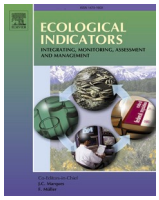
Ecological Indicators 154 (2023) 110684

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Toward broad-scale mapping and characterization of prairie dog colonies from airborne imagery using deep learning

Sean P. Kearney [a,*], Lauren M. Porensky [a], David J. Augustine [a], David W. Pellatz [b]

[a] United States Department of Agriculture, Agricultural Research Service, 1701 Centre Ave., Fort Collins, CO 81526, USA
[b] Thunder Basin Grasslands Prairie Ecosystem Association, 671 Steinle Road, Douglas, WY 82633, USA

ARTICLE INFO

Keywords:
Burrow detection
Deep convolutional neural networks
Cynomys ludovicianus
Prairie dog colony
Rangeland ecology
Remote sensing

ABSTRACT

Monitoring wildlife is fundamental to managing the health of rangelands but challenging due to the extensive and dynamic nature of these ecosystems. The black-tailed prairie dog (*Cynomys ludovicianus*) is considered both a keystone species of conservation concern and an agricultural pest. This animal is an example of a wildlife species for which detailed monitoring is both high priority and difficult to accomplish cost-effectively using ground-based methods. In this study, we conducted a robust evaluation of the potential to use deep learning to detect prairie dog burrows from remotely sensed imagery acquired from unoccupied aerial systems (UAS). We processed UAS imagery to create RGB, topographic position index (TPI) and normalized difference vegetation index (NDVI) products at varying spatial resolutions (2–30 cm). We then evaluated the minimum set of inputs and image resolution required to train a deep convolutional neural network (CNN) for burrow detection and scale this up to identify entire colonies. We validated results at the scale of individual burrows, sub-colony burrow density and range-wide colony area using ground and digitized observations. We found the 2 cm imagery proved computationally impractical for scaling, but performance did not decline between 2 and 5 cm imagery, and models performed well up to 10–15 cm. The top models always included TPI and the combination of RGB + TPI tended to perform best across spatial resolutions. Adding NDVI generally did not improve model performance. At 5 cm resolution, the top models achieved high precision and recall for detecting individual burrows (F-score 0.84–0.87) and burrow density was strongly correlated with validation data ($r = 0.94–0.95$). In pastures with active colonies, overlap between predicted and ground delineated colonies was high (60–94%). The CNN-based approach could not distinguish between currently active colonies and a colony that had recently become inactive due to a sylvatic plague (*Yersinia pestis*) epizootic. However, further analysis showed that CNN-derived burrow density was related to colony age and satellite-derived vegetation conditions in active colonies, and that the plague-affected colony deviated from expected vegetation trends. We conclude that a deep learning algorithm can accurately detect prairie dog burrows from UAS imagery acquired at 5–10 cm resolution, and that scaling from individual burrows to entire colonies is achievable but warrants further research. Combining CNN-derived burrow density maps with satellite-derived vegetation conditions may help identify recent colony abandonment, despite ongoing presence of burrows.

1. Introduction

Monitoring wildlife presence and abundance is a critical and challenging component of extensive land management. In semi-arid ecosystems, such as the rangelands of western North America, southern South America, Australia, central Asia and sub-Saharan Africa, temporally dynamic and spatially heterogeneous conditions make frequent, broad-scale wildlife monitoring essential for effective biodiversity conservation (Durant et al., 2011; Gentle et al., 2018; Ogutu et al., 2016). However, rangeland landscapes are often extensive and difficult to access, which makes repeated pedestrian ground measurements challenging, expensive and prone to bias or error. Rangeland managers are seeking new techniques that can provide frequent, accurate and extensive monitoring coverage to complement traditional ground-based efforts.

The black-tailed prairie dog (*Cynomys ludovicianus*) is a good example of an important and wide-spread wildlife species for which detailed monitoring is both high priority and difficult to accomplish

* Corresponding author at: USDA – ARS, 1701 Centre Ave., Fort Collins, CO 81526 USA.
E-mail address: sean.kearney@usda.gov (S.P. Kearney).

https://doi.org/10.1016/j.ecolind.2023.110684
Received 3 April 2023; Received in revised form 13 July 2023; Accepted 15 July 2023
Available online 22 July 2023

cost-effectively using ground-based methods. Across the North American Great Plains, black-tailed prairie dogs are a keystone species and serve as ecosystem engineers because their burrows provide refugia for small mammals, burrowing owls (*Athene cunicularia*) and reptiles and amphibians (Kretzer and Cully, 2001; Lantz and Conway, 2009), their colonies provide unique breeding habitat for certain bird species (Augustine and Skagen, 2014; Duchardt et al., 2019), and they are an essential prey source for several species, including the endangered black-footed ferret (*Mustela nigripes*) (Duchardt et al., 2023; Roelle and Miller, 2006). Despite their importance for conservation, black-tailed prairie dogs are often viewed as pests due to their competition with livestock for forage (Augustine and Derner, 2021; Crow et al., 2022; Derner et al., 2006) and they are frequently controlled using rodenticides and other methods. In addition to anthropogenic lethal control, black-tailed prairie dogs populations face threats from widespread habitat conversion (Augustine et al., 2021) and frequent epizootics of the non-native sylvatic plague (*Yersinia pestis*) introduced from Europe at the turn of the 20th century (Cully et al., 2010). Thus, a species once widely distributed throughout the North American Great Plains saw a decline in population and distribution of 90–98% from European colonization to the end of the 20th century (Gober, 2000).

In response to concerns about declining populations of black-tailed prairie dogs and associated species, many jurisdictions regularly monitor prairie dog colonies to determine their locations and sizes. Accurate information on colony locations and sizes is also necessary for planning lethal control operations to minimize competition with livestock. Accurate maps are highly desirable to facilitate communication among managers tasked with conserving prairie dogs on the landscape while minimizing competition with livestock and other grazers. Minimum colony area thresholds may be set to trigger plague mitigation, shooting closures or increased monitoring, while maximum area thresholds may also exist to trigger lethal population control when competition with livestock is of concern (e.g., Colorado Parks and Wildlife, 2020; U.S. Forest Service, 2009). Current colony area estimates are often obtained by traversing colony perimeters with a handheld GPS unit or conducting aerial surveys (Colorado Parks and Wildlife, 2020; Davidson et al., 2022; McDonald et al., 2015; Sidle et al., 2012).

In addition to monitoring colony size, rangeland managers would ideally have information related to prairie dog activity and population density across and within colonies. Colonies with higher population densities may experience greater plague transmission (Cully and Williams, 2001) or provide different types of wildlife habitat (Duchardt et al., 2021). Burrow density and temporal history of active burrows are correlated with spatial variation in plant community composition, vegetation structure and bare soil exposure within colonies (Brennan et al., 2020; Duchardt et al., 2019; Duchardt et al., 2018; Duchardt et al., 2021), with implications for forage utilization by large ungulates (Augustine and Derner, 2021; Brennan et al., 2021). From a management perspective, within-colony data on population density could be used to improve the efficacy of both prairie dog population control and plague mitigation measures, and could also inform spatially-explicit forecasts of colony change over time (e.g., Barrile et al., 2023). Such within-colony information is rarely available from existing survey methods.

Remotely sensed images acquired from satellites, occupied aircraft, or unoccupied aerial systems (UAS, i.e., drones) have been widely applied in environmental monitoring and show promise for monitoring prairie dog colonies using manual digitization (McDonald et al., 2015; Sidle et al., 2002) and automated techniques (Delparte et al., 2019). The spatial resolution of imagery likely needs to be fine enough to discern individual burrow entrances (hereafter called burrows) constructed by prairie dogs, typically consisting of a hole surrounded by a mounded patch of bare soil about 1–2 m in diameter. At coarse spatial scales, colonies are characterized by reduced vegetation structure or biomass and increased bare ground exposure associated with burrow mounds and vegetation clipping by individual prairie dogs (e.g., Connell et al., 2018; Duchardt et al., 2019). However, bare ground can be widespread in rangelands due to many other factors, especially at certain times of year or during drought, and in some cases coarse metrics of bare ground exposure or vegetation biomass are not strong indicators of prairie dog presence (e.g., Duchardt et al., 2021). Even at relatively fine spatial scales (e.g., < 5 m), other ground features typified by a circular bare patch, such as anthills, are prevalent in some rangeland landscapes and may look like prairie dog burrows based solely on their spectral reflectance, despite having a very different topographic profile (i.e., mounded hills rather than burrows).

In this study, we addressed research objectives related to supporting prairie dog monitoring programs using remote sensing, with an eye toward developing methodologies for extensive mapping of burrows and colonies in an accurate and repeatable manner. Specifically, we sought to (1) evaluate whether very high-resolution UAS imagery could be used to accurately detect individual black tailed prairie dog burrows, (2) evaluate whether a map of individual burrows could be scaled up to accurately delineate entire colonies and (3) identify the image inputs and minimum spatial resolution needed to achieve acceptable accuracies, since these will have a strong influence on cost and scalability. The approach developed here has direct application at continental scales for the monitoring of small burrowing mammals, but also provides a framework to develop approaches to monitor other spatially extensive, temporally dynamic wildlife populations for which fine-scale landscape features are a good indicator.

We applied state of the art image acquisition and processing technologies and robust validation in this proof-of-concept study. For image acquisition, we used a lightweight fixed-wing UAS equipped with high resolution sensors to acquire imagery and a dense photogrammetric point cloud across a relatively large study area (1,120 ha). We employed deep learning image segmentation techniques to detect individual burrows. Deep neural networks are particularly effective in detecting features within an image based on spatial patterning (Kattenborn et al., 2021), but to our knowledge have not yet been investigated to detect individual prairie dog burrows. We ran training and validation using different input combinations and artificially coarsened imagery to determine how each affected accuracy, and validated colony-scale predictions in two seasons to evaluate the transferability of models. We also evaluated whether estimates of burrow density were related to heterogeneity of colony age (using multi-year ground data of colony boundaries) and vegetation conditions (using satellite-derived maps of vegetation cover and biomass) within prairie dog colonies.

