Integrating dispersal proxies in ecological and environmental research in the freshwater realm

Jani Heino¹, Janne Alahuhta², Terhi Ala-Hulkko², Harri Antikainen², Luis Mauricio Bini³, Núria Bonada⁴, Thibault Datry⁵, Tibor Erős⁶, Jan Hjort², Ossi Kotavaara², Adriano S. Melo³ and Janne Soininen⁷

¹Finnish Environment Institute, Natural Environment Centre, Biodiversity, Paavo Havaksen Tie 3, FI-90570 Oulu, Finland.

²University of Oulu, Geography Research Unit, P.O. Box 3000, FI-90014 Oulu, Finland.

³Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, 74001-970, GO, Brazil.

⁴Grup de Recerca Freshwater Ecology and Management (FEM), Departament d’Ecologia, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Diagonal 643, 08028-Barcelona, Catalonia, Spain.

⁵IRSTEA, UR-MALY, 5 rue de la Doua, BP 32108, 69616 VILLEURBANNE Cedex, France.

⁶Balaton Limnological Institute, MTA Centre for Ecological Research, Klebelsberg K. u. 3., H-8237 Tihany, Hungary.

⁷University of Helsinki, Department of Geosciences and Geography, P.O. Box 64, FI-00014 Helsinki, Finland.
Email: jani.heino@environment.fi
ABSTRACT

Dispersal is one of the key mechanisms affecting the distribution of individuals, populations and communities in nature. Despite advances in the study of single species, it has been notoriously difficult to account for dispersal in multispecies metacommunities, where it potentially has strong effects on community structure beyond those of local environmental conditions. Dispersal should thus be directly integrated in both basic and applied research by using proxies. Here, we review the use of proxies in the current metacommunity research, suggest new proxies and discuss how proxies could be used in community modelling, particularly in freshwater systems. We suggest that while traditional proxies may still be useful, proxies formerly utilized in transport geography may provide useful novel insights into the structuring of biological communities in freshwater systems. We also suggest that understanding the utility of such proxies for dispersal in metacommunities is highly important for many applied fields, such as freshwater bioassessment, conservation planning and recolonization research in the context of restoration ecology. This is because these research fields have often ignored spatial dynamics, and focused mostly on local environmental conditions and changes therein. Yet, the conclusions of these applied studies may change considerably if dispersal is taken into account.

Key words: accessibility, bioassessment, connectivity, conservation, dispersal, freshwater, links, metacommunity, nodes, transport geography.
Introduction

Ever since Charles Darwin, ecologists have been interested in dispersal (Ridley 2004), i.e., the movement of an organism from one location to another. Dispersal is one of the most important mechanisms affecting the distribution of individuals, populations and communities (Baguette et al. 2013; Lowe and McPeek 2014). At the same time, dispersal is also one of the most difficult phenomena to study even for a single individual or a single species in nature (Bilton et al. 2001; Nathan et al. 2008). The problem is exacerbated for dozens to hundreds of species in a metacommunity, i.e., a set of local communities connected by dispersal (Leibold et al. 2004), making it virtually impossible to account for dispersal directly for such large number of entities in natural settings. Ecologists have therefore relied on various proxies, which are assumed to relate to the effects of dispersal on community structure (Jacobson and Peres-Neto 2010; Jones et al. 2015).

Dispersal may mask the importance of purely environmental control of local ecological communities (Palmer et al. 1996; Leibold et al. 2004; Brown et al. 2011; Winegardner et al. 2012). This is because very high or very low dispersal rates may interfere with species sorting, decoupling the otherwise strong relationships between biological communities and local environmental factors (Leibold et al. 2004; Ng et al. 2009; Brown and Swan 2010; Winegardner et al. 2012). For instance, in mass effects, very high dispersal from ‘source’ populations may produce a constant flow of migrants that guarantees the maintenance of populations in unsuitable or ‘sink’ localities (Pulliam 1988), thus interfering with local environmental control (Mouquet and Loreau 2003). On the other hand, species
may be absent from suitable localities owing to dispersal limitation (Heino et al. 2015a), also
contributing to low variation explained by environmental factors in multivariate models.

Multivariate models of community structure can typically explain only a small fraction (adj. 
$R^2 < 50\%$, often varying between 0 and 20\%) of community variation (Beisner et al. 2006;
Nabout et al. 2009; Alahuhta and Heino 2013; Soininen 2014; Heino et al. 2015b), which
may simply be due to unmeasured environmental factors, but also to our inability to
adequately account for dispersal in statistical models (Cottenie 2005; Leibold and Loeuille
2015; Soininen, 2016). An alternative view suggests that statistical models may also
overestimate the spatial component potentially related to dispersal, which may be due to
specifics of the spatial methods used (Gilbert and Bennett 2010; Smith and Lundholm 2010).
Therefore, refining the spatial methods and various proxies for dispersal should aid in taking
dispersal better into account in metacommunity ecology.

