Yang, C.-I., & Li, Y.-P. (2023). Explainable uncertainty quantifications for deep learning-based molecular property prediction. *Journal of Cheminformatics*, *15*(1), 13. https://doi.org/10.1186/s13321-023-00682-3

Zhou, X., Liu, H., Pourpanah, F., Zeng, T., & Wang, X. (2022). A survey on epistemic (model) uncertainty in supervised learning: Recent advances and applications. *Neurocomputing*, *489*, 449–465. https://doi.org/10.1016/j.neucom.2021.10.119

Zurell, D., Elith, J., & Schröder, B. (2012). Predicting to new environments: Tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, *18*(6), 628–634. https://doi.org/10.1111/j.1472-4642.2012.00887.x

Figures and Tables

Figure 1. Conceptual framework for our methodological workflow for both habitat step-selection (HSS) ("where to move?") and movement distance models ("how far to move?). For HSS models, we grouped individual wild pigs in the pre-processing stage according to habitat availability. Then, HSS scores (positive indicates selection for, negative indicates selection against) are displayed separately for each habitat availability set. For movement distance models, we aggregated movement distances during pre-processing by obtaining mean and variance of movement distance distribution at the individual-level, and modeling both distance mean and variances as responses. Then, we compared spatial and random blocking for reduced root mean square error values to optimize predictions and reduce overfitting. Lastly, we disaggregated final predictions to U.S. watersheds by using individual-level mean and variances to bootstrap step-level predictions for each watershed.

