1 # Measuring heterogeneity in soil networks: a network analysis and
2 # simulation-based approach
3

4 Natalie Davis [a, b, *]

5 J. Gareth Polhill [b]

6 M. J. Aitkenhead [b]

7 [a] Lancaster Environment Centre, Lancaster, UK.

8 [b] The James Hutton Institute, Aberdeen, UK.

9 * Corresponding author: ndavis.research@gmail.com

10

11 ## Abstract

12

13 Quantifying soil structural and ecological heterogeneity is crucial for understanding their interactions
14 and their relationships to the resilience and health of the wider ecosystem. However, a clear
15 understanding of how structural heterogeneity affects soil biodiversity is still emerging. Previous
16 work has primarily used expensive, often laboratory-based methods to quantify soil pore network
17 structure, and typically separated study of structural and biological dimensions. Here, we test
18 whether standard network metrics can be used to quantify structural heterogeneity in soil pore
19 networks, and how this network structure, along with characteristics of the consumer and resource
20 populations, affects the heterogeneity of a population of consumers. Specifically, we extract
21 simplified soil pore networks from digital photographs of soil profiles and apply established metrics
22 from network science and transport geography to quantify and compare the networks. The
23 networks are also used as the medium for an agent-based model of generalised consumers, to
24 analyse the effects of consumer and resource parameterisations and network structure. Combining
25 network analysis and simulation modelling in this way can provide insights on the structure,
26 function, and diversity possible in the soil, as well as avenues for exploring the impact of future
27 structural or environmental changes.

28 **Keywords**: soil networks; soil structure; digital soil morphometrics; agent-based model; network
29 analysis; ecological heterogeneity.

30

1. Introduction

The distribution of energetic resources in an ecosystem plays a key role in determining the complexity, quantity, and behaviour of organisms that it can support (e.g. Giller, 1996; Tews *et al.*, 2004; Roshier, Doerr and Doerr, 2008; Stevens & Tello, 2011). To understand these systems more fully, and inform actions to protect those relying on them, we must understand how resource distribution networks develop and function. For example, resource location and movement can create heterogeneity that allows species to specialise and differentiate (e.g. Bardgett, Yeates and Anderson, 2009; Tews *et al.*, 2004; Stevens & Tello, 2011), as well as cause inequality among individuals of the same species, topics that are relevant for both biologists and ecologists.

The soil provides a unique and diverse ecosystem in which to study resource distribution, and its effect on organisms. Soil structure can be defined as the collection of soil particles and pore space among them (Oades, 1993). This pore space provides access to nutrients stored on the surface of soil particles, allows for preferential flow of water through the soil matrix, and serves as the resource distribution network through which micro-, meso-, and macrofauna (soil biota) forage. As this structure determines how air, water, and soil biota move through the soil, it allows or impedes the foraging of organisms, regulates the air and water balance in the soil matrix, and affects chemical signals used in foraging, such as those of bacterial decomposition (Young and Ritz, 2009). Furthermore, crevices and niches along soil pores provide habitats for smaller microbes to avoid predation, and the overall spatial and temporal heterogeneity of the soil environment allows for resource partitioning and habitat specialisation that limits the effect of competitive exclusion (Bardgett, Yeates and Anderson, 2009). This is similar to the hypothesised effect of heterogeneity in aboveground habitats (e.g. Tews *et al.*, 2004; Stevens and Tello, 2011).

Soil biota in turn can increase the porosity of soil, through burrowing and consuming organic matter, and releasing gases during decomposition, which create or expand soil pores (Kravchenko and Guber, 2017). Additionally, there is evidence of feedbacks between the soil biota and aboveground plant communities (e.g. Baer *et al.*, 2005; Wijesinghe, John and Hutchings, 2005; García-Palacios *et al.*, 2012), which alter soil structure as their roots burrow in pore networks, and roots and hyphae bind and stabilise soil particles (Vezzani *et al.*, 2018). Through regulating movement and diffusion of water and energy resources, gases, and fauna in the soil matrix; providing habitat; and mediating biological feedbacks; soil structure is the foundation of all earth systems.

Past efforts to quantify and model soil structure have primarily focussed on measuring the stability of soil, by utilising soil aggregate size distribution as a measure of structure. While this does represent the spatial distribution in the soil, it is not a complete representation of physical properties (see e.g. Young, Crawford and Rappoldt, 2001). Methods for visualising the pore network within a soil sample include CT scans and X-ray tomography, NMR, and SPECT scanning, mostly for the purposes of measuring solute flow and transport processes (see review in Young, Crawford and Rappoldt, 2001). Gas diffusion and solute flow have also been examined with modelling approaches, including neural networks, Boolean models, and cellular automata. Additionally, fractal modelling has also been used successfully to quantify the degree of connectivity, tortuosity, and heterogeneity of the soil pore network (Crawford, Ritz and Young, 1993), three characteristics that have also been associated with a higher level of heterogeneity of resource distribution in generalised networks (Davis *et al.*, 2020).

Overall, past work has highlighted the important connections between soil function and structure, especially of the pore network. Much of this work has been done from a geometric or hydrological perspective, however, rather than an energetic one, leading to criticisms of unrealistic separation of

76 soil physics and biology, and emphasis on the importance of integrating these spatially explicit
77 approaches in future soil ecology research (Bardgett, Yeates and Anderson, 2009). Additionally,
78 much of the imaging equipment required for the techniques above is large and expensive, requiring
79 soil samples to be brought back to the laboratory. Even if disruption to the soil structure during
80 extraction and transport is minimised, these methods are more suitable for intensive analyses of
81 individual samples and smaller areas.

82 In contrast, some previous work has focussed on quantifying the structure of soil networks through
83 image morphology techniques applied to a photograph of a sample, in order to extract the relevant
84 network (e.g. Velde, Moreau and Terribile, 1996; Gargiulo, Mele and Terribile, 2013; Hartemink and
85 Minasny, 2014). This method will not reveal the network at the same level of detail as CT scans or X-
86 ray tomography, and may require use of resins and dyes to highlight the underlying structure
87 (Hartemink and Minasny, 2014). Good arguments have also been raised regarding the importance of
88 analysing soil structure from a three-dimensional perspective, as it reveals considerably more about
89 the habitat of the soil (Young and Ritz, 2009). However, if rotational invariance is assumed,
90 connectivity and structure of a two-dimensional sample can be assumed representative of any
91 random two-dimensional plane taken through the system. This inference does not consider lateral
92 flow, which would undoubtedly play an important influence in sloping areas by transporting
93 nutrients laterally through the soil. In areas where the surface is flat and lateral flow effects are
94 negligible, standard network metrics could usefully approximate soil structure and provide insights
95 into its effect on biotic and abiotic processes within an environment.

96 Moreover, the two-dimensional techniques are considerably more portable and feasible than the
97 three-dimensional techniques, and processing time can be significantly faster. Image analysis
98 methods, particularly those that can be performed entirely in the field, could potentially be
99 incorporated into software for use by farmers and researchers who may otherwise not have access
100 to the equipment necessary for the more costly and lab-intensive methods of quantifying structure
101 (e.g. Aitkenhead *et al.*, 2016). These methods could also act as preliminary investigations to highlight
102 potential areas of future exploration using more intensive analyses.

103 In this paper, we test whether standard network metrics can be used to quantify structural
104 heterogeneity in soil pore networks, and how this network structure, along with characteristics of
105 the consumer and resource populations, affects the heterogeneity of a population of consumers.
106 Specifically, we develop a method for extracting approximate soil networks from digital photographs
107 using image morphology techniques, then apply metrics from network science and transport
108 geography to quantify and compare the networks. The networks are also used as the medium for an
109 agent-based model (ABM, which in ecology is more typically known as an individual-based model,
110 e.g. Grimm *et al.*, 2006), where the agents represent generalised consumers who explore the
111 network and consume food resources. The variation in population size and resource consumption is
112 compared across simulations, to evaluate how both the network structure and simulation
113 parameters affect outcomes of the biotic community. This methodology is applied to a case study
114 using soil images from two test sites in Aberdeenshire, United Kingdom.

115

2. Methods

2.1 Soil image collection

Images were taken at two field sites in Aberdeenshire, United Kingdom. The first site had a brown forest soil, or Cambisol (Fig A1a); photographs were taken from seven locations in both forested and converted agricultural areas. The second site had a sandy beach soil, or Arenosol (Fig A1b); photographs were taken at five locations across a dune area, with sparse grass and shrub cover. Neither Cambisols nor Arenosols are highly developed, but Cambisols have some diagnostic features, while Arenosols are lacking diagnostic features and are defined only on the basis of being coarse (sandy) textured (FAO, 2015). The known difference between the two soils therefore provides a basis for preliminarily evaluating the methodology. Additionally, both soil types can be assumed to show limited profile variation with depth on the scale of the observed soil profile sections under study (FAO, 2015), such that a uniform network extraction method and analysis can be applied across the image. The specific sampling sites were also chosen as they provided easy access to multiple sampling locations for both soil types. As this work is an exploratory proof-of-concept, an exhaustive sampling regime across different soil, land use, and geographic regions was not undertaken.

The methodology for taking pictures was replicated from Aitkenhead *et al.* (2016). In summary, the photographs were taken of the soil profile of shallow (30 cm) pits in flat areas, using an angle that provided maximum natural light and minimum shadow (Fig A1a, b). No artificial lighting was required during photography. Additionally, each photograph included a 10 cm x 6 cm colour correction card within the frame. Colour correction has been used in past work (e.g. Aitkenhead *et al.*, 2016) to correct colour variation in ambient lighting. However, in this work we were only interested in overall intensity, rather than light balance, so the cards were inserted into the image to provide a spatial scale reference for future work.

In Fig A1a and A1b, the white area is an excised section of the image that is larger than the correction card. The imaging was taken with the card viewed straight on, without distortion, so the image distortion and impact on length of edges is not an issue. Extracting an area larger than the correction card also attempted to eliminate shading effects around the card. This may not have been done sufficiently to eliminate all the shading, possibly introducing some additional dark pixels and error into the network metric calculations. However, taking multiple pictures within the same profile can provide some robustness against this. Future work should attempt to remove this effect from near the correction card.

In total, seven Cambisol profiles and five Arenosol pits were used for each soil type, with several images taken of the profile of each pit. In taking multiple images from each soil pit, we moved the camera slightly to present different viewing angles and thus generate different images. This was done to compare the robustness of extracted networks from each pit (see Section 2.2), and replication within pit was considered in all statistical analyses.

2.2 Network extraction

To extract the approximate soil network structure from the photographs, the photographs were converted to text files containing the red, green, and blue (RGB) triplet values for each pixel. All non-soil pixels were then identified as those whose triplet values exceeded the ranges expected for soil particles, based on the average of the rest of the image. Using the average to determine this threshold customised it slightly for each image, so that outliers such as roots and rocks specific to that sample were captured, but samples having an overall more reddish tone were not stripped

160 completely. The identified non-soil particles were removed, and variations in brightness across the
161 remaining pixels were standardised using the mean pixel intensity.

162 As soil structure and porosity are only loosely related, soils of the same porosity can have different
163 structural properties. A common assumption made is that soils, unless compressed/compacted,
164 have up to 50 % pore space. As the pixel resolution of the images here is between 0.3 – 0.5 mm, and
165 therefore much higher than the smallest pore space possible (sub-micron scale), it follows that the
166 pore space actually visible is less than this 50 %. An evaluation of the distribution of pixel values
167 showed that for soil profile images used in this study, the greatest change in the distribution
168 occurred around a pixel intensity where 30 – 40 % of the pixels were below this value (Fig A2). We
169 have therefore assumed that 30 % of the soil is 'void' (i.e. dark pixels). Therefore, the darkest 30 % of
170 the soil pixels were retained as pores, and the image was inverted to convert these darker pixels to
171 white, and vice versa (Fig A1e, f). The images from the same profile were visually compared after
172 thresholding and showed a high degree of agreement in the pores identified (e.g. Fig A3). Network
173 outlines were then drawn through a process known in image morphology as 'skeletonization,' where
174 lines of white pixels were iteratively stripped down until they were all one pixel in width (Fig. A1g, h).
175 We then mapped the networks to a list of links, which were series of pixels that were more than one
176 pixel long, and nodes, defined as junction points between two or more links. Redundant links
177 between nodes were removed.

178 For simplicity, all links in the final networks were represented with straight lines along the shortest
179 distance between two nodes. This lost some of the details of the topology, such as pore size and
180 shape. However, this work intended to create an abstraction of the network taken from the soil,
181 rather than replicate and analyse the exact soil structure itself. This emphasised overall soil
182 structural characteristics and heterogeneity, rather than modelling how specific transport processes
183 and biological activities would occur. Replicating the exact soil network would also have markedly
184 increased the computational burden, as link lengths would have had to be calculated through pixel-
185 counting rather than the Euclidean geometry measuring shortest distances. As many of the links as
186 represented were quite short (see Results), the difference between the true link length and the
187 shortest distance between nodes was assumed to be negligible. Currently, we assume that the
188 method requires further validation and improvement to provide a measure of soil structure that can
189 be used in soil science or pedological characterisation of the soil. We also assume however, that the
190 method, while not perfect in its current form, provides sufficient quality of network data to allow
191 simplified networks to be extracted and analysed, and used as the basis for simulations.

192 The process of rendering the network also identified which sections of the network were fully
193 connected, and which nodes were part of disconnected subnetworks (Fig. A1i, j). An outline of the
194 image morphology process, and images of each step, are available in Appendix 1.

195 2.3 Network analysis
196 Two types of analysis were used to quantify the heterogeneity present in the soil network images.
197 The first involved applying metrics adapted from network science and transport geography to
198 measure structural characteristics of the abstracted networks, which allows for easy comparison
199 among soil types. These were calculated using R v4.0.2 (R Core Team, 2020), including the packages
200 igraph, qgraph, and sp (Pebesma and Bivand, 2005; Csardi and Nepusz, 2006; Epskamp *et al.*, 2012;
201 Bivand, Pebesma and Gomez-Rubio, 2013). All additional data analysis and visualisations were also
202 done in R, using the packages ARTool v0.10.7 (Kay and Wobbrock, 2020; Wobbrock *et al.*, 2011),
203 emmeans v1.5.0 (Lenth, 2020), lmerTest v3.1.2 (Kuznetsova, Brockhoff, and Christensen, 2017),
204 dunn.test v1.3.5 (Dinno, 2017), rcompanion v2.3.25 (Mangiafico, 2020), dplyr v1.0.0 and ggplot2

205 v3.3.2 packages (Wickham, 2016; Wickham *et al.*, 2019). The scripts for calculating network metrics
206 are available at (Davis, 2020).

207 A brief description of each of the metrics chosen is given in Table 1. These were chosen to measure
208 the size, connectivity, and structural heterogeneity of the networks from a range of node-centric,
209 link-centric, and global perspectives, to obtain a broad picture how the networks may differ. The
210 metrics chosen also minimised assumptions about inaccessibility of the soil matrix between pores:
211 for example, the convex hull area was chosen over the concave hull area as the former is a more
212 generous estimate of the spatial area.

