


RESEARCH ARTICLE

Remote sensing-supported mapping of the activity of a subterranean landscape engineer across an afro-alpine ecosystem

Luise Wraase^{1a} , Victoria M. Reuber^{2a}, Philipp Kurth^{1,3}, Mekbib Fekadu^{3,4}, Sebsebe Demissew⁴, Georg Miehe⁵, Lars Opgenoorth^{3,6}, Ulrike Selig¹, Zerihun Woldu⁴, Dirk Zeuss¹, Dana G. Schabo², Nina Farwig² & Thomas Naus¹

¹Department of Geography, Environmental Informatics, Philipps-Universität Marburg, Deutschhausstraße 12, 35032, Marburg, Germany

²Department of Biology, Conservation Ecology, Philipps-Universität Marburg, Karl-von-Frisch-Straße 8, 35034, Marburg, Germany

³Department of Biology, Plant Ecology and Geobotany, Philipps-Universität Marburg, Karl-von-Frisch-Straße 8, 35034, Marburg, Germany

⁴Department of Plant Biology and Biodiversity Management, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia

⁵Department of Geography, Vegetation Geography, Philipps-Universität Marburg, Deutschhausstraße 10, 35032, Marburg, Germany

⁶Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903, Birmensdorf, Switzerland

Keywords

afro-alpine ecosystems, ecosystem engineer, machine learning, remote sensing, species distribution, subterranean animals

Correspondence

Luise Wraase, Department of Geography, Environmental Informatics, Philipps-Universität Marburg, Deutschhausstraße 12, 35032 Marburg, Germany. Tel: +49 6421 28-25334; Fax: +49 6421 28-28932. E-mail: luise.wraase@geo.uni-marburg.de

^aThese authors contributed equally to this study.

Editor: Temuulen Sankey

Associate Editor: Jin Wu

Received: 8 December 2021; Revised: 19 August 2022; Accepted: 23 August 2022

doi: 10.1002/rse2.303

Remote Sensing in Ecology and Conservation 2023; **9** (2):195–209

Abstract

Subterranean animals act as ecosystem engineers, for example, through soil perturbation and herbivory, shaping their environments worldwide. As the occurrence of animals is often linked to above-ground features such as plant species composition or landscape textures, satellite-based remote sensing approaches can be used to predict the distribution of subterranean species. Here, we combine in-situ collected vegetation composition data with remotely sensed data to improve the prediction of a subterranean species across a large spatial scale. We compared three machine learning-based modeling strategies, including field and satellite-based remote sensing data to different extents, in order to predict the distribution of the subterranean giant root-rat GRR, *Tachyoryctes macrocephalus*, an endangered rodent species endemic to the Bale Mountains in southeast Ethiopia. We included no, some and extensive fieldwork data in the modeling to test how these data improved prediction quality. We found prediction quality to be particularly dependent on the spatial coverage of the training data. Species distributions were best predicted by using texture metrics and eyeball-selected data points of landscape marks created by the GRR. Vegetation composition as a predictor showed the lowest contribution to model performance and lacked spatial accuracy. Our results suggest that the time-consuming collection of vegetation data in the field is not necessarily required for the prediction of subterranean species that leave traceable above-ground landscape marks like the GRR. Instead, remotely sensed and spatially eyeball-selected presence data of subterranean species could profoundly enhance predictions. The usage of remote sensing-derived texture metrics has great potential for improving the distribution modeling of subterranean species, especially in arid ecosystems.

Introduction

Subterranean animals act as ecosystem engineers as they shape and maintain grassland ecosystems worldwide. By burrowing, they rework sediments, redistribute nutrients in the soil and change vegetation patterns and when taken

together, create and modify habitats for other organisms (Corenblit et al., 2011; Gabet et al., 2003; Hastings et al., 2007; Jones et al., 1994, 1997; Reichman & Seabloom, 2002). At the same time, the distribution of subterranean animals and the species' functions for biodiversity and ecosystems are affected by human-induced

habitat modification and degradation, such as changes in vegetation cover caused by livestock grazing (Bakker et al., 2009; Keesing, 1998; Vial et al., 2011). Studies on the distribution and abundance of subterranean animals remain rare despite their importance for ecosystems and the increasing demand for detecting and predicting ecosystem changes.

The giant root-rat (GRR; *Tachyoryctes macrocephalus*, Rüppel, 1842 by Yalden, 1985) is a prime example of an animal ecosystem engineer. By creating extensive underground burrows and tunnel systems, it has a strong impact on the environment, particularly on the surrounding soil structure and vegetation (Šklíba et al., 2017; Yalden, 1985). The rodent is endemic to the afro-alpine ecosystem of the Bale Mountains of southeast Ethiopia, stretching between elevations of 3,000 and 4,150 m a.s.l., in lawns of *Alchemilla abyssinica*, which is their preferred diet (Yaba et al., 2011; Yalden, 1985; Yalden & Largen, 1992). Over time, the subterranean activity of the GRR changes the vegetation and soil structure of the prevailing open dwarf shrublands of *Helichrysum splendidum*, the Cyperaceae swamps and grasslands into discrete open-soil mounds (Miehe & Miehe, 1994; Yalden & Largen, 1992). Due to soil texture changes, these discrete mounds are distinguishable from the characteristically flat areas marked by an absence of GRR activity. Further, the species affects plant species richness, composition and biomass in its direct vicinity due to soil perturbation and above-ground herbivory (Sillero-Zubiri et al., 1995; Šklíba et al., 2017; Yalden & Largen, 1992). The changes in soil structure and vegetation patterns underpin the GRR's influence on the landscape and ultimately, their strong impact on an ecosystem and its processes. The ecological role of the species is even broader as GRRs are the predominant prey of the endangered Ethiopian wolf *Canis simensis* (Sillero-Zubiri & Gottelli, 1995). Moreover, the GRR is currently listed as endangered and is especially sensitive to increasing human-induced habitat degradation in the Bale Mountains (Lavrenchenko & Kennerley, 2016). Therefore, assessing the distribution of GRRs on the landscape scale is equally important for estimating their effectiveness as an ecosystem engineer as it is for evaluating the endangerment of the species. Here, we used satellite-based remote sensing to upscale the ecosystem engineering signs and quantify ground burrows across large extents (i.e., assessing the species distribution across the Bale Mountains).

Many satellite-based remote sensing studies have predicted particularly large animal species that are easily detectable from space either via direct observations across large extents or indirectly by predicting discrete structures or habitat changes caused by these animals, for example, elephants, penguins, or lions (Barber-Meyer et al., 2007;

Fretwell et al., 2012, 2014; Hollings et al., 2018; Kellenberger et al., 2018; LaRue et al., 2014; Loarie et al., 2013; Wang et al., 2019). To predict smaller-sized or species with low visibility such as birds, invertebrates, or subterranean animals, remote sensing approaches must rely on vegetation cover and composition, or geomorphological properties as proxies for habitat suitability and probability of the species' presence (Culbert et al., 2012; Estes et al., 2010; Farwell et al., 2021; Gabet et al., 2003; Grigusova et al., 2021; Koshkina et al., 2020; Porcasi et al., 2005). Generally speaking, predicting the presence of an animal species becomes increasingly difficult as its visibility decreases, for example, due to small body size or fast movement; this is particularly true for subterranean animals in homogenous grasslands. Hence, in order to predict GRR distribution, the focus should be on changes in vegetation patterns or soil structure (i.e., on the mounds created by the species) (Grigusova et al., 2021; Koshkina et al., 2020).

Machine learning has recently been used to bridge the gap between grain and extent. In our approach, we opted for statistical classification models, which can be trained on the local field data that link spectral remote sensing observations or other area-wide data sets, such as digital elevation models, with the occurrence of GRRs. This approach is regularly used to scale surveys across large areas (Pöyry et al., 2018; Wakulińska & Marcinkowska-Ochtyra, 2020). However, the performance of a given model depends on numerous factors, such as the observed environment and the species it aims to predict (Aguirre-Gutiérrez et al., 2014; Fiedler et al., 2008). Some studies show that including plant species composition as an additional independent variable can improve the predictability of animal species distribution or diversity (Schaffers et al., 2008; Wallis et al., 2017). In this specific context, if plant species composition and its change over space (i.e., turnover) are caused by a subterranean engineer, the turnover indicates the distribution of the species. However, this requires a labor-intensive field campaign to collect sufficient training data to directly predict the species composition in space as a surrogate for the animals' distribution. Thus, comparing remote sensing models that integrate field survey data to different extents with a model solely trained on readily available data could determine whether integrating time-intensive field observations improves the prediction of a subterranean ecosystem engineer (e.g., GRRs) across large extents.

In this study, we map GRRs distribution across the Bale Mountains using three different modeling strategies that require no (method 1, M1), some (M2) and very intensive (M3) fieldwork to collect training data. First, we hypothesize that in-situ collected GPS coordinates (M2), which accurately depict GRR presence and absence, improve the classifications exclusively based on training

areas selected with the eyeball method (M1). Second, we hypothesize that using plant species composition data (M3) and remotely sensed observations improve the classification outcome. Our comparative approach demonstrates methodological tools for mapping the current distribution of a subterranean species. Therewith, we provide insights into how GRRs shape the afro-alpine Bale Mountains ecosystem across their entire distribution range.