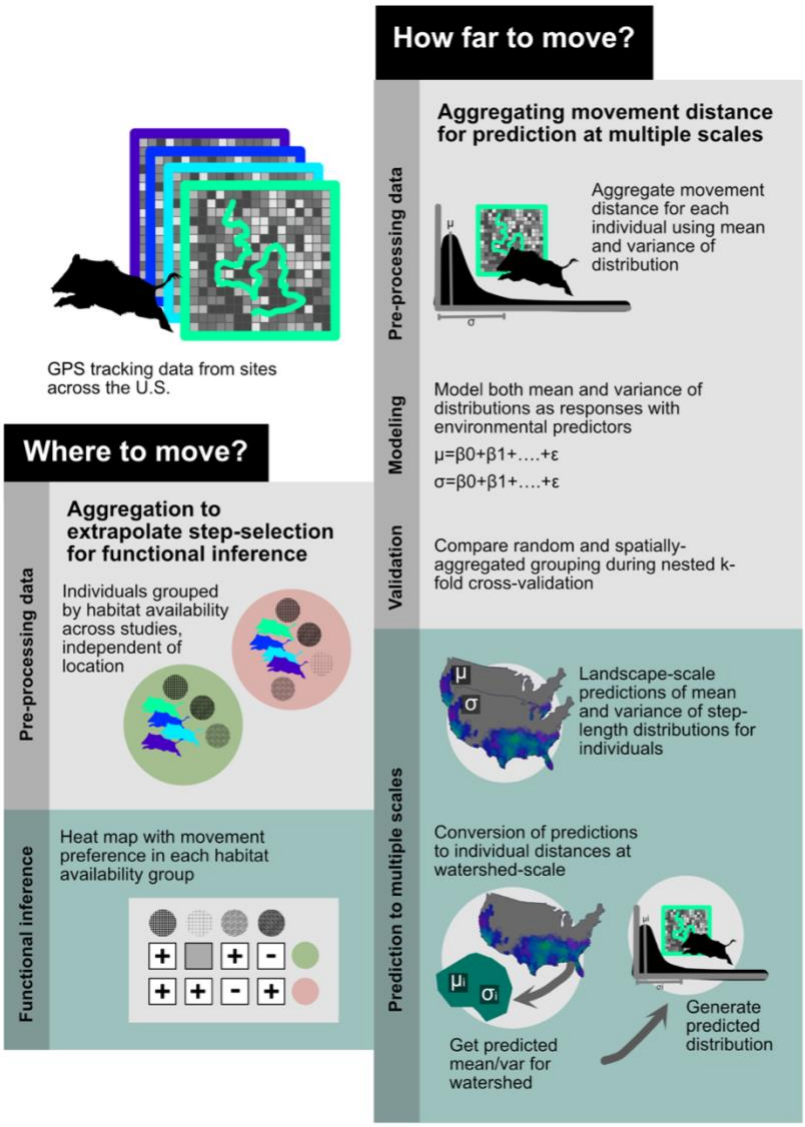
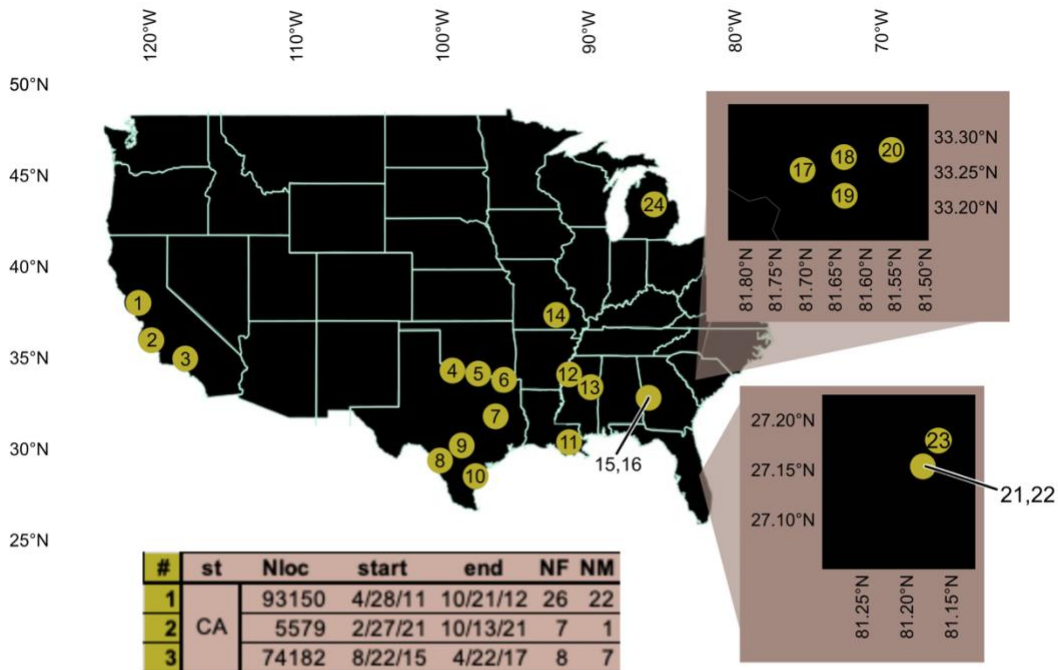


Figure 2. Locations of U.S. wild pig (*Sus scrofa*) GPS collaring studies included in this analysis with total numbers of wild pigs per study, total numbers of females (NF), males (NM) with number of geolocations (Nloc) and start and end dates of geolocation data use in analyses.



#	st	Nloc	start	end	NF	NM
1	CA	93150	4/28/11	10/21/12	26	22
2		5579	2/27/21	10/13/21	7	1
3		74182	8/22/15	4/22/17	8	7
4	TX	117797	1/14/18	3/13/18	23	13
5	OK	198036	8/31/17	4/7/20	25	15
6	TX	106292	1/25/19	11/11/19	18	11
7		61	9/11/04	1/15/05	0	2
8		19461	12/9/05	7/25/06	3	6
9		378395	1/30/16	7/14/16	24	16
10		56945	1/13/08	5/29/11	17	29
11	LA	67960	8/4/11	1/6/16	5	22
12	MS	38338	2/4/16	1/31/17	5	10
13		12647	11/25/20	2/7/22	10	0
14	MO	35429	10/2/10	5/29/13	9	13
15	GA	787	6/2/04	6/19/06	8	0
16		52256	7/28/07	6/2/10	0	39
17	SC	3977	3/28/14	8/19/16	3	0
18		20518	12/9/13	11/12/15	14	10
19		6487	11/14/15	12/4/16	4	9
20		56911	2/3/17	11/4/17	11	8
21	FL	53963	1/12/13	4/25/13	9	9
22		316192	5/9/15	5/9/17	31	25
23		145851	2/23/17	4/18/18	16	7
24	MS	43178	7/29/14	8/23/16	6	2

