TITLE:

Inconsistencies of empirical ecological network inference are governed by considerations of statistical approaches and dimensions of input data

AUTHORS:

Erik Kusch[1,2,3] (0000-0002-4984-7646)

Malyon D. Bimler[4] (0000-0003-0059-2360)

F. Guillaume Blanchet[5,6,7] (0000-0001-5149-2488)

James A. Lutz[8] (0000-0002-2560-0710)

Alejandro Ordonez[2,3,9] (0000-0003-2873-4551)

AFFILIATIONS:

[1]Machine Readable Nature Research Group, Natural History Museum, University of Oslo, Oslo, Norway, Norway

[2]Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Biology, Arhus University, Aarhus, Denmark

[3]Section for Ecoinformatics & Biodiversity, Department of Biology, Arhus University, Aarhus, Denmark

[4]School of Biosciences, University of Melbourne, Parkville, Victoria, Australia

[5]Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

[6]Département de mathématiques, Université de Sherbrooke, Sherbrooke, Québec, Canada

[7]Département des sciences de la santé communautaire, Université de Sherbrooke, Sherbrooke, Québec, Canada

[8]Wildland Resources, S. J. & Jessie E. Quinney College of Natural Resources and the Ecology Center, Utah State University, Utah, United States of America

[9]Center for Sustainable Landscapes under Global Change (sustainscapes), Department of Biology, Arhus University, Aarhus, Denmark

CORRESPONDING AUTHOR:

Erik Kusch; erik.kusch@nhm.uio.no

Office K07, Lids Hus, Botanisk Hage, Sars' gate/Monrads gate, 0562 Oslo

STATEMENTS & DECLARATIONS

COMPETING INTERESTS

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

ACKNOWLEDGEMENTS

This research was supported by the Aarhus University Research Foundation Start-up Grant (grant no. AUFF-2018-7-8) to AO. AO considers this a contribution to the Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), funded by Danish National Research Foundation (grant DNRF173). Additionally, this work was supported by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101057437 (BioDT, https://doi.org/10.3030/101057437).

AUTHOR CONTRIBUTIONS

EK and AO conceptualised this study. EK curated data and led the analysis. EK created all R scripts necessary for the analyses. NDD-RIM integration was aided by MB. Results were interpreted by EK and AO. YFDP data was provided by JAL. FGB provided critical feedback and aided writing efforts. Manuscript writing was led by EK. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY

This submission uses novel code, provided at https://github.com/ErikKusch/Ecological-Network-Inference-Across-Scales. Download queries for raw FIA and ERA5-Land data are specified in the R scripts (FIA data at https://apps.fs.usda.gov/fia/datamart/datamart.html; ERA5-Land data at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means). Raw YFDP data was downloaded from the Smithsonian ForestGEO data portal using this form http://ctfs.si.edu/datarequest/index.php/request/form/6 and specifying the following query: “full data predating fire-event in 2013”. All subsequent handling of raw data is detailed in the R scripts.

ABSTRACT

Identifying the most suitable method of ecological network inference in line with individual research considerations is a non-trivial task, which significantly hinders adoption of network approaches to forest management applications. To advance the study of ecological networks and better guide their use in managing forest ecosystems, we propose a framework that aligns pairwise species-association inference methods with specific research questions, biological interaction types, data availability, and spatial scales of study. We motivate the adoption of this framework through an empirical comparison of multiple inference methods, highlighting substantial inconsistencies that arise across scales and methodologies. Using data on species distributions and attributes at local, regional, and continental scales for temperate conifer forests in North America, we show that network inference varies significantly depending on whether occurrence, abundance, or performance data are used and the degree to which confounding factors are accounted for. Across four widely used and/or cutting-edge inference methods (COOCCUR, NETASSOC, HMSC, NDD-RIM), we find notable disparities in both whole-network metrics and pairwise species associations, particularly at continental scales. These findings underscore that no single method is likely to universally outperforms others across scales, emphasizing the importance of choosing an inference approach that aligns with specific ecological and spatial contexts. Our framework aids in interpreting network topologies and interactions in light of these method- and datatype-driven variances, providing a structured approach to more reliably infer ecological associations and address complex network dynamics in forest management practices.

KEYWORDS:

Cooccurrence, Ecological Networks, Ecological Network Inference, Method Comparison, Spatial Scale, Species Associations

1. INTRODUCTION

Climate change, habitat loss, and biological invasions are reshaping Earth's biosphere (Blowes et al. 2019), including global forest communities (Allen et al. 2010), where altered species interactions (Kunstler et al. 2012) challenge ecosystem functioning (Woodward et al. 2010), forest resilience (Verbesselt et al. 2016; Liu et al. 2019) and contemporary management practices (Tang and Gustafson 1997; Allen et al. 2010; Rist and Moen 2013). Consequently, ecological and forestry research increasingly emphasizes the need to consider species interactions when predicting the emergence of novel communities and changes in ecosystem behaviour and functioning (Gilman et al. 2010). Elemental to this paradigm shift are ecological networks, which encode species identities as nodes and pairwise species interactions as links (Morales-Castilla et al. 2015) and predictions of how these nodes and links would be altered following changes in composition of ecosystems and biological communities (Kusch and Ordonez 2023).

Traditionally, studies of ecological networks have focused on local-scale interactions (Lavorel and Garnier 2002), but recent efforts have expanded to regional (Hackett et al. 2019) and even global meta-networks (Fricke et al. 2022), revealing broader-scale interaction patterns critical for forest ecosystems. In response to this trend, various inference approaches have been developed to predict interspecies interactions from species-specific input data, even where direct interaction data from field measurements or observations are unavailable (e.g., Griffith et al. 2016, Blonder and Morueta-Holme 2017, Tikhonov et al. 2020, Momal et al. 2020). However, these inference approaches encapsulate different ecological interpretations of interactions - such as mutualistic, competitive, or trophic links between species. Whereas trophic networks typically infer biological interactions as directed links (i.e., individuals of different species are influencing each other), plant-plant networks usually focus on biological association expressed as undirected links (i.e., two species occur at the same location but it is not clear whether they are interacting). While many inference approaches have been suggested for undirected link inference (e.g., Griffith et al. 2016, Blonder and Morueta-Holme 2017, Tikhonov et al. 2021), only few have been proposed for the inference of directed links (Freilich

et al. 2018; Bimler et al. 2023). Consequently, forest researchers and managers need to be aware of inference approach specific link-types when selecting ecological network inference approaches for their purposes.

In addition to the distinction between direct and undirect links, methods of network inference can also be categorised along two additional data-driven axes: (1) the type of biodiversity data that is used to represent species characteristics and (2) the capacity to integrate additional data sources such as environmental conditions (Figure 1). First, biodiversity inputs and attributes which inform the inference of species interactions/associations may encode relevant information either as species occurrence, abundance, or performance – forming a spectrum of information content ranging from occurrence to performance. This spectrum establishes three distinct categorisations. Here, network inference methodologies focus on (1) occurrence (Veech 2013) or performance either implicitly through (2) abundance (Ulrich and Gotelli 2010) or (3) explicitly through biomass or fitness measurements (Bimler et al. 2023). This occurrence-performance axis arguably represents one of the most important philosophical and methodological discriminating characteristics between network inference approaches. Second, the capacity to account for different covariates such as environmental gradients, phylogenetic relatedness, and functional traits of species carries great implications for network inference performance (Morales-Castilla et al. 2015; Kusch and Vinton 2023). Approaches along this spectrum can be sorted based on covariates, like environmental conditions, being either absent or present (again, either implicitly or explicitly) when inferring the interactions/associations in a network. Most ecological network inference methodology does not incorporate such information (Blanchet et al. 2020), most likely due to data availability constraints, concerns of collinearity, or complexity of analyses. Despite the proliferation of network inference approaches for various purposes, there thus exists grounds for classification and investigation of these statistical methods to gain a better understanding of their applicability for the study and management of forest environments which has, thus far, remained unexplored. While ultimately network inference approaches should be selected based on their accuracy, comparing the accuracy of different methods requires extensive parameterisation of complex

and computationally expensive simulation-validation efforts (Kusch and Vinton 2023), precluding universally applicable choices of network inference approaches at this point in time. In lieu of such wide-ranging quantitative comparisons of inference approach accuracy, a framework to support judgement of inference approach suitability and applicability for specific study purposes is prudent, timely, and within reach.

