A framework to bridge scales in distribution modelling of soil microbiota


(1) Plant and Ecosystems Research Group (PLECO), University of Antwerp, Belgium.
(2) UR 'Ecologie et Dynamique des Systèmes Anthropisées' (EDYSAN, UMR 7058 CNRS-UPJV), Université de Picardie Jules Verne, Amiens, France.

*Corresponding author. jonas.lembrechts@uantwerpen.be

Orcid ID J JL: orcid.org/0000-0002-1933-0750.
Orcid ID JL: orcid.org/0000-0003-0638-9582.
Orcid ID EV: orcid.org/0000-0001-7015-1515.

Keywords: species distribution models, soil microbes, biotic interactions, joint distribution models, niche modelling, microclimate

1 Corresponding author. Jonas Lembrechts, CDE C.007 Universiteitsplein 1, University of Antwerp, 2610 Wilrijk, Belgium, +3232651727, jonas.lembrechts@uantwerpen.be.
Abstract

Creating accurate habitat suitability and distribution models (HSDMs) for soil microbiota is far more challenging than for aboveground organism groups. In this perspective paper, we propose a conceptual framework that addresses several of the critical issues holding back further applications. Most importantly, we tackle the mismatch between the broad-scale, long-term averages of environmental variables traditionally used, and the environment as experienced by soil microbiota themselves. We suggest using nested sampling designs across environmental gradients and objectively integrating spatially hierarchical heterogeneity as covariates in HSDMs. Secondly, to incorporate the crucial role of taxa co-occurrence as driver of soil microbial distributions, we promote the use of joint species distribution models, a class of models that jointly analyze multiple species’ distributions, quantifying both species-specific environmental responses (i.e. the environmental niche) and covariance among species (i.e. biotic interactions). Our approach allows incorporating the environmental niche and its associated distribution across multiple spatial scales. The proposed framework facilitates the inclusion of the true relationships between soil organisms and their abiotic and biotic environment in distribution models, which is crucial to improve predictions of soil microbial redistributions as a result of global change.

Introduction

Habitat suitability and distribution models (HSDMs) have turned into irreplaceable assets to study the spatial distribution and redistribution of species under given environmental circumstances, and changes thereof (Elith & Leathwick, 2009). This broad group of statistical models relates known species occurrences (or presence-absence) with information about the environmental conditions at these locations (Guisan & Thuiller, 2005, Elith & Leathwick, 2009, Jiménez-Valverde et al., 2011). Based on statistically or theoretically derived response curves (Guisan et al., 2017), they aim to define the environmental niche and related geographical range in which organisms can operate. Applications range from studies on the effects of anthropogenic climate change to predictions of biological invasions, and they have proven valuable across a wide range of organism groups both above and below the soil surface (Martiny et al., 2006, Elith & Leathwick, 2009, Sato et al., 2012, Bahram et al., 2018). For belowground organisms in particular, however, the successful application of HSDMs needs to overcome important limitations concerning the scale, resolution and accuracy that can reliably and routinely be obtained. Indeed, many of the available global, high-resolution gridded environmental variables cover the above-ground world.
only, with many of the key drivers of soil microbial distributions (like soil pH and soil organic matter) insufficiently recorded, or at coarse resolutions only (Fierer & Jackson, 2006, Mod et al., 2016). This spatial mismatch between the environment as experienced by organisms on the one hand, and the data we have available to model their distribution on the other, is particularly pronounced for soil microbes. Their distribution, more than for most aboveground species groups, indeed depends largely on processes happening at a small sub-meter scale in the hard-to-measure belowground environment.