2. Materials and methods

2.1. Study area

This study was carried out at the Central Plains Experimental Range (CPER) in NE Colorado (40.8417, −104.7162), a US Department of Agriculture Long Term Agroecosystem Research (LTAR) site. This study focused on four pastures (named: 5W, 22EW, CN, 29-30) encompassing 1,120 ha. Lethal prairie dog control is not used in these pastures, and they contained active colonies in 2020 and 2021 (Fig. 1). Pasture 5W is co-dominated by a mixture of cool-season (C3) perennial mid-height grasses and warm-season (C4) perennial shortgrasses with a sparse shrubland layer dominated by four-wing saltbush. The colony in pasture 5W contracted dramatically due to a plague epizootic between 2015 and 2016, and then has been expanding since that time. Ground mapping indicated the colony continued expanding from 2020 to 2021. Pastures 22EW and CN are separated by a gravel road, but share similar characteristics, being dominated by C4 perennial shortgrasses. These colonies contracted due to plague between 2014 and 2015, and then expanded from 2015 to 2020. The colony boundaries of the colony spanning 22EW and CN were essentially unchanged between 2020 and 2021, however ground mapping indicated infilling on the north side of pasture CN. Pasture 29–30 is characterized by a mosaic of four-wing saltbush shrublands in the lowlands, C3 perennial mid-height

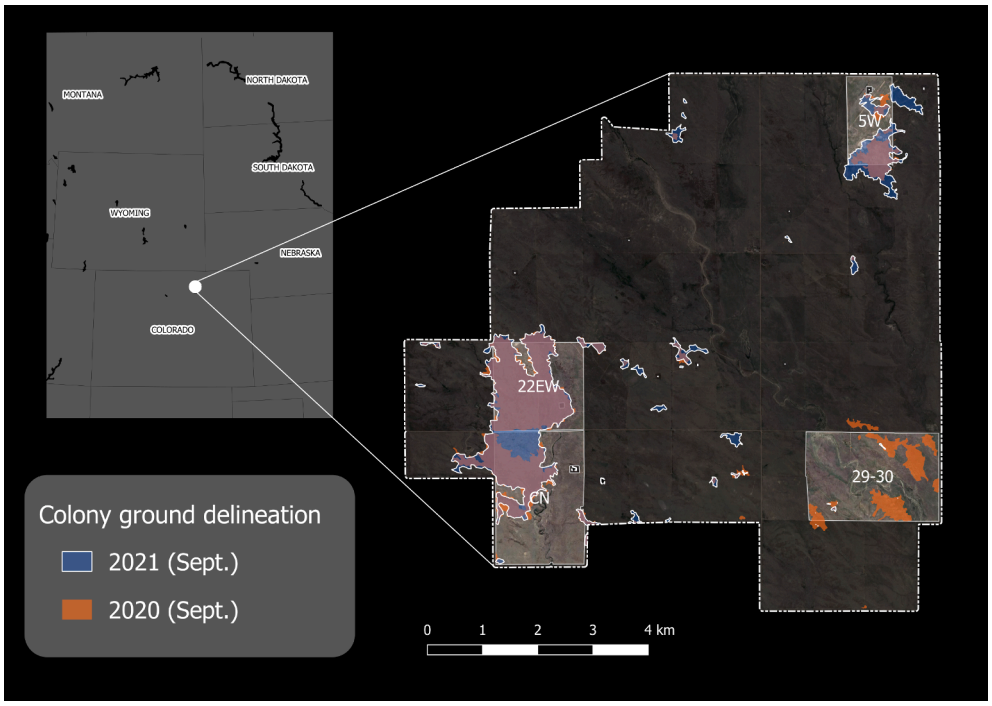


Fig. 1. The four study area pastures and ground delineated prairie dog colonies within the Central Plains Experimental Range (CPER) in NE Colorado. Dashed line shows the extent of CPER, and the four highlighted pastures were flown with a UAS in July and September of 2021. Orange polygons show colony delineation in September 2020 and blue polygons with white outlines show delineation in September 2021. Note that colonies were also delineated in April and July of 2021 (not shown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

grasslands on sandy soils, and C4 perennial shortgrass on ridges and loamy soils. This colony contracted dramatically due to plague in 2017, expanded during 2018–2020, and then contracted dramatically again during the 2021 growing season; by September 2021, less than 1 ha was delineated as active based on ground mapping.

2.2. Overview of methods

We iteratively tested the ability of a deep convolutional neural network (CNN) to detect prairie dog burrows from UAS imagery by using different combinations of UAS image inputs (RGB, NDVI, TPI) at varying spatial resolutions (pixel sizes: 2–30 cm) for training the neural network. We conducted training at the pixel scale using semantic image segmentation and performed independent test validation at multiple scales. We first validated at the scale of individual burrows. Then, we converted predicted burrows to burrow density, which was independently validated at the scale of 30 × 30 m tiles. Finally, we used burrow density to classify entire prairie dog colonies, which we compared to ground-delineated colony perimeters to assess final area estimates and overlap of image-predicted vs ground-measured colonies. A visual overview of our training and validation workflow can be found in Fig. A.1. Additionally, we analyzed heterogeneity within colonies by comparing variability of predicted burrow density to the variability of observed colony age and vegetation.

2.3. Ground data

2.3.1. Colony data

Ground mapping procedures followed methods employed at CPER annually beginning in 2000. Each year in September, all active prairie dog colonies at CPER were delineated using a handheld GPS receiver (Trimble, Sunnyvale, CA). Colony boundaries were delineated based on 1) the locations of burrow entrances and surrounding mounds that showed evidence of recent prairie dog presence (digging–soil disturbance and/or presence of green or brown fecal pellets near the burrow entrance), and 2) sharp transitions between prostrate or recently cropped vegetation and zones of taller vegetation that did not show signs of cropping by prairie dogs. In cases where a colony site contained both inactive and active burrow entrances, the observer walked or drove slowly with an all-terrain vehicle along systematic transects across the colony, with transects spaced at 50-m intervals. The observer examined burrow entrances for activity and used pin flagging to demarcate burrows with signs of activity as well as zones of vegetation height transitions in spaces between the outermost burrows showing signs of activity. With this method, individual burrows that lack signs of recent activity can be (and frequently are) included within the mapped colony boundary because other nearby burrows show signs of activity and/or other signs (vegetation clipping, scat) indicating the use of the area by prairie dogs.

For the purpose of this study, we utilized existing ground-based surveys of September colony boundaries from 2016 to 2020 (see also Augustine and Derner, 2021) and collected new data in 2021. During the 2021 study year, we surveyed colony boundaries in April and July, in addition to September, resulting in three separate colony boundary maps. The additional surveys were conducted to better understand change in colony activity throughout the season and enable validating our models in multiple seasons across varying vegetation conditions.

2.3.2. Plot data

During the summer of 2021, we established 11 plots (50 m × 50 m) within the study area where we geolocated individual active prairie dog burrows and burrow-like feature (e.g., anthills, badger burrows, old and inactive prairie dog burrows) using a high precision GPS unit capable of sub-meter accuracy (99% of positions). After differential correction using Global Navigation Satellite Systems (GNSS) and nearby base stations, 72% of GPS positions were accurate to within 30 cm and 45% of positions to within 15 cm. For all active burrows, we also ranked (0–3) the size, height, activity and degree of vegetation cover associated with each geolocated burrow. These ground data served three objectives. First, they provided reference points to train personnel for manual digitization of burrows from the imagery, which served as the training data for the deep neural network (see below). Second, a holdout of these ground data provided a validation dataset for assessing burrow-scale accuracy of both the manual digitization and the neural network output. Third, the holdout ground data allowed us to assess the rate of false positives associated with burrow-like features (e.g., anthills) and

the characteristics (e.g., burrow size, activity level) of false negatives.

2.3.3. UAS data acquisition and pre-processing

Images were acquired by the US Forest Service's Geospatial Technology and Applications Center (GTAC) in late July and early September 2021. The two acquisitions used a Trinity F90+ fixed-wing UAS equipped with a downwelling light sensor. During each acquisition, the entire area was covered twice, once with a Sony RX1 RII 42 MP camera (3-band RGB) and once with a Micasense RedEdge MX and MX Blue dual camera system (10-band multispectral). Flight altitude was limited to a maximum of 400 ft (122 m), resulting in a spatial resolution of ~1.7 cm for the RGB product and ~ 7 cm for the multispectral product. Georectification was performed in the application QBase using location data from the UAS platform corrected with data from a nearby station of the NOAA Continuously Operating Reference Stations (CORS) Network. Image control points were used for mosaicking multiple datasets. Visual observation of burrow centers in the imagery tended to be within ~30 cm of the GPS-located burrow centers, indicating that orthoimage accuracy was high. Agisoft Metashape software was used for spectral calibration, orthomosaicking and photogrammetric point cloud generation. The RGB flights had a 75% endlap and 70% sidelap, which produced a point cloud with about 220 points m^{-2}, which was used to generate a smoothed 3D mesh, and subsequently converted to a 6 cm digital surface model (DSM).

From the DSM, we calculated a topographic position index (TPI) for two reasons: (1) to standardize these data for input into the deep neural network and (2) to highlight the burrow features, typified by a 10–30 cm hole surrounded by a mound up to 2.5 m in diameter (Cincotta, 1989). We created the TPI by applying a moving window function using the Xarray-spatial package in Python. Specifically, for each pixel, we took the pixel's elevation and from it subtracted the mean elevation of pixels in an annulus (i.e., doughnut) around the pixel. We set the annulus to have a minimum radius of 25 cm and a maximum radius of 75 cm, with the intention of capturing the mound surrounding the burrow. Therefore, negative TPI values indicate the surrounding elevation is higher than that of the pixel (i.e., the pixel is a local low point), positive values indicate the surrounding elevation is lower than that of the pixel (i.e., the pixel is a local high point) and values near zero indicate the pixel is in a locally flat area.

We calculated the normalized difference vegetation index (NDVI) from the multispectral imagery as

$$\frac{NIR - Red}{NIR + Red}$$

using bands from the red Micasense camera, where *NIR* is the near-infrared band centered at 842 nm, and *Red* is the red band of the visible spectra centered at 668 nm. NDVI ranges from −1.0 to 1.0, with high values capturing green vegetation and low values indicative of bare soil. Negative values are rare, so we reset all negative values to zero, resulting in a possible range of 0.0 to 1.0. We expected the NDVI to help highlight bare soil associated with active burrows and to help separate out old, inactive burrows that may have experienced recent revegetation.

2.4. Burrow detection

2.4.1. Data preparation for training and validation

To create a robust training and validation dataset, we manually digitized individual burrows by drawing polygons around burrows visible in the September imagery. We first created 30 × 30 m image tiles covering all the ground plots (n = 45 tiles) with geolocated burrows as well as a random sample of tiles generated at a rate of one per 10 ha across the study area (n = 110). We then digitized all burrows visible in 50% of the tiles covering the ground data (n = 23) and in all the randomly generated tiles, resulting in a total of 133 tiles. Of the randomly assigned tiles, we kept 80% (n = 88 tiles) for training and set aside the other 20% (n = 22), along with the digitized tiles covering the ground data, for testing (i.e., independent validation; n = 45 tiles). To train ourselves on how burrows and non-burrow features (e.g., anthills) appeared in the RGB, NDVI and TPI images, at the start of digitization we inspected approximately 25% of the tiles co-located with ground plots, with the ability to also visualize the ground data overlayed on the image layers. This left the remaining digitized tiles co-located with ground data (n = 13 tiles) available for independent ground validation of our manual digitization.