Materials and Methods

Field survey

The field survey was conducted in October and November 2017 across the Sanetti Plateau, and Web Valley of the Bale Mountains National Park in Ethiopia (6°29'N–7°10'N and 39°28' E–39°57' E; Fig. 1B). The highest peak of the study area was the Mount Tullu Dimtu at 4,377 m a.s.l. The climate of the study area is characterized by two consecutive wet and warm seasons (April–September) and a dry and colder season (November–March) with annual rainfall of approximately 1,000 mm (station records of

the DFG Research Unit Bale Exile at 10 sites). The slopes of the Bale Mountains are covered in moist mountain forests merging into the Ericaceous Belt around 3,000 m a.s.l. At the upper reaches, these ericaceous shrubland and dwarf forests of *Erica trimera* again merge in an extended ecotone into the afro-alpine ecosystem between 3,800 and 4,000 m with the highest *Erica* outposts at approximately 4,250 m a.s.l. The afro-alpine ecosystem consists of species-poor open dwarf shrublands of *Alchemilla haumannii* and *Helichrysum citrispinum* in the northern part and *H. splendidum* in the southern highlands. *Lobelia rhynchopetalum*, a giant rosette plant, is scattered across the entire afro-alpine ecosystem (Chala et al., 2016). In addition to the GRR, other wild herbivores such as mountain nyalas *Tragelaphus buxtoni*, bohor reedbucks *Redunca redunca* and rock hyraxes *Procavia capensis capillosa* feed upon vegetation (Mekonen, 2020; Teklehaimanot & Balakrishnan, 2018). Furthermore, growing human presence in the Bale Mountains National Park followed by livestock and associated grazing, human settlements, grass collection or frequent fire and bush encroachment impact the landscape of the Bale Mountains National Park (Mekonen, 2020). The impact GRRs

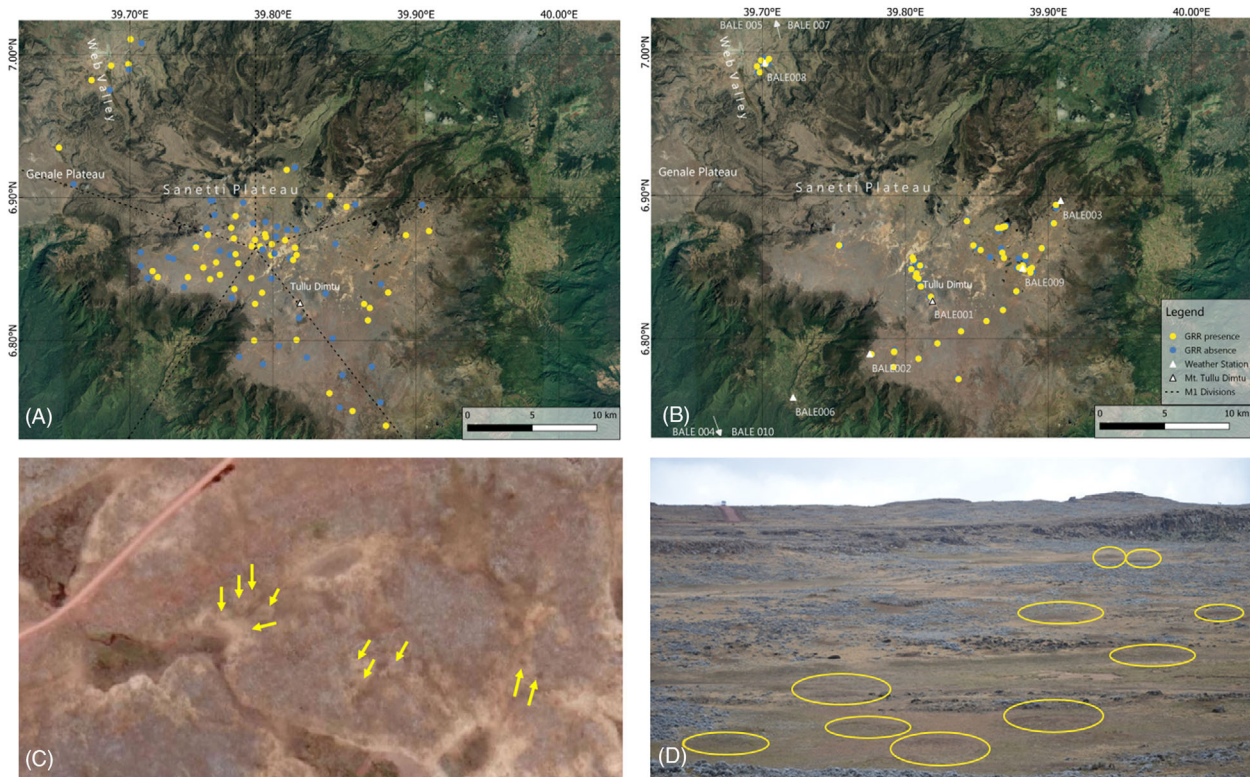


Figure 1. Locations of training areas with presence (yellow dots) and absence (blue dots) of giant root-rat (GRR) activity used in the different model strategies based on (A) Google Earth (M1) and (B) local GPS records (M2 and M3). Black dotted lines in (A) show sectors, in which training areas were selected, and white triangles in (B) show the location of climate stations; for details, see the methods section. (C) Shows GRR mounds as observable in Google Earth and (D) in the field.

have on the landscape can be clearly separated from the impact of other species as a result of the pronounced effect they have on soil structure creating discrete mounds and a “spongy” ground (Fig. 1C and D) (Miehe & Miehe, 1994; Sillero-Zubiri et al., 1995).

In the field survey, we selected 94 GPS points and sampled plant species composition data in a pairwise plot design at the GPS points, in areas with and without GRR activity for later machine learning analyses (47 presence and 47 absence plots; GPS: Garmin eTrax30, precision 3 m). The survey was conducted along the main track of the Sanetti Plateau, the northern and northeastern parts of the Sanetti Plateau, and the Web Valley. The track was unpaved and infrequently used, and thus, the impact on plant species composition or GRR activity was presumably insignificant. Plots following the main track were established in a 2 km interval with a minimum of 100 m distance to the main track and a minimum of 50 m between presence and absence plots. The presence and absence plots were selected and carefully distinguished by the observers in GRR presence or absence areas (two persons with each 25 days and 10 h observations per day; PK and MF). Areas of GRR presence were clearly detectable as mound structures, which showed altered vegetation patterns in comparison to areas where GRR were absent and thus, areas without GRRs were also clearly identifiable. One mound was approximately 20 m in diameter (personal observation), with several burrow openings scattered across one mound. The immediate surrounding of GRR burrow openings was characterized by bare vegetation, while herbaceous vegetation covered the rest of the mound (Miehe & Miehe, 1994). For each mound, GRR activity could further be identified by fresh burrow openings (Leyer & Wesche, 2007), whereby one GRR individual used several burrow openings. The burrow openings from other subterranean species were smaller in diameter and thus distinguishable from GRR burrows, wherefore presence and absence of GRR activity areas were clearly specifiable. Nonetheless, past GRR activity at absence points could theoretically not be ruled out entirely; however, the impact on our analyses should be negligible, as the aim was to map current GRR presence and absence. We documented the plant species composition on each plot of 5 × 5 m size, and estimated the cover fraction in intervals of 5%, following a typical Mueller-Dombois and Ellenberg (1974) design. In addition, specialists identified plant species that could not be determined in the field (MF and SD). As a result, we found 60 (63) different plant species on GRR presence (absence) plots and 79 different plant species in total. The most abundant plant species for both areas with and without GRR activity was *A. abyssinica*. The abundance of other plant species varied between both areas.

Pre-processing

Species composition analysis

To compare the plant species composition between plots with and without GRR activity, we used constrained correspondence analysis (CCA) as an ordination method, based on the correlation matrix of the 79 plant species (rare species included) across all 94 plots (Fig. 1B). In general, ordination techniques can be used for assessing the main environmental gradients driving plant composition turnover across sites, using raw data of species richness and abundance. The gradients are projected into axes and displayed in a multidimensional ordination space with the first axis explaining the largest variance in the changes of plant composition by a gradient, with decreasing variance explained by the subsequent axes (Feilhauer & Schmidlein, 2009; Leyer & Wesche, 2007). A constrained ordination method assumes that the variation in the vegetation data is displayed by a priori chosen environmental gradient (i.e., constraints) (Leyer & Wesche, 2007), which are included in the first constrained axis. Further, a CCA assumes that a species' response to environmental gradients is unimodal and not linear. CCAs can deal with zero-inflated data and are therefore suited for our species composition data set (Leyer & Wesche, 2007; ter Braak, 1987). As temperature is a main driver for plant species composition (Keller et al., 2000; Nottingham et al., 2018), the satellite-predicted mean air temperature of 2017 was used as a constraining variable on the first CCA axis (CCA: Fig. S1A; for Landsat-8 temperature prediction see Appendix S1; Supplementary Method 1). By constraining species compositions on temperature, we removed the effect of temperature on species composition from the remaining unconstrained axes. The ordination scores of the first two unconstrained axes (CA1 and CA2 axes; Fig. S1B) were used in subsequent machine learning steps, herein referred to as CA1 and CA2 scores. Thus, the CA scores present plant species composition across survey sites corrected for the effects of temperature on vegetation composition. Temperature as a constraining variable explained 5% of the plant species composition. The package *vegan* (Oksanen et al., 2020) was used for the analysis in R version 4.0.2 (R Core Team, 2021).