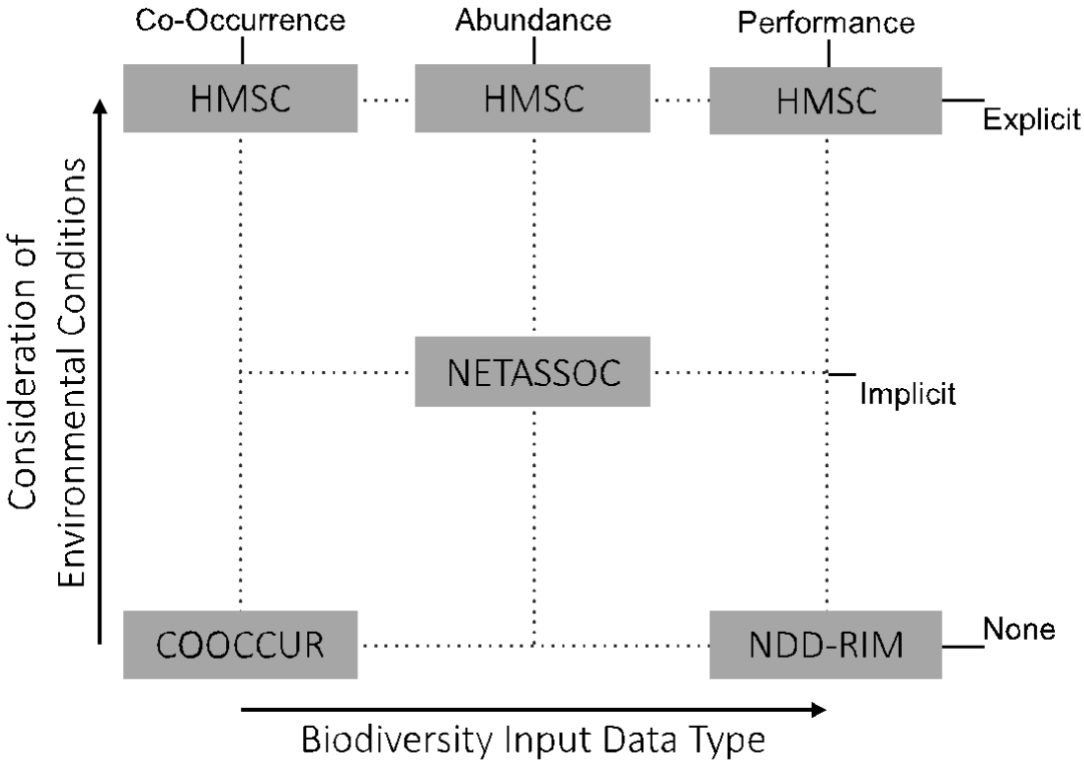


FIGURE 1 – PERFORMANCE-ENVIRONMENT ORDINATION OF ECOLOGICAL NETWORK INFERENCE APPROACHES – Network inference methodology can be distinguished with regards to their biodiversity input data type (corresponding to the occurrence-performance spectrum, see Figure S1) and their consideration of environmental conditions. The network inference approaches we selected for our study were chosen due to their comprehensive coverage of the resulting spectrum of data type use.

Delineating such a framework for choice of network inference approach with respect to link-type and data axes classifications would prove useful for the global community of forest ecosystem researchers and managers facing the decision of how to infer ecological networks for their diverse study systems at varying scales. Firstly, forest ecosystems contain both biological interactions as well as biological associations (Bello et al. 2017; Damgaard and Weiner 2021), requiring network inference approaches capable of rendering these. Secondly,

forest ecosystems are understood and managed vastly differently when being assessed for species richness (e.g., management for biodiversity), biodiversity as weighted by abundance of individual species, or through measures of wood production or biomass (e.g., management for forestry and production; Rist and Moen 2013). Thirdly, the impacts of including additional covariates in analyses of ecological networks has been demonstrated by climatological conditions altering the persistence of competing species (Usinowicz and Levine 2021) and the role of functional traits in governing interactions between forest tree species (Kunstler et al. 2012). Consequently, this framework will help answering the pressing question: *How comparable are different methods of ecological network inference in real-world applications*? The necessary comparisons of modern ecological network inference approaches at scales relevant to forest management remain absent so far.

Here we establish such a framework for inference method choice in forest research and management to facilitate deeper understanding of how to best implement ecological network approaches to forestry. To do so we compare and quantify similarities between four methods of ecological network inference when inferring pairwise-associations among woody forest species across local, regional, and continental scales. Our analysis focuses on two main areas: (1) assessing consistency in network metrics across inference methods at the same scale and (2) evaluating consistency across spatial scales using the same method. We hypothesize that networks inferred using occurrence versus performance data will show structural differences, with greater similarity among methods using similar input data types. Furthermore, we anticipate that differences between approaches that account for environmental gradients and those that do not will increase at broader scales, where these gradients are primary drivers of species distribution (Stephan et al. 2021). Finally, we expect network structures to vary across scales, with spatial changes in species pools contributing to network dissimilarity but unable to fully explain the observed discrepancies (Carstensen et al. 2014). Synthesising the outcomes of this empirical consistency analysis, we conclude with a framework to guide choice of network inference approach for the study of plant-plant components of forest systems across scales.

2. MATERIAL & METHODS

2.1. ECOLOGICAL NETWORK INFERENCE METHODOLOGY

We inferred horizontal undirected (i.e., non-trophic) ecological plant-plant networks using four methods (COOCCUR, NETASSOC, HMSC, and NDD-RIM) that encapsulate a considerable variability in statistical principles, data requirements, and ecological interpretations/linktypes (Table 1). A verbose description of each inference method is provided in the supplementary material. These methods have been used primarily to infer ecological associations between species-pairs at scales ranging from local to continental applications. While other inference methods have been established (e.g.: Damgaard et al. 2018, Swain et al. 2021, Lamonica et al. 2021), we have selected this set of methods as distinct placements across the performance-environment ordination we propose for classification of such approaches (Figure 1).

Data handling, network inferences, and analyses were carried out in R (R Core Team 2021). The packages cooccur (Griffith et al. 2016), netassoc (Blonder and Morueta-Holme 2017), and hmsc (Tikhonov et al. 2021) were used for the COOCCUR, NETASSOC, and HMSC methods, respectively. NDD-RIM was implemented via custom R code (see data availability statement). The workflow diagrams of data handling and network inference are presented in Figures S2 and S3.

TABLE 1 - METHODS OF NETWORK INFERENCE - Description of statistical principles, data requirements, and ecological interpretation of outputs of the four methods of network inference used in this study. The evaluated methods are: **Probabilistic Species Co-Occurrence Analysis in R** (COOCCUR), **Inference of Species Associations from Co-Occurrence Data** (NETASSOC), **Hierarchical Model of Species Communities** (HMSC), and **Neighbour-Density Dependent - Response Impact Model** (NDD-RIM). In our implementation of NETASSOC, range estimates for local-scale data were not used as they did not represent informative null models. In our implementation of HMSC**,** trait means for each species were only used locally as functional trait expressions become too diverse at regional scales to be expressed with a mean value.