For all tiles, we extracted smaller image windows to serve as the final training and validation images for the neural network. This was done to keep image sizes computationally reasonable (i.e., 320 × 320 pixels or smaller) and to allow for random image augmentation (e.g., flipping/mirroring, blurring/sharpening), which can improve the performance and generalization of neural networks (Kattenborn et al., 2021). The size of the windows depended on the resolution of the imagery created during the artificial coarsening procedure (see Table A.1). One window was extracted for each digitized burrow present in each tile, along with five additional windows extracted for random locations within the tile. This process resulted in a total of 641 training images and 455 testing images.

All windowed image values were rescaled to the range 0–1 prior to training. The RGB image values were rescaled using the data range 0–255, the TPI values were rescaled using the data range −0.10–0.40, and the NDVI values already had a data range of 0–1, and thus did not require rescaling.

Image preparation was repeated at five spatial resolutions: 2, 5, 10, 15, and 30 cm. We resampled all image inputs to each resolution using the nearest neighbor technique. For the TPI, we first resampled the DEM and then recalculated TPI at the new resolution.

2.4.2. Training the deep neural network

Our neural network was created using the Segmentation Models Pytorch package (Iakubovskii, 2019) and used the DeepLabV3+ architecture with a Resnet34 encoder initialized with pretrained weights using the imagenet dataset. We also tried the Unet++, FPN and MAnet architectures in early training and validation tests with very similar results but chose to use only the DeepLabV3+ for full analysis for simplicity. We used the Mathew's Correlation Coefficient loss function (Abhishek and Hamarneh, 2021), a sigmoid activation function and stochastic gradient descent optimization with a triangular cyclic learning rate. The learning rate cycled between a base rate of 0.025 and an upper limit of 0.10, completing a full cycle back to the base rate every 4 epochs (one epoch is a complete iteration of all training images). We determined the learning rate range by running initial tests at a range of learning rates (1.0×10^{-5} to 0.5×10^{-1}) with all inputs (RGB, TPI, NDVI) and setting the base learning rate to the value where model accuracy first increased and the upper learning rate to the value where accuracy leveled off (Smith, 2015).

Out of the 641 training images, we randomly held out 20% (n = 129) for training validation. We fed the remaining images (n = 512) into the neural network in batches of 8. The images were randomly augmented using the albumentations package (Buslaev et al., 2020). Augmentations included cropping, flipping, mirroring, blurring, motion blurring and sharpening (see Table A.2). At the end of each learning rate cycle (i.e., every 4 epochs), the updated model was validated by predicting burrows in the held-out training data and calculating the F-score – our measure of accuracy – at the burrow scale, based on whether the centroid of the predicted burrow was within the digitized burrow polygon. The F-score can range between 0.0 and 1.0 and balances precision and recall, calculated as:

$$Fscore = 2^{*}\frac{precision^{*}recall}{precision + recall}$$

If the F-score improved by at least 0.005 over the previous learning rate cycle, we saved the model. We ran the training for a minimum of 2 learning rate cycles and a maximum of 16 learning cycles (i.e., 8–64 epochs), however if no improvement in F-score was observed for 2 consecutive learning rate cycles (i.e., 8 epochs), we stopped the training early and used the last saved model for all subsequent analysis.

2.4.3. Independent validation

We validated all the saved models at the scale of burrows and tiles using the independent testing data. At the burrow scale, we calculated precision, recall and the F-score based on whether the centroid of the predicted burrow was within the digitized burrow polygon. At the tile scale, we assessed the Pearson correlation and mean percent error (MPE) between predicted and observed burrow density.

2.5. Colony classification and comparison to ground delineation

We applied each model to the entire study area and extracted the centroid to generate a map of individual burrows. To scale the individual burrow predictions to delineate colonies, we used an approach similar to the ground-based method used to delineate colonies with handheld GPS units. First, we generated a 5-m grid over the study area and assigned to each grid cell the sum of burrows within a 25 m radius of each grid-cell center. Then, we created a binary grid of all cells with a value greater than 5, thus representing all locations in the study area with more than 5 burrows within 50 m of each other. To identify broad groupings of cells that meet this criteria, which we expect to represent active colonies, we adapted a method developed by Riitters et al. (2002) to classify forest connectivity. We calculated the density and connectivity of cells in the binary grid using a 11 × 11 cell (55 × 55 m) moving window. We then classified grid cells into five classes: Core (density $\geq$ 0.90), Perforated Core (0.60 $\geq$ density $<$ 0.90 and density $\geq$ connectivity), Edge (0.60 $\geq$ density $<$ 0.90 and density $<$ connectivity), Transition (0.40 $\leq$ density $<$ 0.60) and Non-colony (density $<$ 0.40). Additionally, any grid cells completely surrounded by Core or Perforated Core were assigned to the Core classes, and any remaining grid cells completely surrounded by any combination of Core, Edge or Transition classes were assigned to the Transition class. We expected the Core classes to be the most likely to be part of an active colony, the Edge class to be uncertain and the Transition classes to be least likely.

We compared the model-derived colony classification to the ground-based colony delineation for each of the four pastures in the study area. This was done based on total hectares and the Jaccard score, which is calculated as the intersection of predicted and observed colonies divided by the union of the two (i.e., IOU score), and ranges between 0.0 (no overlap) and 1.0 (perfect overlap). We did this using the September imagery and ground data, from which the models were derived, and for the July imagery and ground data. The latter gave us an indication of how well a model might perform when vegetation conditions are different, in this case earlier in the growing season when vegetation is greener and cover and biomass are greater.

2.6. Within-colony heterogeneity

2.6.1. Colony age

We evaluated whether burrow density was related to the age of colonies using historical ground delineation of colonies from 2015 to 2021. For each grid cell within colonies delineated as active in 2021, we summed the number of consecutive years since 2015 the grid cell had been mapped as active. If the grid cell had not been delineated as active in 2021, but had been delineated as active in 2020, it was given a value of –1. If the grid cell had not been delineated in 2021 or 2020, it was assigned N/A (not active). We then created boxplots of model-predicted burrow density within each age class.

2.6.2. Vegetation

We evaluated whether burrow density was related to ground vegetation using satellite-derived estimates of standing biomass and bare ground cover. The satellite-derived estimates were created for a separate study using the Harmonized Landsat-Sentinel (HLS) dataset, produced at 30 m spatial resolution and interpolated at daily temporal resolution (see Kearney et al., 2022). We extracted the average burrow density within each 30 m HLS pixel and compared it to average biomass and bare ground values within 15 days of the UAV flights, as well as the 50-day change in biomass and bare ground leading up to the UAV flight. All biomass and bare ground values were converted to z-scores by subtracting the mean and dividing by the standard deviation of the region (i.e., the entire CPER area). We then created scatterplots of burrow density versus each metric, analyzed their linear relationships and plotted the mean and standard deviation within each predicted colony class (i.e., Core, Edge, Transitional and Not active).

3. Results

3.1. Burrow detection

3.1.1. Fine-scale (2 cm) resolution

At the finest resolution of 2 cm, training validation performed well (burrow-scale F-score 0.84–0.93) for all the input combinations we tested, except when using only NDVI (F-score 0.55; Table A.3). We therefore did not consider the NDVI-only model in further analysis. For independent testing validation at the 2 cm resolution, the model using only TPI performed best (F-score 0.87) and including additional inputs along with TPI did not improve the F-score (Table 1). Performance was notably lower in models that did not include TPI; the RGB and RGB + NDVI models had F-scores of 0.73 and 0.66, respectively.

The ground dataset of geolocated burrows (n = 273) showed that at 2-cm resolution, omitted burrows (false negatives) tended to be smaller, less active and more vegetated compared to detected burrows (true positives) regardless of the inputs used (Fig. 2). The detected burrows were marginally larger, more active and less vegetated than average, and this trend was slightly more pronounced for models that did not include TPI. The ground dataset of non-burrow features (n = 108) showed that commission error (false positives) of anthills was lowest for TPI, however the models with both RGB and NDVI tended to have lower commission error for most, but not all, of the other burrow-like features (Table 2).

Burrow densities predicted by models were strongly and significantly correlated to the density of manually digitized burrows for the 30 × 30 m validation tiles (Fig. 3). Again, models that included TPI performed best, with correlation coefficients of 0.94–0.95 and low bias (MPE = −0.01–0.13). The RGB model had a correlation coefficient of 0.88 and MPE of 0.03. The RGB + NDVI model had a strong correlation (r = 0.84) but tended to overpredict burrow density (MPE = 0.46).

We were able to compare the validation of predictions to a validation of the manual digitization for a small subset of ground geolocated burrows (n = 47) that overlapped the manually digitized 30 × 30 m tiles (n = 13). We found that accuracy of manual digitization was high, and that 2 cm models with TPI slightly outperformed manual digitization. For the

Table 1
Independent burrow-scale validation of 2 cm resolution models using the test images derived from the 20% hold-out tiles and tiles covering ground data (n = 455 images).

Inputs	Precision	Recall	F-score
TPI	0.90	0.84	0.87
RGB + TPI	0.86	0.84	0.85
RGB + TPI + NDVI	0.88	0.81	0.84
RGB	0.70	0.75	0.73
RGB + NDVI	0.59	0.76	0.66