To understand and predict soil functioning in a rapidly changing world, we need the ability to accurately model the spatial distribution patterns of soil microorganisms. To achieve this, we need to incorporate local-scale drivers and processes relevant to these species groups in our HSDMs, in order to improve predictions of large-scale distributional patterns, especially in response to global change. For example, accurately attributing local variance to abiotic drivers versus biotic interactions can aid our understanding of the importance of species interactions involving microbes as drivers of spatial distributions (de Mesquita et al., 2016, Ovaskainen et al., 2017). Such an approach can help improve our predictions of the faith of ecologically important, but so far elusive groups including for example plant pathogens, mycorrhizal fungi and nitrogen fixing bacteria and their effects on ecosystem processes (Martiny et al., 2006). While many of these questions have been covered in microbial distribution studies before, a clear framework is still lacking to incorporate the relevant processes at the local scale of sub-centimeters to meters – in which microbes operate – with the large-scale spatial patterns across several kilometers that are the key focus in spatial ecology, without the need for an ever-increasing higher resolution of the used spatial data.

**Belowground microbial distributions: the importance of the local scale**

It has long been assumed that many soil microbiota have cosmopolitan distributions. This theory stems from the observation that they have large population sizes and short generation times resulting in high dispersal, as well as the capacity to disperse over long distances if they get into the air or water streams (Frey, 2015). Additionally, the actual drivers of their occurrence might vary more on a sub-centimeter scale than on the kilometer-scale traditionally used, as is the case for e.g. pore size, or biotic interactions with other soil microbiota. If soil microbial distributions would indeed relate only little to global-level environmental gradients, the validity of HSDM-approaches would be hampered, as it would be impossible to identify an environmental niche at
the regional or global scale. However, recent research has demonstrated that spatial patterns of microbial diversity qualitatively similar to those observed for plants and animals exist, suggesting that soil microbes indeed “have a biogeography” (Martiny et al., 2006, Hanson et al., 2012, Sato et al., 2012, Peay et al., 2016). While the existence of such spatial structuring should not come as a surprise for host-dependent soil microbes like plant-associated mycorrhizae, the presence of a spatial distributional pattern is more telling for microbes like saprotrophs that operate independent of a plant host. Nevertheless, their high dispersal flexibility implies that – even more so than for most aboveground organisms - the range of environmental conditions and the nature of biotic interactions at the local scale is likely to be of particular importance for their distribution (Sato et al., 2012, Vos et al., 2013), and an efficient way to include this local environmental variation in HSDMs is thus crucial.

While existing HSDMs often rely heavily on air temperature averages obtained from global climate models (e.g. Aguilar & Lado, 2012, Tedersoo et al., 2014), the use of disconnected environmental variables overlooking the small-scale nature of the community processes involved is unlikely to provide meaningful descriptions of the regional and global distribution of soil biota, and thus not fully exploit the potential of HSDMs for soil organisms (Fierer & Jackson, 2006). Recent HSDMs studies indeed increasingly acknowledge the need for more environmental data at the scale and especially the location relevant to the studied organisms, as opposed to the traditional use of coarse spatiotemporal averages of environmental conditions (Lembrechts et al., 2018, Lembrechts et al., 2019, Zellweger et al., 2019). This issue becomes particularly important when one aims to use HSDMs predictively to investigate the effects of global (or local) change on species (re)distributions. Incorporating local conditions is for example critical in regard to microrefugia, where isolated populations can persist in a favorable microclimate amid deteriorating climatic conditions until the latter become favorable again (Hannah et al., 2014, Lenoir et al., 2017). For many soil microbiota, these microrefugia might exist within centimeters to meters from where the species currently occurs due to the large local heterogeneity in conditions (Veresoglou et al., 2015), and HSDMs averaging out these local gradients are thus likely to largely overestimate local extinction incidences, as has been shown for other organism groups (Lenoir et al., 2013). Similarly, this local heterogeneity can create stepping stones of favorable conditions that facilitate range expansion beyond that predicted at a coarse scale (Lembrechts et al., 2017). The relevant data to counter these biased estimates is nevertheless rarely available.
The way forward to improve HSDMs for soil microbiota