Sentinel-2 observations for spatial prediction of GRR presence

An almost cloud-free (3.5% cloud cover) Sentinel-2 scene from December 15, 2017 at 07:54:34 UTC was retrieved from the USGS Earth Explorer repository and used as a basis for predicting GRR mounds in the study area. The

data were atmospherically corrected using the Sen2cor algorithm (Filippini, 2018). We used Sentinel-2 observations of the red, green, blue, red edge, near-infrared bands – each with 10 m resolution – and short-wave infrared satellite bands with 20 m resolution to include multiple indices with different foci on vegetation, soil and water, which best represent the grassland habitat of the GRR. The 10 m bands (three visible and one near-infrared) were resampled to match the 20 m resolution of the near- and short-wave infrared bands, to characterize the GRR presence in the study area. Spectral indices that highlight different vegetation and environmental characteristics (e.g., soil wetness) were computed from the individual band observations using the RStoolbox (Leutner et al., 2017). In addition, the k-mean distance from the centroid (KMDC) was computed in each case on every band and generated an index raster image (Table S1). Since GRR activity leads to hill-like structures, such as mounds, a group of grey-level co-occurrence matrix features (GLCM) and Rao's Q (Rocchini et al., 2018) were also compiled. For the GLCM, the entropy, homogeneity, and second moment (Haralick et al., 1973) for three different window sizes (3×3 , 11×11 and 31×31 pixels) were derived separately for both KMDC indices that were reduced to 32 grey levels. Different texture metrics are helpful to detect similarities and differences, and also patterns in the topography that simple satellite bands and vegetation indices cannot depict as accurately (Kupidura, 2019; Mishra et al., 2019). For depicting topographical differences, especially in soil texture and soil type from space, a pre-processing step with *k*-means cluster analysis was conducted using the Hartigan & Wang algorithm, set center = 1 and squared the fitted result which was multiplied by 2, in R-version 4.0.2 (Brus, 2019; Brus et al., 2006; Hartigan, 1975). Rao's Q (Rao, 1982) was determined based on the combined set of original bands, the scaled vegetation indices, and the results from KMDC (Table S1).

Modeling workflow

We followed three different modeling strategies to map the presence and absence of GRR activity, that is, by identifying locations with or without mound occurrences (Fig. 2; Table S2). All three strategies, hereafter referred to as M1, M2 and M3 utilized the Sentinel-2-based spectral indices and texture metrics as predictor variables. M1 was based on 94 training data points (i.e., 47 presence and 47 absence points) that were visually chosen from Google Earth. The visually chosen training data points (i.e., mounds for GRR presence and no mounds for GRR absence) were taken in the middle of a mound (presence) and in landscapes visually without mounds (absence). Mounds (Fig. 1C and D) were selected visually by one

person (LW) and cross-checked by a second person (TN) in a two-day process from Google Earth imagery, using their specific non-edge landscape characteristics and often repetitive appearance. M2 used in-situ collected GPS points as training data instead, with 94 training data points in total that were composed of 47 presence and 47 absence points. The GPS points (i.e., the training data) in M2 were taken in the middle of each mound and in areas without mounds. Finally, M3 employed the same in-situ collected training data points as in M2 (i.e., 94 points in total, 47 presence and 47 absence points); however, the CA1/CA2 scores derived from vegetation composition (CA prediction) into space were first predicted using the Sentinel-2-based variables. Next, the GRR activity was classified into a binary presence and absence map by a second model trained on the CA1/CA2 maps (M3, CA classification).

Our approach for M1 and M2 included using a random forest classifier (Breiman, 2001) in a forward feature selection (FFS), a 5-fold external leave location-out (LLO) and 10-fold internal cross-validation (CV). Within each of the five iterations, 70% of the data was used for training and 30% data was withheld for independent testing. For M3, the CA1/CA2 scores were first predicted into space using a random forest regression model (CA prediction step) using the same FFS, LLO and CV settings as in M1 and M2. For the second step, the variables CA1 and CA2 were added as predictors. Lastly, for the M3 approach, the same model settings were used as for M1 and M2. The FFS used in M1, M2 and M3 always starts by identifying the two best-performing variables based on the LLO error estimates. Subsequently, the algorithm incrementally increases the number of predictor variables and tests for each additional predictor variable if it is improving the current model further. The model stops training when adding another variable no longer increases the overall performance. For details on FFS, see Meyer et al. (2018). Previous studies identified the random forest classifier robust (Kuhn & Johnson, 2013; Meyer et al., 2018). The classification model performance was measured by Cohen's Kappa (Cohen, 1960), while the root mean square error (RMSE) was used for the regression models. Finally, the GRR presence and absence was classified after receiver operator characteristic (ROC) analysis for each of the three modeling strategies M1, M2 and M3 (Fig. S5). Since models are generally restricted to the information dimension of the input data sets (i.e., the spectral range of the Sentinel-2 data across the 94 extracted training areas) and a certain level of uncertainty generally remains after the final prediction, it is necessary to consider those (Jansen et al., 2022). Here, the area located outside of the actual area of applicability (AOA) of the model was masked and not further considered, following the approach from Meyer et al. (2018) and