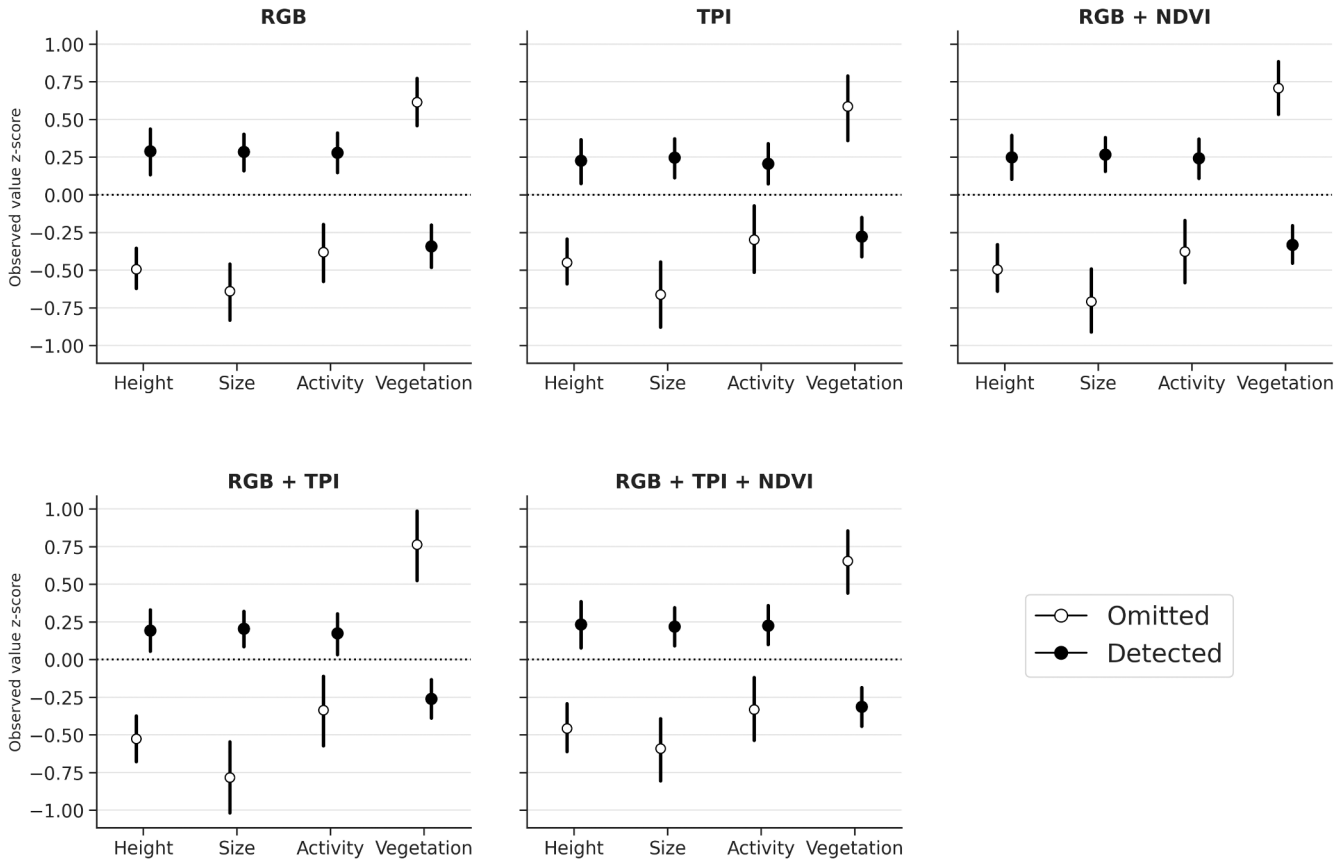


Fig. 2. Patterns of omission error (false negatives) for 2 cm resolution models compared to detected burrows (true positives) for ground-mapped active burrows (n = 273). Y-axis represents the z-score (standard deviations away from the mean) of the group compared to the entire dataset.

Table 2
Commission error (false-positive) rates of 2 cm resolution models for ground-mapped features expected to appear similar to active burrows.

Feature type	n	RGB	TPI	RGB + NDVI	RGB + TPI	RGB + TPI + NDVI
Anthill	67	0.07	0.01	0.18	0.03	0.03
Den	8	0.38	0.50	0.12	0.50	0.12
Digging	6	0.33	0.17	0.17	0.17	0.17
Old burrow	27	0.07	0.11	0.07	0.11	0.15
Overall	108	0.06	0.05	0.08	0.05	0.04

47 burrows, manual digitization achieved an F-score of 0.77, while the predictions from models that included TPI had F-scores of 0.78–0.82, and models without TPI had F-scores of 0.67–0.69 (Table A.4). The correlation between burrow density derived from manual digitization vs ground geolocation for the 13 tiles was 0.93 (Fig. A.2), whereas for the model predictions vs. ground geolocation, correlation ranged from a low of 0.84 for the RGB model to 0.94 for the RGB + TPI model (Fig. A.3).

3.1.2. Coarser resolutions

When imagery was coarsened to 5 cm, test validation performance was little changed compared to 2 cm. Performance improved slightly for models that included NDVI as an input, remained the same for models with TPI as an input, and declined slightly for the RGB only model (Fig. 4; Table A.5). In general, performance began to decline slightly at 10 cm resolution and decreased more sharply when imagery was coarsened to 15 and 30 cm, though the rate of decline varied across models. At 10 cm, the three models with TPI continued to perform best with F-scores of 0.77–0.79 (Fig. 4a) and burrow density correlations of 0.88–0.93 (Fig. 4b). The F-scores and burrow density correlations of the RGB and RGB + NDVI models fell below 0.70 and 0.80, respectively, at 10 cm. At 15 cm the RGB + TPI model performed best (F-score = 0.67). It was the only model with a burrow density correlation greater than 0.80 (r = 0.85) and recall greater than 0.50 (recall = 0.53). Precision remained moderate or high for all models at 15 cm (0.68–0.92), indicating that, up to 15 cm image resolution, models were not spurious, rather they were simply omitting more burrows. However, at 30 m resolution, performance degraded for all models as a result of both poor recall and poor precision (Table A.5), indicating more spurious performance.

3.2. Colony delineation

3.2.1. Prediction for season of model training (September)

We focused our colony-scale evaluation on models using the 5, 10 and 15 cm imagery. This was done for two reasons. First, predicting burrows across the entire study using the 2 cm image inputs proved computationally challenging, largely due to the processing time required to generate the TPI layer. Second, since burrow detection of the 30 cm imagery was poor, we did not consider it further at the colony scale. We also dropped the RGB + NDVI model since burrow detection performance tended to be poorer compared to the RGB only model.

The areas predicted as Core colony (including Perforated Core) from the 5 cm models had moderate to high overlap with the ground delineation of active colony for the three pastures with large active colonies in September (22EW, CN, 5 W), and overlap generally varied more by pasture than by model (Figs. 5–7; Fig. A.4). Of the three active pastures, the highest Jaccard scores were in pasture 22EW (0.89–0.94) and the lowest in pasture 5 W (0.58–0.61; Fig. A.4). For pasture CN, the three models with TPI had Jaccard scores of 0.82–0.85 and the RGB model had a score of 0.55 (Fig. A.4). For these three pastures combined, the RGB

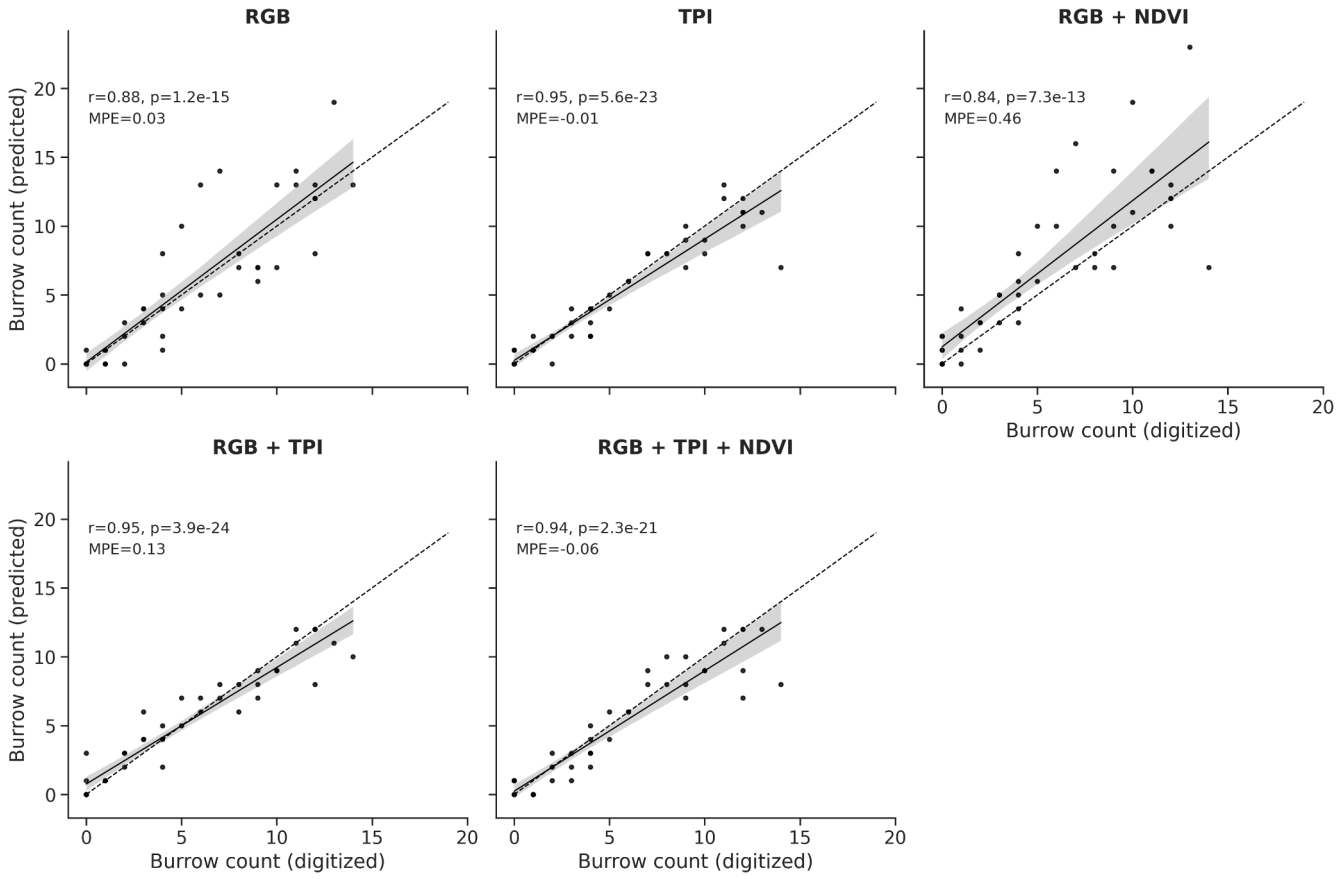


Fig. 3. Correlation between model predicted and manually digitized (observed) burrow densities at the scale of 30 × 30 m testing tiles (n = 45). 'r' is the Pearson correlation coefficient, 'p' is the alpha (significance) of the correlation and 'MPE' is the mean percent error of predictions, an indication of bias.

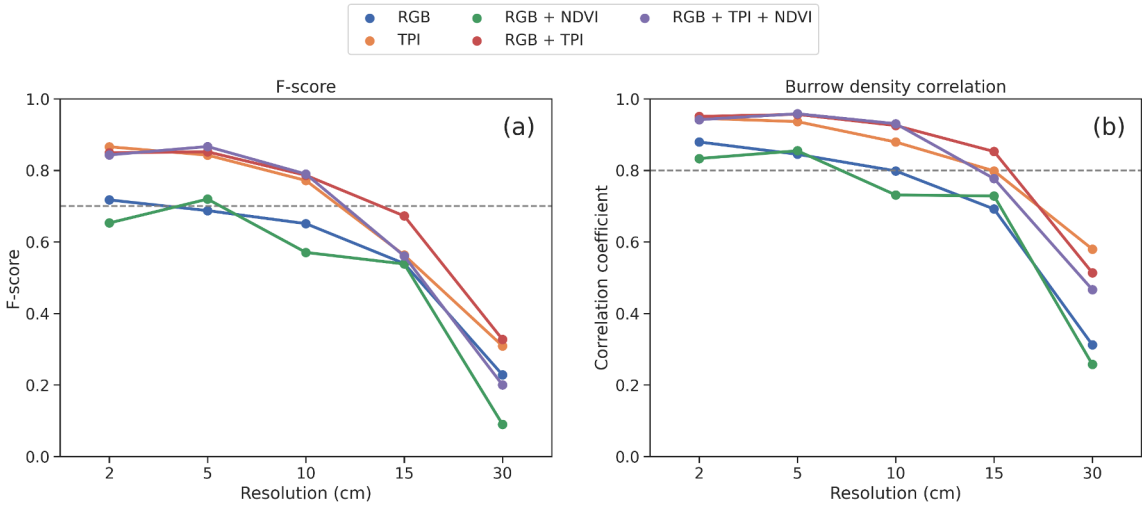


Fig. 4. Model test validation at increasingly coarse spatial resolution of imagery for each combination of model inputs. F-score (a) is calculated from individual burrows across all validation images and burrow density correlation (b) is calculated from the 30 × 30 m validation tiles.

model had the closest predicted total Core colony area compared to the total ground delineated active colonies. However, this was due to relatively large offsetting differences in different colonies (Table A.6), and when we looked at Jaccard score and the relative absolute difference (which accounts for colony area), we found that RGB + TPI and RGB + TPI + NDVI models performed best across the active pastures, with pasture-scale absolute differences relative to the ground delineation of 11%, compared to a difference of 15% for the TPI model and 24% for the RGB model (Table A.7).