213 **Table 1. The name and description of the metrics used to analyse the soil networks.**

Metric name	Description	Type of measure	Reference
Mean and standard deviation (SD) of link length	Quantifies the typical length and variability of lengths included within the network.	Size	N/A
Beta index	The ratio of links to nodes.	Connectivity	Rodrigue, 2017
Gamma index	Number of observed vs. possible links: nLinks / (nNodes * (nNodes – 1))	Connectivity	Rodrigue, 2017
Diameter	The length of the longest geodesic (shortest path between two nodes) in the network – the shortest path between the two most distant nodes.	Size	Rodrigue, 2017
Node count	The number of nodes in the network.	Size	Barabási, 2016
Edge count	The number of edges (links) in the network.	Size	Barabási, 2016
Mean node degree	Mean number of links per node.	Connectivity	Barabási, 2016
Cost	The total length of the network measured in real transport distances.	Size	Rodrigue, 2017
Global reach centrality (GRC)	The difference between the maximum and average local reach centrality (LRC), where the LRC is the nodes that a given node can connect to, weighted by distance (here, spatial distance).	Structure, connectivity	Adapted from Mones, Vicsek and Vicsek (2012)
Mean convex hull area	The area of a polygon that minimally encompasses every node in the network.	Size, connectivity	Rockafellar, 1970
Network density	The ratio of the number of nodes to the convex hull area.	Structure	N/A

214 As introduced, the imaging method and metrics used here are two-dimensional (2D), and we have
215 been unable to find literature describing characterisations of three-dimensional (3D) soil structure
216 metrics based on two-dimensional imaging. Aitkenhead *et al.* (1999) derived 3D models of soil pore
217 systems based on 2D metrics but did not compare the two sets of structural metrics. Future work
218 would be necessary to determine the extent to which 3D variation in soil structural metrics
219 correlates to the variation seen in 2D. Here, we are assuming that it does correlate, and that this
220 allows 2D imaging to provide structural metrics representative of different soil types.

221 We calculated each metric for each of the networks, which contained all nodes and links in the
222 image, hereon called 'main networks.' We also calculated each metric for each of the disconnected
223 subnetworks within the main networks, hereon called 'subnetworks.' As the distributions of metrics
224 in the main soil networks had similar variance across soil types and relatively normal distributions,
225 these were compared with nested ANOVA, using profile ID as a random effect to account for
226 replication. The distribution of metrics across the subnetworks did not meet the assumptions for
227 classical ANOVA, so non-parametric Aligned-Ranks Transformation (ART) ANOVAs were used instead,
228 also with profile ID as a random effect.

229 ## 2.4 Agent-based model overview
230 The second analytical method used a simulated population of consumers to explore each network,
231 using the resulting heterogeneity in consumer resource stocks to further elucidate the heterogeneity
232 of the network. This provided a more functional perspective, alongside the structural quantification
233 of the network metrics. The purpose was to investigate the structure's generalised impact, rather
234 than test the precision of this model in predicting outcomes for real species. Therefore, rather than
235 using parameterisations that reflected specific species or groups, five generic model species with
236 different sets of values for each trait were used, similarly to e.g. Polhill and Gimona (2014).

237 The same approach was taken for resources, with three sets of resource bases of different
238 combinations of maximum capacity and maximum growth rates. Resources were assumed to be
239 located at nodes within the network, as identified during the extraction process (see Section 2.2).
240 Food resources in real soil networks are located throughout the soil matrix, but are often
241 concentrated in 'hotspots' such as those created by plant roots and decomposition processes
242 (Ettema and Wardle, 2002), which would be represented in the networks here as nodes. As
243 exploring the effect of size of the generic species was not in scope for the work here, only the most
244 accessible areas of the network were treated as potential resources.

245 A brief description of the model purpose, variables, and processes is presented below, following the
246 Overview, Design concepts, and Details (ODD) protocol (Grimm *et al.*, 2006, 2010). The full ODD
247 document, including description of design concepts, initialisation, input data, and sub-models, is
248 available in Appendix 2. The model source code, written in NetLogo 6.1, is available in the Modelling
249 Commons repository as "Soil network simulation" (Davis and Polhill, 2020).

250 **2.4.1 Overview section of Overview, Design Concepts, and Details (ODD)**

251 I.   Model purpose
252   a. The model is designed to be an analytical tool to explore the heterogeneity in
253   resource supply potential of a network by populating it with idealised energy-
254   consuming agents, and to quantify the effects of consumer, resource, and network
255   characteristics on resulting consumer population outcomes.
256 II.   Entities, state variables, and scales
257   a. Consumer entities

258   i.   State variables

Property	Description
Location	The resource on which the consumer is located
Target location	The resource to which the consumer will move next
Active?	Whether a consumer is active (or dead)

259   ii.   Parameters

Property	Description
Basal metabolism	How much resource an agent needs per day to stay alive
Active metabolism	How much resource an agent uses with each step
Resource stock	How much resource an agent has consumed but not metabolised
Consumption rate	Maximum number of resource units that an agent takes from a resource it visits, per timestep
Spawn energy	How much energy an agent requires to spawn (depletes this quantity from stocks and passed to offspring as starting quota)

260

261   b.   Resource entities
262   i.   State variables

Property	Description
Current supply	The current quantity of resource at this point

263
264   ii.   Parameters

Property	Description
Resource capacity	How much energy is stored in a resource when it is full
Regrow rate	The amount the resource regrows each timestep

265

266   c.   Link entities
267   i.   State variables

Property	Description
Length	The length of the link - determines energy and time required to traverse it

268

269   d.   Scales

Property	Description
Timestep	A single unit of time in the model, defined as that which is required for consumers to move 1 pixel (approximately 0.3 – 0.5 mm), and for which they require `basal-metabolism` units of energy.
World size	400 x 500, determined by the size of the soil networks used as the environment.

270

271   III.   Sequence of events

272   a. Consumers start on random nodes around a pre-specified network, where nodes are
273   resource patches.
274   b. Consumers move around the network randomly following links. If they find a
275   resource patch, they consume as much as they can from it, and the patch depletes.
276   i. Consumers require `basal-metabolism` units of resource per timestep. If
277   they do not consume this resource, they die.
278   ii. Consumers can stay put on a resource and consume it (`consumption-`
279   `rate` units consumed per timestep), but it depletes, and if there is no more
280   resource there then they move on.
281   iii. Consumers metabolise `active-metabolism` units of resource per patch
282   of link that they cross.
283   iv. If there is more than one agent on a resource patch, they each take
284   `consumption-rate` units per timestep, or split the remainder if there is
285   not enough resource remaining for them to each get `consumption-rate`
286   units.
287   c. If consumers have twice as much energy as the set `spawn-energy`, they can
288   spawn new consumers (who take the same amount of `resource-stock` from
289   their parent that the parent started with, so now parent and offspring both have the
290   same `resource-stock`).
291   d. Resources regrow at a constant rate (`regrow-rate`) per timestep, up to their
292   maximum capacity (`resource-capacity`).

2.4.2 Sensitivity analysis

To determine the sensitivity of the ABM to input parameters, and the robustness of any emergent
patterns of heterogeneity, we performed an extensive sensitivity analysis following
recommendations in the agent-based modelling literature. This is detailed in Appendix 3. Table 2
shows the final parameter values used for the consumer populations, resource populations, and
general model. In the actual simulation runs, each combination of the five consumer parameter sets,
and three resource parameter sets, was tested against each network architecture, resulting in 8700
total runs including replicates.

Table 2. Final values for (a) consumer, (b) resource, and (c) general simulation parameters.

     a. Consumer parameters

Parameter	Consumer type				
	High metabolism, high consumption, high spawn energy (HHH)	Low metabolism, low consumption, low spawn energy (LLL)	Low metabolism, moderate consumption, low spawn energy (LML)	Low metabolism, moderate consumption, moderate spawn energy (LMM)	Moderate metabolism, moderate consumption, moderate spawn energy (MMM)
Basal metabolism	3	1	1	1	2
Active metabolism	3	1	1	1	2
Consumption rate	10	5	7	7	7
Spawn energy	100	50	50	75	75
Initial resource stock	30	30	30	30	30

     b. Resource parameters

	Resource type		
Parameter	High capacity, low growth (HL)	Moderate capacity, moderate growth (MM)	Low capacity, high growth (LH)
Maximum resource capacity	50	35	20
Maximum regrow rate	10	15	20

305

306   c.   General parameters

Parameter	**Value**
Initial population size	500 consumers
Length of simulation	2000 timesteps

307

308 **2.4.3 Analytical method**

309 At each time step, the ABM calculated five metrics (Table 3), including measures of centre and
310 spread of consumer resource stocks, the final population size, and two additional inequality metrics:
311 the Gini coefficient and a modified form of the Shannon entropy. The latter estimates the
312 differential entropy of a continuous variable, by discretising the distribution into bins (Appendix 4).
313 These metrics were chosen to include measures of absolute and relative inequality, and a measure
314 of evenness common to ecology. As the distributions of each metric across the soil types did not
315 meet assumptions of most parametric tests, mixed-effects ART ANOVAs with profile ID as a random
316 effect were again used to quantify how the outcome metrics differed, for each combination of
317 resource and consumer population parameters and soil type. As the final population size and the
318 entropy of consumer resource stocks both showed variance not fully explainable by consumer or
319 resource population parameters, these were also tested with Kruskal-Wallis tests comparing them
320 across profile IDs and soil types. The significantly different pairs of profiles were identified with Dunn
321 post-hoc analysis. All data processing, analysis, and visualisation was done in R, using the packages
322 listed previously, as well as the entropy v1.2.1 (Hausser and Strimmer, 2014) and ineq v0.2.13
323 packages (Zeileis, 2014).

324

325 **Table 3. The name and description of outcome variables calculated for the agent-based model**
326 **(ABM).**

Variable name	Description
Mean consumer resource stock	The mean of the resource stocks held by all active consumers. Units are the same as those of the quantity measured.
Standard deviation consumer resource stock	The square root of the sum of squared absolute differences between each observation and the mean, normalised by the number of observations (minus one, to allow for sample estimation). Units are the same as those of the quantity measured. $$s = \sqrt{\frac{\sum |x_i - \bar{x}|^2}{n - 1}}$$
Gini coefficient consumer resource stock	Measures the deviation of a population from perfect equality. Mathematically, it can be calculated as half the relative mean absolute difference, or half the average absolute difference between all pairs of the population, divided by the average of the population to normalise. Unitless.
Entropy consumer resource stock (Shannon index)	Measures the amount of information that would be needed to represent the state of the system. Specifically, it is the negative sum of the probability of a consumer's resource stock occurring within a given range, and the log of that probability, normalised by the maximum value (log n). This is the discretised formula for entropy. The units depend on the base of the log: here we use base 2 (units: bits). $$H(X) = \frac{-\sum f(x_i) \log_2 \frac{f(x_i)}{w(x_i)}}{H_{max}}$$
Final population size	Count of currently active ('alive') consumers.

327

3. Results

3.1 Network metrics

The network metrics showed several significant differences between the Cambisols and Arenosols, with the Cambisols having higher values for most metrics measuring size and structure. These are summarised in Table 4 and Fig 1.

Table 4. (a) Estimated marginal means, standard errors, and outcomes for mixed-effect nested ANOVAs comparing network metrics between Cambisol and Arenosol main soil networks, and (b) medians and 95 % confidence intervals and results of mixed-effect nested Aligned-Ranks Transformation (ART) ANOVAs comparing Cambisol and Arenosol subnetworks. Shown in (a) are the Type II Wald Chi-square statistic and p-values for models comparing each network metric across soil types. Profile ID was included as a mixed effect; its log-likelihood ratio test (LRT) statistic and *p*-value are also shown. Both the Chi-square and LRT used one degree of freedom to compare soil types. Estimated marginal means and standard errors were calculated from ANOVAs. In (b) ART ANOVAs were used as the data were non-normal; profile ID was also included as a mixed effect. Shown are Type III Wald F tests with Kenward-Roger degrees of freedom. The asterisks designate level of significance: $p < 0.1$: ·, $p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***. Descriptions of the metrics are in Table 1.

345   a.   **Main networks**

346

	Arenosols (n = 25)		Cambisols (n = 25)		Soil type		Profile ID (random effect)	
	Est. marginal mean	SE	Est. marginal mean	SE	χ^2	*p*	LRT	*p*
No. of nodes	2670.000	9.570	3326.000	10.070	7.809	0.005 **	21.384	< 0.001 ***
No. of links	4225.000	344.000	5581.000	293.000	9.058	0.003 **	19.280	< 0.001 ***
Mean node degree	3.140	0.055	3.350	0.047	9.098	0.003 **	9.005	0.003 **
Mean link length	3.570	0.023	3.700	0.022	17.667	0.000 ***	0.414	0.520
SD link length	2.110	0.025	2.210	0.025	9.815	0.002 **	0.000	1.000
Gamma index	0.001	0.000	0.001	0.000	4.838	0.028 *	20.231	< 0.001 ***
Beta index	1.570	0.028	1.680	0.024	9.098	0.003 **	9.005	0.003 **
Diameter	154.000	11.100	201.000	10.700	10.017	0.002 **	0.141	0.708
Cost	15180.000	1257.000	20600.000	1072.000	10.840	0.001 ***	16.153	< 0.001 ***
Global reach centrality	0.002	0.000	0.002	0.000	0.049	0.825	10.923	< 0.001 ***
Convex hull area	166940.000	2949.000	167086.000	2514.000	0.001	0.970	11.730	< 0.001 ***
Network density	0.016	0.001	0.020	0.001	8.037	0.005 **	16.547	< 0.001 ***
No. of subnetworks	163.000	3.382	158.000	3.236	0.009	0.924	6.302	0.012 *

347

348

349   **b. Subnetworks**

350

	Arenosols (n = 3906)			Cambisols (n = 3834)			ANOVA	
	Median	Lower CI	Upper CI	Median	Lower CI	Upper CI	F	p
Number of nodes	11.000	11.000	11.000	11.000	11.000	11.000	$F(1, 8.614) = 1.542$	0.247
Number of links	15.000	14.000	15.000	15.000	15.000	16.000	$F(1, 8.443) = 3.793$	0.085 ·
Mean node degree	2.670	2.640	2.670	2.710	2.670	2.750	$F(1, 7.971) = 9.239$	0.016 *
Mean link length	3.190	3.170	3.220	3.250	3.230	3.280	$F(1, 7.011) = 11.322$	0.012 *
SD link length	1.620	1.600	1.640	1.670	1.640	1.690	$F(1, 8.334) = 3.648$	0.091 ·
Gamma index	0.132	0.127	0.136	0.133	0.128	0.136	$F(1, 8.720) = 0.355$	0.567
Beta index	0.133	0.132	0.133	0.136	0.133	0.138	$F(1, 7.971) = 9.239$	0.016 *
Diameter	19.000	18.600	19.500	19.300	18.800	19.700	$F(1, 7.957) = 1.664$	0.233
Cost	46.700	45.200	48.700	49.200	47.500	51.600	$F(1, 8.161) = 4.717$	0.061 ·
Global reach centrality	0.058	0.057	0.059	0.058	0.057	0.059	$F(1, 7.645) = 0.554$	0.479
Convex hull area	63.500	59.800	67.000	62.500	58.500	66.000	$F(1, 8.132) = 0.385$	0.552
Network density	0.171	0.167	0.176	0.172	0.168	0.179	$F(1, 8.117) = 0.013$	0.913