Method	**Principle**	**Input**	**Output**	**Interpretation of Links**
COOCCUR (Griffith et al. 2016**)**	Probability of species-pair co-occurrence being higher or lower than random.	▪ Site-by-Species occurrence	Standardised effect sizes (differences between expected and observed frequency of co-occurrence)	Associations, Spatial aggregation/disaggregation of species-pairs
NETASSOC (Blonder and Morueta-Holme 2017**)**	Frequency of species-pair co-occurrence compared to null model.	▪ Site-by-Species occurrence or abundance ▪ Site-by-Species expectation of range estimates as occurrence	Standard partial correlations between species-occurrence pairings rescaled by mean and standard deviation of random generations or site-by-species expectations	
HMSC (Tikhonov et al. 2021**)**	Hierarchical modelling of community assembly processes were residual species-species correlations are treated as association metrics.	▪ Site-by-Species occurrence, abundance, or performance ▪ Environmental conditions at each site ▪ Phylogeny of species ▪ Trait means for each species	Residual correlations between species-pairing in joint distribution model	Associations, Aggregation/disaggregation (in space, abundances, or performance) of species-pairs that aren't explained through other community assembly processes
NDD-RIM (Bimler et al. 2023**)**	Neighbour-density dependent regression and deconstruction of species interactions into response and effect coefficients for inference of rare interactions.	▪ Individual/species performance at each site ▪ Site-by-Species matrix of abundance	Regression model coefficients describing how a change in abundance of one species affects the performance of another species	Interactions, Change in performance of interaction subject with increase in abundance of interaction actor

2.2. DATA SOURCES & DATA HANDLING

We inferred networks based on two datasets of North American woody plants: the Forest Inventory Analysis Database (FIA, Burrill et al. 2018) and the Yosemite Forest Dynamics Plot (YFDP, Lutz et al. 2012). These were aggregated to describe species at three scales: local (0.256 km^2), regional (3,030.17 km^2), and macro (1,508,451 km^2) representing existing scale-consideration in forest management of stand and landscape-scales (Tang and Gustafson 1997).

The local scale pool of species was defined using the YFDP data, which is a 25.6 ha study area within which all trees ≥1 cm diameter at breast height (dbh) are measured according to the protocols of the Smithsonian ForestGEO network (Davies et al. 2021). We used a subset of the YFDP data set containing 34,589 records of dbh of 11 woody plant species measured in a grid of 640, 20 m × 20 m quadrats which we subsequently treated as individual plots. For consistency with the FIA, we considered only those stems ≥2.5 cm dbh (Gray et al. 2012), retaining the records of 34,444 individuals and their dbh, as a measurement of resource allocation to growth, for network inference. These data were further aggregated to express species' abundances and occurrence in the 640 quadrants of the YFDP.

Species pools at the regional and macro scales were defined by aggregating FIA-plot data. This dataset summarizes information on tree species' presence, abundance, and size/biomass across the forested portion of the United States of America (Gray et al. 2012). To remove possible biases in inferred association matrices introduced by rare species, we removed all species present in less than 10% of all locations for our analyses. We used the FIA-contained records of biomass per species per plot as proxies for species' performances. The regional scale species pool was defined by limiting FIA-plot data to the region encompassing Yosemite National Park as defined by the U.S. National Park Service (National Park Service 2009). This selection resulted in 291 observations (i.e., biomass and abundance of a species at an FIA-plot) at 101 plots containing 13 species. The macro scale species pool was defined by limiting FIA-plot data to the Temperate Conifer Forests Biome (World Wildlife Fund 2020). This masking resulted in 96,169 observations across 46,328 plots containing 15 species.

2.2.1. Functional Trait Data

The YFDP dataset also includes species-specific measurements of multiple functional traits. We selected specific leaf area, leaf carbon content, and leaf nitrogen content for our analyses. These traits describe important trade-offs in plant investment in photosynthetic organs and storage of resources (Wright et al. 2004). The FIA dataset does not record functional trait measurements for its respective scales. While functional trait data is available for our target species from functional trait repositories like BIEN (Enquist et al. 2009), functional leaf trait expression is modulated by the environment (Wright et al. 2005), potentially leading to an inaccurate representation of local realities at large scales (Violle et al. 2012). We therefor chose to use functional trait means exclusively when implementing HMSC at local scale (Table 1).

2.2.2. Environmental Data

To quantify plot-level environmental conditions experienced by species/individuals, we matched air temperature, soil moisture in a depth of 0-7 cm, total precipitation, and potential evaporation to the spatial scale of our data using the KrigR package (Kusch and Davy 2022). We chose these climate parameters due to their link to vegetation performance (Kusch et al. 2022). To reduce biases introduced by climate regime changes, we aggregated environmental conditions to 10-year climatologies leading up to the sampling date of each plot.

2.2.3. Phylogenetic Relatedness

We established phylogenetic relatedness by pruning a phylogenetic tree supplied with the V.PhyloMaker package (Jin and Qian 2019). Species not represented in this backbone phylogeny were removed from our analyses which was in part why at the local, regional, and macroecological scale the above-mentioned 11, 13, and 15 species were respectively selected.

2.2.4. Species Distributions

We obtained species range maps (for use with NETASSOC, Table 1) using the BIEN package (Maitner et al. 2018) to inform null expectations of species co-occurrences at FIA scales (i.e., regional and macro). At the YFDP scale, null models for NETASSOC were established by resampling the observed sites-by-species matrix.

2.3. Network Consistency Quantification

To quantify dissimilarities of inferred ecological networks at the level of individual links (interactions/associations) and node sets (species composition), we calculated association beta-diversity (a relative measure of how many present species-species links and/or species identities are shared between two networks) using the betalink (Poisot et al. 2012) R package. This comparison evaluates whether link patterns or species turnover drive changes in the identity and sign of links. We also implemented a comparison of inferred network matrices accounting links that are absent and present links thus calculating an additional inference dissimilarity score which takes into account absent associations between species shared across scales (which the betalink implementation of network beta-diversity does not). For this analysis, we reduced link sets to statistically significant links (p-value < 0.05 for COOCCUR and NETASSOC; 90% credible interval not overlapping 0 for HMSC and NDD-RIM) in the inferred weighted ecological networks. To ensure comparability across methods, we collapsed opposing links in NDD-RIM networks into mean links when computing topological metrics and assigned statistical significance only when both underlying directed links were identified as statistically significant. Additionally, we changed link weights to -1 and 1 based on their original sign to mimic large-scale decision-making information for forest inventory management of whether species-pairs suggest facilitation or competitive exclusion. The pairwise associations common to all inferred networks owing to the shared species are highlighted with black outlines in Figures 2 and 3. Subsequently, we were able to assess network inference consistency of only significantly inferred non-absent links as well as entire link-sets including non-statistically significant (thus absent) inferred links across diverging species pools and scales of assessment.

3. RESULTS

To assess comparability and consistency of inferred ecological networks across the performance-environment ordination, we focus on plot scale YFDP data (Figure 2) because most data types are available at this scale (Table 1). The corresponding trends observed at plot-scale are mirrored at regional and macro-scales (Figure S4 & Figure S5).