In pasture 29–30, where colonies experienced a plague epizootic during 2021, ground surveys delineated only 0.70 ha as active colony in September of 2021. Overlap between model-predicted Core colony and ground delineated active colony was essentially zero for all 5 cm models, which predicted between 37.25 ha (RGB) and 120 ha (TPI) of Core colony (Table A.6). In this case, models that included TPI had lower performance due to detection of recently uninhabited burrows. Fig. 5

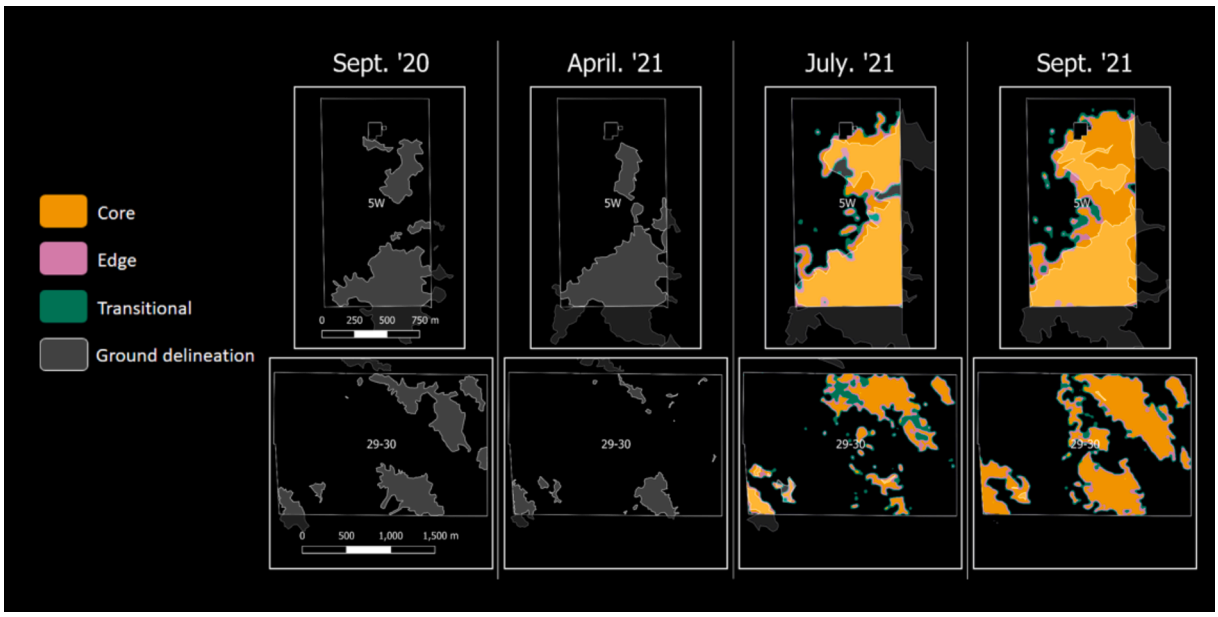


Fig. 5. Ground delineation (grey with white outline) from all colony surveys in 2020 and 2021 for pastures 5 W and 29–30 compared with predicted colony classes from the 5-cm RGB + TPI model for July and September of 2021. Note that these pastures are not spatially adjacent and are each shown at different scales (scale bar in left-most panel for each pasture).

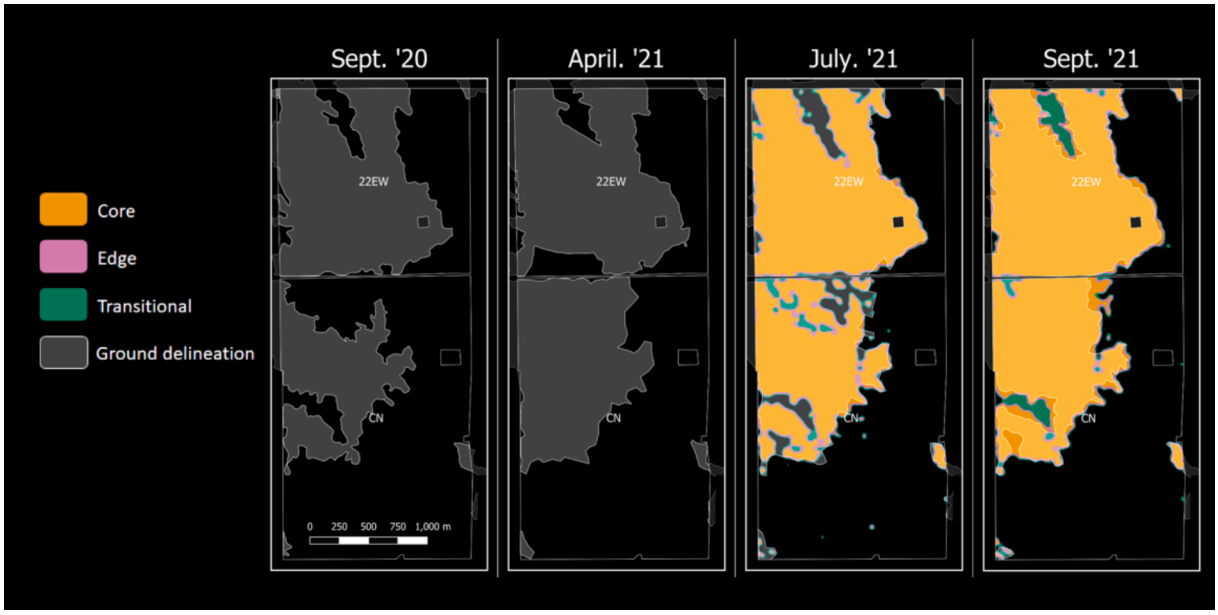


Fig. 6. Ground delineation (grey with white outline) from all colony surveys in 2020 and 2021 for pastures 22EW and CN compared with predicted colony classes from the 5-cm RGB + TPI model for July and September of 2021. Note that these pastures are spatially adjacent, separated by a gravel road.

shows that much of the area predicted as Core colony by the RGB + TPI model corresponds to areas delineated as active in September 2020 but not active due to plague throughout 2021. We note that results were similar for the other models. When evaluated across all four pastures, absolute differences relative to the ground delineation were: RGB (35%), RGB + TPI (37%), RGB + TPI + NDVI (40%), TPI (50%) (Table A.7).

At 10 cm resolution, Jaccard scores only changed marginally for most models and pastures, with some increasing slightly, some decreasing slightly and some remaining essentially unchanged (Fig. A.4). However, the area predicted as Core colony in pasture 29–30 was much closer to zero for the three models that included TPI (Table A.6). This improvement, combined with only a marginal reduction in performance for the active pastures, resulted in an improvement of the overall Jaccard score when evaluated across all pastures for the three models with TPI (Table A.6). At 15 cm, Jaccard scores decreased for most models and pastures, although to a lesser degree for the TPI only and RGB + TPI models (Fig. A.4).

3.2.2. *Prediction in a new season (July)*

When we applied models derived from the September data to the imagery collected in July (a season for which the model was not trained) overlap of model-predicted Core colony and ground delineated active colony was similar to what we observed for September. All 5 cm models had moderate to high overlap (0.66–0.88) with the ground delineation of active colony for the three pastures with large active colonies, except for the RGB model in pasture CN (0.25), which predicted substantially

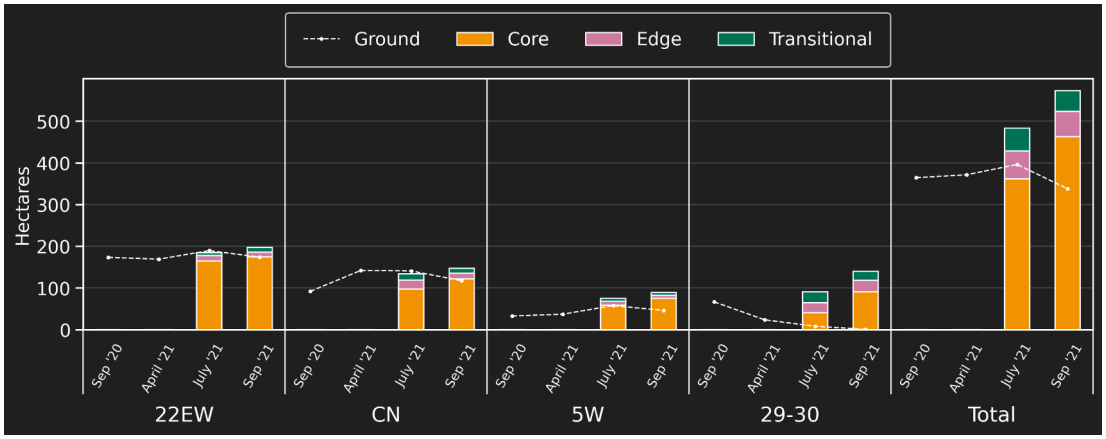


Fig. 7. Trends in 2020–2021 ground delineated area of active prairie dog colony (dotted lines) compared to the model-predicted area from the 5-cm RGB + TPI model in the three potential colony classes (bars) for July and September of 2021, the two months for which UAS imagery was available.

less area as Core colony compared to ground delineation (Figs. 5–7; Figs. A.4-5). As with September, the models that included TPI tended to have higher overlap than the RGB model for July. The three models with TPI had absolute relative differences in area of 11–18 % compared to the ground delineation in the active pastures, while the RGB model had a difference of 38% (Table A.7).

Overlap was slightly higher in plagued pasture 29–30 (0.08–0.22) for the July predictions compared to September. This resulted in a slightly lower overall absolute relative difference in colony area across all pastures, ranging from 20% for the RBG + TPI + NDVI model to 39% for the RGB model (Table A.7).