351

352   **a.**
353

b.
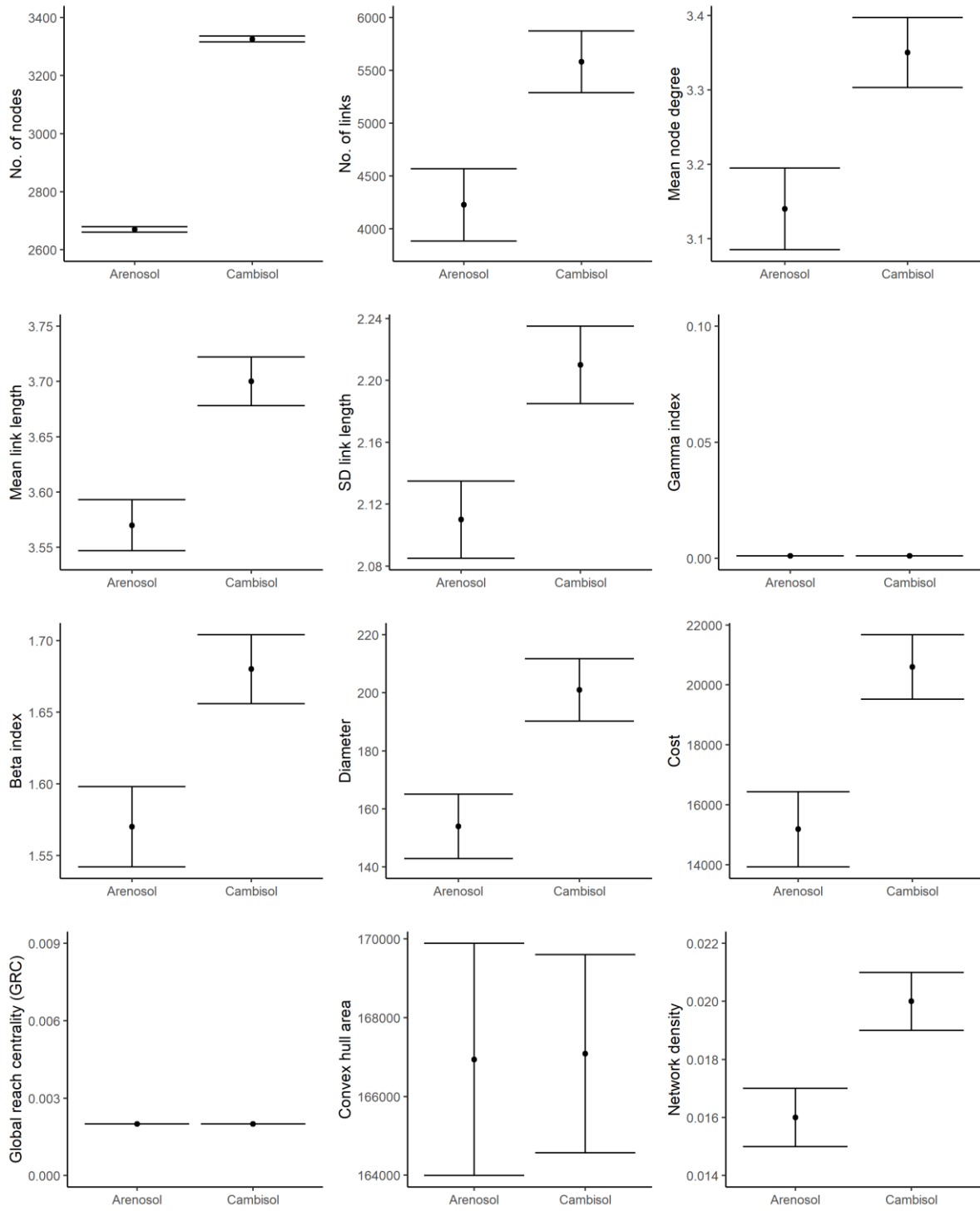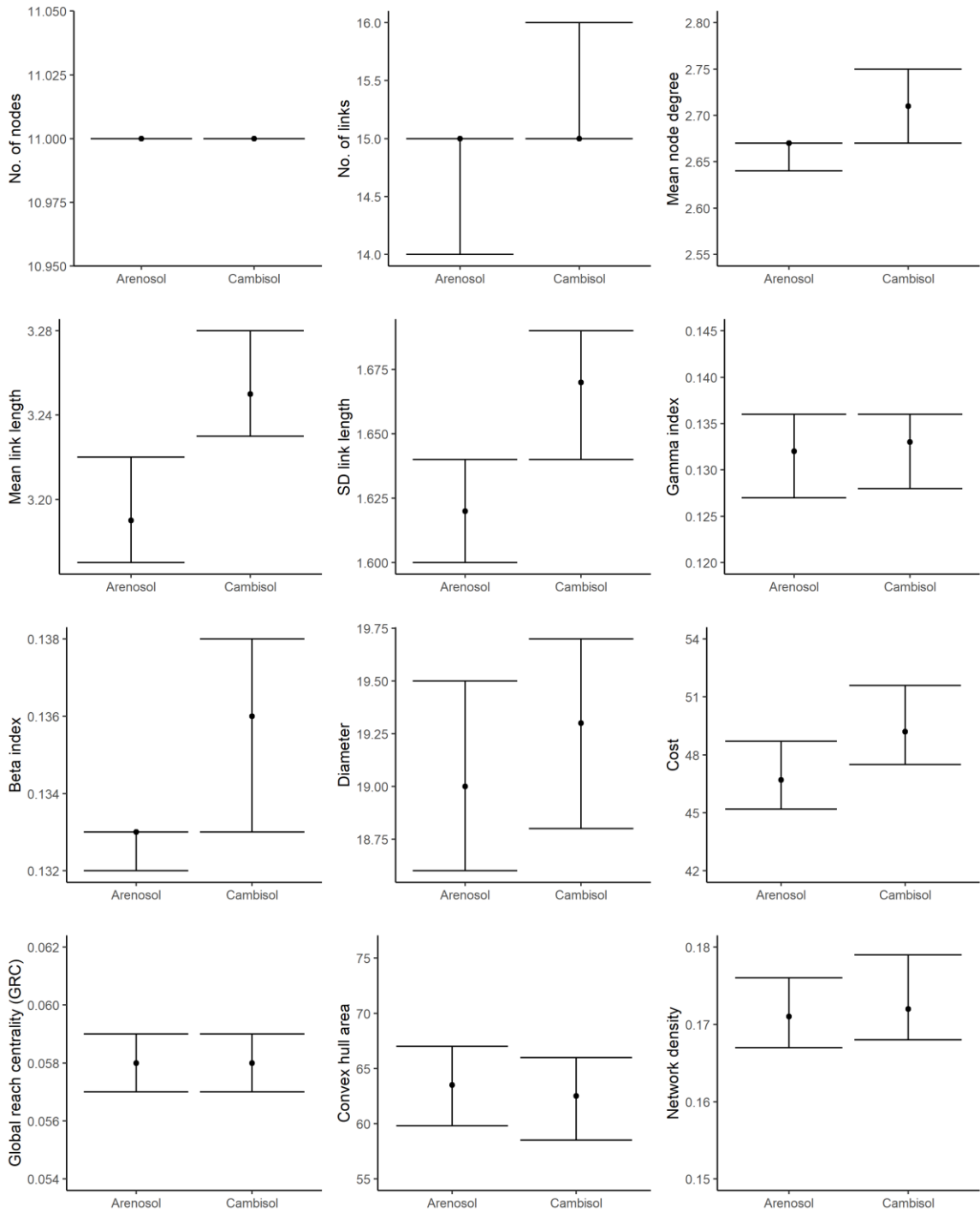
Figure 1. The distribution of each network metric by soil type, for (a) main soil networks and (b)
subnetworks. The point and error bars in (a) represent the estimated marginal mean and standard
error for that network type and soil type, as determined by the ANOVAs (Table 4a), and the point
and error bars in (b) represent the median and upper and lower 95 % confidence intervals,
respectively. Descriptions of the network metrics are in Table 1.

362 At the main network level, the networks extracted from the Cambisols had significantly more nodes
363 and links, a larger mean node degree and standard deviation of link length, and longer mean link
364 length (Table 4a).These networks also had a higher beta index, higher cost, and higher density. While
365 the main networks of the two soil types had significantly different gamma indexes, the absolute
366 difference in the estimated marginal means between the two soil types was negligible ($< 10^{-3}$) (Table
367 4a, Fig 1a). At the subnetwork level, Cambisol networks had longer mean link length, and higher
368 mean node degree and beta index (Table 4b). While not significant, Cambisol subnetworks also had
369 noticeably larger number of links and standard deviation of link length, and higher cost (Fig 1b).

370 To control for the effect of replication on the significance, the profile ID was included in the ANOVAs
371 as a mixed effect. This was significant for all metrics except mean and standard deviation of link
372 length and diameter. Most profiles within each soil type at the main network level showed low
373 absolute variation across the networks extracted from each however, and noticeably higher metric
374 values for Cambisols than Arenosols (Fig A7a). At the subnetwork level, the distributions were quite
375 similar across all profiles, but the Cambisol profiles showed more frequent and higher outliers.
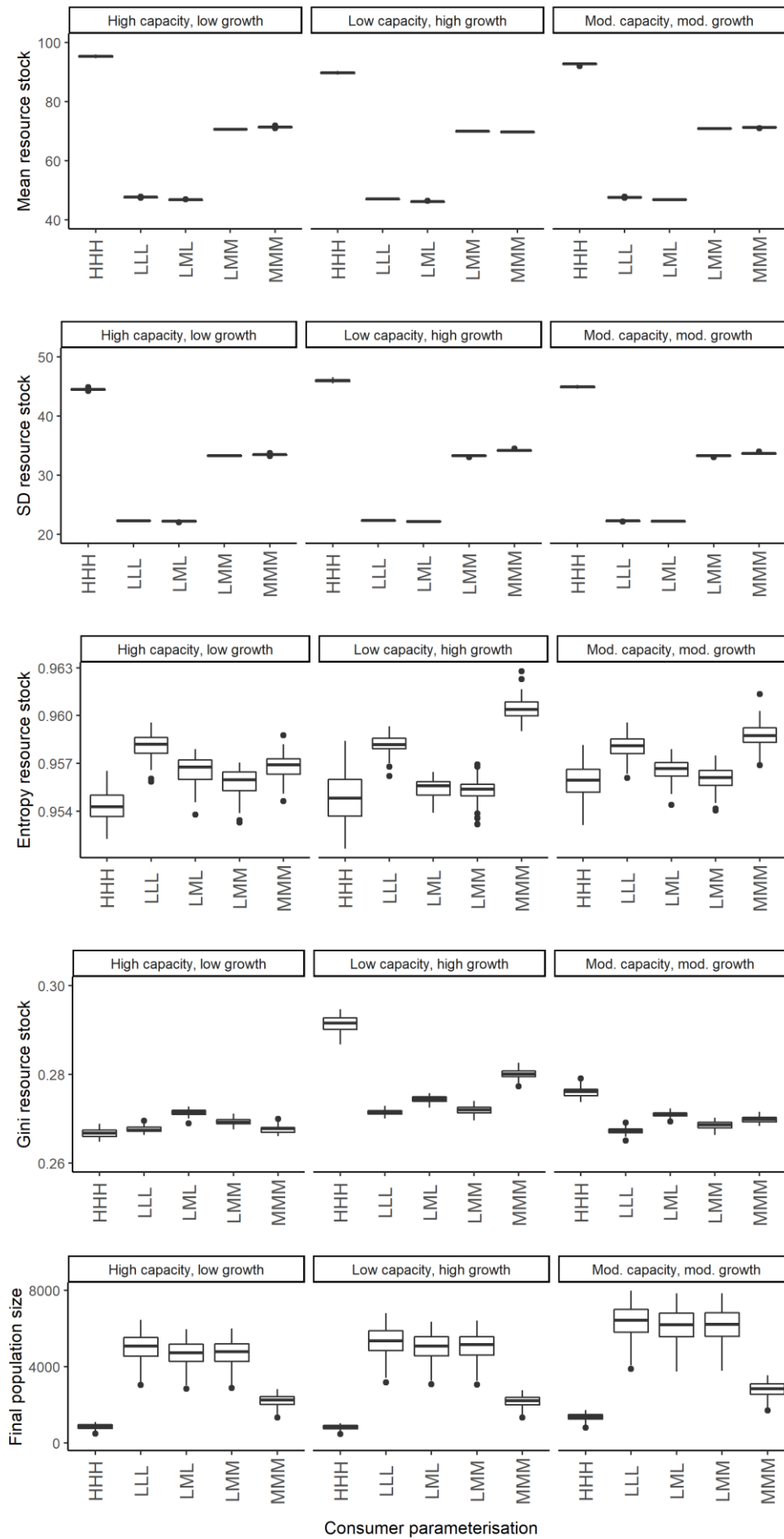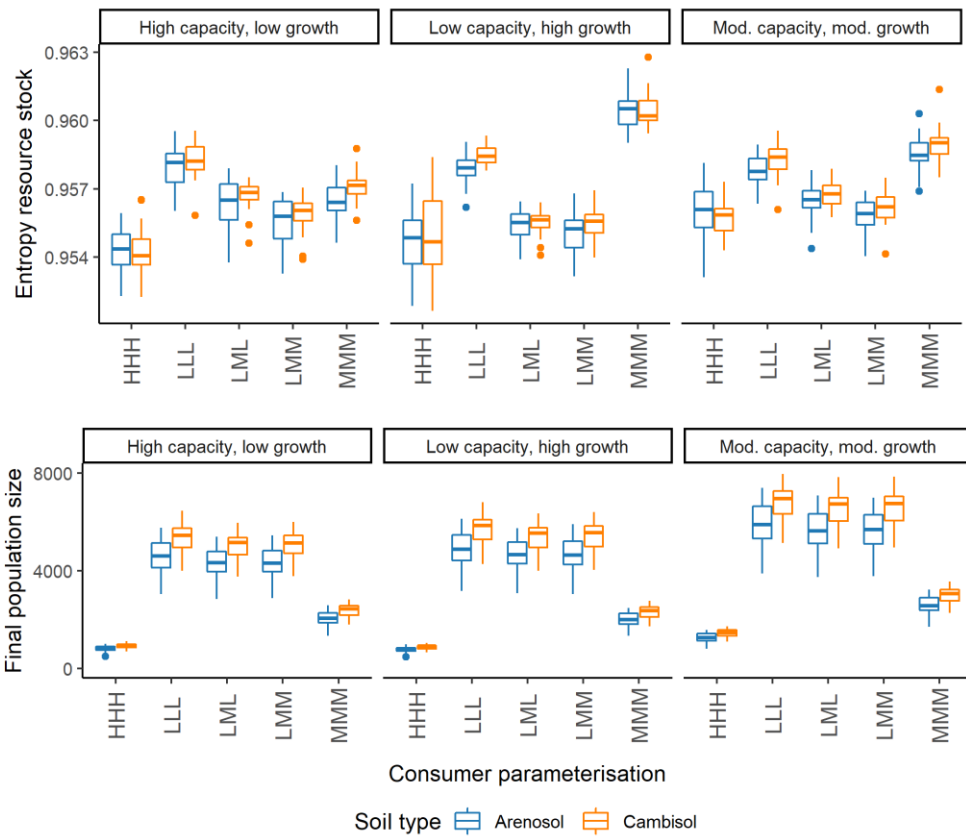
376 ## 3.2 Agent-based model
377 The ABM results showed significant differences across the different combinations of
378 parameterisations and soil types, summarised in Tables 5 and 6 and Fig 2. The simulations run on the
379 Cambisol networks had significantly higher final population sizes (Tables 5 and 6, Fig 2b), and
380 interactions between soil type and consumer and resource parameterisation were significant for
381 several outcome variables (Table 6).