In this paper, we introduce a conceptual framework that provides the missing link between the local-scale (abiotic and biotic) belowground environment and large-scale distribution patterns, building on the recent advancements in tools, measurement techniques and models (Fig. 1, see also Lembrechts (2020) for a simple overview of the followed steps in R). Keeping the strengths and limitations of HSDMs in mind, this framework builds on 1) smart in-situ measurements of local environmental heterogeneity and community distributions, 2) recent developments in spatial modelling of environmental conditions, and 3) the hierarchical integration of the local with the regional scale. It thus aims for a relatively simple way to incorporate the complexity of the belowground world into statistical HSDMs without dramatically increasing the sampling effort, and is especially suited for distributions driven by a complex interaction between local environmental conditions and biotic interactions which mechanistic distribution models cannot (yet) handle.

Before expanding further, it is noteworthy that the definition of a ‘species’, as traditionally used in HSDMs, is far less straightforward for soil microbes than it is for aboveground organisms, and that approaches to identify micro-organism identities differ for different microbial groups. The following framework focuses specifically on soil micro-organisms that are routinely identified as operational taxonomic units (OTUs) or amplicon sequence variants (ASVs). These OTUs are commonly used as proxies for species (e.g. Bahram et al., 2016), yet recently, delineation of taxa has often been abandoned in favor of ASVs (Knight et al., 2018) as the unit of analysis, in particular when sequence heterogeneity is not too high. The latter avoids the arbitrary binning of sequences and allows comparisons across different studies, yet it would increase complexity of the models due to a higher number of identities many of which would be intraspecific, as would be the case for e.g. fungal ITS-sequences (Heeger et al., 2018). If needed, meaningful HSDMs can also still be produced by focusing on keystone taxa or higher taxonomic resolutions only.

Nested sampling of abiotic conditions

Global databases and gridded datasets are currently being developed for many soil variables (e.g. SoilGrids, Hengl et al., 2017). Even though in most cases database resolutions are still relatively coarse for application to the world of microbiota, the accuracy of the available data on many local environmental drivers is rapidly rising. For soil moisture, for instance, the first attempts to obtain global remotely sensed data at a coarse scale (3 and 9 km) are now appearing (He et al., 2015). For soil microbial distributions, however, large-scale gridded datasets are still several
orders of magnitude too coarse in spatial resolution. For many other relevant variables, e.g. soil climate, users often still have to rely on the above-ground equivalent for higher resolutions (e.g. free-air temperatures from CHELSA at 1 km² resolution (Karger et al., 2017) vs soil temperatures from ERA5 at 9 km² resolution (Copernicus Climate Change Service, 2019)).

The success of any application of HSDMs to soil microbial distributions thus fundamentally needs to rely on a sampling design that allows linking the local heterogeneity, inherently present in soils, with the microbial distribution at regional scales (Fig. 2). We therefore recommend a nested sampling approach, measuring the relevant environmental variables right where it matters for each specific study organism (e.g. measuring soil temperatures in-situ instead of relying on free-air large-scale interpolations (Lembrechts et al., 2018). The idea of a nested sampling design is to capture the spatial autocorrelation signal that may exist both at a very fine spatial resolution as well as across larger spatial extents, due to the modifying impact of both climatic and biotic interactions on soil conditions (King et al., 2008, King et al., 2010, Ovaskainen et al., 2017). Such nested sampling can not only provide a better mechanistic understanding of drivers of the local and regional distribution of a focal study organism, it is also a crucial requirement to use HSDMs for predictions under global change (Mateo et al., 2019).

These measurements should cover both the local (e.g. meters or even sub-centimeters apart, or at different depths in the soil) and regional scales (e.g. up to kilometers apart) (King et al., 2010). Importantly, one should not aim for the finest possible resolution of environmental variables across the whole landscape. Such a run for an increasing refinement in the spatial resolution would greatly reduce the efficiency of HSDMs over larger spatial extents, while not necessarily improving their accuracy (Bennie et al., 2014, Lembrechts et al., 2018). Instead, we suggest to explicitly keep local-scale heterogeneity as a variable in regional-scale distribution models when aggregating to a coarser resolution. This can for example be achieved by using environmental data with a fine spatiotemporal resolution from a selection of sites to estimate the variation around the mean within a certain measurement location both in space and time (Fig. 2a). For example, one can take into account the differences in the variability in soil moisture levels occurring within different environments along large-scale environmental gradients (e.g in mountains vs. flatlands).