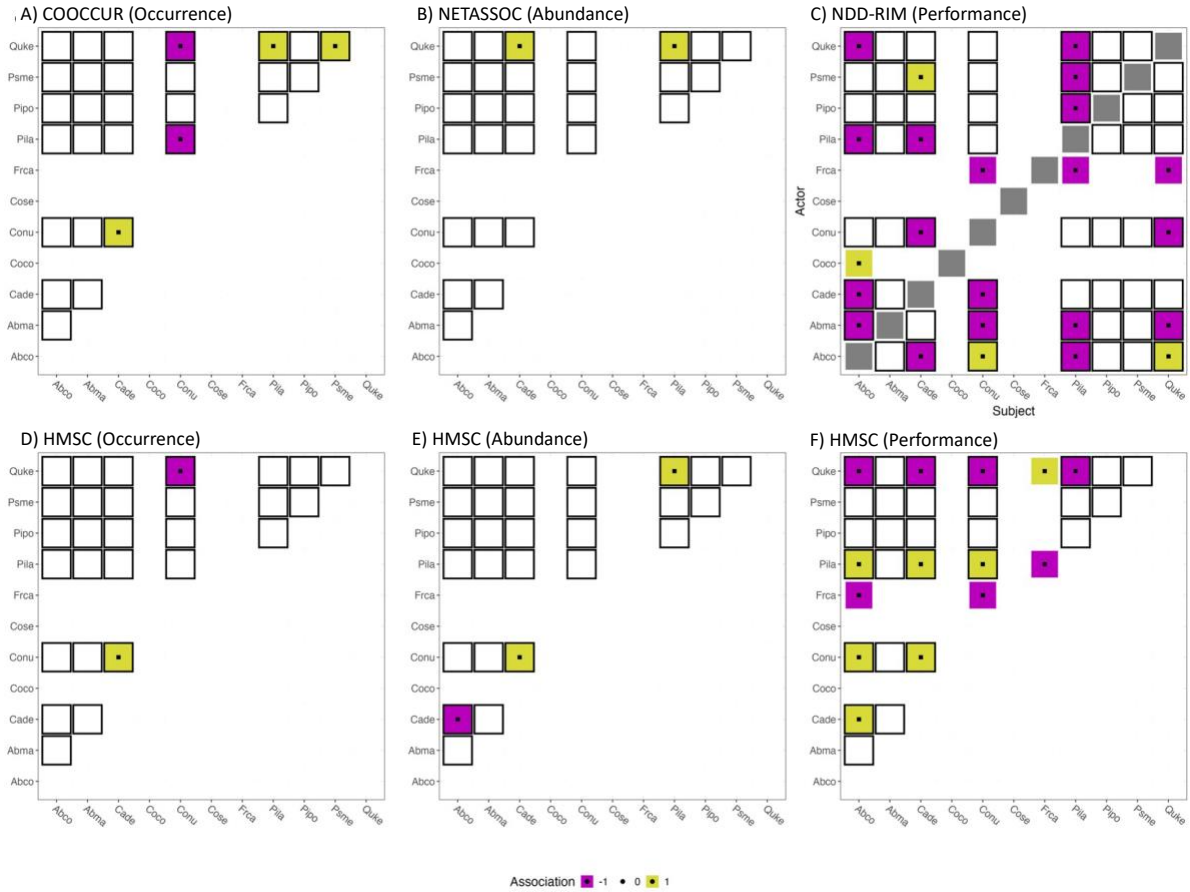
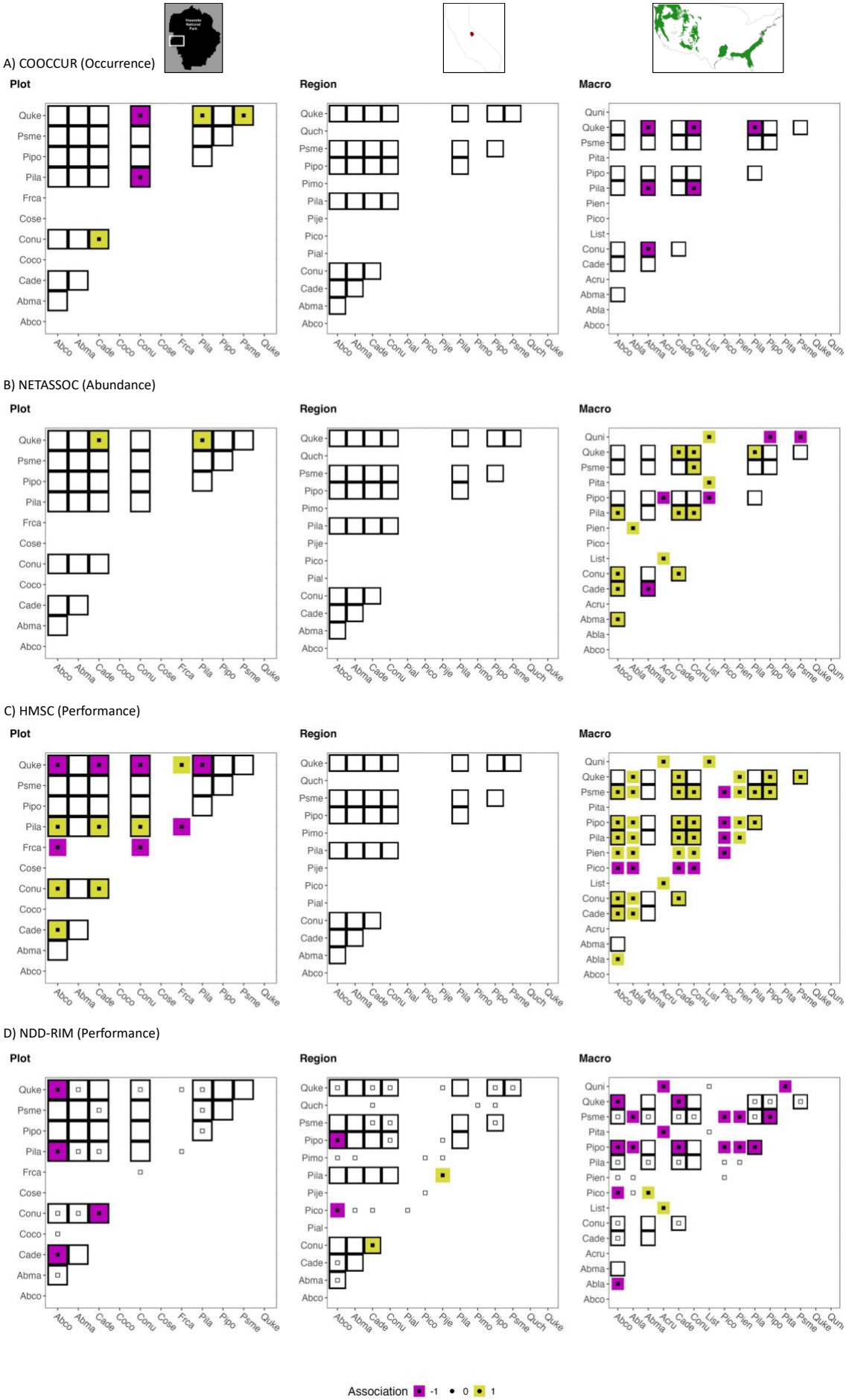


FIGURE 2 - MODEL SPECIFICATION DRIVES NETWORK INFERENCE – Ecological networks inferred at plot scale using the four methods of network inference with differing specifications of biodiversity products (i.e., occurrence, abundance, performance). COOCCUR (A), NETASSOC (B), NDD-RIM (C) represent use-cases for occurrence, abundance, and performance data. HMSC (D-F) can incorporate occurrence (D), abundance (E), or performance (F) data. Inferred associations/interactions have been set to +1 (facilitative), -1 (competitive) or 0 (no link inferred) to ensure comparability. Black outlines highlight species pairings shared across all scales of assessment. Black squares in the network matrices signify statistically significant inferred connections between species. See table S1 for an overview of four-letter abbreviations and corresponding species identities. Note that NDD-RIM (C) infers interactions rather than associations hence rendering non-symmetrical network matrices.

Pairwise inferred links showed consistent patterns across methods for the YFDP dataset (Figure 2). This pattern holds even for different specifications of the HMSC approach (Figure 2D to F). Inference methods focused on performance data (NDD-RIM and HMSC in Figure 2F) identify more statistically significant links between species pairs than those focused only on occurrence or abundance data (Figure 2A, B, D, E) despite notable similarities between occurrence- and abundance-informed HMSC inference at the plot scale.

Inferred network matrices for the eight-species pairings shared across all scales (black outlined tiles in Figure 2 and 3) show no scale consistency even when using the same inference approach. The sign, magnitude, and statistical significance of links change across scales. In the case of HMSC (Figure 3C), we infer a heterogenous network matrix at the plot scale (containing negative and positive statistically significant links) which changes to an exclusively positive, statistically significant link matrix at the macro scale. This scale-inconsistency can also be seen for COOCCUR (Figure 3A) and NETASSOC (Figure 3B), where we fail to detect any statistically significant associations at the regional scale. Lastly, using NDD-RIM, there is little consistency in the sign and magnitude of statistically significantly identified directed links at the plot scale (Figure 2C), regional scale (Figure S4C) and macro scale (Figure S5C). These inconsistencies remain when collapsing opposing directed links inferred by NDD-RIM into undirected links (Figure 3D).