Figure 3. Spatial gradient maps of predicted individual-level mean daily step lengths (m) for wild pigs (*Sus scrofa*) across the United States for a) January-March; b) April-June; c) July-September; d) October-December. Histograms compare step-level relative % frequencies for each binned step-length value between predicted (orange) and observed (gray) values subset across each season. QQ-plots demonstrate similarity between observed and predicted step length distributions for individual mean pig-level step length distributions (blue) and step-level bootstrapped (red) step length distributions. For qq-plots, mean and step-level distributions were scaled, where respective maximum values for each set were scaled to 1.

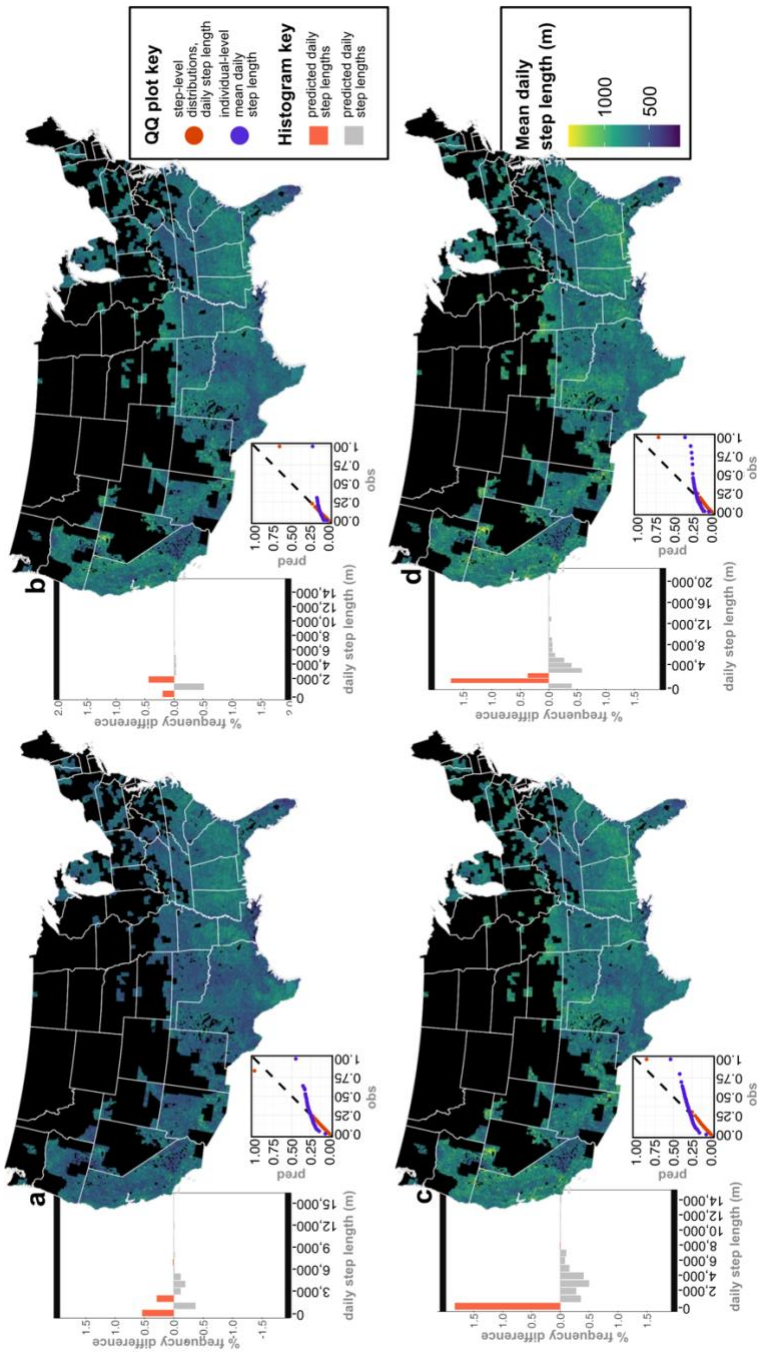


Figure 4. Spatial gradient maps of the predicted wild pig-level mean daily displacement (m) for individual wild pigs (*Sus scrofa*) across the United States for a) January-March; b) April-June; c) July-September; d) October-December. Histograms compare step-level relative % frequencies for each binned displacement value between predicted (orange) and observed (gray) values subset across each season. QQ-plots demonstrate similarity between observed and predicted displacement distributions for individual mean wild pig-level displacement distributions (blue) and step-level bootstrapped (red) displacement distributions. For qq-plots, mean and step-level distributions were scaled, where respective maximum values for each set were scaled to 1.