Such estimates of within-pixel heterogeneity can be obtained with relatively simple aggregating techniques, for example by using the local standard deviation in addition to the mean, or by fitting correlative models that include within-pixel values as a covariate to incorporate uncertainty. If desired, more complex approaches can be used, for example using mechanistic models that
describe the spatiotemporal variability in abiotic conditions in the system, or by calibrating HSDMs at the local scale using Hierarchical Niche Models (HNMs) – distribution models that allow formally including information from different spatial scales (Mateo et al., 2019).

**Converting nested sampling into gridded datasets**

Importantly, measured variables at the sampling locations are only one part of the story. To allow spatial interpolations of a species’ distribution between the sampling locations, gridded environmental datasets – both of the means and the variation around these means – are necessary. The absence of such gridded datasets for belowground environmental variables with sufficient resolution and reliability is one of the fundamental issues currently hampering further improvements in HSDMs, especially so for belowground organisms. In what follows, we describe how the measured environmental conditions can be converted into gridded products with a spatial resolution far finer than what is currently used, including the even finer-scaled local heterogeneity at the sub-centimeter to meter scale.

Recent work has shown that remotely sensed drivers and proxies can accurately be linked as covariates to measured environmental variables in the soil by using hybrid models combining statistical correlations with mechanistic knowledge of the drivers of environmental variation (Fig. 2, Lembrechts et al., 2018, Zellweger et al., 2019). For example, soil moisture correlates strongly with local topography, for which highly accurate data at meter to centimeter-resolution can be obtained through remote sensing (satellite- or LiDAR-based DEMs) (Sörensen et al., 2006). By combining such remotely-sensed data with *in-situ* measurements, soil moisture can be modelled with increasingly high accuracy, e.g. up to 1 m² resolution across a 3 km² regional extent (Kemppinen et al., 2017). Often, a resolution of 20 x 20 m up to 1 x 1 km is achievable using freely available satellite-based remotely-sensed datasets, like digital elevation models. If one applies these interpolation techniques to both the local averages and the within-pixel heterogeneity, both can be converted into gridded datasets with the desired spatial resolution (Fig. 2b). To accomplish this interpolation of the means and variation in these in-situ measurements, one can either use simple statistical modelling methods - making sure that model residuals are not spatially autocorrelated - such as generalized linear mixed-effects models, generalized additive mixed-effects models, boosted regression trees, or random forests (e.g. Kemppinen et al., 2017), as well as spatially explicit geostatistical approaches such as spatial kriging or geographically weighted regressions (GWR) (Fotheringham et al., 2003, Lembrechts et al., 2019). The latter technique
estimate model parameters at each geographical location by implementing a kernel and weighing explanatory variables by distance. Although such geostatistical approaches have a more explicit way of incorporating spatial structure than traditional generalized linear models or generalized linear mixed-effects models for spatial interpolation of environmental variables, they cannot be used to extrapolate microclimate outside the spatiotemporal extent covered by the data (Lembrechts et al., 2018). While doing this at the scale of single regional studies has thus proven productive, globally coordinated efforts measuring and compiling such data on belowground conditions – preferably using a nested framework - are needed more urgently than ever (Robock et al., 2000, Slessarev et al., 2016, Hengl et al., 2017). Such efforts are indeed required to make a wider range of soil environmental variables available for interpolation at large spatial scales, and to calibrate emerging mechanistic models that allow predicting variables at every relevant scale (Kearney et al., 2014).