A) COOCCUR (Occurrence)
B) NETASSOC (Abundance)
C) HMSC (Performance)
D) NDD-RIM (Performance)
Plot
Region
Macro
Quke Psme Pipo Pila Frca Cose Conu Coco Cade Abma Abco
Abco Abma Cade Coco Conu Cose Frca Pila Pipo Psme Quke
Quke Quch Psme Pipo Pimo Pila Pije Pico Pial Conu Cade Abma Abco
Abco Abma Cade Conu Pial Pico Pije Pila Pimo Pipo Psme Quch Quke
Quni Quke Psme Pita Pipo Pila Pien Pico List Conu Cade Acru Abma Abla Abco
Abco Abla Abma Acru Cade Conu List Pico Pien Pila Pipo Pita Psme Quke Quni
Association -1 0 1

317

FIGURE 3 - NETWORK INFERENCE IS NOT CONSISTENT ACROSS APPROACHES OR SCALES – Contrasting the ecological networks inferred using four separate methods (A-D) reveals a lack of scale-consistency when assessing species-species links across geographic scales (i.e., plot scale – left column, region scale – centre column, and macro scale – right column). Inferred associations have been set to +1 (facilitative), -1 (competitive) or 0 (no link inferred) to ensure comparability. Black outlines highlight species pairings shared across all scales of assessment. Filled dots in the network matrices highlight statistically significant inferred connections between species. Empty dots indicate in NDD-RIM visualisation (D) indicate that only one of the underlying directed links was identified statistically significantly. See table S1 for an overview of four-letter abbreviations and corresponding species identities.

Quantifying inference dissimilarity via decomposition of network-dissimilarities into node-set and link-pattern changes reflects the overall lack of inference consistency across scales and approaches. Dissimilarities of present (positive or negative in sign and non-zero in magnitude) links within scales are high and cannot be traced back to species turnover (Figure 4A) as each network of the same scale consists of the same node set. The same is true for link sets including absent (i.e., non-statistically significant inferred links) as well as present links but to a lesser degree due to sparse inference of present associations (Figure 4B). Instead, these must be attributed to each inference approach's input data or statistical philosophy. In line with our expected differences due to the use of occurrence, abundance, and performance data, the lowest dissimilarity scores of present links within-scales can be found between HMSC and NDD-RIM outputs. This is likely due to similar information being captured in abundance and performance records (Figure S7). As for abiotic factors governing ecological interactions and their inference, we find comparatively low dissimilarity of present links between HMSC and NETASSOC at the macro scale, for which we hypothesised climatological gradients to play crucial roles. Considering also absent links no such insights can be found. Instead, we identify HMSC as inferring vastly different link motifs as compared to the other approaches. This may be an effect of HMSC's ingestion of additional data (i.e., functional trait expressions and phylogenetic relatedness).

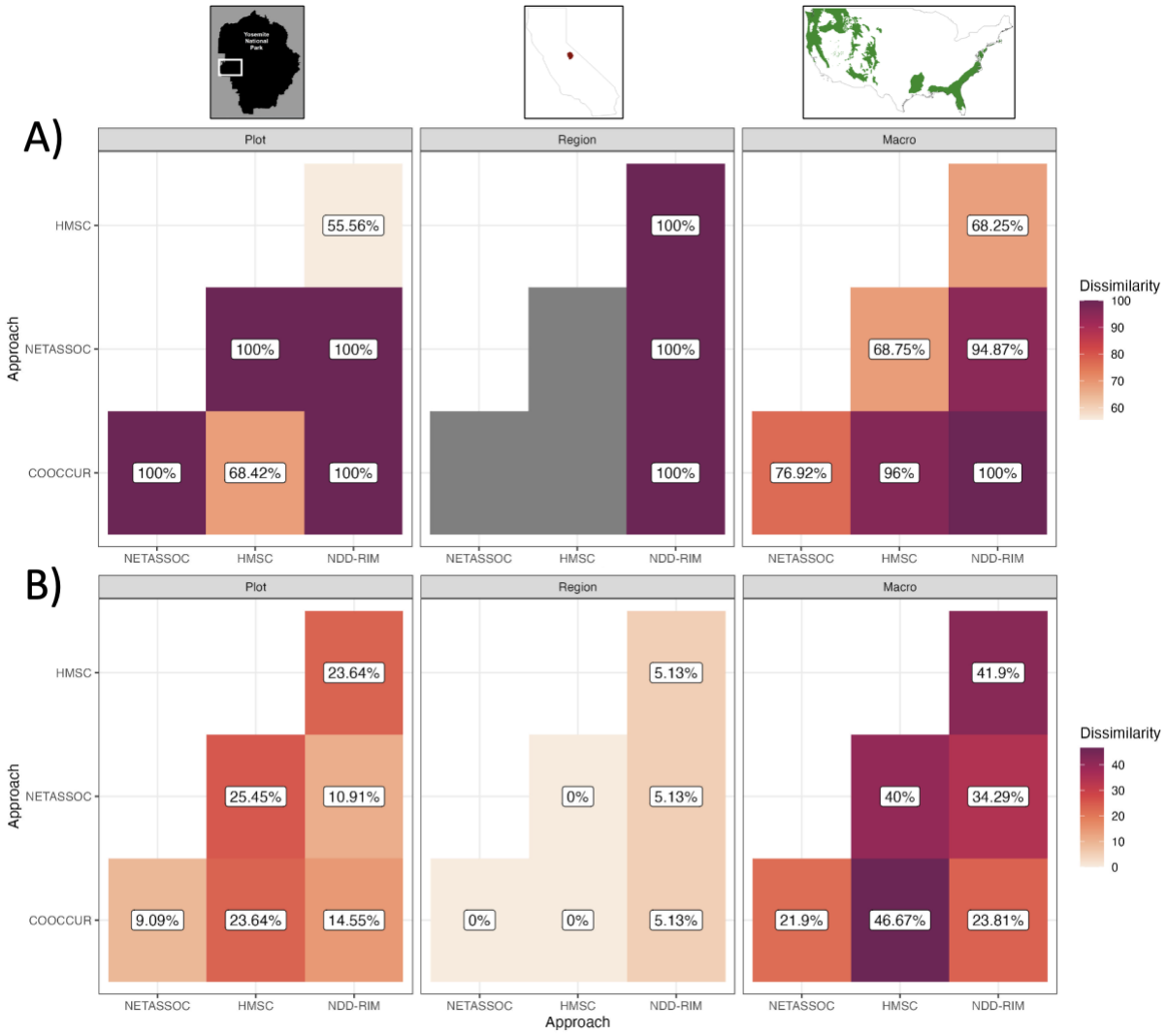


FIGURE 4 – NETWORK DISSIMILARITY ACROSS APPROACHES WITHIN SCALES OF ASSESSMENT – There exists considerably dissimilarity between ecological networks made up of A) statistically significant links and B) statistically significant links and links inferred as absent by the target approaches within each scale of assessment. Notably, inference at region scale (centre panel) was held back by low data availability.

When investigating cross-scale consistencies, we find that dissimilarity of only present links as well as present and absent links is driven by changes in links between species present across the local, regional, and macroecological scale instead of added or removed link structures introduced by species turnover. This dissimilarity may be reduced by large amounts of data (Figure S6). This finding is consistent for all inference methods assessed here with some approaches rendering significantly more similar inferences across scales of assessment than others (Figure S6). Finally, these inconsistencies extend beyond comparisons of individual link

structures and lead to a lack of consistency in network topology as described by modularity and node centrality (see supplementary material and Figures S8 & S9).

4. DISCUSSION

We have identified a lack of consensus among networks inferred for forest ecosystems using COOCCUR, NETASSOC, HMSC, and NDD-RIM. This discrepancy could be attributed to (1) differences in philosophy of methods (e.g., performance-environment ordination placement or targeted link-types), (2) scale-dependant applicability of different methods and data availability/richness, (3) realised biological interactions change across scales due to differences in species pools, and (4) changes in the ecological processes driving community assembly across scales.

4.1. SCALE-CONSISTENCY & BIOLOGICAL PROCESSES

Biological processes revealed by ecological networks are conditional on the choice and scale of the inference method. The approaches of ecological network inference assessed here focus on representing networks emerging from different ecological processes. While COOCCUR and NETASSOC aim to identify species associations and their outcome on space-utilisation, HMSC is primarily aimed at disentangling community assembly processes with link-inference as a side-product of residuals. In contrast, NDD-RIM considers asymmetries in species associations and aims to identify the effect of one association partner on the performance of the other. In addition, the response-impact model component of NDD-RIM may prove helpful in predicting novel interactions as communities reshuffle, thus giving a valuable set of testable hypotheses for experimental macroecology (Alexander et al. 2016) and management.

Figure 3 and Figure S6 illustrate that 1-1 associations are inconsistent across geographic scales when inferred using the same approach. These scale-inconsistencies are not necessarily a flaw in the methods of network inference but instead a feature of (1) the inference method applicability at varying scales, (2) the scale-dependence of species-interactions, and (3) which biological processes are being captured at specific geographic scales.