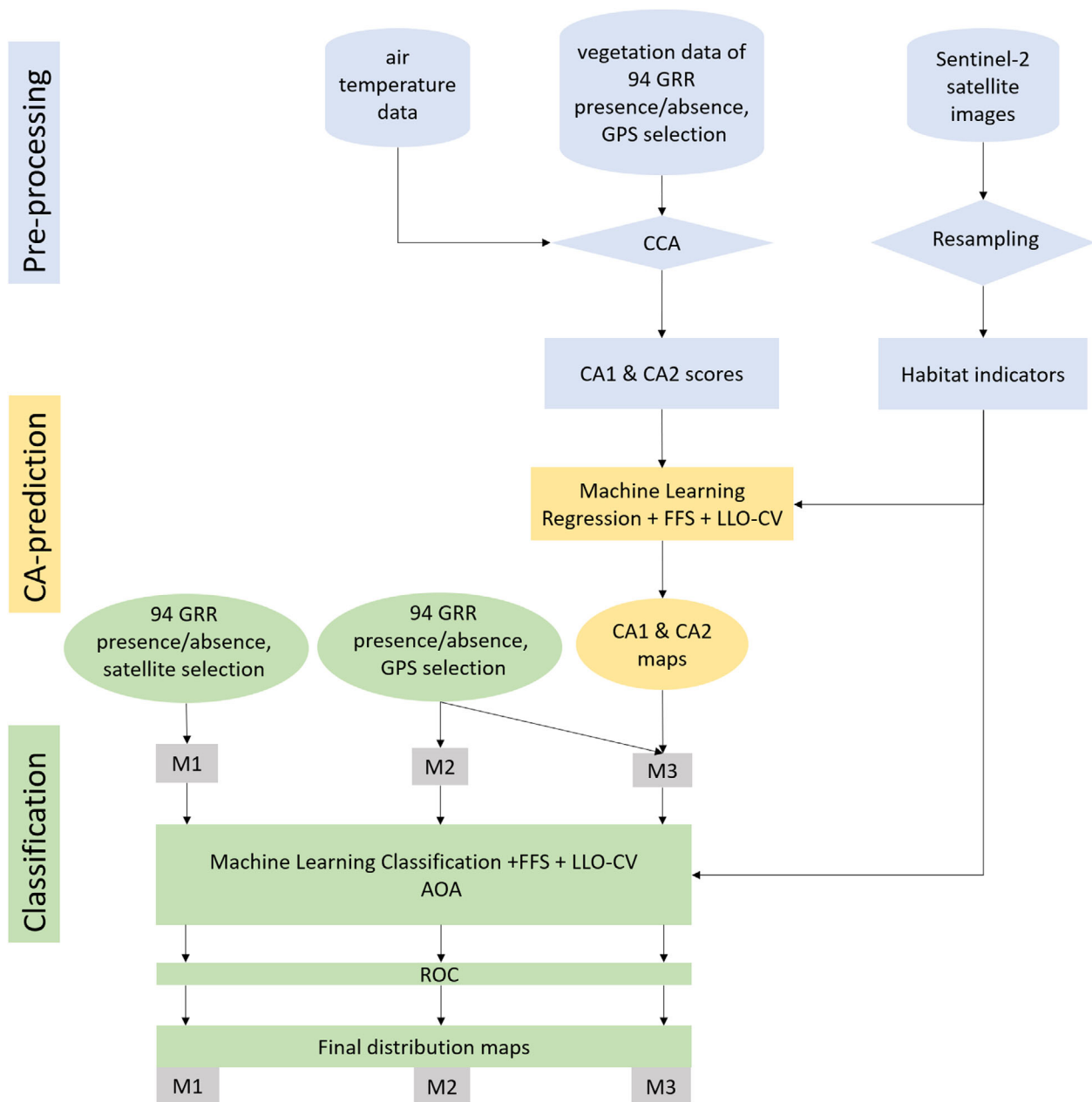


Figure 2. Modeling workflow showing the pre-processing (blue), the CA prediction (yellow, M3 only) and the classification (green). Pre-processing: Air temperature and habitat indicators were processed for downstream analyses; field-based plant species composition data was used in constrained correspondence analysis (CCA) to retrieve CA scores; CA-prediction: A machine learning regression model was used to predict CA1/CA2 scores into space with Sentinel-2 variables. Using forward feature selection (FFS) and a 10-fold leave-location-out (LLO) cross-validation (CV) to compile the CA1/CA2 maps, which were used in subsequent classification for M3; Classification: Machine learning models were applied, using FFS and 10-fold LLO-CV for predicting species distribution maps. The boxes with the numbers M1, M2 and M3 depict the three model strategies compared in this paper, with M1 supplied by image-selected classification categories, M2 using in-situ collected GPS coordinates and M3 using the same settings as M2 and also additionally the predicted CA scores as a predictor for the modeling process (see section Pre-processing). The area of applicability (AOA) method was used to calculate the validity of prediction error in space.

Meyer and Pebesma (2021) as implemented in the *CAST* package. Besides the random forest-based methods M1–M3, a Maximum Entropy (MaxEnt) (Phillips et al., 2006) machine learning model software accessed by the R-Package *dismo* (Hijmans et al., 2021) was conducted to further evaluate and cross-check the results of our three machine learning modeling strategies. In short, we used the same model settings and input data including the predicted CA variables as in M2 and M3, with the exception of the absence data (also 47 points) in MaxEnt which was selected randomly within the raster extent. The MaxEnt and random forest models were compared by the area under a receiver operating characteristic curve (AUC) value. If available, true presences and absences should be considered first in any modeling of species distributions (Elith et al., 2011; Guillera-Arroita et al., 2014; Zaniwski et al., 2002). Thus, the random forest model we used was the preferred method over the MaxEnt model, as it used true presences and true absences in our data. In contrast, absence points (or “background” points) in MaxEnt were randomly sampled across the whole area (Massada et al., 2012; Oppel et al., 2012) and may, by chance, also depict single presence points, which might influence the model results. Nevertheless, in other comparative studies, both machine learning-based methods tend to perform similarly (Acharya et al., 2019; Bektas et al., 2022; Kaky et al., 2020; Kaky & Gilbert, 2016; Mi et al., 2017; Zhao et al., 2022). As also in our case, the results of the MaxEnt and the modeling strategy M1 (the best of the random forest models) were qualitatively similar, we only present the results of the random forest machine learning approaches in the results section. The model performance, evaluation and final predictor variable selection of the MaxEnt model can be found in Appendix S1, Supplementary Method 2.

Results

CA prediction

Predicting CA1/CA2 values in M3 resulted in relatively high RMSE values with CA2 (0.882) exceeding CA1 (0.715); the R^2 (coefficient of determination) values provided a slightly contrary perspective (CA1: 0.188 and CA2: 0.129). The difference between the mean absolute error (MAE) values in CA1 (0.445) and CA2 (0.770) was more pronounced than for the RMSE and R^2 . The predicted CA1 scores overlapped more with the original scores than CA2 (Fig. 3). The two most important predictors for CA1 were the Normalized Difference Vegetation Index (NDVI) and Land Surface Water Index (LSWI); for CA2 they were the Sentinel-2 band 3 (blue) and GLI (Table S3; Fig. S2). Three of the six (CA1) and five of the eight (CA2) variables were composed based on the pixels' surroundings (Table S3; Fig. S2).

GRR classification

In terms of predicting the presence and absence of GRR activity, the model accuracies showed considerable differences with M1 (Cohen's Kappa: 0.777) and M2 (0.494) outperforming M3 (0.375) (Table 1b). Models M1 and M2 shared similarities, for example, they each included at least two texture metrics, representing information computed on window sizes of 3×3 or 11×11 pixels, which cover an extent of 60 and 220 m, respectively (Fig. 4). In M1, the most important predictor variable was the KMDC indices stack (k-means distance from the center of indices stack) and the second most important predictor was the texture metric of Rao's Q KMDC indices stack. The texture metrics were the most important predictor

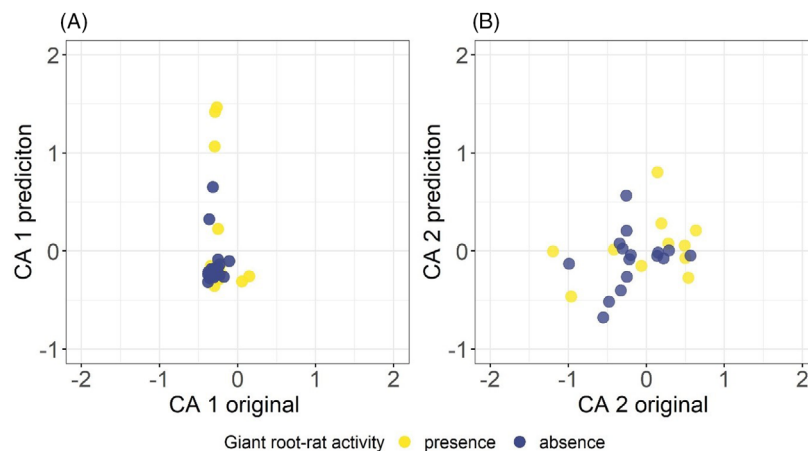


Figure 3. Correlation between original and predicted CA scores of (A) CA1 and (B) CA2 by M3 for each of the 47 GRR presence and absence locations. This figure displays a randomly chosen test set of 27 data points.

Table 1. Comparison of the three model strategies (M1–M3); Google Earth (M1) and local GPS records (M2 and M3 with M3 additionally including CA scores as predictors, see section Modeling workflow for details) by depicting the (a) error matrix and (b) accuracy values. (a) shows which response, that is, presence or absence of giant root-rat (GRR) activity, explains how much percentage of its own or the other classes. (b) shows each model's performance for the applied classifications (ROC threshold as graph, see Figure S5).

	(a) Error matrix		(b) Accuracy values				
	Prediction	Presence	Absence	Accuracy	Cohen's Kappa	AUC	ROC threshold
M1	Presence	49.3	6.0	0.895	0.777	0.849	0.555
	Absence	4.5	40.3				
M2	Presence	41.8	13.4	0.750	0.494	0.640	0.575
	Absence	11.9	32.8				
M3	Presence	38.8	16.4	0.691	0.375	0.654	0.581
	Absence	14.9	29.9				

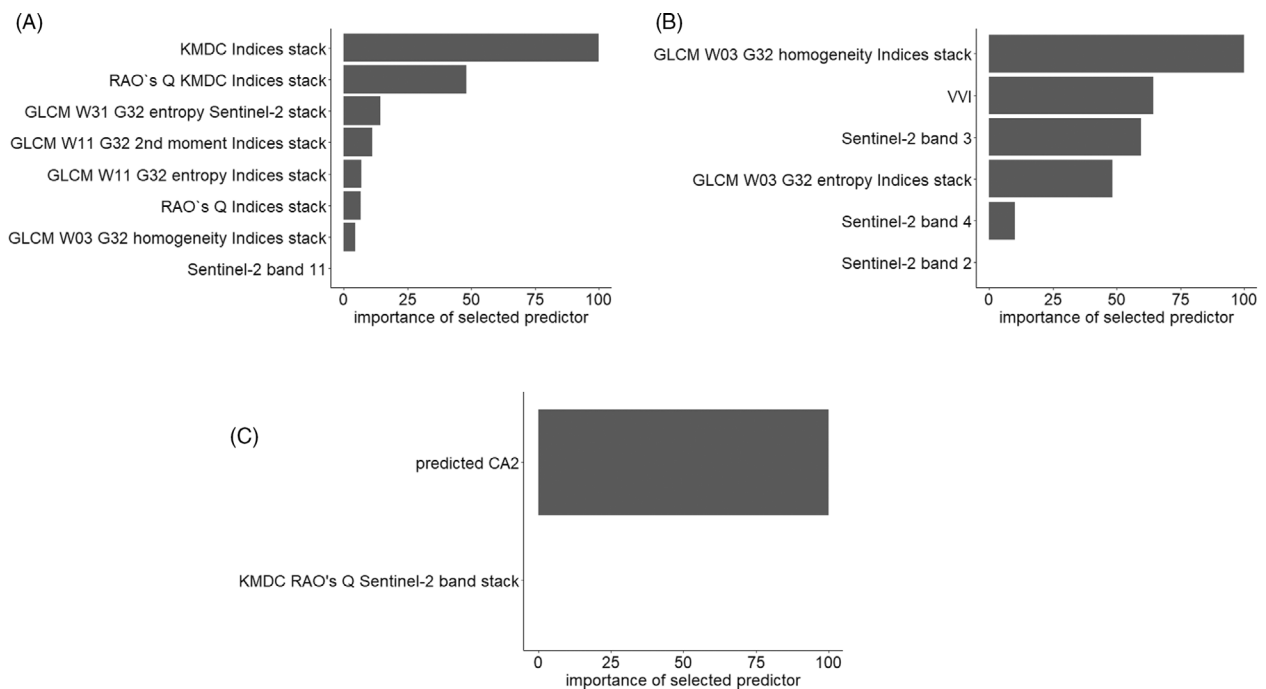


Figure 4. Selected predictors in order of importance and their explanatory power in percent (%) for each of the three model strategies M1–M3 (A–C); training areas based on Google Earth (M1; A), local GPS records (M2; B) and (M3; C). M3 additionally included CA scores as predictors (see section Modeling workflow for details; predictor names: Table S1).