382

383 **Table 5. The medians, first and third quantiles for agent-based model (ABM) outcome variable**
384 **values across the two soil types.** These values represent the overall results across all consumer and
385 resource parameterisations. Descriptions of the variables are in Table 3.

	Arenosols (n = 375)			Cambisols (n = 375)		
	Median	1st Quantile	3rd Quantile	Median	1st Quantile	3rd Quantile
Mean resource stock	69.970	47.086	71.391	69.940	47.077	71.397
SD resource stock	33.266	22.262	34.123	33.300	22.276	34.112
Entropy resource stock	0.956	0.955	0.958	0.957	0.956	0.958
Gini resource stock	0.271	0.268	0.273	0.271	0.268	0.274
Final population size	4049.728	1967.921	5102.043	4890.158	2279.1560	5842.834

386   a.

387   **b.**

388   **Figure 2. Distributions of (a) each agent-based model (ABM) outcome variable, grouped by**
389   **resource parameterisation (columns, labelled at top) and consumer parameterisation (x axis**
390   **within columns), across both soil types, and (b) ABM outcome variables that were significantly**
391   **affected by soil type (represented by colour), grouped by resource parameterisation (columns)**
392   **and consumer parameterisation (x axis within columns).** The three-letter consumer
393   parameterisation codes refer to the metabolism, consumption rate, and spawning threshold,
394   respectively, where H is high, M is medium, and L is low. Descriptions of the resource and consumer
395   parameterisations are in Table 2, and descriptions of the outcome variables are in Table 3.

396

397   The ART ANOVAs showed that measured outcomes all differed significantly across consumer
398   parameterisation, resource parameterisation, and consumer-resource parameterisation interactions.
399   Final population size differed significantly by soil type, soil type-resource parameterisation
400   interaction, and soil type-consumer parameterisation interaction. Mean resource stock also differed
401   significantly by soil type-resource parameterisation interaction.

402

Table 6. Overview of Aligned Ranks Transformation ANOVA models of consumer population

outcomes by consumer and resource parameterisation and soil type. The tests were Type III Wald F

tests with Kenward-Roger degrees of freedom. Profile ID was included as a random effect. The

asterisks designate level of significance: $p < 0.1$: ·, $p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***.

Descriptions of consumer and resource parameterisations are in Table 2 and descriptions of

response variables are in Table 3.

Response variable	Predictors	F	Significance
	Consumer population	$F(4, 710.063) = 3585.924$	< 0.001 ***
	Resource population	$F(2, 710.063) = 2585.400$	< 0.001 ***
	Soil type	$F(1, 9.692) = 0.328$	0.560
	Consumer pop. x resource pop.	$F(8, 710.094) = 998.152$	< 0.001 ***
	Consumer pop. x soil type	$F(4, 710.106) = 1.460$	0.213
	Resource pop. x soil type	$F(2, 710.103) = 3.513$	0.030 *
Mean consumer resource stock	Consumer pop. x resource pop. x soil type	$F(8, 710.102) = 1.020$	0.419
	Consumer population	$F(4, 710.185) = 3629.337$	< 0.001 ***
	Resource population	$F(2, 710.292) = 1137.215$	< 0.001 ***
	Soil type	$F(1, 9.212) = 0.555$	0.475
	Consumer pop. x resource pop.	$F(8, 710.233) = 677.837$	< 0.001 ***
	Consumer pop. x soil type	$F(4, 710.315) = 2.164$	0.071 ·
	Resource pop. x soil type	$F(2, 710.32) = 0.538$	0.584
SD consumer resource stock	Consumer pop. x resource pop. x soil type	$F(8, 710.306) = 1.123$	0.345
	Consumer population	$F(4, 710.030) = 586.700$	< 0.001 ***
	Resource population	$F(2, 710.036) = 59.661$	< 0.001 ***
	Soil type	$F(1, 9.861) = 3.105$	0.109
	Consumer pop. x resource pop.	$F(8, 710.025) = 67.989$	< 0.001 ***
	Consumer pop. x soil type	$F(4, 710.037) = 2.364$	0.052 ·
	Resource pop. x soil type	$F(2, 710.037) = 0.241$	0.786
Entropy consumer resource stock	Consumer pop. x resource pop. x soil type	$F(8, 710.037) = 0.949$	0.475
	Consumer population	$F(4, 710.677) = 1296.640$	< 0.001 ***
	Resource population	$F(2, 711.086) = 2004.095$	< 0.001 ***
	Soil type	$F(1, 7.791) = 2.445$	0.158
	Consumer pop. x resource pop.	$F(8, 710.847) = 1005.251$	< 0.001 ***
	Consumer pop. x soil type	$F(4, 711.364) = 0.470$	0.758
	Resource pop. x soil type	$F(2, 711.281) = 2.502$	0.083 ·
Gini consumer resource stock	Consumer pop. x resource pop. x soil type	$F(8, 711.287) = 0.614$	0.766
	Consumer population	$F(4, 710.001) = 1361.66$	< 0.001 ***
	Resource population	$F(2, 710.001) = 604.376$	< 0.001 ***
	Soil type	$F(1, 9.998) = 9.239$	0.012 *
	Consumer pop. x resource pop.	$F(8, 710.001) = 33.651$	< 0.001 ***
	Consumer pop. x soil type	$F(4, 710.001) = 41.516$	< 0.001 ***
	Resource pop. x soil type	$F(2, 710.001) = 5.039$	0.007 **
Final population size	Consumer pop. x resource pop. x soil type	$F(8, 710.001) = 0.282$	0.972

409

410 The entropy of consumer resource stocks and the final population size both showed considerable
411 variation in the initial boxplots that was not explained by the consumer and resource
412 parameterisation (Fig 2a), and the ANOVA results suggested that soil type was influential on final
413 population size. Therefore, these were further explored with Kruskal-Wallis tests, first with profile ID
414 as the grouping variable, then soil type (Table 7, also Fig 2b). Significant differences in profile ID were
415 explored with Dunn post-hoc analysis. This showed that entropy differed significantly between
416 profiles D and H, which were Cambisol and Arenosol, respectively, while final population size
417 differed significantly between several pairs of profiles, including both intra- and inter-type profile
418 pairings.

419

420 **Table 7. Results of Kruskal-Wallis tests and Dunn post-hoc analysis comparing entropy of**
421 **consumer resource stocks and final population size by soil profile ID and soil type.** The degrees of
422 freedom for the Chi-square statistics were 11 and 1 for profile ID and soil type, respectively. Profile
423 IDs A – G correspond to Cambisols, while profile IDs H – K correspond to Arenosols. Significant pairs
424 of profiles were identified at the level of $\alpha/2$, where $\alpha = 0.05$. Profile pairings in italics denote inter-
425 type pairs.

Response variable	Grouping variable	Significance		Significantly different pairs ($p <$ 0.025)
		χ^2	p	
Entropy consumer resource stock	Profile ID	25.824	$p = 0.007$ **	*D : H*
	Soil type	4.965	$p = 0.026$ *	Cambisol : Arenosol
Final population size	Profile ID	65.167	$p < 0.001$ ***	A : C, *A : H, A : I, A : K, A: L,* B : H, B: I, B: K, *D : H, D : I, D : K, G : H, G : I, G : K,* H : J, I : J, J: K
	Soil type	21.974	$p < 0.001$ ***	Cambisol : Arenosol

426

427

428

4. Discussion

4.1 Network analysis

Given the known characteristics of the two soil types, the results of the network analysis suggest that the methodology developed here captures overall trends of soil structural development. Cambisols typically have more soil structure, higher porosity, higher levels of biotic activity, and greater stability than Arenosols (FAO, 2015). Correspondingly, the abstracted Cambisol soil networks analysed here showed higher values for the metrics measuring size, structure, and connectivity than the abstracted Arenosol soil networks did.

Specifically, the Cambisol soil networks had significantly more nodes and links, longer mean and standard deviation of link lengths, and higher total cost, density, and diameter (Table 4). This suggests more pore-creating activities modifying the soil, and a soil structural matrix that can support longer pores. This would also lead to higher water holding capacity, and increased internal drainage, both of which are commonly associated with Cambisols (FAO, 2015). In contrast, the smaller and less connected Arenosol networks have a low water-holding capacity, and the weaker coherence of their matrix material prevents longer pores from being stable, making them prone to erosion (FAO, 2015). Cambisols are also classified as more structurally developed than Arenosols, and contain more organic matter (FAO, 2015), both of which further validate the increased structure seen in the Cambisol networks here.

The global reach centrality, gamma index, and convex hull area were not as clearly differentiated between the Cambisol and Arenosol soil networks, however. The global reach centrality values were small and functionally identical, with an estimated marginal mean of 0.001 and 0.058 for both soil types at the main and subnetwork level, respectively (Table 4). Similarly, the estimated marginal mean gamma index for main networks of both soil types was 0.001. This is likely due to the presence of a similar number of disconnected subnetworks within each soil network, limiting the total number of nodes that any given node can reach. The Cambisol main networks also had a slightly smaller range of convex hull areas, although the opposite trend emerges at the subnetwork level (Fig 1, Table 4). When this is decomposed by profile, the Cambisols show more variation and outliers across and within profiles for several metrics, including convex hull area (Fig A7), suggesting that soil type includes a greater heterogeneity of network sizes and structures. As with the other metrics, further work is required to establish ranges across different soil types and geographical regions, and to compare these metrics with those more commonly used in soil analysis. Overall, however, the differences between the Cambisols and Arenosols as captured in this analysis broadly reflects those expected, given the known differences in their properties.

The improved profile development and heterogeneity of Cambisols highlights their potential for agriculture and forestry, and in underpinning the diversity of a range of ecosystems. It is vital to manage them in a way that preserves and enhances their soil structure, however, to maintain their porosity and biodiversity, and resulting stability, drainage, and aeration. Similarly, Arenosols should be managed in a way that minimises their propensity for erosion and soil loss. In both cases, this can be accomplished through limiting or eliminating tillage (e.g. Young and Ritz, 2000; Helgason, Walley and Germida, 2010; Kravchenko *et al.*, 2011), and increasing cover crops and native species (e.g. Fernández *et al.*, 2019; Kravchenko *et al.*, 2011). These provide additional organic inputs to the soil to promote an active and diverse soil biota, and therefore the positive feedback between biota, and structural development and stability (e.g. Oades, 1993; Young and Ritz, 2009; Crawford *et al.*, 2012). The feasibility of the measurement and analysis methods presented here could provide a basis for estimating changes in structure over time and under different management strategies or

474 environmental changes. This would help inform actions taken to preserve or improve the soil
475 structure. However, further work is required to standardise the approach and demonstrate its
476 application over multiple soil types.

477 As introduced in the Methods, the networks analysed here represent abstractions of the true soil
478 structure present in the samples. This simplification is reasonable for analysing overall structural
479 characteristics and heterogeneity and made the computation of the network metrics feasible.
480 Although the short link lengths (Table 4) suggest that using Euclidean distance is likely negligibly
481 different than measuring the path through the pixels, it does limit the interpretation of the findings
482 we present. Specifically, the absolute values of the metrics cannot be taken to characterise the
483 precise soil structure, but rather suggest general trends in structural development. As the exact size
484 and shape of the pores was not preserved, many of the finer distinctions between networks may
485 also be lost. This could cause the magnitudes of differences found between soil samples here to
486 appear lower than they are. As discussed above, the relatively rapid, low-cost, and lightweight
487 approach used here for estimating soil structure should be compared against more established
488 approaches and metrics to determine its effectiveness. This methodology provides simplified and
489 potentially inaccurate measurements of soil structure, but with further improvement it could be a
490 suitable approach for rapid assessment of soil structure in the field. The results presented suggest
491 that the methodology can still capture general known trends of heterogeneity within soil networks,
492 meriting further refinements and application.

4.2 ABM analysis

494 The ABM evaluated the effects of and interactions between consumer and resource characteristics,
495 and the structure of the abstracted soil networks, on the measured consumer outcomes. Overall,
496 the results showed that the size and energetic heterogeneity of the consumer population was
497 heavily influenced by the parameterisation of the consumer population and resource base, and their
498 interactions. Moreover, while outcome variables were less directly affected by soil network
499 structure, they were more influenced by the interactions between this network structure and
500 consumer or resource parameterisations.

501 Across all simulations, measured outcomes varied most strongly across consumer and resource
502 characteristics, and their combinations as overall consumer and resource parameterisations or types
503 (Fig 2a, Table 6). Specifically, the mean, standard deviation, and entropy of consumer resource
504 stocks, as well as the final population size, were most different across consumer types. These
505 differences in outcome variables resulted from how each consumer population responded to the
506 provided resource base. For example, the consumer populations with low metabolisms, low
507 consumption rate, and a low energy requirement for spawning had a lower mean resource stock,
508 and a higher final population size, for any given resource base. The consumers with high
509 metabolisms, high consumption rate, and a high energy requirement for spawning had a lower final
510 population size, but higher mean resource stock. This is similar to the distinction between r-
511 strategists and K-strategists. In these simulations, the threshold for spawning and the active and
512 basal metabolic rates appeared to have the largest impact on the measured outcome variables (Fig
513 2a). This is likely due to these parameters balancing one another in determining energy allocation
514 between maintenance and reproduction (e.g. Brown *et al.*, 2004; Kooijman, 2009).

515 In addition to consumer and resource characteristics, the soil type, and therefore soil network
516 structure, also affected population size and diversity (Table 6). Specifically, the mean consumer
517 resource stock and final population size showed significant differences across resource and soil type
518 interactions, and final population size also showed significant differences between soil types (Table

519   5). While the final population size and entropy also differed significantly across profiles (Table 6),
520   post-hoc analysis revealed that for entropy this was only significant for inter-type profile pairings,
521   and a slight difference was visible between groups when plotted (Fig 2b). This entropy is also known
522   as the Shannon Index or Shannon-Wiener Index, and here measures the diversity or 'evenness' of
523   the distribution of consumer resource stocks (Hill, 1973; Spellerberg and Fedor, 2003). Higher
524   entropy therefore meant that given quantities of resource stock were represented in equal
525   proportional abundance across the population. This is typically caused by groups of consumers
526   emerging, where group members each have the same quantity of resource stock, but these
527   quantities differ among groups. Over time, adaptations in this context could drive the system toward
528   speciation. In these simulations, the larger populations supported by the larger Cambisol soil
529   networks were more likely to have higher entropy, through different quantities of consumer
530   resource stocks represented with equal proportional abundance.

531   The relatively low Gini coefficients (Table 5, Fig 2a) can also suggest the emergence of distinct
532   groups of consumers with equal resource stocks, with similar numbers of consumers across the
533   groups. As the Gini coefficient measures relative inequality, both inequality in resource stocks across
534   groups, and more groups, cause it to increase. Equal group sizes can somewhat counter this. In both
535   soil types however, as the consumers in a given simulation had identical characteristics, it is
536   reasonable that they would have similar outcomes, slightly differing based on the subnetwork in
537   which they found themselves, and the resource base available to them there. The similarity among
538   subnetworks of the two soil types (Table 4b) suggests that the heterogeneity between soil types is
539   more apparent at the main network level. As the consumers in these simulations were unable to
540   move between subnetworks, they likely did not experience the full range of environmental
541   heterogeneity between the soil types, which would have limited its effect on the measured
542   outcomes.