Inferred species links can be considered data-driven hypotheses about realised biological interactions (Ovaskainen et al. 2016), which have been demonstrated to change in strength and sign when additional species are introduced/removed (Blanchet et al. 2020; Bartomeus et al.

2021). Our analyses of network link beta diversity suggest that scale-inconsistencies of 1-1 associations are partially driven by variation in realised interactions but cannot be explained by changes in species pools alone. Therefore, while there is a precedent for using ecological networks measured/inferred at one scale to draw conclusions at other scales (Guimarães 2020), our findings indicate that neither 1-1 associations nor whole-network topology measures should be transferred between scales and management decision based on such procedures may prove to be erroneous and detrimental in forest management. For example, we show that ecological networks inferred for individual forest stands such as the YFDP plots would lead to vastly different forest management decisions of temperate conifer forests than decisions based on inferences at whole-biome level. We recommend ecological networks be inferred and interpreted at the scales at which they are to be used for further analyses and decision-making. Additionally, we caution against overconfidence in inferred ecological networks as even within distinct scales can networks vary vastly across approaches and under different parameterisations of the same approach (HMSC inferences in Figures 2, S4, and S5).

4.2. Choice of Network Inference Approach Across Scales

Since quantifications of inference performance for methods of ecological network inference under relevant study settings are absent and will likely remain so, we propose to orient choice, use, and interpretation of ecological network inference for forest ecosystems within a framework accounting for the sources of inference discrepancies we have identified. Choice of network inference method should be driven by a combination of criteria. Most importantly, (1) the type of ecological association/interaction which is targeted should be the first selection criterion to be examined. Additionally, our results highlight two additional method selection criteria: (2) biodiversity inputs available to analyses and (3) data integration and availability which we find to be of particular relevance as the scale of study increases. We present these selection criteria and their relation to the assessed network inference approaches in Figure 5.

Firstly, ecological network inference approaches have typically not been designed to disentangle asymmetric links. Such directed links characterise biological interactions where

each species is influenced differently by the link (e.g., commensalism, predation). In truth, most network inference approaches assume ecological networks are undirected thus suggesting that an ecological network gather either beneficial (e.g., mutualism) or harmful (e.g., reciprocal competition) associations. Due to the ubiquity of these approaches, we have focussed here almost exclusively on such association-based approaches. However, novel approaches such as NDD-RIM have been shown to move past this constraint and infer directed networks, opening the possibilities to study a broader range of association among species (Bimler et al. 2023). Such approaches should be able to account for the entire spectrum of biological associations and interactions (see Figure 5). Thus, while association-based methods may be perfectly suitable to address questions regarding the spatial distribution and co-occurrences of species, performance-based methods are more appropriate for studying the consequences of species interactions/associations on demographic processes. While this capability of inference of directed links lends itself to a wider range of ecological applications, inference of directed links also require more detailed data.

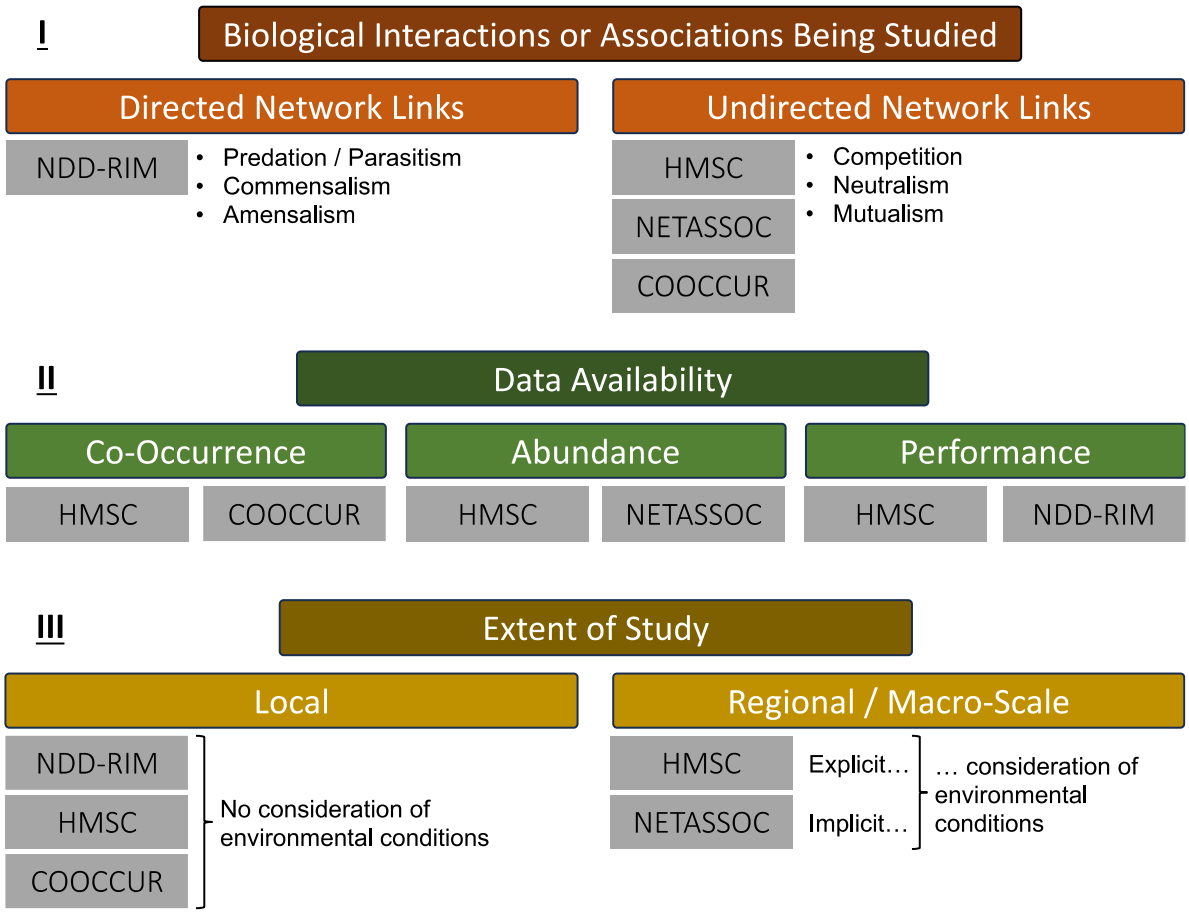


FIGURE 5 – CRITERIA FOR CHOICE OF NETWORK INFERENCE METHOD – Choosing an appropriate approach to study network inference is a complex task that requires careful consideration of three fundamental criteria: I) Which biological connection (association or interaction) is the main interest of the study? Whether inferred networks ought to be directed or undirected should help define the choice of method. II) What data is available? When performance or abundance data is available, it is likely more prudent to use a method of network inference designed to work with this data. III) What is the scale of interest? At regional and macroecological scales, methods of network inference which incorporate environmental information are likely to be the most accurate. Any method of network inference which satisfies the above three criteria for a specific study setting is likely to be more applicable for the desired purpose than a method which does not.

Secondly, while directed-link inference methodology requires abundance/fitness data, inference of undirected links can be carried out using species co-occurrence patterns. This traditional reliance on occurrence data for network inference has been criticised as being too simplistic to measure ecological process outcomes (Blanchet et al. 2020). As seen in Figure 2 and Figure 3, methods of network inference that leverage abundance or performance data identify more statistically significant associations than their occurrence-driven counterparts

and thus create more actionable knowledge for application in forest management. We postulate that differences in identified statistically significant ecological associations is partially driven by the amount of biologically relevant information in abundance and performance data compared to occurrence datasets (i.e., quantification of species performances implicitly and explicitly vs. qualitative information). Indeed, performance-based present-links reported the lowest dissimilarity scores (Figure 4). Such differences in information content have already resulted in null model analyses to discriminate community assembly rules being augmented with abundance data (Ulrich and Gotelli 2010). In addition, a recent simulation-based network inference validation framework suggests that the increasing information content along the occurrence-performance spectrum translates directly into higher accuracy of inferred ecological networks (Kusch and Vinton 2023). We suggest that network inference approaches should be subjected to this paradigm shift, thus leveraging performance or abundance data wherever available despite the considerable data collection efforts this entails, particularly at broad geographic scales.