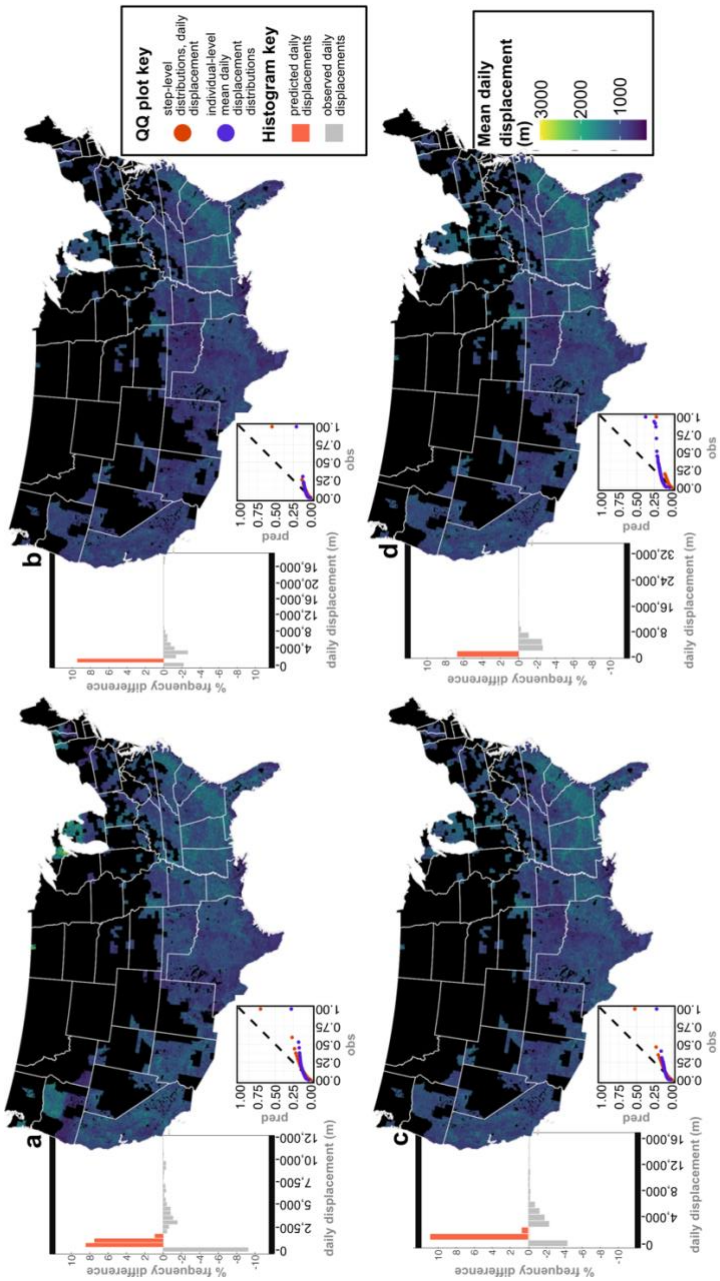


Figure 5. Heat map of proportional relative influence (Rel. Inf. Prop; each predictor scaled to highest relative influence for each model, highest relative influence=1.00) for top models of environmental predictors on wild pig (*Sus scrofa*) movement distances including total mean displacement (Displ.;m), displacement dispersion (Displ. var), step length (S.L.; m), and step length dispersion (S.L. var; m).

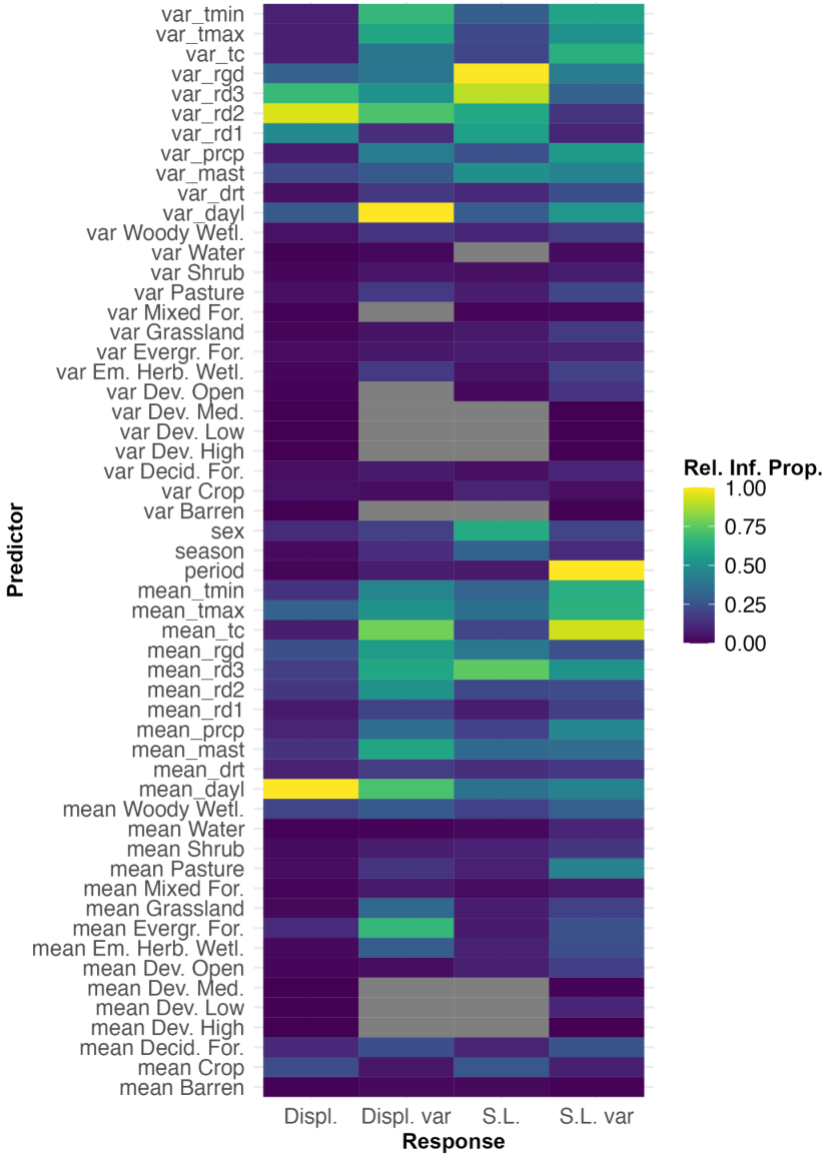


Figure 6. Dot plot showing mean and 95% confidence interval of estimates from habitat step-selection (HSS) model for different land types, grouped according to access to wetlands, urban areas, forests, and cropland. All 12 groups are represented for which of the 12 land types they had access to.