543   Overall, the simulations highlight the differences in population size and diversity across consumer
544   and resource parameterisations and interactions, soil and resource type interactions, and to a lesser
545   extent, soil type on its own. Spatial heterogeneity, through both resource and network structural
546   heterogeneity, can increase the microhabitat diversity (Anderson, 1978; Giller, 1996; Ettema and
547   Wardle, 2002; Nielson *et al.*, 2010), which was shown here through the increased evenness of
548   consumer groups with different resource stocks. Similarly, the heterogeneous habitat of soils can
549   limit competitive exclusion by providing structural and resource niches for different species
550   (Bardgett, Yeates and Anderson, 2009), such that more structurally heterogeneous Cambisols have
551   larger and more diverse populations (FAO, 2015). This was reproduced by the larger populations that
552   emerged in the Cambisol simulations here, although speciation was not explicitly modelled. As with
553   the findings of the network analysis, this emphasises the importance of preserving soil structure and
554   providing adequate substrate for maintaining an active soil biota (e.g. Young and Ritz, 2009;
555   Crawford *et al.*, 2012; Fernández *et al.*, 2019).

556   While the parameterisations presented here were limited, they revealed interesting effects of
557   consumer and resource characteristics and interactions. The programming of the model itself,
558   however, may also have had an impact on the outcome of consumer populations. For example,
559   consumers moved randomly among resources rather than following any sort of search strategy, and
560   there was no energetic penalty imposed for turning, which are simplifying assumptions based on the
561   limited sensory and processing capabilities of most soil biota. This eliminated free parameters that
562   would have to be tuned and analysed or sourced from limited data about specific soil biota
563   metabolism and cognition. It also eliminated any effect that tortuosity of the network would have on
564   consumer resource stocks, though. This may not be a correct assumption if turning has a higher

565 burden physically, cognitively, or both. Furthermore, as consumers were not able to extend the
566 network or move between subnetworks, they were unlikely to experience the full difference
567 between soil networks, as discussed above. This may have led to a smaller effect of soil type on
568 measured consumer outcomes.

569 Additionally, the extraction and simplification process used to create the soil networks may have
570 affected the outcomes of the ABM. As the details of pore size and shape were not maintained, the
571 consumers' ability to forage or hide in crevices was not intended to mimic the true range of
572 consumer sizes and behaviours. Since predation was not included in the model, however, we did not
573 intend to explore the hypothesised effect of physical niches on populations by limiting competitive
574 exclusion and predation. While this would be an interesting future extension, and these changes
575 could increase the observed effect of the soil network structure on consumer population outcomes,
576 it would require refining the network extraction process as discussed above, as well as estimating
577 ranges of consumer sizes and predation dynamics. The model presented here instead focussed on
578 exploring the overall trends that might emerge in a population of consumers, rather than attempting
579 to predict how specific populations might evolve. While its design limits the precision of the
580 implications, it maintains the level of realism and generality assumed within the overall
581 methodology (Levins, 1966).

582

583 # 5. Conclusion

584 This work has explored how analysing abstracted soil networks using standard network metrics,
585 combined with simulations, can quantify the underlying structural and functional differences
586 between soil types. We showed that networks derived from a brown forest soil, or Cambisol, were
587 significantly larger, more connected, and more spatially heterogeneous than the networks derived
588 from a less developed sandy beach soil, or Arenosol. These larger and more structured networks
589 were in turn able to support larger populations of simulated consumers in an agent-based model
590 (ABM). The ABM also demonstrated how the size and heterogeneity of the simulated population
591 were significantly different across consumer and resource parameterisations, and interactions
592 between these parameterisations and soil type.

593 In conclusion, standard network metrics applied to images can be a useful way to quickly assess the
594 structure of networks within a soil profile, by capturing the broad structural differences between
595 distinct soil types, in a way that can suggest functional differences as well. These initial estimates can
596 be used on their own to survey an area more extensively or affordably, or coupled with more
597 intensive analyses, such as three-dimensional imaging techniques. Agent-based modelling can also
598 be used, when seeded with networks obtained from images or scans, to evaluate interactions
599 between consumer and resource characteristics and network structure, and to quantify the impact
600 these and other environmental factors have on the outcomes of simulated populations. Overall,
601 combining network analysis and simulation modelling can provide unique insights on the structure,
602 function, and diversity of an area of soil, and provide avenues for exploring the impact of future
603 management, structural, or environmental changes.

604

605

Competing interests

The authors declare no competing interests.

6. References

Aitkenhead, M. J. *et al.* (1999) 'Modelling water release and absorption in soils using cellular automata', Journal of Hydrology 220(1-2), pp. 104-112. doi: 10.1016/S0022-1694(99)00067-0.

Aitkenhead, M. *et al.* (2016) 'Automated soil physical parameter assessment using smartphone and digital camera imagery', *Journal of Imaging*, 2(4). doi: 10.3390/jimaging2040035.

Anderson, J. M. (1978) 'Inter- and intra-habitat relationships between woodland cryptostigmata species diversity and the diversity of soil and litter microhabitats', *Oecologia*, 32, pp. 341-348. doi: https://doi.org/10.1007/BF00345112.

Baer, S. G. *et al.* (2005) 'Soil Heterogeneity Effects on Tallgrass Prairie Community Heterogeneity: An Application of Ecological Theory to Restoration Ecology', *Restoration Ecology*, 13(2), pp. 413–424. doi: 10.1111/j.1526-100x.2005.00051.x.

Barabási, A.-L. (2016) *Network Science*. Cambridge, UK: Cambridge University Press.

Bardgett, R. D., Yeates, G. W. and Anderson, J. M. (2009) 'Patterns and determinants of soil biological diversity', in *Biological Diversity and Function in Soils*, pp. 100–119.

Bivand, R. S., Pebesma, E. and Gomez-Rubio, V. (2013) *Applied spatial data analysis with R*. 2nd edn. New York, New York, USA: Springer. Available at: http://www.asdar-book.org.

ten Broeke, G., van Voorn, G. and Ligtenberg, A. (2016) 'Which Sensitivity Analysis Method Should I Use for My Agent-Based Model?', *Journal of Artificial Societies and Social Simulation*, 19(1), pp. 1–35. doi: 10.18564/jasss.2857.

Brown, J. H. *et al.* (2004) 'Toward a metabolic theory of ecology', *Ecology*, 85(7), pp. 1771–1789. doi: 10.1890/03-9000.

Crawford, J. W. *et al.* (2012) 'Microbial diversity affects self-organization of the soil - Microbe system with consequences for function', *Journal of the Royal Society Interface*, 9(71), pp. 1302–1310. doi: 10.1098/rsif.2011.0679.

Crawford, J. W., Ritz, K. and Young, I. M. (1993) 'Quantification of fungal morphology, gaseous transport and microbial dynamics in soil: an integrated framework utilising fractal geometry', *Geoderma*, 56(1–4), pp. 157–172. doi: 10.1016/0016-7061(93)90107-V.

Csardi, G. and Nepusz, T. (2006) 'The igraph software package for complex network research', *InterJournal*, Complex Sy, p. 1695.

Davis, N. *et al.* (2020). 'Trajectories toward maximum power and inequality in resource distribution networks', *PLoS ONE*, 15(3), e0229956. doi: https://doi.org/ 10.1371/journal.pone.0229956.

Davis, N. and Polhill, J. G. (2020). 'Soil network simulation (Version 1.0.0)', *Zenodo*. http://doi.org/10.5281/zenodo.4001622

Davis, N. (2020) 'Soil network analysis functions (Version 1.0.0).' *Zenodo*. http://doi.org/10.5281/zenodo.4001702

Dinno, A. (2017) 'dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.' R package version 1.3.5. https://CRAN.R-project.org/package=dunn.test

Epskamp, S. *et al.* (2012) 'qgraph: Network Visualizations of Relationships in Psychometric Data', *Journal of Statistical Software*, 48(4), pp. 1–18.

Ettema, C. H. and Wardle, D. A. (2002) 'Spatial soil ecology', *Trends in Ecology and Evolution*, 17(4), pp. 177–183. doi: 10.1016/S0169-5347(02)02496-5.

657 FAO (2015) *World reference base for soil resources 2014: International soil classification system for*
658 *naming soils and creating legends for soil maps, World Soil Resources Reports No. 106*. Rome, Italy:
659 Food and Agriculture Organization of the United Nations. doi: 10.1017/S0014479706394902.

660 Fernández, R. *et al.* (2019) 'Pore morphology reveals interaction of biological and physical processes
661 for structure formation in soils of the semiarid Argentinean Pampa', *Soil and Tillage Research*.
662 Elsevier, 191(January), pp. 256–265. doi: 10.1016/j.still.2019.04.011.

663 García-Palacios, P. *et al.* (2012) 'Plant responses to soil heterogeneity and global environmental
664 change', *Journal of Ecology*, 100(6), pp. 1303–1314. doi: 10.1111/j.1365-2745.2012.02014.x.

665 Gargiulo, L., Mele, G. and Terribile, F. (2013) 'Image analysis and soil micromorphology applied to
666 study physical mechanisms of soil pore development: An experiment using iron oxides and calcium
667 carbonate', *Geoderma*. Elsevier B.V., 197–198, pp. 151–160. doi: 10.1016/j.geoderma.2013.01.008.

668 Giller, Paul S. (1996) 'The diversity of soil communities, the 'poor man's tropical rainforest'',
669 *Biodiversity and Conservation*, 5, pp. 135-168. doi: https://doi.org/10.1007/BF00055827.

670 Grimm, V. *et al.* (2006) 'A standard protocol for describing individual-based and agent-based
671 models', *Ecological Modelling*, 198(1–2), pp. 115–126. doi: 10.1016/j.ecolmodel.2006.04.023.

672 Grimm, V. *et al.* (2010) 'The ODD protocol: A review and first update', *Ecological Modelling*, 221(23),
673 pp. 2760–2768.

674 Hartemink, A. E. and Minasny, B. (2014) 'Towards digital soil morphometrics', *Geoderma*. Elsevier
675 B.V., 230–231, pp. 305–317. doi: 10.1016/j.geoderma.2014.03.008.

676 Hausser, J. and Strimmer, K. (2014) 'entropy: Estimation of Entropy, Mutual Information and Related
677 Quantities.' R package version 1.2.1. https://CRAN.R-project.org/package=entropy

678 Helgason, B. L., Walley, F. L. and Germida, J. J. (2010) 'No-till soil management increases microbial
679 biomass and alters community profiles in soil aggregates', *Applied Soil Ecology*. Elsevier B.V., 46(3),
680 pp. 390–397. doi: 10.1016/j.apsoil.2010.10.002.

681 Hill, M. O. (1973) 'Diversity and evenness: A unifying notation and its consequences', *Ecology*, 54(2),
682 pp. 427–432.

683 Kay M., Wobbrock J. (2020) 'ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs.'
684 doi: 10.5281/zenodo.594511

685 Kooijman, S. A. L. M. (2009) *Dynamic Energy Budget Theory for Metabolic Organisation*. 3rd edn.
686 Cambridge, UK: Cambridge University Press.

687 Kravchenko, A. N. *et al.* (2011) 'Long-term Differences in Tillage and Land Use Affect Intra-aggregate
688 Pore Heterogeneity', *Soil Science Society of America Journal*, 75(5), pp. 1658–1666. doi:
689 10.2136/sssaj2011.0096.

690 Kravchenko, A. N. and Guber, A. K. (2017) 'Soil pores and their contributions to soil carbon
691 processes', *Geoderma*. Elsevier B.V., 287, pp. 31–39. doi: 10.1016/j.geoderma.2016.06.027.

692 Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017) 'lmerTest Package: Tests in Linear Mixed
693 Effects Models', *Journal of Statistical Software*, 82(13), pp. 1-26. doi: 10.18637/jss.v082.i13

694 Lenth, R. (2020). 'emmeans: Estimated Marginal Means, aka Least-Squares Means', R package
695 version 1.5.0. https://CRAN.R-project.org/package=emmeans

696 Levins, R. (1966) 'The strategy of model-building in population biology', *American Scientist*. Sigma Xi,
697 54(4), pp. 421-431. www.jstor.org/stable/27836590.

698   Lorscheid, I., Heine, B.-O. and Meyer, M. (2012) 'Opening the "black box'' of simulations: increased
699   transparency and effective communication through the systematic design of experiments"',
700   *Computational and Mathematical Organization Theory*. Springer US, 18(1), pp. 22–62. doi:
701   10.1007/s10588-011-9097-3.

702   Mangiafico, S. (2020) 'rcompanion: Functions to Support Extension Education Program Evaluation.' R
703   package version 2.3.25. https://CRAN.R-project.org/package=rcompanion

704   Mones, E., Vicsek, L. and Vicsek, T. (2012) 'Hierarchy measure for complex networks', *PLoS ONE*,
705   7(3), pp. 1–10. doi: 10.1371/journal.pone.0033799.

706   Nielson, U. N. *et al.* (2010) 'The enigma of soil animal species diversity revisited: The role of small-
707   scale heterogeneity', *PLoS ONE*, 5(7), e11567. doi: 10.1371/journal.pone.0011567.

708   Oades, J. M. (1993) 'The role of biology in the formation, stabilization and degradation of soil
709   structure', *Geoderma*, 56(1–4), pp. 377–400. doi: 10.1016/0016-7061(93)90123-3.

710   Pebesma, E. J. and Bivand, R. S. (2005) 'Classes and methods for spatial data in R', *R News*, 5(2).
711   Available at: https://cran.r-project.org/doc/Rnews/.

712   Polhill, G. and Gimona, A. (2014) 'Using genetic algorithms to fit species and habitat parameters for
713   modelling the effect of climate change on species distributions with stochastic patch occupancy
714   models', *Proceedings - 7th International Congress on Environmental Modelling and Software: Bold
715   Visions for Environmental Modeling, iEMSs 2014*, 3.

716   R Core Team (2020) 'R: A Language and Environment for Statistical Computing', *R Foundation for
717   Statistical Computing*. Vienna, Austria, p. https://www.R-project.org. Available at: http://www.r-
718   project.org.

719   Rockafellar, R. T. (1970) *Convex Analysis*. Princeton, NJ: Princeton University Press.

720   Rodrigue, J. P. (2017) *The Geography of Transport Systems*. Hofstra University, Department of Global
721   Studies & Geography. Available at: https://transportgeography.org.

722   Roshier, D.A., Doerr, V. A. J. and Doerr, E. D. D. (2008) 'Animal movement in dynamic landscapes:
723   interaction between behavioural strategies and resource distributions', *Oecologia*, 156, pp. 465-477.
724   doi: 10.1007/s00442-008-0987-0.

725   Spellerberg, I. F. and Fedor, P. J. (2003) 'A tribute to Claude Shannon (1916–2001) and a plea for
726   more rigorous use of species richness, species diversity and the "Shannon–Wiener" Index', *Global
727   Ecology and Biogeography*, 12, pp. 177–179. doi: 10.1046/j.1466-822X.2003.00015.x.

728   Stevens, R. D. and Tello, J. S. (2011) 'Diversity begets diversity: relative roles of structural and
729   resource heterogeneity in determining rodent community structure', *Journal of Mammalogy*, 92(2),
730   pp. 387–395. doi: 10.1644/10-mamm-a-117.1.