Using these techniques shown to be effective in studies on aboveground organisms, in combination with a nested sampling approach can help solve the persisting mismatch between the available gridded environmental data products and the environment as it is perceived by soil organisms themselves, and will greatly enhance the reliability of soil microbial HSDMs. It would indeed allow to significantly improve predictive distribution models, as it permits for microrefugia to be identified within a local pixel through the variation in observed values (Hattab et al., 2014). The number of samples needed to achieve this heavily depends on the local heterogeneity present in the studied variable (e.g. a higher sampling density might be needed for soil moisture in mountainous environments than in flat terrain).

Even if we obtain sufficiently detailed environmental variables to allow the calibration of accurate and fine-grained HSDMs, the question remains which of these environmental variables are driving the regional distribution of soil microbes. Importantly, which variables are relevant for a particular group of microbes should be decided on a case-by-case basis, and requires a thorough understanding of the microbial group under study. Indeed, these drivers – and the scale on which they operate – will depend among others on the size of the studied organism groups. For example, Birkhofer et al. (2012) showed that abiotic soil properties explained significant amounts of variation in fungal diversity, but not in yeasts or bacteria. The use of HSDMs should thus always be seen in parallel with other experimental approaches in which the mechanisms behind the spatial drivers are further disentangled (Ettema & Wardle, 2002).
Nevertheless, we argue here that many available datasets of soil microbial diversity would be suitable for our approach. The idea of nested sampling *per se* is not new in HSDMs (Diez & Pulliam, 2007, Elith & Leathwick, 2009), and applications of similar sampling designs have been used in studies of soil microbial diversity (i.e. repeated local sampling along environmental gradients, or hierarchically structured designs to capture variation at different scales) (e.g. King *et al.*, 2008, Bahram *et al.*, 2016, Zhou *et al.*, 2016). The current scale of most sampling campaigns (collecting soil diversity and environmental information with an accuracy of meters) would indeed already provide sufficient detail to answer a myriad of questions regarding the true relationship between soil microbial spatial distributions and the local environment, especially if one would refrain from the common practice to pool all samples out of a plot together to create an ‘average’ sample, and thus allow objective integration of local-scale heterogeneity into models (Manter *et al.*, 2010). Global database efforts compiling data on soil microbial distributions and/or their abiotic drivers can for example explicitly keep this hierarchy in their structure and, as much as possible, link up the diversity data with in-situ measured local-scale environmental conditions. Emerging efforts to synchronize such sampling efforts across the globe (e.g. Kao *et al.*, 2012, Thompson *et al.*, 2017) are thus pivotal to move this forward, as are efforts to standardize sampling protocols between studies (Halbritter *et al.*, 2020) to reduce uncertainties when merging studies (Ramirez *et al.*, 2018).

**Pooling soil microbial communities**

For the sampling of soil microbial communities, practical restrictions (time and money) often preclude the analysis of large numbers of samples. However, as summarized in Fig. 2c, repeated sampling of local communities within a plot is not a prerequisite for the successful application of our framework. Indeed, soil microbial studies traditionally measure the pooled community structure by analyzing a sample of mixed soil from a random set of locations within a plot and expressing OTUs by their (normalized) read number after sequencing (Staddon *et al.*, 1997, Cleary *et al.*, 2016). This approach on its own already provides an integration of the variability in local communities driven by the local environmental heterogeneity. Linking the pooled community data – which is thus weighed based on the local distribution of environmental conditions - to the average and heterogeneity in local environmental conditions effectively takes into account the necessary fine-scale variation. It is however recommended to validate the local relationship between microbial communities as displayed in Fig. 2b and local environmental conditions in a few sites.
Importantly, note that OTU abundances are semi-quantitative relative abundances only. They are indeed only meaningful in relation to the community in which they are measured, but they do give information on shifts in dominance of a certain OTU between different samples. Additionally, the proposed approach still suffers from false absences in the community data, as do all spatial sampling schemes balancing output with resource input (Jiménez-Valverde & Lobo, 2006). The latter issue is not solved here, and would not necessarily benefit proportionally from increasing sample sizes.