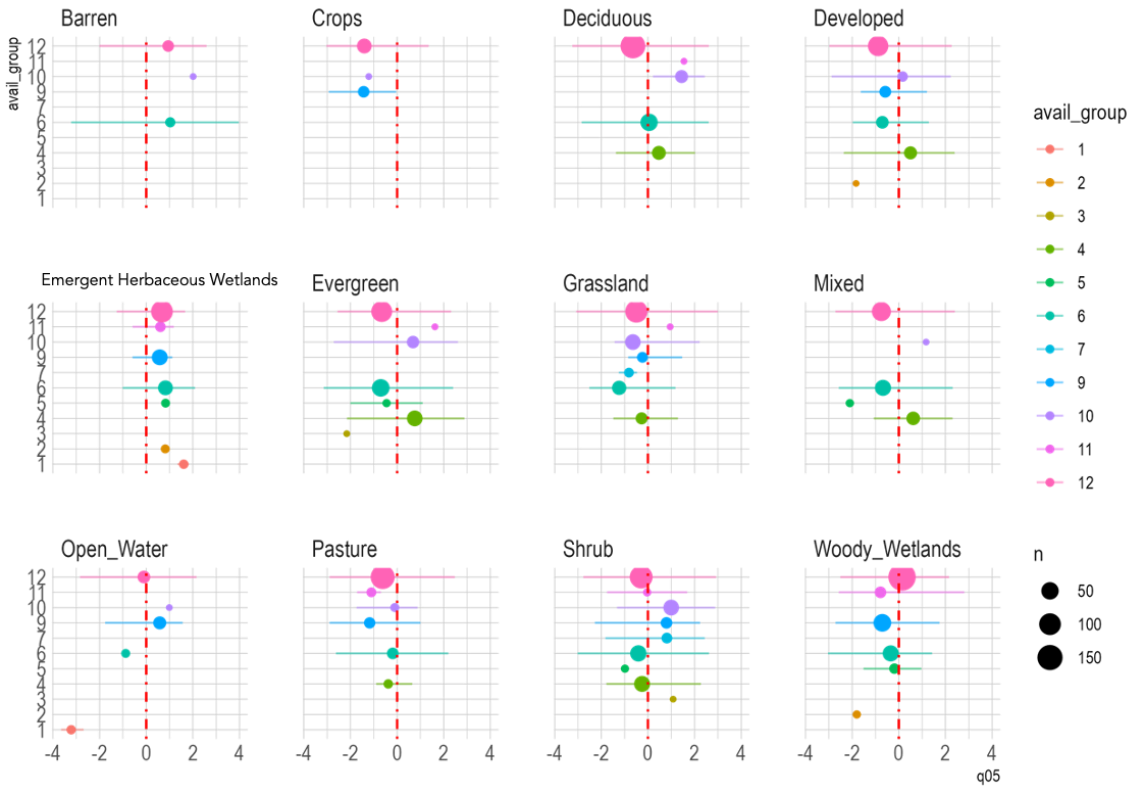


Table 1. Name, description, type, and source (with resolution, for raster sources) for each variable used in this study. m- meters, s- seconds, no.- number, deg. C- degrees Celsius, Temp.- temperature, Tmax- maximum temperature, Tmin- minimum temperature., res- resolution, as- arc-seconds.