731   Sturges, H. A. (1926) 'The choice of a class interval', *Journal of the American Statistical Association*,
732   21(153), pp. 65-66. doi: 10.1080/01621459.1926.10502161.

733   Tews, J. *et al.* (2004) 'Animal species diversity driven by habitat heterogeneity/diversity: The
734   importance of keystone structures', *Journal of Biogeography*, 31(1), pp. 79–92. doi: 10.1046/j.0305-
735   0270.2003.00994.x.

736   Velde, B., Moreau, E. and Terribile, F. (1996) 'Pore networks in an Italian vertisol: Quantitative
737   characterisation by two dimensional image analysis', *Geoderma*, 72(3–4), pp. 271–285. doi:
738   10.1016/0016-7061(96)00033-X.

739   Vezzani, F. M. *et al.* (2018) 'The importance of plants to development and maintenance of soil

740 structure, microbial communities and ecosystem functions', *Soil and Tillage Research*. Elsevier,
741 175(March 2017), pp. 139–149. doi: 10.1016/j.still.2017.09.002.

742 Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
743 Available at: http://ggplot2.org.

744 Wickham, H. *et al.* (2019) 'dplyr: A Grammar of Data Manipulation'.

745 Wijesinghe, D. K., John, E. A. and Hutchings, M. J. (2005) 'Does pattern of soil resource heterogeneity
746 determine plant community structure? An experimental investigation', *Journal of Ecology*, 93(1), pp.
747 99–112. doi: 10.1111/j.1365-2745.2004.00934.x.

748 Wobbrock, J., Findlater, L., Gergle, D., Higgins, J. (2011) 'The Aligned Rank Transform for
749 Nonparametric Factorial Analyses Using Only ANOVA Procedures.' In *Proceedings of the ACM
750 Conference on Human Factors in Computing Systems (CHI '11)*', pp. 143-146.

751 Young, I. M., Crawford, J. W. and Rappoldt, C. (2001) 'New methods and models for characterising
752 structural heterogeneity of soil', *Soil and Tillage Research*, 61(1–2), pp. 33–45. doi: 10.1016/S0167-
753 1987(01)00188-X.

754 Young, I. M. and Ritz, K. (2000) 'Tillage, habitat space and function of soil microbes', *Soil and Tillage
755 Research*, 53(3–4), pp. 201–213. doi: 10.1016/S0167-1987(99)00106-3.

756 Young, I. M. and Ritz, K. (2009) 'The habitat of soil microbes', in *Biological Diversity and Function in
757 Soils*, pp. 31–43.

758 Zeileis, A. (2014) 'ineq: Measuring Inequality, Concentration, and Poverty.' R package version 0.2-13.
759 https://CRAN.R-project.org/package=ineq

760

Appendices

Appendix 1. Network extraction process

For all images:

1. Convert the image to a text file containing RGB triplets
2. Identify and eliminate all non-soil pixels (set to -1)
3. Calculate mean pixel intensity at all points
4. Adjust pixel intensity to remove variations in brightness across image
5. Threshold the image to retain the darkest 30 % soil pixels
6. Carry out erosion and thinning operators
7. Clean image to produce skeletal pixels
8. Identify networks
9. Remove redundant pathways
10. Calculate distances between nodes
11. Save the network
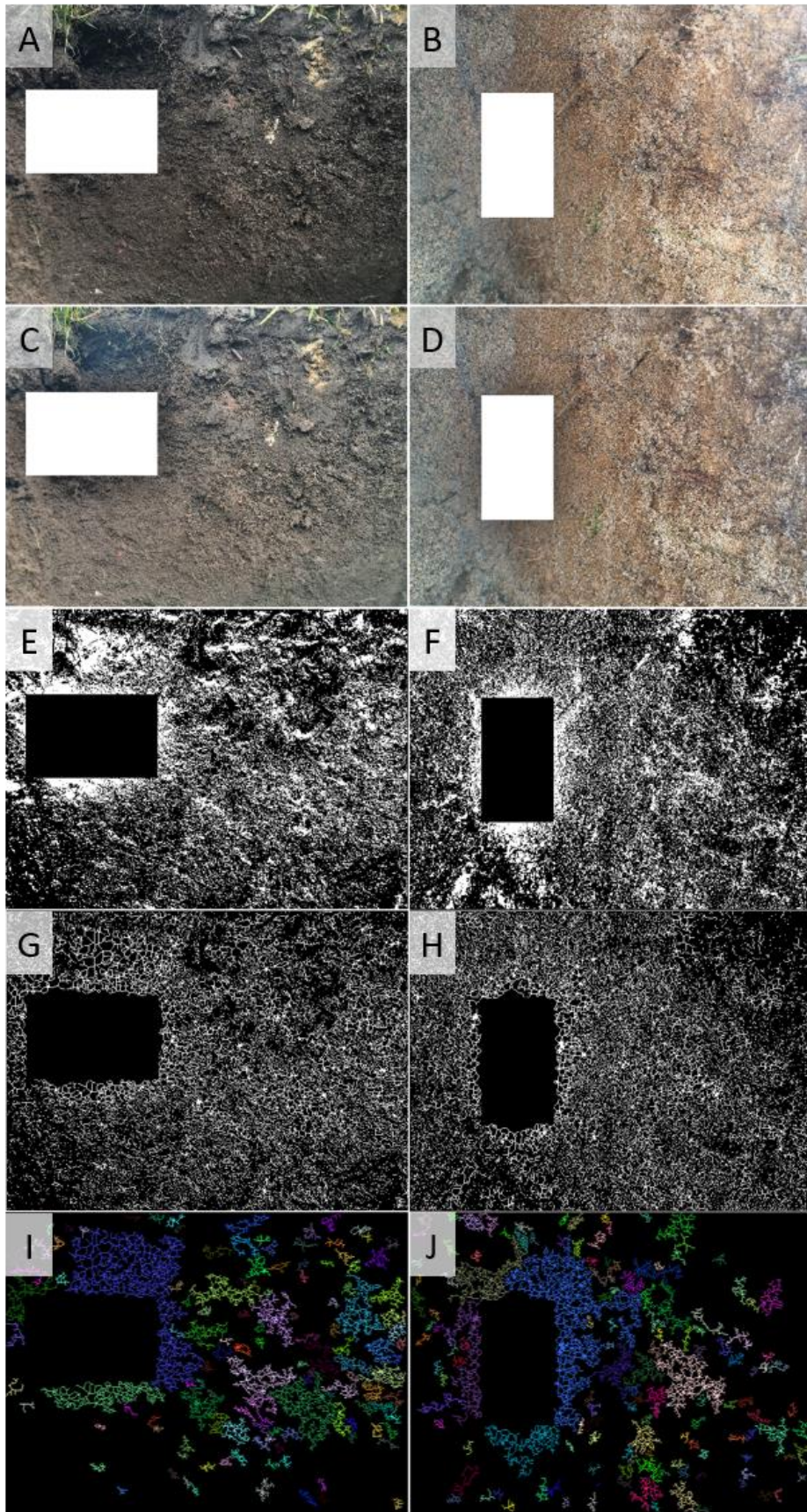
Figure A1. Soil image morphology process for a Cambisol (a) and Arenosol (b) profile image. Steps show include (c-d) colour correction, (e-f) thresholding, (g-h) erosion and thinning operations, and (i-j) subnetwork identification. White areas represent colour correction cards, which were excised.

780
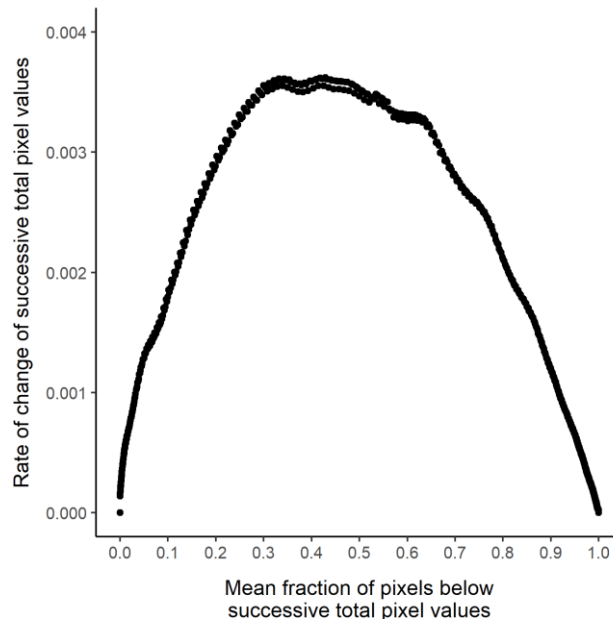
Figure A2. The rate of change of mean fraction of pixels for each mean fraction below a given threshold value. The plot starts on the left with pixel values of 0, with no pixels below this value, and ends on the right with pixel values of 755 (with correction card removed from image). The y-axis shows the rate of change of the mean fraction of pixels below each value.

785

786
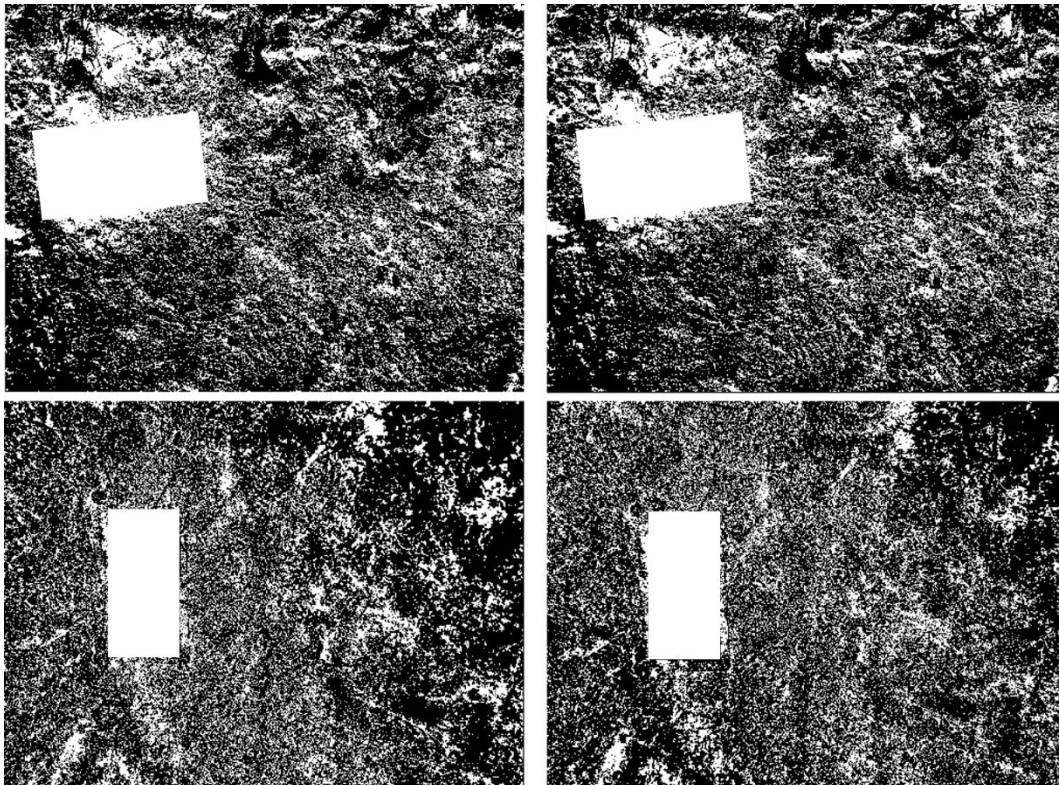
Figure A3. Comparison of different images from the same pit after thresholding. The two pairs of images from each pit are arranged horizontally. The white rectangle is the correction card. The thresholding process was the same as described in Section 2.2, where the darkest 30 % of pixels have been retained as pores, and other pixels removed.

791 Appendix 2. ODD design concepts, initialisation, input data, and submodels

792
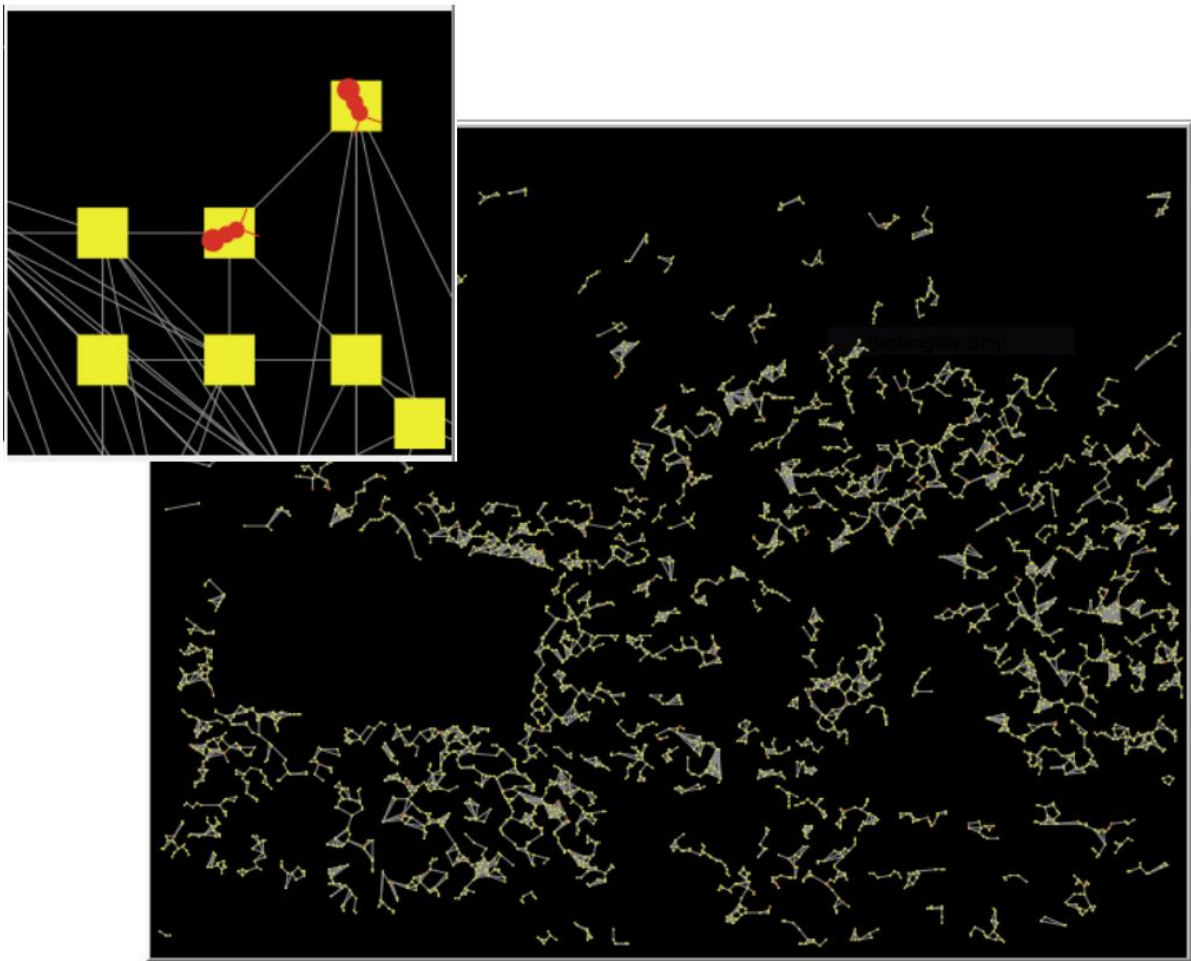
794 **Figure A4. A screenshot of the model.** The inset at the top left shows an enlarged version of some of
795 the resource nodes (yellow squares) and agents (red ants).