At coarser spatial resolutions, Jaccard scores dropped off more quickly for July compared to September, and were below 0.70 in all pastures (Fig. A.4). At 10 cm, overall absolute relative differences in area were 45–88%, and at 15 cm they were 46–97% (Table A.7).

3.3. Burrow density and colony heterogeneity

To assess heterogeneity of burrow density and its relationship to colony age and vegetation metrics, we used burrow density predictions from the 5 cm RGB + TPI model, shown in Fig. 8. We chose this model since it performed well in our burrow-scale evaluation for September and at the colony scale for both seasons.

For a given age class, model-predicted burrow density was generally higher for September compared to July (Fig. 9). Burrow density tended to increase with colony age for both July and September. Regardless of history, areas delineated as active in September of either 2020 or 2021 contained 11-fold to 20-fold greater predicted burrow density than areas not mapped as active in either 2020 or 2021. We found little difference in burrow density between the areas delineated as active in 2021 but not in 2020 (colony age of 1 yr.) compared to areas delineated as active in 2020 but not in 2021 (i.e., colony age of −1 yr.), especially for September, but a clearer difference between the areas with a colony age of −1 versus ≥2 years (Fig. 9).

Burrow density was negatively correlated with standardized biomass for July (Pearson r = −0.43) and September (r = −0.46) and positively correlated with bare ground exposure for both seasons, respectively (r = 0.27 and r = 0.41; Fig. 10). Burrow density was also negatively correlated with the change in biomass between July and September (r = −0.27) and positively correlated with the change in bare ground cover

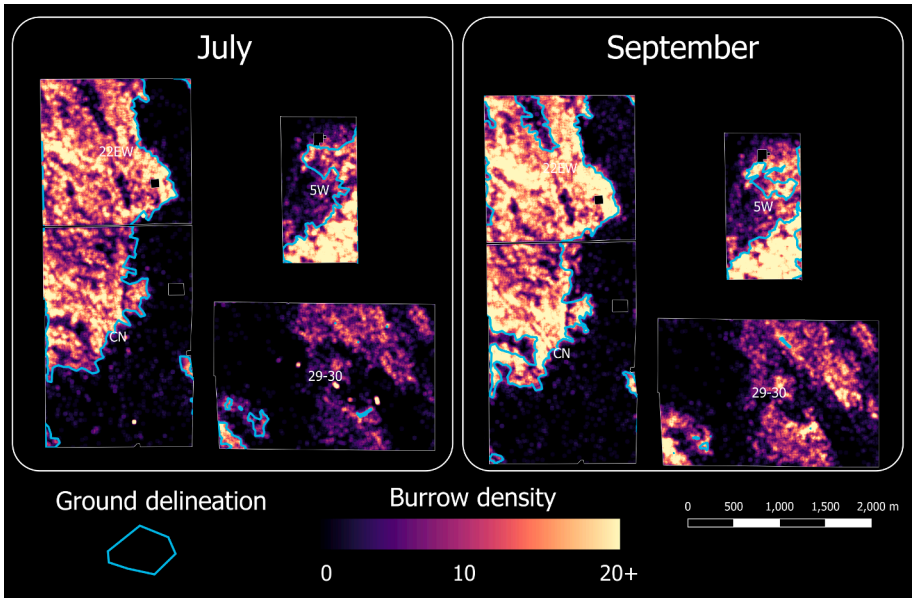


Fig. 8. Maps of model-predicted prairie dog burrow density (heatmap, warmer colors indicate higher density) overlayed with ground delineated boundary of active colonies (light blue polygons). Note that not all pastures are actually spatially adjacent (see Fig. 1) but are all shown at the same scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

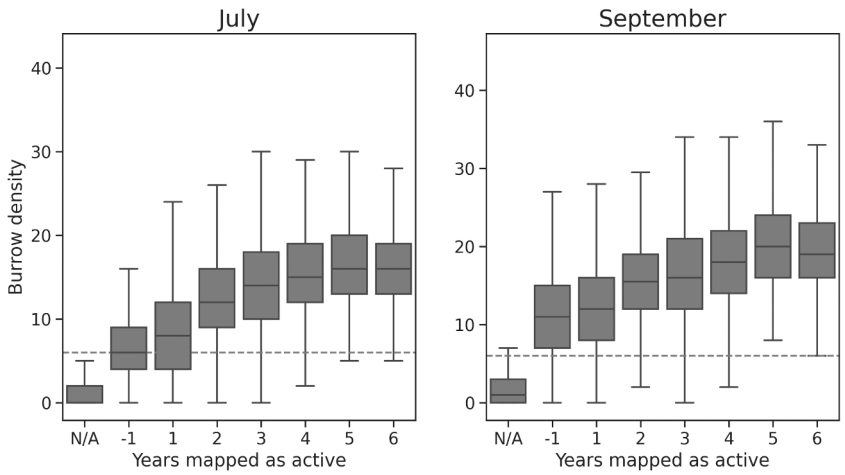


Fig. 9. Variation in September 2021 model-predicted prairie dog burrow density (within 25 m radius) compared to the number of years an area has been mapped as an active colony using ground delineation. "N/A" refers to an area that was not mapped as active in 2020 or 2021, and "−1" refers to areas mapped as active in 2020 but not in 2021. The dotted line shows a burrow density of 6, the threshold used in the model-predicted colony delineation procedure.

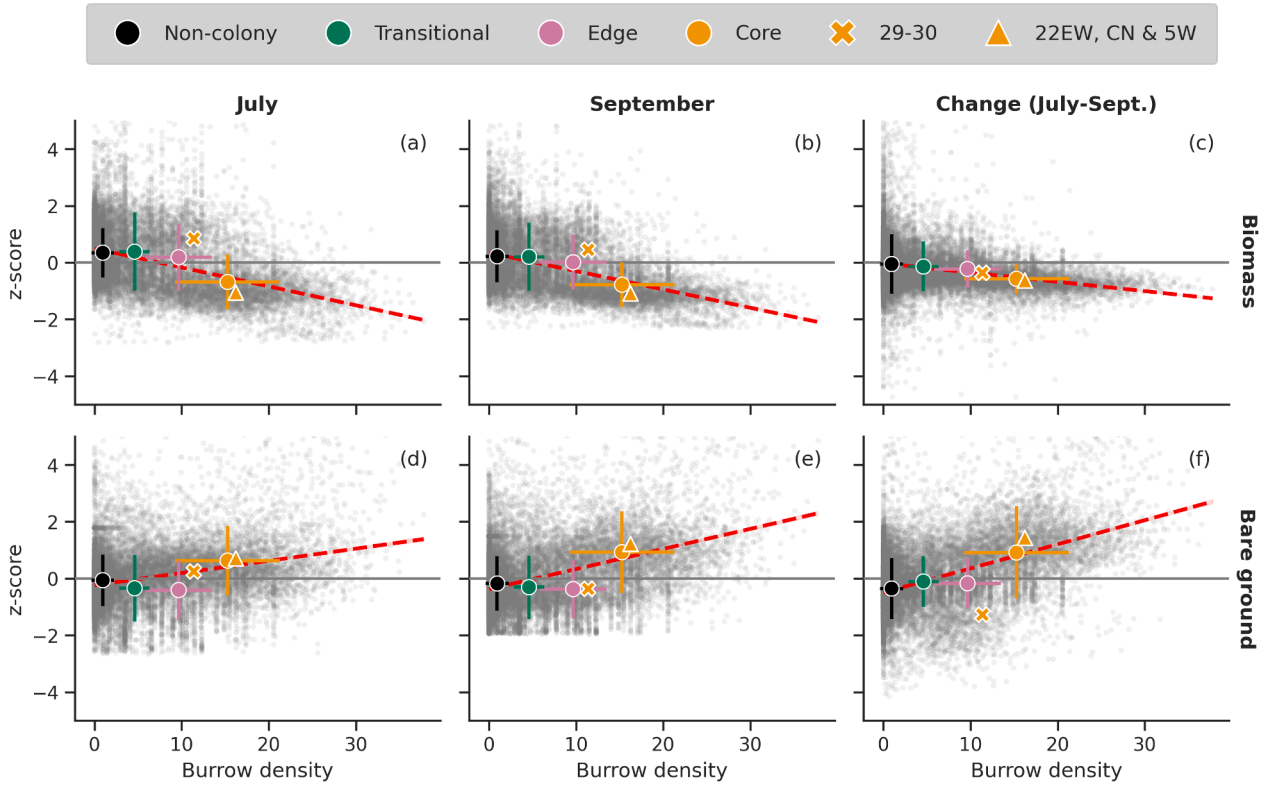


Fig. 10. Scatterplots of burrow density (within a 25 m radius) and satellite-derived metrics for July, September and the slope of the change in metrics over the July-September period. Grey dots represent individual 30 m pixels for the entire Central Plains Experimental Range (CPER) and are z-score standardized by subtracting the mean and dividing by the standard deviation of the entire CPER. Red dotted lines show the linear relationship for all CPER for each variable pair. Means (circles) and standard deviations (lines) are shown for the four model prediction classes (Non-colony, Transitional, Edge, Core) across the four study pastures. Means (symbols) are also shown separately for the Core areas in pasture 29–30, which experienced plague in 2020/21 ('X' symbol), and for the other three pastures with large, active colonies ('Δ' symbol for mean of 22EW, CN, 5 W). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(r = 0.46). Across all pastures, areas predicted as Core colony had higher burrow density, lower biomass and higher bare ground cover compared to areas not predicted to be part of a colony (Fig. 10). For areas predicted as Edge or Transitional, burrow density was higher than non-colony areas, but biomass and bare ground cover were similar to non-colony areas (and to the overall mean of the region). For pasture 29–30, which experienced plague during the 2020–2021 period, areas predicted as Core colony had lower burrow density, higher biomass, and less bare ground cover than Core colony areas for the three active pastures. Moreover, the change in bare ground cover from July to September was markedly lower for pasture 29–30 than for the three active pastures or the regional mean (Fig. 10f). It was also much lower than would be expected based on the linear trend between burrow density and July-September bare ground cover change.

4. Discussion

4.1. Burrow detection and colony classification

Overall, agreement between burrow-scale and colony-scale model predictions and the corresponding validation datasets were high. However, the ground-based validation datasets have their own sources of uncertainty and no one dataset should be considered perfect truth. For example, accuracy of the ground-based mapping method utilized here has not been evaluated, and mapping of colony boundaries could vary based on the field personnel's interpretation of whether a particular burrow is active or inactive.