	Variable	Abbrev	Units	Description	Type	Source
Responses	Step length	S.L.	m/day	mean daily step length averaged for each individual.	geolocation data	national wild pig geolocation data
	Step length dispersion	S.L. Var.	-	dispersion parameter of a gamma distribution fit to distribution of daily step lengths for each individual.	geolocation data	national wild pig geolocation data
	Displacement	Displ.	m/day	mean daily displacement from starting point based on average location over first seven days of each season.	geolocation data	national wild pig geolocation data
	Displacement dispersion	Displ. Var.	-	dispersion parameter of a gamma distribution fit to distribution of daily displacements for each individual.	geolocation data	national wild pig geolocation data
Predictors	period	per	-	period of day, split into quarters (12A-6A; 6A-12P; 12P-6P; 6P-12A)	geolocation data	national wild pig geolocation data
	season	season	-	season of year, split into quarters (Jan-Mar; Apr-Jun; Jul-Sep; Oct-Dec)	geolocation data	national wild pig geolocation data
	count	ct	-	number of daily averaged step lengths for each individual	geolocation data	national wild pig geolocation data
	daylight	dayl mn; dayl var	s	number of seconds of daylight per day	raster, 1km res.	Thornton et al. 2020
	drought	drt mn; drt. var	-	drought monitor index	shapefile	National Drought Mitigation Center et al. (2022)
	mast	mast mn; mast var.	-	number of mast-producing species	raster, 1km res.	Miller et al. 2021
	precip	prcp mn; prcp var	mm/day	amount of precipitation in mm per day	raster, 1km res.	Thornton et al. 2020
	highway density	rd1 mn, rd1 var	m/km2	length of highways for every 10 km2	raster, 10km res.	Meijer et. Al (2018)
	primary road density	rd2 mn, rd1 var	m	length of primary roads for every 10 km2	raster, 10km res.	Meijer et. Al (2018)

	secondary road density	rd3 mn, rd1 var	m	length of secondary roads for every 10 km2	raster, 10km res.	Meijer et. Al (2018)
	ruggedness	rgd mn, rgd var	arc-seconds	three-dimensional complexity of terrain surface of landscape accessed daily	Raster, 30 arc-seconds	Nunn and Puga (2012)
	sex	sex	-	sex of individual	geolocation data	national wild pig geolocation data
	tree cover	tc mn, tc var	-	mean percent of tree cover	Raster, 30m	USDA Forest Service (2019)
	tmax	tmax mn, tmax var	degrees Celsius	maximum daily temperature	raster, 1km res.	Thornton et al. 2020
	tmin	tmin mn, tmin var	degrees Celsius	minimum daily temperature	raster, 1km res.	Thornton et al. 2020

Table 2. Model selection table of root mean squared error (RMSE) values for each cross-validation (CV) method, feature selection method, and response variable. The random CV method used randomly selected groups for the nested k-fold validation, whereas the spatially-blocked ("spatial") CV used data from each study as groups for the nested k-fold validation. Full- all parameters were used in the model; Drop 01- variables having less than 1% influence were removed from the model; Lasso- lasso variable selection used; Null- intercept-only model. s.l- mean daily step length (m); s.l. var- daily step length dispersion; displ.- mean daily displacement (m); displ. var.- daily displacement dispersion. * indicates the lowest RMSE value for each response variable.

	Random CV			**Spatial CV**		
	full RMSE random	**drop 01 RMSE random**	**null RMSE random**	**full RMSE region**	**drop 01 RMSE region**	**null RMSE region**
S.L.	214.37	213.35*	445.01	388.51	389.04	453.27
S.L. var	0.42*	0.42	0.58	0.56	0.56	0.59
Displ.	420.12*	420.53	1107.40	660.67	652.82	778.99
Displ. Var	0.18	0.18*	0.30	0.35	0.36	0.30

Table 3. Komogorov-Smirnov (KS) statistics for step length and displacement distributions. Each KS-test compared observed and predicted distributions for the individual mean wild pig-level (u) and step-level movement distributions for each season (q1, Jan-Mar; q2, Apr-Jun; q3, Jul-Aug; q4, Sep-Dec). Lower KS values indicate that the predicted distribution is more similar to the observed distribution. P-values were <2e-16 for all KS values.

Response	**q**	**u D**	**step D**
sl	1	0.31	0.067
	2	0.31	0.032
	3	0.29	0.036
	4	0.38	0.056
disp	1	0.28	0.094
	2	0.38	0.12
	3	0.32	0.096
	4	0.39	0.13