group of M1, with seven texture metrics selected in total (Fig. 4A). Compared to M1, the difference in the importance of the top four predictors in M2 was more gradual. The final model of M2 included two predictor variables less than M1, including the Visible Vegetation Index (VVI), two Sentinel-2 bands and two texture metrics based on a 3×3 window size of 60 m^2 as predictor variables for the GRR distribution (Fig. 4B). In M3, only the predicted CA2 axis scores were chosen in combination with texture Rao's Q KMDC (Sentinel-2 band stack) as the second most important predictor (Fig. 4C). Model tuning resulted in five (M1), three (M2) and two (M3) variables used for splitting at each tree node of the

random forest classifier (i.e., mtrys, Fig. S3). The error matrix, which depicts how much of each class is assigned correctly (i.e., GRR presence or absence), determined that M1 outperformed M2 and M3.

Distribution of the giant root-rat

The three model strategies predicted GRR presence and absence across the Sanetti Plateau and Web Valley. M1 and M2 both predicted a strong concentration of GRR presence on the central Sanetti Plateau and the north-western descent toward the Web Valley and the lower plateau area (Fig. 5). While M1 predicted a higher

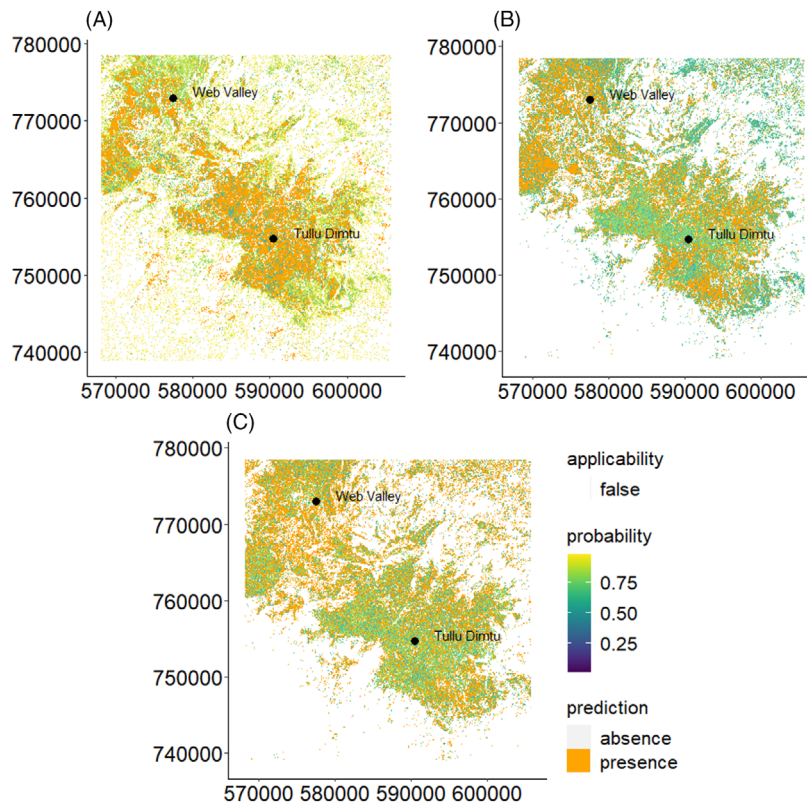


Figure 5. Spatial predictions of giant root-rat (GRR) presence across the Bale Mountains. Each map depicts the distribution of the GRR with a different model strategy M1 (A), M2 (B) and M3 (C). The prediction layer of GRR distribution shows presence in orange and absence in dark grey. The probability of GRR presence in an area is indicated from 1 (presence, yellow) to 0 (absence, dark purple). This is based on the best fitting threshold from each model strategy defined by a receiver operating characteristic (ROC) curve analysis (see Table 1 and Fig. S4 for plotted ROC curves). The validity of the prediction result for GRR presence using the area of applicability method (AOA) is displayed with a transparent mask; white areas lie outside the AOA – and predictions in this area should not be considered.

presence of GRR activity in the middle of the Sanetti Plateau and resulted in an area of 28,366 ha for GRR presence (i.e., 19% of the total area of 147,963 ha). M2 showed a higher concentration around the upper north-western parts and less concentration on the Sanetti Plateau with 55,589 ha of GRR presence. The classification map of M3 predicted 84,859 ha of GRR presence centered around the plateau area and with extra parts in the north–northwestern area. These values were based on prediction probability ROC thresholds of 0.555 (M1), 0.575 (M2) and 0.581 (M3) for delineating presence and absence locations (Fig. S5). In general, the plateau area was more likely to be predicted for the three methods (Fig. 5A–C). These values were also corrected for the models AOA (Fig. 5). For M1, 41.35% of the entire satellite image extent fell into the valid area; this percentage was much greater for M2 (62.73%) and M3 (86.82%). Areas in white were not considered applicable for the model results and areas in transparent displayed overlap. For M1, the AOA partly aligned with the prediction results (orange) but excluded some distinct, concentrated

parts in the southern section. For M2, the AOA also aligned with the prediction result with a greater AOA at an area west adjacent to the Sanetti Plateau and northern parts the Web Valley. In general, the AOA for M3 covered more area than M1 and M2, excluding larger areas in the south and southwestern parts adjacent to the Sanetti Plateau. The AOA also eliminated all areas that did not exhibit GRR mound structures.

Discussion

Comparing subterranean species prediction methods that include field data to a different extent helps to elucidate whether integrating detailed plant species composition data improves predictions across space. Contrary to our first and second hypotheses, our study showed that in-situ collected GPS coordinates of GRR presence and absence and additional plant species composition information did not improve the landscape-scale prediction of the distribution of a subterranean rodent in a homogeneous afro-alpine environment. Remotely sensed textural

metrics and vegetation indices significantly improved models for predicting the presence of the subterranean GRR. The model strategy based on training areas visually selected on Google Earth images (M1), outperformed training areas using in-situ collected GPS coordinates (M2). In the overall comparison, the complex model M3, including CA scores for plant species composition as a model predictor, had the lowest accuracy. Hence, detailed vegetation surveys are superfluous for predicting the distribution of the GRR, a species that leaves distinct above-ground, and remotely distinguishable landscape marks therefore it is advisable to focus on remote sensing analyses.

Model performance

Overall, M1 had a considerably higher accuracy (Cohen's Kappa) compared to M2 and M3. Comparisons revealed that the spatial coverage of the training data was decisive in improving the quality of the models. As such, M1 performed better given that the training data points covered the entire extent of the Sanetti Plateau and were possibly more heterogeneous. Previous studies have shown that the more spatial representative the training data, the better the resulting models (Berhane et al., 2019; Hengl, 2007; Warren et al., 2014). Future field campaigns should focus on covering the entire study area to generate a heterogeneous data set. Furthermore, M1 only used training points that were clearly identifiable as locations with GRR presence or absence in the Google Earth images. This selection of training points may skew towards large or characteristic features, which is concomitant with deficits in mapping the potential variability of the GRR locations. The visual selection of training data (M1) works if the species makes a tremendous impact on its landscape, like the GRR. In contrast, the training data in M2 and M3 were restricted to in-situ collected GPS coordinates and plant species composition data of GRR presence and absence areas (only M3). In-situ sampling was subject to human labor and temporal constraints on the extensive and partly difficult to access mountain plateau. An extended field period in remote areas at high elevations (4,000 m a.s.l.) would require several months of physically challenging work conditions to collect a sample comparable to M1, for which the data was retrieved within a few office days.

Predictor importance

M1 and M2 shared similarities as both methods chose texture metrics as predictors of 3×3 and 11×11 pixel window sizes. The texture metrics as the most frequently chosen predictors, likely reflect the topographic pattern of the GRR (i.e., the mounds created by GRR activity). These mounds are distinct from the relatively flat surroundings and, hence,

detectable from space. Numerous studies have demonstrated the use of environmental structural heterogeneity for predicting a broad range of plant and animal species (Bellis et al., 2008; Farwell et al., 2021; Tuanmu & Jetz, 2015; Wood et al., 2013). The impact of the GRR on soil structure is extreme compared to other species; however, we would recommend using texture metrics in future predictions of subterranean rodents, for example, the east African root-rat, north American pocket gophers or Mongolian marmots (Gabet et al., 2003; Huntly & Reichman, 1994; Koshkina et al., 2020), which also leave above-ground marks.