**Including biotic interactions in HSDMs of soil microbiota**

A recurring pattern in analyses of the spatial distribution of soil microbes is that the local spatial distribution of many soil microbiota is not only determined by the abiotic environment, but also by the presence or absence of other taxa (Ettema & Wardle, 2002, Wardle et al., 2004, de Vries et al., 2012, Bahram et al., 2018). Indeed, for many soil microbes spatial distribution largely depends on where competitors, facilitators, mutualists, predators and/or other interactors reside, especially at the local scale. This is for example the case for mycorrhizal fungi, which rely on the presence of relevant host plants (and vice versa), but also strongly interact with each other, other fungi, bacteria, and protists (Sato et al., 2012). To get reliable predictions of their spatial distributions, it will thus be important to include these local biotic interactions, within and between trophic levels, in our HSDMs.

Again, there are a variety of ways in which one can integrate these community effects. One could for example include plot-level metrics of the cohesion of interaction networks (Herren & McMahon, 2017) or metrics for food web complexity (Pellissier et al., 2013) as covariates in HSDMs. Alternatively, a recently developed species distribution modelling tool now also allows to integrate the need for both an environmental and community-based approach in our HSDM-framework: joint species distribution modelling (jSDMs, see e.g. Ovaskainen et al., 2017, Tikhonov et al., 2017, Fig. 1). The so-called jSDMs are a class of models that jointly analyze multiple species’ distributions, quantifying both species-specific environmental responses (i.e. the environmental niche of the focal species) and covariance among species (i.e. positive or negative co-occurrences that can potentially capture biotic interactions). The use of jSDMs has proven a major step forward regarding the modelling of the distribution of a wide range of aboveground species groups, especially when incorporating a hierarchical study design (Ovaskainen et al., 2017). It is important to remember, however, that jSDMs are not suitable for highly diverse communities of
more than 100 species (with example studies commonly using up to around 50 co-occurring
species, Ovaskainen et al., 2017, Tikhonov et al., 2017). One should thus focus on specific groups
of soil microbes, like nitrogen fixers, mycorrhizae or pathogens, and/or restrict analyses to the
most common OTUs only. The latter is realistic, as the structure of belowground communities is
often very uneven, with the top 2% of microbial taxa worldwide shown to make up 41% of microbial
abundance (Delgado-Baquerizo et al., 2018). If one does want to model whole communities of
thousands of taxa, other network based approaches, e.g. using plot-level interaction network
metrics like the cohesion factor of interaction networks as mentioned above can be
recommended.

For optimization of HSDMs for soil microbiota, we strongly recommend this integration of both
abiotic and biotic drivers. Interestingly, this can work both for the interactions between soil
organisms among each other, and between soil organisms and the aboveground world (e.g.
between plant species and their mycorrhiza, if such data is available, see also Fig. 1). Vice versa,
jSDMs can also be used to include the role of soil microbes as drivers of the distribution of
aboveground organisms like plants (de Mesquita et al., 2016). The result is a hybrid model that
allows giving equal importance to specific species and local environmental conditions as drivers
of species distributions.

Such integration of biotic interactions could also help bypass the issue of functional redundancy
in soil microbial ecology, i.e. the idea that many soil microbial taxa would share the same function
and thus are often only absent (or rare) because their redundant counterpart is common (Allison
& Martiny, 2008, Mori et al., 2016). This would result in an apparent unfilling of their ecological
niche, i.e. the absence of the species in an environment with suitable environmental conditions.
However, if the presence of a certain redundant species affects the distribution of the focal
species, this will be picked up by the species correlation matrix used in the jSDM and result in a
high relative importance of biotic co-occurrences.