Thirdly, macroecological research has long recognised the relevance of environmental conditions on species presence (Wilkinson et al. 2019), abundance (Johnson and Sinclair 2017), and performance (Coutts et al. 2016). Thus, the scale applicability of ecological network inference depends on whether or not climatological gradients ought to be considered in a given study. Of the four evaluated methods, COOCCUR and NDD-RIM do not account for climatological gradients, while NETASSOC does so implicitly through range expectations, and HMSC does so explicitly through direct use of climate data to quantify contributions of bioclimatic niches and environmental conditions to species performance. By accounting for environmental gradients, which alter species co-occurrence potential, NETASSOC and HMSC address one recurring criticisms of ecological network inference (Blanchet et al. 2020) and may drastically improve network inference accuracy even when leveraging occurrence data (Kusch and Vinton 2023). The effect of climatological data integration into ecological network inference manifests itself in lower network dissimilarity scores for statistically significant inferred links with NETASSOC and HMSC at macroecological scales. The ability to account

for climatic data is thus of particular importance when inferring networks at scales for which species' performances and distributions vary meaningfully as a result of environmental conditions. Note that this scale may be much finer than the macroecological scale depending on the organisms involved.

In addition to including climatological patterns, there are calls to account for the shared evolutionary history of species pairs as well as their functional trait expressions (Ovaskainen et al. 2017; Blanchet et al. 2020). This information is likely to be of particular importance to infer interactions which are evolutionarily conserved (e.g. parasitism) or which can be reliably matched to easily-measurable functional traits (e.g. predation and body size traits). Meta-analyses of networks across geographic scales, regions, link types, and trophic levels have revealed that a combination of these information criteria is required to uncover mechanisms explaining network structure (Xing and Fayle 2021). However, for our study this information was unavailable at some geographic scales (Table 1). Our results clearly show that data sparsity can also affect network inference at the regional scale (Figure 3 and Figure 4), where NDD-RIM was the only capable method. This capability is likely due to the response-impact model component of the method, designed to infer rare/unobserved links. While complex methods like HMSC are seemingly the only ones fit for inferring ecological networks at macroecological scales (although demonstrating markedly dissimilar inference of present and absent links within scales compared to other approaches, Figure 4B) with respect to integration of additional confounding variables, this scale is also where selective data sparsity problem are the strongest, and inference may still be limited as the sampling of relevant information is costly at such scales.

Finally, to illustrate these criteria or our selection framework, consider two study settings necessitating inference of ecological networks: (1) a study of plant-plant relationships within a singular forest stand to evaluate how exotic plants have integrated the native community and (2) a study of plant-frugivory networks at several locations around the Globe to predict range shifts in response to global warming. For study (1), it is of pivotal relevance to select a method

of network inference which is capable of rendering directed links as competition between plants is rarely symmetrical (e.g., NDD-RIM, for example). Since the study system is small in extent, climatological gradients will likely play a minor role in altering occurrence, abundance or performance of forest-stand organisms over their biological interactions. Thus, choosing an approach without consideration of environmental conditions is likely adequate (e.g., COOCCUR, NDD-RIM, or HMSC and NETASSOC without environmental inputs). Finally, if abundance or performance data is available or can be sampled, it would be prudent to select an approach capable of using such data (e.g., NDD-RIM, NETASSOC or HMSC). Ultimately then, for this study purpose, one should select (from the pool of approaches demonstrated here) NDD-RIM for network inference as it satisfies all selection criteria and, most importantly, infers directed links which are required to express interaction effects from natives on exotic plants and vice versa. As for study (2), we are dealing with a broad scale of study at which consideration of climatological gradients will likely be highly relevant for accurate network inference and answering our research questions. Therefore, we may elect to use either HMSC or NETASSOC for inference in this setting. This aligns with the requirement of needing to infer mutualistic interactions which may be expressed as undirected links (although a case could be made that directed links would contain more information even in such a setting) which would also predispose a choice of HMSC or NETASSOC. The final choice between which of these two approaches to use will likely be informed by whether phylogenetic relatedness of species or their functional trait expressions are known which would subsequently give an advantage to HMSC because it can readily use such data. Ultimately, when arriving at an impasse in choosing between competing methods of network inference, one may wish to employ novel simulation-based network inference validation frameworks with study specific parameterisation to quantify network inference performance a priori (Kusch and Vinton 2023). Nevertheless, it is unlikely that any single network inference approach will render a fully accurate counterpart to real-world ecological interactions/associations. Instead, to maximise inference accuracy at the cost of a deeper understanding of factors structuring ecological networks, it may be prudent to combine different inference approaches in an ensemble

framework to establish more robust network inference while also accounting for the different ecological processes which disparate inference approaches are centred on.

5. CONCLUSION

Methods of network inference provide a valuable tool to gain a deeper understanding of ecological systems. They can suggest potential biological interactions and associations that may be worth exploring more completely depth and that would not have been considered otherwise (e.g., because of the size of communities that can include many potential and realised interactions). We show however, that scale inconsistencies are considerable for pairwise species links. In the absence of careful simulation-based assessment of network inference performance for specific study criteria, we argue that the choice of method needs to consider: (1) biological mechanisms which are identifiable using network inference at the scale of study, (2) alignment of link-interpretation with research questions, (3) data availability and information content at the geographic scale of study, and (4) applicability of the network inference method at the geographic scale of study.

REFERENCES

Alexander JM, Diez JM, Hart SP, Levine JM (2016) When Climate Reshuffles Competitors: A Call for Experimental Macroecology. Trends Ecol Evol 31:831–841. https://doi.org/10.1016/j.tree.2016.08.003

Allen CD, Macalady AK, Chenchouni H, et al (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. For Ecol Manage 259:660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Bartomeus I, Saavedra S, Rohr RP, Godoy O (2021) Experimental evidence of the importance of multitrophic structure for species persistence. Proceedings of the National Academy of Sciences 118:e2023872118. https://doi.org/10.1073/pnas.2023872118

Bello C, Galetti M, Montan D, et al (2017) Atlantic frugivory: a plant-frugivore interaction data set for the Atlantic Forest. Ecology 98:1729–1729. https://doi.org/10.1002/ecy.1818

Bimler MD, Mayfield MM, Martyn TE, Stouffer DB (2023) Estimating interaction strengths for diverse horizontal systems using performance data. Methods Ecol Evol 14:968–980. https://doi.org/10.1111/2041-210X.14068

Blanchet FG, Cazelles K, Gravel D (2020) Co-occurrence is not evidence of ecological interactions. Ecol Lett ele.13525. https://doi.org/10.1111/ele.13525

Blonder B, Morueta-Holme N (2017) netassoc: Inference of Species Associations from Co-Occurrence Data. 1

Blowes SA, Supp SR, Antão LH, et al (2019) The geography of biodiversity change in marine and terrestrial assemblages. Science (1979) 366:339–345. https://doi.org/10.1126/science.aaw1620

Burrill EA, Wilson AM, Turner JA, et al (2018) The Forest Inventory and Analysis Database: database description and user guide version 8.0 for Phase 2.

Carstensen DW, Sabatino M, Trøjelsgaard K, Morellato LPC (2014) Beta diversity of plant-pollinator networks and the spatial turnover of pairwise interactions. PLoS One 9:112903. https://doi.org/10.1371/journal.pone.0112903

Coutts SR, Salguero-Gómez R, Csergő AM, Buckley YM (2016) Extrapolating demography with climate, proximity and phylogeny: approach with caution. Ecol Lett 19:1429–1438. https://doi.org/10.1111/ele.12691

Damgaard C, Ehlers BK, Ransijn JCG, et al (2018) Insights on plant interaction between dominating species from patterns of plant association: expected covariance of pin-point cover measurements of two species. Environ Ecol Stat 25:221–235. https://doi.org/10.1007/S10651-018-0400-Z/FIGURES/6

Damgaard C, Weiner J (2021) The need for alternative plant species interaction models. Journal of Plant Ecology 14:771–780. https://doi.org/10.1093/jpe/rtab030

Davies SJ, Abiem I, Abu Salim K, et al (2021) ForestGEO: Understanding forest diversity and dynamics through a global observatory network. Biol Conserv 253:108907. https://doi.org/10.1016/j.biocon.2020.108907

Enquist BJ, Condit R, Peet RK, et al (2009) The Botanical Information and Ecology Network (BIEN): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity.