In direct comparison, M1 selected two more predictors for the final model than M2. The importance of the predictors was less abrupt for M2 than for M1. For M1, the most interpretable variable is Rao's Q, which can be linked to plant functional types (i.e., grouping plant species with similar structural features) (Botta-Dukát, 2005; Rocchini et al., 2018). The presence of GRRs is related to grasslands where *Alchemilla* is the predominant species, whereas the species' absence is characterized by other habitat features, such as denser shrubland or *Erica*-thickets (Miehe & Miehe, 1994).

For M2, the VVI and Sentinel-2 bands were chosen along with texture metrics. VVI is typically used to predict biomass and can be linked to GRR presence and absence because the species keeps vegetation low and in pioneer stages through soil perturbation, its herbivorous diet (Miehe & Miehe, 1994) and reduced vegetation covers at the top of burrows. Furthermore, the abundance and activity of herbivorous rodents are affected by vegetation as their primary food resource. Hence, they are typically more abundant in areas with higher plant productivity (Eldridge & Whitford, 2014; Huntly & Reichman, 1994; Zhang et al., 2003). Further, the reduced food supply during the dry season causes GRRs to change their home range to food-rich periodic wetlands in the Bale Mountains (Šklíba et al., 2020; Vlasatá et al., 2017). Thus, selecting a vegetation index describing biomass is in accordance with the ecology and habitat preferences of the target species and emphasizes our findings that GRR distribution can be predicted using remote sensing.

In M3, the predicted CA2 scores describing species composition were chosen as a predictor for GRR presence and absence, indicating the effect that the GRR has on species composition. For instance, the species reduces *A. abyssinica* but fosters *Salvia merjame* (Šklíba et al., 2017). In general, the vegetation and texture metrics selected for predicting the CA scores can be related to the rodent's ecological function. The selected vegetation index is sensitive to chlorophyll concentration, the reflectance of which can be related to the distribution of herbivorous species as they correlate to plant biomass (Gitelson et al., 2009; Olofsson et al., 2012) and, hence, available food resources (Reichman

& Seabloom, 2002). However, including vegetation data in the form of CA scores did not explain GRR presence as accurately as the purely remote sensing-based approach in M1, even when combined with remote sensing-based predictors. The prediction of the area of M2 and M3 only approximates to the final distribution map of M1 when the AOA layer is included, which masks areas where the prediction should not be considered.

Despite the well-known impact of GRRs on vegetation (Miehe & Miehe, 1994; Sillero-Zubiri et al., 1995; Šklíba et al., 2017; Yalden, 1985), the in-situ collected raw plant species composition data, surprisingly did not enhance the prediction of GRR distribution. One explanation might be that the spectral difference of plant species between areas with and without current GRR activity was not pronounced enough to be detected remotely and to be a reliable predictor for GRR distribution. In fact, the most abundant plant species in both presence and absence areas was *A. abyssinica*. The less abundant plant species differed between GRR presence and absence areas; however, their spectral signature was presumably not pronounced enough or masked by the spectral signature of the dominant plant species *A. abyssinica*. Our results demonstrate that textural metrics can reliably predict the presence and absence of the species *via* their impact on soil structure. However, more subtle impacts of the GRR on its biotic environment, such as the vegetation composition, could not be assessed from space. Hence, field assessments are indispensable if subtle impacts of a species on ecosystem functionality are the primary focus. Yet, we emphasize that remote sensing is a promising tool for predicting the presence of a subterranean species based on texture metrics, while field-based knowledge about plant species composition is not required to predict the distribution of the GRR. Predicting the current distribution with minimized effort is particularly relevant considering the endangerment of the species and its ecological role as ecosystem engineer.

To summarize, our aim to predict the spatial distribution of the focal subterranean animal species was best conducted with textural and vegetation indices; detailed knowledge of the vegetation composition around the mounds was not required. As such, remote sensing and machine learning approaches can facilitate spatial modeling of subterranean species that create distinctive above-ground landscape structures. In this study, we examined tools to meet the complex challenge of predicting less visible species. This could be particularly valuable for spatial modeling in remote areas and environments with low structural heterogeneity. Our approach may be applicable to other arid ecosystems, where vegetation stands are low and sparse. Here, subterranean rodents are present frequently with important implications for ecosystem processes (Contreras et al., 1993; Desmet &

Cowling, 1999; Kerley et al., 2004; Lacey & Wiczorek, 2003; Miranda et al., 2019), which is particularly critical if the ecosystems are difficult to access. Yet, further studies need to confirm if our approach is applicable for other subterranean species, for instance for the GRR's sister species *T. splendens*, subterranean mammals in the Tibetan and Mongolian grasslands, or other cryptic species such as social-insect colonies that create vegetation patterns like the Namibian and Australian fairy circles.

We appreciated funding of the Research Unit 2358 ("The Mountain Exile Hypothesis") by the German Research Foundation (NA 783/12-2, OP 219/5-2, FA-925/14-1 und SCHA-2085/3-1, MI271/33-2).

Author contributions

LW and VR contributed equally to this work. LW, VR, PK, MF, SD, GM, LO, ZW, DZ, DS, NF, and TN designed the research concept, PK and MF mapped the GRR mounds and sampled the vegetation on the GRR presence and absence plots. LW, VR and PK processed the vegetation survey data that was collected. VR conducted the statistical vegetation analysis. US did the temperature modeling of the Landsat-8 images. LW and TN performed the statistical and spatial modelling and prediction. LW and VR drafted the manuscript and figures with contributions from all authors.

Data Availability Statement

The corresponding data set is available here: [10.5281/zenodo.5704707](https://zenodo.org/record/5704707)

Acknowledgements

This research was funded by the German Research Council (DFG) within the framework of the joint Ethio-European DFG Research Unit 2358 "The Mountain Exile Hypothesis. How humans benefited from and re-shaped African high-altitude ecosystems during Quaternary climate changes" (NA 783/12-2, OP 219/5-2, FA-925/14-1 und SCHA-2085/3-1, MI271/33-2). We thank the Ethiopian Wildlife Conservation Authority, the College of Natural and Computational Sciences (Addis Ababa University), the Department of Plant Biology and Biodiversity Management (Addis Ababa University), the Philipps-Universität Marburg, the Frankfurt Zoological Society, the Ethiopian Wolf Project and the Bale Mountains National Park for their cooperation and kind permission to conduct fieldwork. We are grateful to Awol Asefa, Wege Abebe, Katinka Thielsen, Tiziana Li Koch, Kevin Frac, Terefe Endale, Geremew Mebratu for helping to prepare the fieldwork and this manuscript. Open Access funding enabled and organized by Projekt DEAL. WOA