Conclusion

In this perspective paper, we argued that we should aim to model the current and projected
distribution of soil microbiota at a relevant scale, resolution and measurement location, and with
the relevant variables that drive their distributions. We propose using a nested sampling and
hierarchical modelling approach for the spatial niche and associated distribution, with the
possibility of including different drivers of the distribution of the studied microbiota at different scales, and accounting for (above- and belowground) biotic interactions, and refer to the recent advancements in high-resolution spatial environmental modelling. Applying this framework to existing datasets of soil microbial distributions - and keeping it in mind when setting up sampling networks and globally coordinated efforts - promises to resolve many remaining questions about the local, regional and global distribution of soil microbes. Eventually this approach will allow answering fundamental questions concerning microbial distribution and community assembly, but will also be relevant in applied work such as predicting abundance and resilience of soil-dwelling pests or beneficial soil microbes in agriculture. Importantly, this framework paves the way towards the prediction of changes in soil microbial distributions as a result of different global change drivers, for which first the true relationships between organismal distributions and the environment need to be unraveled.

Funding

This work was supported by a postdoctoral fellowship from the Research Foundation – Flanders (FWO) to JJL.

Acknowledgements

We gratefully acknowledge the support from the Research Foundation - Flanders (FWO) to JJL and the comments from the 3 reviewers.

References


Cleary DW, Bishop AH, Zhang L, Topp E, Wellington EM & Gaze WH (2016) Long-term antibiotic exposure in soil is associated with changes in microbial community structure and prevalence of class 1 integrons. *FEMS microbiology ecology* **92**.


Figure legends

Fig. 1
Figure 1. General overview of the proposed roadmap for HSDMs of soil microbiota at relevant spatial scales. The strength of the approach lies in the nested sampling of above- and belowground communities and environmental data, and the consequent inclusion of those at relevant scales in HSDMs. Microbial OTUs and environmental conditions are sampled both at a high resolution in a selection of plots, and at a coarse resolution across a region (see Fig. 2). The OTU-distribution data is then used to obtain information on the distribution of a specific focal species of interest, as well as on its co-occurrences with other microbes in the microbial community. The environmental variables can provide information on interactions with aboveground organisms (e.g. between plants and their associated mycorrhizas), and on the species’ abiotic niche. The latter can be converted into gridded data, both for coarser-scale averages and their fine-scale heterogeneity, by interpolating them using remotely sensed gridded environmental proxies. Then, the regional distribution of specific focal species of interest can be modelled using habitat suitability and distribution models (HSDMs), as a function of both these biotic interactions and the abiotic drivers, for example using joint species distribution models (see main text). Note that displayed details regarding sample sizes, resolution and spatial distribution of sample points are examples only, with actual requirements depending on the study system, study question and practical limitations.
Figure 2. Detailed view of how the proposed nested sampling approach can be used to improve HSDMs, here exemplified for the link between microbial communities and local abiotic conditions (e.g. soil moisture).
(a) Fine-scale measurements of the abiotic conditions of interest in a selection of small plots, distributed evenly along large-scale environmental gradients (e.g. in similar vegetation types, across topographic gradients in landscapes of 100 x 100 km), can be used to get an estimate of the local heterogeneity in environmental conditions. (b) These local relationships in a selection of plots (red dots) can be extrapolated across a whole region using coarser-grained averages of the same environmental variables (measured in the sites represented by black dots) and the relationships between these local conditions (averages and heterogeneity) and the large-scale environmental gradients. For example, one can extrapolate the
relationships in (a) using high-resolution digital elevation models and, if repeated in different vegetation
types, gridded vegetation maps. (c) The locally present microbial community depends on the local range of
abiotic conditions and the specific abiotic niche of each OTU (visualized here using the same color gradient
as for the environmental gradient). A pooled community composition sample will then have OTU
abundances dependent on the relative frequency of each local environmental condition within a plot. As
this pooled community data thus inherently contains the underlying local-scale variation in the microbial
community, one can calibrate distribution models using pooled community data measured in the red and
black dots depicted in (b). Finally, the distribution of each OTU can be modelled as a function of the
predicted regional averages and heterogeneity in abiotic conditions using HSDMs.