At the burrow scale, our results show that, for this study area, we can expect around 80–85% harmonic accuracy (F-score) of burrow detection using a deep CNN model trained with a moderately sized digitization dataset from that same imagery. In fact, our results suggest that models may even slightly outperform manual digitization, even when trained with that same imperfectly digitized data (e.g., Table A.4, Figs. A.2-3). Few studies have attempted to detect individual burrows of ground-dwelling mammals from remotely sensed imagery, and, to our knowledge, this is the first attempt to use a CNN for this purpose. We reiterate here that individual burrow detection was not the primary objective of this study, but rather a means to an end (the 'end' being colony-scale delineation), capitalizing on the ability of CNN models to detect multi-pixeled objects with clear spatial patterns. Pixel-based approaches have generally had very poor accuracy due to the strong spectral similarity of burrows to other land cover typified by bare ground cover (e.g., Folluo, 2019; Jost, 2018). A few studies have achieved promising results using object-based image analysis (e.g., Delparte et al., 2019; Folluo, 2019), however these approaches tend to require licensed software and rely on computationally expensive object delineation followed by developing a ruleset to classify objects, both of which require parameterization and may have limited transferability.

At the colony scale, we observed high overlap (60–94%) in September between model predictions and ground mapping for the three pastures with active colonies. It was encouraging that, for these three pastures, overlap remained high for the July (68–86%) predictions, for which no training data were used. This suggests that well-trained models may be robust across seasons and changing vegetation conditions for active colonies. However, further testing across different soil types, plant communities and vegetation conditions (e.g., drought) are needed to determine model transferability over space and time, or to establish if there are ideal vegetation conditions under which to acquire imagery for mapping burrows and colonies. Creating a larger training dataset, perhaps over multiple seasons, may also improve model consistency; our training dataset (n = 512) was relatively small by CNN standards. Additionally, it is possible that more robust techniques to scale from individual burrow detection to colony delineation can be developed.

Our results suggest that identifying recent declines in colony activity (e.g., from plague, lethal control) using image-based burrow detection alone will prove challenging (also see Sidle et al., 2012). Ground assessment of colony activity relies on a variety of indicators beyond burrow presence that are likely impractical to identify from remotely sensed imagery, such as presence of prairie dogs, feces, tracks, digging, cobwebs in burrow entrances and more. Recently uninhabited burrows are likely to persist on the landscape for longer than these other indicators after colony activity ceases. In this study, we were only able to assess results for a plague event that occurred within the same growing season as image acquisition. It is unclear how long the algorithm will continue to detect uninhabited burrows, and this timeframe will likely vary depending on soil type and the amount and intensity of rainfall since colony collapse. Fortunately, our analyses showed that near-real time satellite monitoring of vegetation conditions may help to identify recently uninhabited colonies. For the portion of the recently plague-affected colony in pasture 29–30 that UAS imagery identified as 'Core colony', we found higher vegetation biomass, less bare ground cover and a lower slope of change over time in bare ground cover than expected given the burrow density of that area (Fig. 10). Typifying the relationships between burrow density, vegetation metrics and known active colonies for a given region may enable early detection of colony abandonment, even when burrows persist on the landscape, and provide an indicator of colonies in need of field verification for prairie dog activity. More research is needed to better understand how long it takes for uninhabited burrows to no longer be detected by the CNN, and the degree to which this depends on vegetation conditions and precipitation patterns, as well as whether robust methods can be developed to flag recently uninhabited colonies using satellite imagery time-series.

4.2. Burrow density and colony heterogeneity

It is clear from the September imagery that predicted burrow density is strongly correlated with burrow density derived from manual digitization (Fig. 3) and observed on the ground (Fig. A.3). We are less certain about the accuracy of burrow density predictions when the model was applied to July imagery, since we did not perform burrow-scale validation for July. We observed an increase in predicted burrow density between July and September, which could be the result of (a) new burrow construction, (b) higher detection rates of burrows later in the season when vegetation cover is declining, and/or (c) poorer detection accuracy in July since the model was trained for September conditions (e.g., with less green vegetation). If new burrows are being constructed at a rate faster than uninhabited burrows are demolished (i.e., (a) above), or if we expect to detect more inactive burrows when vegetation cover is low (i.e., (b) above; also see previous section), this could make monitoring active burrow density change or active colony expansion and contraction challenging over short periods of time, for example within a single growing season.

The fact that burrow density increased in the plague-affected colony suggests that causes (b) and (c) above are both contributing factors. Additional testing using data from this study, for example developing a new model for July or with July and September imagery combined – would help to better understand the relative contributions of (b) vs (c). Repeating this study across multiple years and seasons to capture a wider range of vegetation conditions would help to answer this question more definitively.

Fine-scale, extensive mapping of burrow density could open the door to many new monitoring and research opportunities which cannot be addressed using ground-based colony perimeter mapping. Ground mapping typically does not involve intensive surveys of the interior of colonies and may over-estimate colony areas if large interior zones are not colonized. Such low-density interior areas can be easily detected using burrow density maps. Moreover, burrow density has been shown to be strongly correlated with population density (Johnson and Collinge, 2004) and vegetation parameters (e.g., Duchardt et al., 2021). Maps of burrow density could help map habitat for prairie dog associated species or be integrated into models of colony dynamics. For example, Cincotta et al. (1988) showed that burrow density at the edge of a colony influenced the probability that expansion would occur. Other studies suggest that large, well-connected colonies and colonies with higher prairie dog density may have higher plague transmittance rates compared to, smaller, less populated colonies (Cully and Williams, 2001; Davidson et al., 2022; Johnson et al., 2011). Burrow density maps could become inputs for models forecasting plague (Barrile et al., 2023) and subsequent recolonization from nearby colonies (Davidson et al., 2022). Burrow density maps could also help evaluate habitat suitability for associated species. For example, higher burrow densities may indicate support for larger populations of species that that rely on prairie dogs for a large portion of their diet (e.g., ferruginous hawk, swift fox) (Duchardt et al., 2023). Vegetation associated with greater burrow density may be expected to enhance habitat for some species (e.g., mountain plover, burrowing owls) and diminish habitat for other species (e.g., grasshopper sparrow, lark bunting) (Augustine and Derner, 2015; Duchardt

et al., 2019), while also reducing forage availability for livestock (Augustine and Derner, 2021; Brennan et al., 2021). Finally, network and clustering analyses of burrow configuration could potentially help to identify individual coteries and better understand spatial organization within colonies (Alba-Lynn and Detling, 2008; Hasan, 2019).

4.3. Scaling up and out: lessons for monitoring of burrowing mammals using UAS

Our results suggest that having RGB and TPI data at spatial resolutions of 5–15 cm will be ideal for future monitoring of black-tailed prairie dogs and other small-burrowing mammals across the North American Great Plains. Resolutions finer than 5 cm present computational challenges and resolutions greater than 15 cm likely fail to resolve individual burrows. Adding NDVI generally did not improve results at the burrow or colony scale for any spatial resolution, and in some cases decreased performance. TPI was the most important input for consistent burrow detection, likely due to reduced false positive detections for anthills and other features typified by circular bare ground patches but lacking the distinctive burrow mound and entrance feature highlighted by the TPI. Combining RGB with TPI performed well at the colony scale (Table A.6) and improved burrow detection at image resolutions greater than 10 cm (Table A.5, Fig. 4), at which point TPI images likely lose their ability to resolve smaller burrow entrances. Additionally, using only TPI as an input may increase the likelihood of detecting recently uninhabited burrows, leading to over-estimation of colony areas in recently uninhabited colonies (Table A.7).

It is important to note that our original 6-cm DEM, which we resampled to finer or coarser resolutions before creating the TPI's in our study, was initially derived from a point cloud with a density of around 220 points m^{-2} created photogrammetrically from the 1.5–1.9 cm RGB imagery. While we found that models using the coarsened 10–15 cm resolution TPI still performed well, it not clear whether a DEM originally derived from 10 to 15 cm RGB imagery or from airborne laser scanning (a.k.a. LiDAR) would perform equally well. In our tests, we resampled the DEM at a pixel scale using a simple box average but did not resample the original point cloud. Based on visual inspection in a GIS, burrow entrance depressions typically had a diameter of 15–40 cm. Using an equation proposed by Čekada et al. (2010), the smallest theoretical point density required to resolve features of this size ranges from 25 points m^{-2} (for 40 cm graphical accuracy) to 178 points m^{-2} (for 15 cm graphical accuracy). Further testing is needed to better understand whether a TPI derived from point clouds in this range performs as well as our artificially coarsened TPI, especially at the lower end of the point cloud density range which could be achieved with LiDAR. With current technology, creating point clouds at the upper end of this range may be cost-prohibitive over large areas. It is also possible that, given a sufficiently large training dataset, adequate accuracy could be achieved using only RGB imagery, though our results suggest TPI will generally improve results.

Assuming these technical challenges of acquiring a high-quality DSM can be addressed, the combination of RGB and TPI may also be useful for detecting other dynamic wildlife habitat features that involve a combination of topographic and vegetation features at relatively fine scales. This could include the dens of other burrowing mammals, termite mounds, large-mammal wallow features, and ant mounds, as well as the size and abundance of shrubs.

The fixed-wing Trinity F90 + UAS was useful for acquiring imagery across the entire 1,120-ha study area in a timely manner. Additionally, the Trinity has an on board GPS capable of real time kinematics (RTK). This results in fast and accurate orthorectification without the need for ground control points, although we note that high orthoimage accuracy is only needed if matching imagery to ground-referenced burrows or other small objects. One drawback to the Trinity fixed-wing unit is that the platform has limited capabilities to fly in windy conditions and could not be flown at windspeeds greater than 16–20 kph (4.5 ~ 5.5 m/sec). North American prairie dog colonies tend to occur in some of the windiest parts of the continent, which could limit the ability to use this platform for regular monitoring of large areas. Occupied aircraft may be required to monitor vast areas of windy and rugged terrain occupied by prairie dogs. Sharing costs with other monitoring programs (e.g., vegetation, critical habitat, fire fuels) may be necessary to enable widespread image acquisition at the requisite resolution. We expect the semantic segmentation approach used in this study would perform well if trained to simultaneously detect additional fine-scale features with distinct spectral and topographic characteristics (e.g., shrubs, anthills) and may be able to detect more nuanced features (e.g., cacti, invasive annual grasses and other plant communities). Thus, these data could be very useful for other monitoring objectives (e.g., invasive species control) beyond just burrowing mammal detection, which could lead to leveraging of funding or resources.

5. Conclusions

Prairie dogs are a keystone species that exist across an entire continent and have controversial impacts on rangelands. Accurate, timely and frequent colony maps are needed in order to make tough management decisions and arbitrate between those that wish to conserve prairie dog populations and those that wish to minimize or eliminate their presence on the landscape. Managers are currently spending vast resources on ground mapping, with limited success at temporally relevant scales due to dramatic boom-bust population dynamics driven by weather and disease (Davidson et al., 2022; Duchardt et al., 2023).