796   I.   Model purpose
797   a. The model is designed to be an analytical tool to explore the heterogeneity in
798   resource supply potential of a network by populating it with idealised energy-
799   consuming agents, and to quantify the effects of consumer, resource, and network
800   characteristics on resulting consumer population outcomes.
801   II.   Entities, state variables, and scales
802   a. Consumer entities
803   i. State variables

Property	Description
Location	The resource on which the consumer is located
Target location	The resource to which the consumer will move next
Active?	Whether a consumer is active (or dead)

804

805

806   ii.   Parameters

Property	Description
Basal metabolism	How much resource an agent needs per day to stay alive
Active metabolism	How much resource an agent uses with each step
Resource stock	How much resource an agent has consumed but not metabolised
Consumption rate	Maximum number of resource units that an agent takes from a resource it visits, per timestep
Spawn energy	How much energy an agent requires to spawn (depletes this quantity from stocks and passed to offspring as starting quota)

807

808   b.   Resource entities
809   i.   State variables

Property	Description
Current supply	The current quantity of resource at this point
Resource capacity	How much energy is stored in a resource when it is full
Regrow rate	The amount the resource regrows each timestep

810
811   ii.   Parameters

Property	Description
Resource capacity	How much energy is stored in a resource when it is full
Regrow rate	The amount the resource regrows each timestep

812

813   c.   Link entities
814   i.   State variables

Property	Description
Length	The length of the link - determines energy and time required to traverse it

815

816   d.   Scales

Property	Description
Timestep	A single unit of time in the model, defined as that which is required for consumers to move 1 pixel (approximately 0.3 – 0.5 mm), and for which they require `basal-metabolism` units of energy.
World size	400 x 500, determined by the size of the soil networks used as the environment.

817

818   III.   Sequence of events
819   a.   Consumers start on random nodes around a pre-specified network, where nodes are
820   resource patches.
821   b.   Consumers move around the network randomly following links. If they find a
822   resource patch, they consume as much as they can from it, and the patch depletes.
823   i.   Consumers require `basal-metabolism` units of resource per timestep. If
824   they do not consume this resource, they die.

ii. Consumers can stay put on a resource and consume it (`consumption-rate` units consumed per timestep), but it depletes, and if there is no more resource there then they move on.

iii. Consumers metabolise `active-metabolism` units of resource per patch of link that they cross.

iv. If there is more than one agent on a resource patch, they each take `consumption-rate` units per timestep, or split the remainder if there is not enough resource remaining for them to each get `consumption-rate` units.

c. If consumers have twice as much energy as the set `spawn-energy`, they can spawn new consumers (who take the same amount of `resource-stock` from their parent that the parent started with, so now parent and offspring both have the same `resource-stock`).

d. Resources regrow at a constant rate (`regrow-rate`) per timestep, up to their maximum capacity (`resource-capacity`).

IV. Design concepts

a. Basic principles

i. Consumers attempt to consume as much free energy from a resource as they are able, to maximise energy reserves for future movement, and spawning capability.

ii. Conservation equations: energy and matter cannot be created (except at the start of the simulation) or destroyed. In spawning, this is represented by consumers transferring some of their energy to their offspring. Consumers only die when their energy reserves are completely depleted (starvation).

iii. Entropy production: some resource energy is consumed in movement and cannot be recaptured.

b. Emergence

i. The distribution of consumers in space around the network and the distribution of resource stocks across the consumers both emerge from the interactions in the model.

c. Objectives

i. The consumers' objective is to consume as much resource energy as possible, allowing them to stay alive, move, and potentially reproduce.

d. Prediction

i. Consumers do not 'predict' the results of their course of action per se, they are random walkers, but they do 'predict' that they will die if they stay in a non-resource patch, or depleted resource patch, so they keep moving.

e. Sensing

i. Consumers can sense if they are on a resource patch or not, and if it has any resource energy in it. They also know the link-neighbours of the resource patch that they are currently on.

f. Learning

i. Consumers are random walkers; they do not learn in any capacity.

g. Adaptation

i. The population adapts to fill the network in a way that reflects the density of resource availability in that area, as consumers will cluster and reproduce around resources where they can consume what they need.

872			h.	Interaction
873				i.	Consumers interact stigmergically through their consumption of resources.
874					While they do not interact directly in any meaningful way, their
875					consumption of resources affects the availability of resources for others to
876					consume.
877			i.	Collectives
878				i.	There are no collectives present.
879			j.	Stochastic elements
880				i.	Consumers are initialised in random locations and move randomly.
881					Additionally, resources are all initialised with random maximum capacity
882					between 1 and `maximum-resource-capacity` and regrow rates
883					between 1 and `maximum-regrow-rate`.
884			k.	Observation
885				i.	Number of currently active ('alive') consumers at each timestep.
886				ii.	Mean, standard deviation (SD), Gini coefficient, and entropy of the
887					distribution of consumer resource stocks at each timestep.
888				iii.	The resource capacity and regrow rate of each resource at the start of the
889					simulation.
890				iv.	The resource stock and location of each active consumer at 10, 100, 500,
891					1000, and 2000 timesteps.
892	V.				Initialisation
893			a.		The network was supplied as two Comma-Separated Values (CSV) files: one of

872 h. Interaction
873   i. Consumers interact stigmergically through their consumption of resources.
874   While they do not interact directly in any meaningful way, their
875   consumption of resources affects the availability of resources for others to
876   consume.
877 i. Collectives
878   i. There are no collectives present.
879 j. Stochastic elements
880   i. Consumers are initialised in random locations and move randomly.
881   Additionally, resources are all initialised with random maximum capacity
882   between 1 and `maximum-resource-capacity` and regrow rates
883   between 1 and `maximum-regrow-rate`.
884 k. Observation
885   i. Number of currently active ('alive') consumers at each timestep.
886   ii. Mean, standard deviation (SD), Gini coefficient, and entropy of the
887   distribution of consumer resource stocks at each timestep.
888   iii. The resource capacity and regrow rate of each resource at the start of the
889   simulation.
890   iv. The resource stock and location of each active consumer at 10, 100, 500,
891   1000, and 2000 timesteps.
892 V. Initialisation
893 a. The network was supplied as two Comma-Separated Values (CSV) files: one of
894   resource node locations and another of the connections between the resource
895   nodes. The node locations and connections were determined during the process of
896   extracting the soil network from a soil profile image, as described in the main text
897   (Section 2.2). The resource and consumer types and parameters were specified in an
898   Extended Markup Language (XML) file. The models were initialised with 500
899   consumers located on random resource nodes throughout the network. The
900   consumers each began with 30 resource units in their `resource-stock`, and
901   metabolic rates, consumption rate, and spawn energy thresholds as specified in the
902   XML file. Resources were all initialised with random maximum capacity between 1
903   and `maximum-resource-capacity` and regrow rates between 1 and
904   `maximum-regrow-rate` and began the simulation at full capacity.
905 VI. Input data
906 a. This model has no input data.
907 VII. Submodels
908 a. Regrowth of resources: at each timestep, all resources that are less than their
909   maximum capacity regrow by `regrow-rate` units.
910 b. Consuming resources: at each timestep, all consumers currently located on a
911   resource node check whether there is any resource available at that node. If there is
912   enough for each consumer to take `consumption-rate` units, they do, and these
913   are added to their `resource-supply`. If there is not enough, each consumer
914   receives what is at the resource, divided by the number of consumers at the
915   resource. If there is no resource available at that node, the consumer identifies a
916   new `target-node`, selecting randomly from the other resources connected to the
917   first, and moves to the `target-node`.

918
919
920
921
922
923

    c. Spawning new consumers: at each timestep, consumers check whether they have twice the amount of energy specified as `spawn-energy` in their `resource-stock`. If so, they spawn a new consumer who is an exact clone of themselves. The new consumer starts with `spawn-energy` units as their initial `resource-supply`, and the parent consumer loses `spawn-energy` units of resource from their `resource-stock`.

924
925
926
927
928

    d. Check consumer resource stocks: at each timestep, all consumers check whether they have more than `resource-requirement` units, or their basal metabolism, of resource in their `resource-supply`. If they do, they consume `resource-requirement` units, removing them from their `resource-supply`, otherwise they die.

929

## Appendix 3. Sensitivity Analysis

930

931 First, a pre-test was conducted to determine the number of time steps for which to run the
932 simulations, and the number of replicates of each parameter set that were necessary for the outputs
933 to reach equilibrium (ten Broeke, van Voorn and Ligtenberg, 2016). The first set of 500 runs used
934 varied parameter values and a fixed network architecture, determined by Latin Hypercube Sampling
935 from the range of values for global analysis (Table A1). One replicate of each parameter set was run
936 for 3000 timesteps, and the output variables were plotted to determine whether the model reached
937 a stable state, and if so, when. As all runs showed stability in output parameters after 500 – 1000
938 timesteps (Fig. A5), apart from small variations due to stochasticity, the final output variable values
939 for all future runs were calculated as the mean of the values at timesteps 500, 750, and 1000.

940

941 **Table A1. Parameter ranges used for testing to determine length of simulations.** Values shown are
942 the minimum and maximum for that parameter. Latin Hypercube Sampling was used to generate the
943 values, which were then multiplied by the range plus the minimum, to get the value for the
944 parameter for testing.

| Parameter	Value
Initial population size	50, 1000
Consumer basal metabolism	1, 3
Consumer active metabolism	1, 3
Initial consumer resource stock	20, 50
Consumer consumption rate	5, 10
Consumer spawn energy	50, 100
Maximum resource capacity	20, 50
Maximum resource regrowth rate	10, 20

945

946
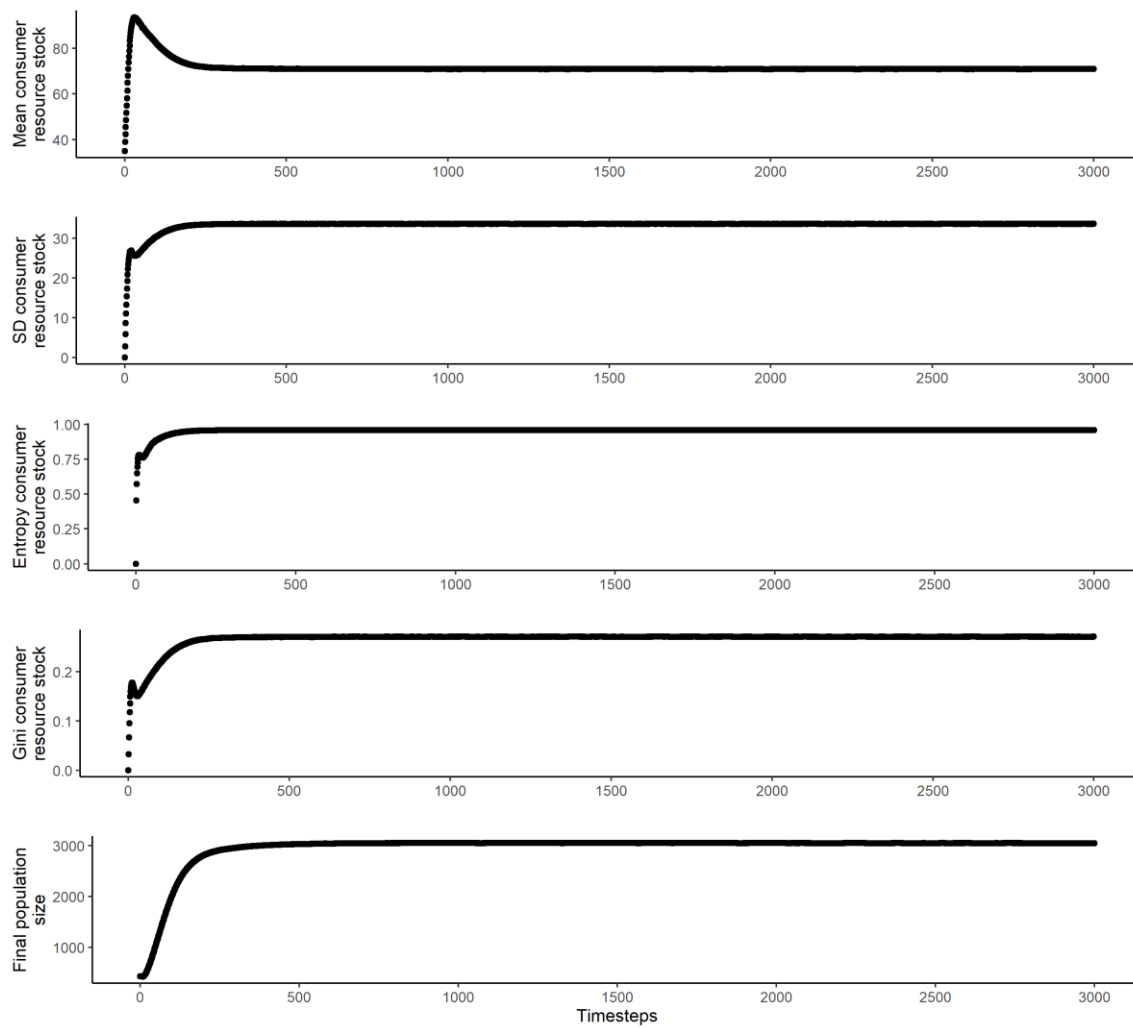
Figure A5. Stability plots from testing to determine the length of simulations. Shown are values averaged for each timestep over 500 runs.

949

950 The second set of pre-test runs used the baseline parameter values for all parameters, and a fixed
951 network, which we repeated 100 times. We then calculated a rolling coefficient of variation for the
952 output variables, including progressively more replicates (Fig. A6). The coefficients of variation for all
953 output variables stabilised around 10 runs. Plotting the distribution of the output variables at that
954 point show approximate normality, such that the mean value across runs is a reasonable measure of
955 centre. Therefore, for all future simulations, the mean of the outcome variables across 10 replicates
956 was used to reduce the effects of stochasticity on the output. As the mean value across replicates
957 was used, there was no effect from replication on the experimental results.