Freilich MA, Wieters E, Broitman BR, et al (2018) Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? Ecology 99:690–699. https://doi.org/10.1002/ecy.2142

Fricke EC, Ordonez A, Rogers HS, Svenning JC (2022) The effects of defaunation on plants' capacity to track climate change. Science (1979) 375:210–214. https://doi.org/10.1126/science.abk3510

Gilman SE, Urban MC, Tewksbury J, et al (2010) A framework for community interactions under climate change. Trends Ecol Evol 25:325–331. https://doi.org/10.1016/j.tree.2010.03.002

Gray A, Brandeis T, Shaw J, et al (2012) Forest Inventory and Analysis Database of the United States of America (FIA). Biodiversity and Ecology 4:225–231. https://doi.org/10.7809/b-e.00079

Griffith DM, Veech JA, Marsh CJ (2016) cooccur: Probabilistic species co-occurrence analysis in R. J Stat Softw 69:1–17. https://doi.org/10.18637/jss.v069.c02

Guimarães PR (2020) The Structure of Ecological Networks Across Levels of Organization. Annu Rev Ecol Evol Syst 51:433–460. https://doi.org/10.1146/annurev-ecolsys-012220-120819

Hackett TD, Sauve AMC, Davies N, et al (2019) Reshaping our understanding of species' roles in landscape-scale networks. Ecol Lett 22:1367–1377. https://doi.org/10.1111/ele.13292

Jin Y, Qian H (2019) V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. Ecography 42:1353–1359. https://doi.org/10.1111/ecog.04434

Johnson DS, Sinclair EH (2017) Modeling joint abundance of multiple species using Dirichlet process mixtures. Environmetrics 28:. https://doi.org/10.1002/env.2440

Kunstler G, Lavergne S, Courbaud B, et al (2012) Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: Implications for forest community assembly. Ecol Lett 15:831–840. https://doi.org/10.1111/j.1461-0248.2012.01803.x

Kusch E, Davy R (2022) KrigR—a tool for downloading and statistically downscaling climate reanalysis data. Environmental Research Letters 17:024005. https://doi.org/10.1088/1748-9326/ac48b3

Kusch E, Davy R, Seddon AWR (2022) Vegetation-memory effects and their association with vegetation resilience in global drylands. Journal of Ecology 110:1561–1574. https://doi.org/10.1111/1365-2745.13891

Kusch E, Ordonez A (2023) Ecological Network Resilience & Extinction Proxies - Updating Projections of Ecological Networks

Kusch E, Vinton AC (2023) A Novel Simulation Framework for Validation of Ecological Network Inference

Lamonica D, Pagel J, Schurr FM (2021) Predicting the dynamics of establishing tree populations: a framework for statistical inference and lessons for data collection. Methods Ecol Evol 2041–210X.13656. https://doi.org/10.1111/2041-210x.13656

Lavorel S, Garnier E (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisting the Holy Grail. Funct Ecol 16:545–556. https://doi.org/Doi 10.1046/J.1365-2435.2002.00664.X

Liu Y, Kumar M, Katul GG, Porporato A (2019) Reduced resilience as an early warning signal of forest mortality. Nat Clim Chang 9:880–885. https://doi.org/10.1038/s41558-019-0583-9

Lutz JA, Larson AJ, Swanson ME, Freund JA (2012) Ecological importance of large-diameter trees in a temperate mixed-conifer forest. PLoS One 7:. https://doi.org/10.1371/journal.pone.0036131

Maitner BS, Boyle B, Casler N, et al (2018) The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods Ecol Evol 9:373–379. https://doi.org/10.1111/2041-210X.12861

Momal R, Robin S, Ambroise C (2020) Tree-based inference of species interaction networks from abundance data. Methods Ecol Evol 11:621–632. https://doi.org/10.1111/2041-210X.13380

Morales-Castilla I, Matias MG, Gravel D, Araújo MB (2015) Inferring biotic interactions from proxies. Trends Ecol Evol 30:347–356. https://doi.org/10.1016/j.tree.2015.03.014

National Park Service (2009) Boundary for Yosemite National Park. In: Data Basin. https://databasin.org/datasets/7ae51cf6-aac8-4636-b1db-eb206ed012d0/

Ovaskainen O, Abrego N, Halme P, Dunson D (2016) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. Methods Ecol Evol 7:549–555. https://doi.org/10.1111/2041-210X.12501

Ovaskainen O, Tikhonov G, Norberg A, et al (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol Lett 20:561–576. https://doi.org/10.1111/ele.12757

Poisot T, Canard E, Mouillot D, et al (2012) The dissimilarity of species interaction networks. Ecol Lett 15:1353–1361

R Core Team (2021) R: A Language and Environment for Statistical Computing

Rist L, Moen J (2013) Sustainability in forest management and a new role for resilience thinking. For Ecol Manage 310:416–427. https://doi.org/10.1016/j.foreco.2013.08.033

Stephan P, Mora BB, Alexander JM (2021) Positive species interactions shape species' range limits. Oikos oik.08146. https://doi.org/10.1111/oik.08146

Swain A, Byrum T, Zhuang Z, et al (2021) Inferring, comparing and exploring ecological networks from time-series data through R packages constructnet, disgraph and dynet. 1–15

Tang SM, Gustafson EJ (1997) Perception of scale in forest management planning: Challenges and implications. Landsc Urban Plan 39:1–9. https://doi.org/10.1016/S0169-2046(97)00072-8

Tikhonov G, Opedal ØH, Abrego N, et al (2020) Joint species distribution modelling with the R-package Hmsc. Methods Ecol Evol 11:442–447. https://doi.org/10.1111/2041-210X.13345

Tikhonov G, Ovaskainen O, Oksanen J, et al (2021) hmsc: Hierarchical Model of Species Communities

Ulrich W, Gotelli NJ (2010) Null model analysis of species associations using abundance data. Ecology 91:3384–3397

Usinowicz J, Levine JM (2021) Climate-driven range shifts reduce persistence of competitors in a perennial plant community. Glob Chang Biol 1890–1903. https://doi.org/10.1111/gcb.15517

Veech JA (2013) A probabilistic model for analysing species co-occurrence. Global Ecology and Biogeography 22:252–260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

Verbesselt J, Umlauf N, Hirota M, et al (2016) Remotely sensed resilience of tropical forests. Nat Clim Chang 6:1028–1031. https://doi.org/10.1038/nclimate3108

Violle C, Enquist BJ, McGill BJ, et al (2012) The return of the variance: intraspecific variability in community ecology. Trends Ecol Evol 27:244–252. https://doi.org/10.1016/j.tree.2011.11.014

Wilkinson DP, Golding N, Guillera-Arroita G, et al (2019) A comparison of joint species distribution models for presence–absence data. Methods Ecol Evol 10:198–211. https://doi.org/10.1111/2041-210X.13106

Woodward G, Benstead JP, Beveridge OS, et al (2010) Ecological Networks in a Changing Climate

World Wildlife Fund (2020) Terrestrial Ecoregions of the World. http://omap.africanmarineatlas.org/BIOSPHERE/data/note_areas_sp/Ecoregions_Ecosystems/WWF_Ecoregions/WWFecoregions.htm. Accessed 25 Apr 2020

Wright IJ, Reich PB, Cornelissen JHC, et al (2005) Modulation of leaf economic traits and trait relationships by climate. Global Ecology and Biogeography 14:411–421. https://doi.org/10.1111/j.1466-822x.2005.00172.x

Wright IJ, Reich PB, Westoby M, et al (2004) The worldwide leaf economics spectrum. Nature 428:821–827. https://doi.org/10.1038/nature02403

Xing S, Fayle TM (2021) The rise of ecological network meta-analyses: Problems and prospects. Glob Ecol Conserv 30:e01805. https://doi.org/10.1016/j.gecco.2021.e01805