Institution: PHILIPPS-UNIVERSITÄT MARBURG Consortium Name : Projekt DEAL

References

- Acharya, B.K., Chen, W., Ruan, Z., Pant, G.P., Yang, Y., Shah, L.P. et al. (2019) Mapping environmental suitability of scrub typhus in Nepal using MaxEnt and random Forest models. *International Journal of Environmental Research and Public Health*, **16**(23), 4845.
- Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. et al. (2014) Fit-for-purpose: species distribution model performance depends on evaluation criteria - Dutch hoverflies as a case study. *PLoS One*, **8**(5), e63708.
- Bakker, E.S., Olff, H. & Gleichman, J.M. (2009) Contrasting effects of large herbivore grazing on smaller herbivores. *Basic and Applied Ecology*, **10**(2), 141–150.
- Barber-Meyer, S.M., Kooyman, G.L. & Ponganis, P.J. (2007) Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology*, **30**(12), 1565–1570.
- Bektas, V., Bettinger, P., Nibbelink, N., Siry, J., Merry, K., Henn, K.A. et al. (2022) Habitat suitability modeling of rare Turkeybeard (*Xerophyllum asphodeloides*) species in the Talladega National Forest, Alabama, USA. *Forests*, **13**(4), 490.
- Bellis, L.M., Pidgeon, A.M., Radeloff, V.C., St-Louis, V., Navarro, J.L. & Martella, M.B. (2008) Modeling habitat suitability for greater rheas based on satellite image texture. *Ecological Applications*, **18**(8), 1956–1966.
- Berhane, T.M., Costa, H., Lane, C.R., Anenkhonov, O.A., Chepinoga, V.V. & Autrey, B.C. (2019) The influence of region of interest heterogeneity on classification accuracy in wetland systems. *Remote Sensing*, **11**(5), 551.
- Botta-Dukát, Z. (2005) Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science*, **16**(5), 533–540.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1), 5–32.
- Brus, D.J. (2019) Sampling for digital soil mapping: a tutorial supported by R scripts. *Geoderma*, **338**(1), 464–480.
- Brus, D.J., de Gruijter, J.J. & van Groenigen, J.W. (2006) Designing spatial coverage samples using the k-means clustering algorithm. *Developments in Soil Science*, **31**(C), 183–192.
- Chala, D., Brochmann, C., Psomas, A., Ehrlich, D., Gizaw, A., Masao, C.A. et al. (2016) Good-bye to tropical alpine plant giants under warmer climates? Loss of range and genetic diversity in lobelia rhynchopetalum. *Ecology and Evolution*, **6**(24), 8931–8941.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46.
- Contreras, L., Gutiérrez, J., Valverde, V. & Cox, G. (1993) Ecological relevance of subterranean herbivorous rodents in semiarid coastal Chile. *Revista Chilena de Historia Natural*, **66**(3), 357–368.
- Corenblit, D., Baas, A.C.W., Bornette, G., Darrozes, J., Delmotte, S., Francis, R.A. et al. (2011) Feedbacks between geomorphology and biota controlling earth surface processes and landforms: a review of foundation concepts and current understandings. *Earth-Science Reviews*, **106**(3–4), 307–331.
- Culbert, P.D., Radeloff, V.C., St-Louis, V., Flather, C.H., Rittenhouse, C.D., Albright, T.P. et al. (2012) Modeling broad-scale patterns of avian species richness across the Midwestern United States with measures of satellite image texture. *Remote Sensing of Environment*, **118**, 140–150.
- Desmet, P.G. & Cowling, R.M. (1999) Patch creation by fossorial rodents: a key process in the revegetation of phytotoxic arid soils. *Journal of Arid Environments*, **43**(1), 35–45.
- Eldridge, D.J. & Whitford, W.G. (2014) Disturbances by desert rodents are more strongly associated with spatial changes in soil texture than woody encroachment. *Plant and Soil*, **381**(1–2), 395–404.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, **17**(1), 43–57.
- Estes, L.D., Reillo, P.R., Mwangi, A.G., Okin, G.S. & Shugart, H.H. (2010) Remote sensing of structural complexity indices for habitat and species distribution modeling. *Remote Sensing of Environment*, **114**(4), 792–804.
- Farwell, L.S., Gudex-Cross, D., Anise, I.E., Bosch, M.J., Olah, A.M., Radeloff, V.C. et al. (2021) Satellite image texture captures vegetation heterogeneity and explains patterns of bird richness. *Remote Sensing of Environment*, **253**, 112175.
- Feilhauer, H. & Schmidtlein, S. (2009) Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science*, **12**(4), 429–439.
- Fiedler, K., Brehm, G., Hilt, N., Süßenbach, D. & Häuser, C.L. (2008) Variation of diversity patterns across moth families along a tropical altitudinal gradient. In: Beck, E., Bendix, J., Kottke, I., Makeschin, F. & Mosandl, R. (Eds.) *Gradients in a Tropical Mountain ecosystem of Ecuador*. Berlin Heidelberg: Springer-Verlag, pp. 167–179.
- Filipponi, F. (2018) BAI2: burned area index for Sentinel-2. *Proceedings*, **2**(7), 364.
- Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B., Ratcliffe, N. et al. (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One*, **7**(4), e33751.
- Fretwell, P.T., Staniland, I.J. & Forcada, J. (2014) Whales from space: counting southern right whales by satellite. *PLoS One*, **9**(2), 1–9.

- Gabet, E.J., Reichman, O.J. & Seabloom, E.W. (2003) The effects of bioturbation on soil processes and sediment transport. *Annual Review of Earth and Planetary Sciences*, **31** (1), 249–273.
- Gitelson, A.A., Gurlin, D., Moses, W.J. & Barrow, T. (2009) A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters*, **4**(4), 45003.
- Grigusova, P., Larsen, A., Achilles, S., Klug, A., Fischer, R., Kraus, D. et al. (2021) Area-wide prediction of vertebrate and invertebrate hole density and depth across a climate gradient in Chile based on uav and machine learning. *Drones*, **5**(3), 86.
- Guillera-Arroita, G., Lahoz-Monfort, J.J. & Elith, J. (2014) Maxent is not a presence-absence method: a comment on Thibaud et al. *Methods in Ecology and Evolution*, **5**(11), 1192–1197.
- Haralick, R.M., Shanmugam, K. & Dinstein, I. (1973) Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **6**(1), 610–621.
- Hartigan, J.A. (1975) *Clustering algorithms (Probability & Mathematical Statistics)*. Hoboken: John Wiley & Sons Inc, p. 366.
- Hastings, A., Byers, J.E., Crooks, J.A., Cuddington, K., Jones, C.G., Lambrinos, J.G. et al. (2007) Ecosystem engineering in space and time. *Ecology Letters*, **10**(2), 153–164.
- Hengl, T.A. (2007) Practical guide to geostatistical mapping of environmental variables.
- Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2021) dismo: Species Distribution Modeling [Internet]. Available at: <https://cran.r-project.org/package=dismo>
- Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T. & Robinson, A. (2018) How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution*, **9**(4), 881–892.
- Huntly, N. & Reichman, O.J. (1994) Effects of subterranean mammalian herbivores on vegetation. *Journal of Mammalogy*, **75**(4), 852–859.
- Jansen, J., Woolley, S.N.C., Dunstan, P.K., Foster, S.D., Hill, N.A., Haward, M. et al. (2022) Stop ignoring map uncertainty in biodiversity science and conservation policy. *Nature Ecology & Evolution*, **6**(July), 828–829.
- Jones, C.G., Lawron, J.H. & Shachak, M. (1997) Positive and negative effects of organisms as physical ecosystem engineers. *Ecology*, **78**, 1946–1957.
- Jones, C.G., Lawton, J.H. & Shachak, M. (1994) Organisms as ecosystem engineers. *Oikos*, **69**(3), 373–386.
- Kaky, E. & Gilbert, F. (2016 Dec) Using species distribution models to assess the importance of Egypt's protected areas for the conservation of medicinal plants. *Journal of Arid Environments*, **135**, 140–146.
- Kaky, E., Nolan, V., Alatawi, A. & Gilbert, F. (2020) A comparison between ensemble and MaxEnt species distribution modelling approaches for conservation: a case study with Egyptian medicinal plants. *Ecological Informatics*, **60**, 101150.
- Keesing, F. (1998) Impacts of ungulates on the demography and diversity of small mammals in Central Kenya. *Oecologia*, **116**(3), 381–389.
- Kellenberger, B., Marcos, D. & Tuia, D. (2018) Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, **216**, 139–153.
- Keller, F., Kienast, F. & Beniston, M. (2000) Evidence of response of vegetation to environmental change on high-elevation sites in the Swiss Alps. *Regional Environmental Change*, **1**(2), 70–77.
- Kerley, G.I.H., Whitford, W.G. & Kay, F.R. (2004) Effects of pocket gophers on desert soils and vegetation. *Journal of Arid Environments*, **58**(2), 155–166.
- Koshkina, A., Grigoryeva, I., Tokarsky, V., Urazaliyev, R., Kuemmerle, T., Hölzel, N. et al. (2020) Marmots from space: assessing population size and habitat use of a burrowing mammal using publicly available satellite images. *Remote Sensing in Ecology and Conservation*, **6**(2), 153–167.
- Kuhn, M. & Johnson, K. (2013) *Applied predictive modeling*. New York: Springer.
- Kupidura, P. (2019) The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. *Remote Sensing*, **11**(10), 1233.
- Lacey, E.A. & Wieczorek, J.R. (2003) Ecology of sociality in rodents: a ctenomyid perspective. *Journal of Mammalogy*, **84** (4), 1198–1211.
- LaRue, M.A., Lynch, H.J., Lyver, P.O.B., Barton, K., Ainley, D.G., Pollard, A. et al. (2014) A method for estimating colony sizes of Adélie penguins using remote sensing imagery. *Polar Biology*, **37**(4), 507–517.
- Lavrenchenko, L. & Kennerley, R. (2016) *Tachyoryctes macrocephalus*, The IUCN Red List of Threatened Species 2016: e.T21293A115161321.
- Leutner, B., Horning, N., Schwalb-Willmann, J. & Hijmans, R.J. (2017). Package “RStoolbox.” R Found Stat Comput. Version 0.1.
- Leyer, I. & Wesche, K. (2007) *Multivariate Statistik in der Ökologie: Eine Einführung*. Berlin: Springer-Verlag, p. 224.
- Loarie, S.R., Tambling, C.J. & Asner, G.P. (2013) Lion hunting behaviour and vegetation structure in an African savanna. *Animal Behaviour*, **85**(5), 899–906.
- Massada, A.B., Syphard, A.D., Stewart, S.I., Radeloff, V.C., Massada, A.B., Syphard, A.D. et al. (2012) Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, **22**(2), 174–183.
- Mekonen, S. (2020) Coexistence between human and wildlife: the nature, causes and mitigations of human wildlife conflict around Bale Mountains National Park, Southeast Ethiopia. *BMC Ecology*, **20**(1), 1–9.