958

959
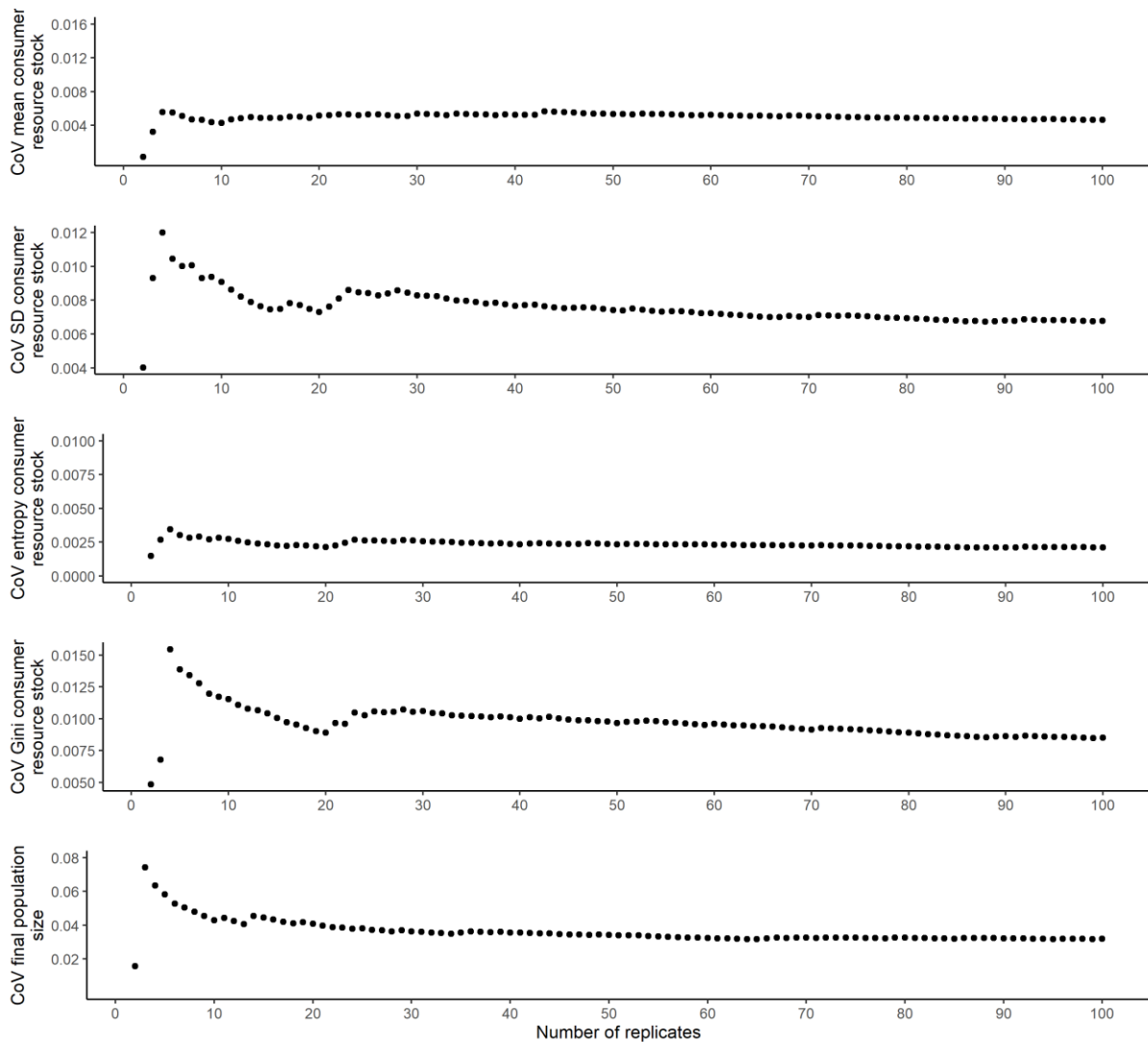
Figure A6. Plots of rolling Coefficient of Variation (CoV) for each outcome variable against the number of replicates included in its calculation. This was used to determine number of replicates needed to average across to minimise stochasticity in output variables.

963

After the pre-test, we used the One-Factor-at-a-Time methodology to identify which of the control variables significantly affected the output variables, and which could be held constant. For this test, the four control variables (initial consumer population, initial consumer resource stock, maximum resource regrowth rate, and maximum resource capacity) were varied across four levels each, changing one variable at a time, while holding all other variables constant at middle values for each. Both the maximum resource regrow rate and maximum resource capacity significantly affected the output variables, while initial consumer resource stock did not (Table A2). The initial consumer population size significantly affected all but the standard deviation of consumer resource stock (Table A2e), but as the magnitude of the effect was quite small, both the initial population size and initial consumer resource stock were held constant at middle values for the rest of the simulations.

974

975 **Table A2. Regression results from One-Factor-at-a-Time analysis.** This was used to identify which
976 control parameters could be fixed, and which significantly affected the outcome variables and
977 needed to be explored. The asterisks designate level of significance: $p < 0.1$: ·, $p < 0.05$: *, $p < 0.01$:
978 **, $p < 0.001$: ***.

a. Mean consumer resource stock

	Estimate	Standard error	t value	p
Intercept	54.360	0.100	543.566	0.000 ***
Initial population size	0.000	0.000	-3.728	0.002***
Initial consumer resource stock	0.000	0.001	0.100	0.920
Maximum resource regrow rate	0.054	0.004	12.874	0.000 ***
Maximum resource capacity	0.060	0.001	41.088	0.000 ***
$F_{(4, 251)} = 467 (p < 0.001)$	$R^2 = 0.88$			

b. SD consumer resource stock

	Estimate	Standard error	t value	p
Intercept	27.110	0.038	711.283	0.000 ***
Initial population size	0.000	0.000	0.948	0.344
Initial consumer resource stock	0.000	0.000	0.333	0.739
Maximum resource regrow rate	0.016	0.002	10.073	0.000 ***
Maximum resource capacity	-0.010	0.000	-17.718	0.000 ***
$F_{(4, 251)} = 104.1 (p < 0.001)$	$R^2 = 0.62$			

979

c. Entropy consumer resource stock

	Estimate	Standard error	t value	p
Intercept	0.953	0.000	2288.704	0.000 ***
Initial population size	0.000	0.000	6.500	0.000 ***
Initial consumer resource stock	0.000	0.000	0.446	0.656
Maximum resource regrow rate	0.000	0.000	19.891	0.000 ***
Maximum resource capacity	0.000	0.000	2.891	0.004 **
$F_{(4, 251)} = 111.6 (p < 0.001)$	$R^2 = 0.63$			

980

d. Gini consumer resource stock

	Estimate	Standard error	t value	p
Intercept	0.283	0.000	431.434	0.000 ***
Initial population size	0.000	0.000	3.472	0.001 ***
Initial consumer resource stock	0.000	0.000	0.154	0.876
Maximum resource regrow rate	0.000	0.000	-2.267	0.024*
Maximum resource capacity	0.000	0.000	-38.517	0.000 ***
$F_{(4, 251)} = 375.2 (p < 0.001)$	$R^2 = 0.85$			

e. Final population size

	Estimate	Standard error	t value	p
Intercept	-2787.749	188.701	-14.773	0.000 ***
Initial population size	2.496	0.118	21.110	0.000 ***
Initial consumer resource stock	0.589	2.738	0.215	0.830
Maximum resource regrow rate	191.341	8.038	23.805	0.000 ***
Maximum resource capacity	61.961	2.738	22.634	0.000 ***
$F(4, 251) = 381.2$ ($p < 0.001$)	$R^2 = 0.86$			

Appendix 4. Calculation of Entropy

The entropy of the consumer resource stocks was calculated as the Shannon index, or Shannon entropy, of the resource stocks held by consumers. As the Shannon entropy is meant to be applied to discrete data, the consumer resource stocks were discretised into a fixed number of 'bins' using Sturges' formula (Sturges, 1926), and the Shannon entropy was calculated for the bins.

Sturges' formula for the number of bins k for a population of size n is

$$k=\lceil \log_2 n \rceil + 1. \qquad \text{Eq. A1}$$

Using a sample of 100 runs from the stability test for run length (Section A3), the normality of the consumer resource stocks at the sampling timesteps T = 500, T = 750, and T = 1000 was tested. Additionally, the entropy was calculated for 5, 10, 15, 20, 25, 50, 75, and 100 bins and compared with the entropy binned using Sturges' formula. By normalising the calculated entropy by the maximum possible entropy for that number of bins, $\log(N)$, the differences in entropy between different numbers of bins were < 0.001. As the data were found to be approximately normally distributed at the sampling timesteps, the assumptions for Sturges' formula was met, and it was chosen to determine the final bin width.

1007   **a.**
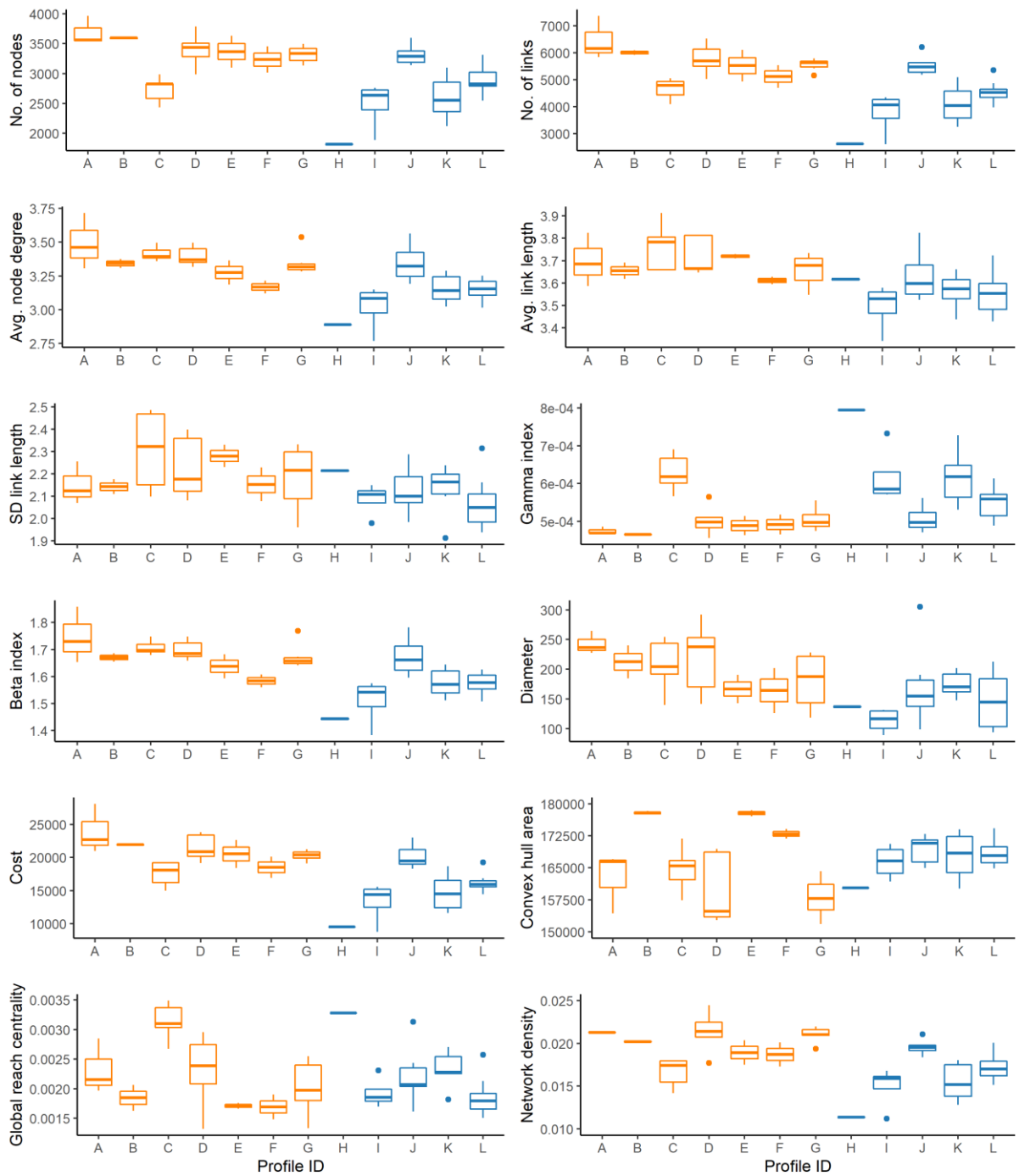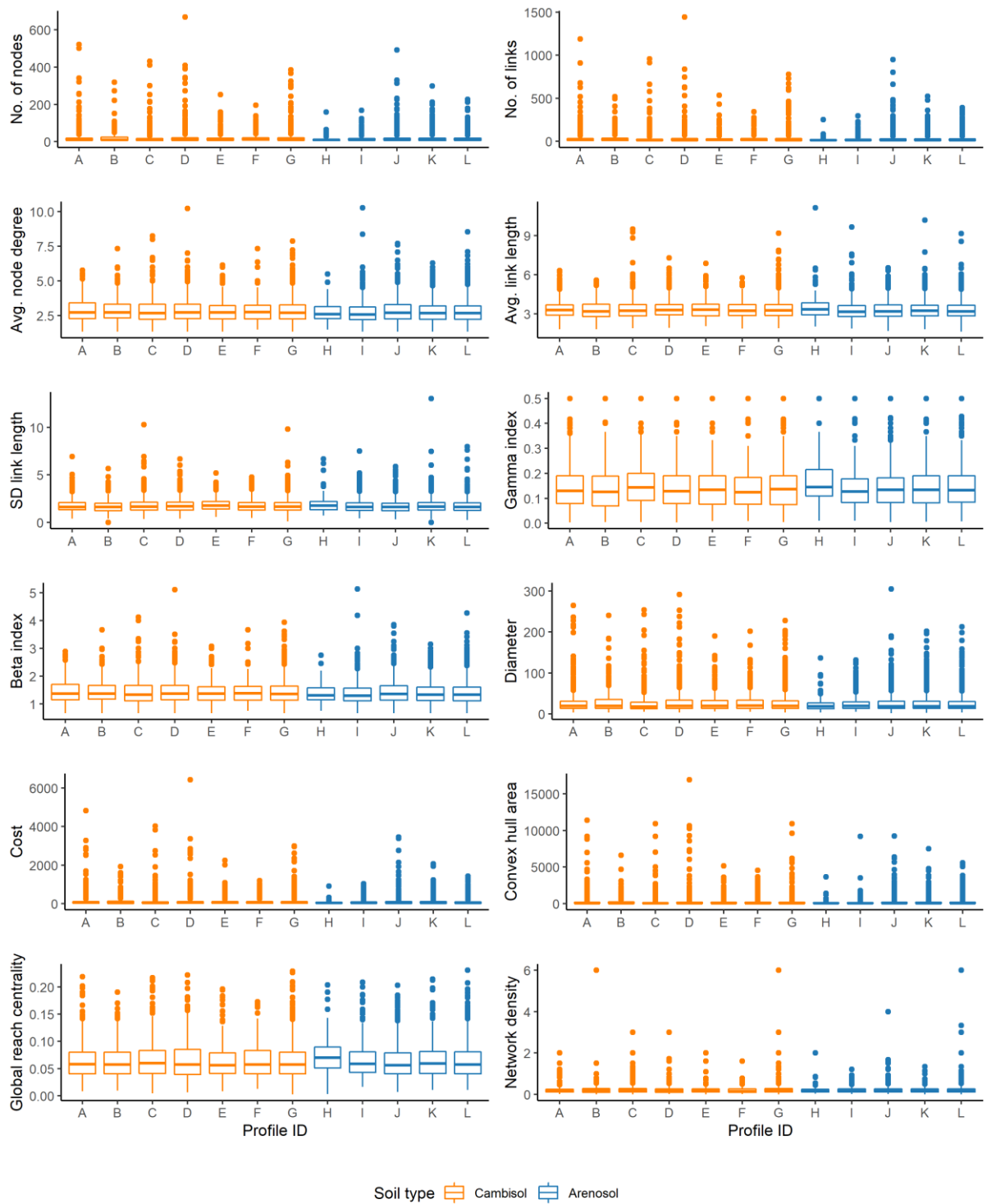
1008　**b.**

Figure A7. Boxplots showing distributions of network metrics across soil profiles for (a) main networks and (b) subnetworks. Profiles A – G correspond to Cambisol soil profiles, and profiles H – L are Arenosol soil profiles. Descriptions of network metrics are in Table 1.