Here we demonstrate that deep CNN's and fine resolution aerial imagery are a powerful potential tool to not only map colonies, but also characterize within and between colony heterogeneity by providing detailed information on individual burrow locations. The possible shortcomings identified in this study (e.g., misclassifying recently unoccupied colonies) can likely be overcome by (1) training the CNN with a larger dataset, (2) integrating satellite-derived vegetation mapping and (3) targeting ground-based mapping to colonies with low detected burrow densities. Although it is unclear how scale-able and cost-effective this approach is at present, airborne platforms are evolving rapidly, becoming exponentially more capable and less costly. The monitoring approach developed here can likely be used in other extensive, dynamic situations where accurate and timely monitoring of fine-scale ecological indicators is needed.

CRediT authorship contribution statement

Sean P. Kearney: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **Lauren M. Porensky:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **David J. Augustine:** Conceptualization, Investigation, Data curation, Writing – review & editing. **David W. Pellatz:** Conceptualization, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the staff at the US Forest Service's Geospatial Technology

and Applications Center (GTAC) for their efforts, resources and expertise to acquire and prepare the UAS imagery for analysis. We sincerely thank Nick Dufek, Stephanie Steele, Greg Paff, and Andrew Scott for collecting the ground-based colony boundary data. Funding for analysis came from the United States Department of Agriculture – Agricultural Research Service (USDA-ARS). This research was a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2023.110684.

References

Abhishek, K., Hamarneh, G., 2021. Matthews correlation coefficient loss for deep convolutional neural networks: application to skin lesion segmentation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, pp. 225-229.

Alba-Lynn, C., Detling, J.K., 2008. Interactive disturbance effects of two disparate ecosystem engineers in North American shortgrass steppe. Oecologia 157 (2), 269–278.

Augustine, D., Davidson, A., Dickinson, K., Van Pelt, B., 2021. Thinking like a grassland: Challenges and opportunities for biodiversity conservation in the Great Plains of North America. Rangel. Ecol. Manage. 78, 281–295.

Augustine, D.J., Derner, J.D., 2015. Patch-burn grazing management, vegetation heterogeneity, and avian responses in a semi-arid grassland. J. Wildl. Manag. 79 (6), 927–936.

Augustine, D.J., Derner, J.D., 2021. Long-term effects of black-tailed Prairie dogs on livestock grazing distribution and mass gain. J. Wildl. Manag. 85 (7), 1332–1343.

Augustine, D.J., Skagen, S.K., 2014. Mountain plover nest survival in relation to prairie dog and fire dynamics in shortgrass steppe. J. Wildl. Manag. 78 (4), 595–602.

Barrile, G.M., Augustine, D.J., Porensky, L.M., Duchardt, C.J., Shoemaker, K.T., Hartway, C.R., Derner, J.D., Hunter, E.A., Davidson, A.D., 2023. A big data–model integration approach for predicting epizootics and population recovery in a keystone species. Ecol. Appl. 33 (4).

Brennan, J.R., Johnson, P.S., Hanan, N.P., 2020. Comparing stability in random forest models to map Northern Great Plains plant communities in pastures occupied by prairie dogs using Pleiades imagery. Biogeosciences 17, 1281–1292.

Brennan, J., Olson, K., Johnson, P., Block, J., Schauer, C., 2021. Grazing behavior, forage quality, and intake rates of livestock grazing pastures occupied by Prairie Dogs. Rangel. Ecol. Manage. 76, 12–21.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: Fast and Flexible Image Augmentations.

Čekada, M.T., Crosilla, F., Fras, M.K., 2010. Theoretical LiDAR point density for topographic mapping in the largest scales. Geodetski vestnik 54, 403–416.

Cincotta, R.P., 1989. Note on mound architecture of the black-tailed prairie dog. The Great Basin Naturalist 49, 621–623.

Cincotta, R.P., Uresk, D.W., Hansen, R.M., 1988. A statistical model of expansion in a colony of black-tailed prairie dogs. In: Uresk, D.W., Schenbeck, G.L., Cefkin, R. (Eds.), Eighth Great Plains wildlife damage control workshop proceedings. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, pp. 30-33.

Colorado Parks and Wildlife, 2020. Colorado Black-tailed Prairie Dog Range-wide Monitoring.

Connell, L.C., Scasta, J.D., Porensky, L.M., 2018. Prairie dogs and wildfires shape vegetation structure in a sagebrush grassland more than does rest from ungulate grazing. Ecosphere 9 (8).

Crow, L., Porensky, L.M., Augustine, D., Ritten, J., Bastian, C.T., Paisley, S.I., 2022. Evaluating prairie dog-cattle competition from the perspective of a ranching enterprise in the Western Great Plains: Economic analysis of potential effects on long-term profitability. Rangel. Ecol. Manage. 85, 56–65.

Cully, J.F., Williams, E.S., 2001. Interspecific comparisons of sylvatic plague in prairie dogs. J. Mammal. 82 (4), 894–905.

Cully, J.F., Johnson, T.L., Collinge, S.K., Ray, C., 2010. Disease limits populations: plague and black-tailed prairie dogs. Vector-borne Zoonot. Dis. 10 (1), 7–15.

Davidson, A.D., Augustine, D.J., Jacobsen, H., Pellatz, D., Porensky, L.M., McKee, G., Duchardt, C., Smyser, T., 2022. Boom and bust cycles of black-tailed prairie dog populations in the Thunder Basin grassland ecosystem. J. Mammal. 103, 1112-1126.

Delparte, D.M., Bly, K., Stone, T., Olimb, S., Kinsey, M., Belt, M., Calton, T., 2019. sUAS for wildlife conservation-assessing habitat quality of the endangered black-footed ferret. In: Sharma, J.B. (Ed.), In Application of Small Unmanned Aircraft Systems: Best Practices and Case Studies. CRC Press, Taylor and Francis Group, New York.

Derner, J.D., Detling, J.K., Antolin, M.F., 2006. Are livestock weight gains affected by black-tailed prairie dogs? Front. Ecol. Environ. 4 (9), 459–464.

Duchardt, C.J., Porensky, L.M., Augustine, D.J., Beck, J.L., 2018. Disturbance shapes avian communities on a grassland-sagebrush ecotone. Ecosphere 9 (10), e02483.

Duchardt, C.J., Augustine, D.J., Beck, J.L., 2019. Threshold responses of grassland and sagebrush birds to patterns of disturbance created by an ecosystem engineer. Landsc. Ecol. 34 (4), 895–909.

Duchardt, C.J., Porensky, L.M., Pearse, I.S., 2021. Direct and indirect effects of a keystone engineer on a shrubland-prairie food web. Ecology 102, e03195.

Duchardt, C.J., Augustine, D.J., Porensky, L.M., Beck, J.L., Hennig, J.D., Pellatz, D.W., Scasta, J.D., Connell, L.C., Davidson, A.D., 2023. Disease and weather induce rapid shifts in a rangeland ecosystem mediated by a keystone species (Cynomys ludovicianus). Ecol. Appl. 33, e2712.

Durant, S.M., Craft, M.E., Hilborn, R., Bashir, S., Hando, J., Thomas, L., 2011. Long-term trends in carnivore abundance using distance sampling in Serengeti National Park, Tanzania. J. Appl. Ecol. 48, 1490–1500.

Folluo, J.N., 2019. An Evaluation of Remote Sensing Methods for Ecologlical Management in Thoedore Roosevelt National Park, Graduate Faculty. University of North Dakota, Grand Forks, North Dakota.

Gentle, M., Finch, N., Speed, J., Pople, A., 2018. A comparison of unmanned aerial vehicles (drones) and manned helicopters for monitoring macropod populations. Wildl. Res. 45 (7), 586.

Gober, P., 2000. 12-Month Finding for a Petition to List the Black-Tailed Prairie Dog as Threatened, in: Fish and Wildlife Service, I. (Ed.).

Hasan, E., 2019. Comparative Analysis of Prairie Dog Colony Spatial Structure, Ecology and Evolutionary Biology. University of Colorado, Boulder.

Iakubovskii, P., 2019. Segmentation Models Pytorch. GitHub.

Johnson, W.C., Collinge, S.K., 2004. Landscape effects on black-tailed prairie dog colonies. Biol. Conserv. 115 (3), 487–497.

Johnson, T.L., Cully, J.F., Collinge, S.K., Ray, C., Frey, C.M., Sandercock, B.K., 2011. Spread of plague among black-tailed prairie dogs is associated with colony spatial characteristics. J. Wildl. Manag. 75 (2), 357–368.

Jost, R., 2018. Detection of Black-tailed Prairie Dogs Through the Use of High Resolution Aerial Imagery and LiDAR Data, Graduate Faculty. Texas Tech University.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49.

Kearney, S.P., Porensky, L.M., Augustine, D.J., Gaffney, R., Derner, J.D., 2022. Monitoring standing herbaceous biomass and thresholds in semiarid rangelands from harmonized Landsat 8 and Sentinel-2 imagery to support within-season adaptive management. Remote Sens. Environ. 271, 112907.

Kretzer, J.E., Cully, J.F., 2001. Effects of black-tailed prairie dogs on reptiles and amphibians in Kansas shortgrass prairie. Southwest. Nat. 46 (2), 171.

Lantz, S.J., Conway, C.J., 2009. Factors affecting daily nest survival of burrowing owls within black-tailed prairie dog colonies. J. Wildl. Manag. 73 (2), 232–241.

McDonald, L., Mitchell, J., Howlin, S., Goodman, C., 2015. Range-wide monitoring of black-tailed prairie dogs in the United States: pilot study. Technical Report.

Ogutu, J.O., Piepho, H.-P., Said, M.Y., Ojwang, G.O., Njino, L.W., Kifugo, S.C., Wargute, P.W., Paiva, S.R., 2016. Extreme wildlife declines and concurrent increase in livestock numbers in Kenya: What are the causes? PLoS One 11 (9), e0163249.

Riitters, K.H., Wickham, J.D., O'Neill, R.V., Jones, K.B., Smith, E.R., Coulston, J.W., Wade, T.G., Smith, J.H., 2002. Fragmentation of Continental United States Forests. Ecosystems 5 (8), 815–822.

Roelle, J.E., Miller, B.J., Godbey, J.L., Biggins, D.E., 2006. Recovery of the black-footed ferret—progress and continuing challenges: U.S. Geological Survey Scientific Investigations Report 2005–5293. USGS, p. 288.

Sidle, J.G., Johnson, D.H., Euliss, B.R., Tooze, M., 2002. Monitoring black-tailed prairie dog colonies with high-resolution satellite imagery. Wildl. Soc. Bull. 30, 405–411.

Sidle, J.G., Augustine, D.J., Johnson, D.H., Miller, S.D., Cully, J.F., Reading, R.P., 2012. Aerial surveys adjusted by ground surveys to estimate area occupied by black-tailed prairie dog colonies. Wildl. Soc. Bull. 36 (2), 248–256.

Smith, L.N., 2015. Cyclical Learning Rates for Training Neural Networks. arXiv.

U.S. Forest Service, 2009. Record of Decision: Thunder Basin National Grassland Prairie Dog Management Strategy and Land and Resource Management Plan Amendment #3.