- Meyer, H. & Pebesma, E. (2021) Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, **12**(9), 1620–1633.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Naus, T. (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, **101**, 1–9.
- Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. (2017) Why choose random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, **5**(1), e2849.
- Miehe, S. & Miehe, G. (1994) Ericaceous forests and heathlands in the Bale Mountains of South Ethiopia. *Ecology and Man's Impact*. p. 206.
- Miranda, V., Rothen, C., Yela, N., Aranda-Rickert, A., Barros, J., Calcagno, J. et al. (2019) Subterranean Desert rodents (genus *Ctenomys*) create soil patches enriched in root endophytic fungal propagules. *Microbial Ecology*, **77**(2), 451–459.
- Mishra, V.N., Prasad, R., Rai, P.K., Vishwakarma, A.K. & Arora, A. (2019) Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*, **12**(1), 71–86.
- Mueller-Dombois, D. & Ellenberg, H. (1974) Community sampling: the relevé method. In: *Aims and methods of vegetation ecology*. New York: John Wiley and Sons, 45–66.
- Nottingham, A.T., Fierer, N., Turner, B.L., Whitaker, J., Ostle, N.J., McNamara, N.P. et al. (2018) Microbes follow Humboldt: temperature drives plant and soil microbial diversity patterns from the Amazon to the Andes. *Ecology*, **99**(11), 2455–2466.
- Oksanen, J., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J. & Suggests, M. (2020) Vegan: Community Ecology Package. R package version 2.5-7. Vol. 10, Community ecology package. pp. 631–7.
- Olofsson, J., Tommervik, H. & Callaghan, T.V. (2012) Vole and lemming activity observed from space. *Nature Climate Change*, **2**(12), 880–883.
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I. et al. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, **156**, 94–104.
- Phillips, S., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distribution. *Ecological Modelling*, **190**(3–4), 231–259.
- Porcasi, X., Calderón, G., Lamfri, M., Gardenal, N., Polop, J., Sabatini, M. et al. (2005) The use of satellite data in modeling population dynamics and prevalence of infection in the rodent reservoir of Junin virus. *Ecological Modelling*, **185**(2–4), 437–449.
- Pöyry, J., Böttcher, K., Fronzek, S., Gobron, N., Leinonen, R., Metsämäki, S. et al. (2018) Predictive power of remote sensing versus temperature-derived variables in modelling phenology of herbivorous insects. *Remote Sensing in Ecology and Conservation*, **4**(2), 113–126.
- R Core Team. (2021) *R: a Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.r-project.org/>
- Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, **21**(1), 24–43.
- Reichman, O.J. & Seabloom, E.W. (2002) The role of pocket gophers as subterranean ecosystem engineers. *Trends in Ecology & Evolution*, **17**(1), 44–49.
- Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N. et al. (2018) Measuring β -diversity by remote sensing: a challenge for biodiversity monitoring. *Methods in Ecology and Evolution*, **9**(8), 1787–1798.
- Schaffers, A.P., Raemakers, I.P., Sykora, K.V. & Ter, B.C.J.F. (2008) Arthropod assemblages are best predicted by plant. *Ecology*, **89**(3), 782–794.
- Sillero-Zubiri, C. & Gottelli, D. (1995) Diet and feeding behavior of Ethiopian wolves (*Canis simensis*). *Journal of Mammalogy*, **76**(2), 531–541.
- Sillero-Zubiri, C., Tattersall, F.H. & Macdonald, D.W. (1995) Habitat selection and daily activity of giant molerats *Tachyoryctes macrocephalus*: significance to the Ethiopian wolf *Canis simensis* in the Afroalpine ecosystem. *Biological Conservation*, **72**(1), 77–84.
- Šklíba, J., Vlasatá, T., Lövy, M., Hrouzková, E., Meheretu, Y., Sillero-Zubiri, C. et al. (2020) The giant that makes do with little: small and easy-to-leave home ranges found in the giant root-rat. *Journal of Zoology*, **310**(1), 64–70.
- Šklíba, J., Vlasatá, T., Lövy, M., Hrouzková, E., Meheretu, Y., Sillero-Zubiri, C. et al. (2017) Ecological role of the giant root-rat (*Tachyoryctes macrocephalus*) in the Afroalpine ecosystem. *Integrative Zoology*, **12**(4), 333–344.
- Teklehaimanot, G. & Balakrishnan, M. (2018) Distribution and population status of the endemic rock hyrax sub-species distribution and population status of the endemic rock hyrax sub-species (*Procavia capensis capillosa*) of the Bale Mountains, Ethiopia. *International Journal of Ecology and Environmental Sciences*, **44**(1), 1–10.
- ter Braak, C.J.F. (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, **69**(1–3), 69–77.
- Tuanmu, M.-N. & Jetz, W. (2015) A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, **24**(11), 1329–1339.
- Vial, F., Macdonald, D.W. & Haydon, D.T. (2011) Response of endemic afroalpine rodents to the removal of livestock grazing pressure. *Current Zoology*, **57**(6), 741–750.

- Vlasatá, T., Šklíba, J., Lövy, M., Meheretu, Y., Sillero-Zubiri, C. & Šumbera, R. (2017) Daily activity patterns in the giant root rat (*Tachyoryctes macrocephalus*), a fossorial rodent from the afro-alpine zone of the Bale Mountains, Ethiopia. *Journal of Zoology*, **302**(3), 157–163.
- Wakulińska, M. & Marcinkowska-Ochtyra, A. (2020) Multi-temporal Sentinel-2 data in classification of mountain vegetation. *Remote Sensing*, **12**(17), 2696.
- Wallis, C.I.B., Brehm, G., Donoso, D.A., Fiedler, K., Homeier, J., Paulsch, D. et al. (2017) Remote sensing improves prediction of tropical montane species diversity but performance differs among taxa. *Ecological Indicators*, **83**, 538–549.
- Wang, D., Shao, Q. & Yue, H. (2019) Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): a review. *Remote Sensing*, **11**(11), 1308.
- Warren, S.D., Alt, M., Olson, K.D., Irl, S.D.H., Steinbauer, M.J. & Jentsch, A. (2014) The relationship between the spectral diversity of satellite imagery, habitat heterogeneity, and plant species richness. *Ecological Informatics*, **24**, 160–168.
- Wood, E.M., Pidgeon, A.M., Radeloff, V.C. & Keuler, N.S. (2013) Image texture predicts avian density and species richness. *PLoS One*, **8**(5), e63211.
- Yaba, M., Mekonnen, T., Bekele, A. & Malcolm, J. (2011) Food selection and feeding behavior of giant mole rat (*Tachyoryctes macrocephalus*, Ruppell, 1842) from the sanetti plateau of bale mountains national park, Ethiopia. *Asian Journal of Applied Sciences*, **4**, 735–740.
- Yalden DW. *Tachyoryctes macrocephalus*. *Mammalian Species* 1985;(237):1–3. Available from: <http://www.jstor.org/stable/3503827>.
- Yalden, D.W. & Largen, M.J. (1992) The endemic mammals of Ethiopia. *Mammal Review*, **22**(3–4), 115–150.
- Zaniewski, A.E., Lehmann, A. & Overton, J.M.C. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**(2–3), 261–280.
- Zhang, Y., Zhang, Z. & Liu, J. (2003) Burrowing rodents as ecosystem engineers: the ecology and management of plateau zokors *Myospalax fontanierii* in alpine meadow ecosystems on the Tibetan plateau. *Mammal Review*, **33**(3–4), 284–294.
- Zhao, Z., Xiao, N., Shen, M. & Li, J. (2022) Comparison between optimized MaxEnt and random forest modeling in predicting potential distribution: a case study with *Quasipaa boulengeri* in China. *Science of The Total Environment*, **842**, 156867.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Landsat-8 and ASTER DEM based air temperature predictions for the vegetation composition analysis.

Data S2. Analysis of GRR presence by MaxEnt modelling and comparison to the three random forest-based approaches M1–M3.

Table S1. List of predictors used in the regression and classification modelling processes.

Table S2. Overview of the three model specifications.

Table S3. Selected predictors for classifications of model strategies M1, M2, M3 and for regression models of CA score predictions.

Figure S1. Constrained correspondence analysis with temperature as constraining variable (CCA1), based on the correlation matrix of 79 plant species recorded on plots with (A) presence and absence of GRR activity and (B) unconstrained axes (CA1 and CA2) for presence and absence of GRR activity.

Figure S2. Selected variables and their importance (in %) in the final regression model for CA1 (A) and CA2 (B).

Figure S3. Root mean square error (RMSE) values from the internal model iterations and number of randomly selected predictors.

Figure S4. Root mean square error (RMSE) values from the internal model iterations and number of randomly selected predictors for the classification models M1 (left) and M2 (right), training areas based on Google Earth (M1), local GPS records (M2).

Figure S5. ROC curves and calculated AUC values for all three methods training areas based on Google Earth (M1), local GPS records (M2 and M2) and M3 including CA scores.

Figure S6. ROC curves and calculated AUC values for the MaxEnt model training areas based on local GPS records for presence data and random background point data.

Figure S7. Jackknife tests of AUC (A), of test gain (B) and training gain (C) for the final MaxEnt model's variables.

Figure S8. Spatial predictions of giant root-rat (GRR) presence across the Bale Mountains.