Understanding the utility of proxies for dispersal is also highly relevant for many
applied fields when the focus is on multiple species in freshwater ecosystems. These
ecosystems are all of high priority for bioassessment, restoration and conservation because
they comprise high levels of biodiversity (Dudgeon et al. 2006; Wiens 2015) and provide
crucial ecosystem services to humans (Vörösmarty et al. 2010; García-Llorente at al. 2011;
Holland et al. 2011). At the same time, freshwater ecosystems are strongly threatened by
anthropogenic impacts such as eutrophication and habitat fragmentation (Dudgeon et al.
2006; Erős and Campbell Grant 2015). We emphasize that different types of freshwater
ecosystems (e.g. ponds, lakes, streams, rivers, springs) show different interactions among
dispersal, anthropogenic impacts and natural environmental factors. Owing to lower
connectivity, it may be that organisms in isolated freshwater ecosystems (e.g. ponds and
springs) are more severely impacted by the interactions of limited dispersal and
anthropogenic effects than those in more continuous ones (e.g. large rivers and large lake
similarity among dispersal, fragmentation and unexpected effects of stressors may occur in all freshwater, marine and terrestrial ecosystems. Therefore, the use of proxies for dispersal will be essential for applied research in all ecosystems. For example, our typical reasoning is that the success of restoration projects (e.g. recovery from acidification) may be delayed due to dispersal limitation because tolerant species may be absent from ecosystems simply because they have not been able to reach the site. Similarly, biomonitoring programs may be less effective in detecting impaired sites when dispersal from pristine to impacted sites is high.

Our aim is to review current use of proxies for dispersal in freshwater ecosystems. Individual sites in freshwater ecosystems are often inherently connected (Tonn and Magnuson 1982; Palmer et al. 1996; Magnuson et al. 1998; Jackson et al. 2001; Olden et al. 2001; Grant et al. 2007; Altermatt 2013). It can be assumed that most of the dispersal of obligate freshwater organisms, such as fish, is restricted to the network comprising running and standing waters (Matthews 1998; Olden et al. 2001). However, for other freshwater organisms, such as aquatic insects, dispersal within the network is not the only option, as insect adults may show active and passive out-of-network dispersal (Malmqvist 2002; Smith et al. 2009). Yet other groups of species, such as aquatic macrophytes, algae, mollusks and crustaceans, may disperse passively through waterways, or their seeds, whole cells, fragments or resting stages are carried by winds or animals for long distances (Kristiansen 1996; Bilton et al. 2001; Bohonak and Jenkins 2003; Riis and Sand-Jensen 2006).

Variation in dispersal mode and ability among groups of organisms is also exacerbated by the fact that even within a single group, dispersal distances vary greatly among species. Rather than being intimidated by such high degrees of variation, we propose that it actually provides a number of possibilities for basic and applied research. However,
better understanding of dispersal in diverse organisms inhabiting freshwater ecosystems is dependent on the better use of existing proxies and the development of new approaches. Here, we claim that while some traditional proxies are still useful, some proxies applied in transport geography are promising tools for basic and applied metacommunity research. Testing the utility of these proxies is, however, still in its infancy, and further case studies are needed. One of the aims of this review is to provide motivation for such further studies.

**Past, present and future proxies for dispersal**

*The distance effect: “…near things are more related than distant things”*

According to Tobler’s (1970) first law of geography, “Everything is related to everything else, but near things are more related than distant things”. Although this law is certainly accurate in geography and ecology (Nekola and White 1999; Hubbell 2001; Soininen et al. 2007), it has an inherent emphasis on Euclidean distances between sites. Nature and organisms are, however, more complex. What we define as “near” or “distant” should be understood in the context of ecological, but not necessarily geographical, distances between sites. Ecological distance takes into account structural (e.g. landscape features) and functional (e.g. animal movements) aspects as related to dispersal (McRae 2006; Sutherland et al. 2015). Hence, by necessity, those distances are much more complex than linear distances between sites (Wang et al. 2009; Graves et al. 2014). Also, organisms differ from each other in their dispersal ability (i.e. capacity to move long distances), although we can also state that all
organisms are different from other organisms, but phylogenetically closely-related organisms are, on average, more similar than distantly-related organisms. Organisms thus also have morphological (e.g. wing morphology in insects) and behavioural (e.g. tendency to fly long distances) characteristics related to dispersal (Hoffsten 2004; Rundle et al. 2007), which are typically phylogenetically conserved (Dijkstra et al. 2014). Below, we will consider pros and cons of organismal, physical and transport geography (i.e. graph-based) proxies for dispersal distances in a multi-species metacommunity context in freshwater systems (Table 1).

Organismal-based proxies

Typical organismal-based proxies for dispersal include separation of species into more homogeneous groups according to body size (Jenkins et al. 2007; De Bie et al. 2012; Datry et al. 2016a), wing size or wingspan (Hoffsten 2004; Sekar 2012), dispersal mode (active vs passive, aquatic vs aerial) and dispersal ability (Thompson and Townsend 2006; Göthe et al. 2013a, 2013b; Grönroos et al. 2013; Heino 2013b; Cañedo-Argüelles et al. 2015; Heino et al. 2015a). First, the use of body size divisions typically assumes that very small organisms are easily carried long distances passively by water currents, wind or by animals, and that increasing body size decreases the possibilities for passive long-distance dispersal (Fenchel and Finlay 2004; Shurin et al. 2009). While this idea is partly supported by empirical findings (De Bie et al. 2012; Padial et al. 2014; Datry et al. 2016a), some studies have also found little support for it (Jenkins et al. 2007). Second, unless the dispersal mode is taken into account, body size is likely to be a poor predictor of dispersal distances. It is likely that very small passively dispersing organisms, such as bacteria, microfungi and microalgae, are able to
disperse passively across very long distances (Baas-Becking 1934; Kristiansen 1996).

However, intermediate-sized and actively dispersing organisms, such as many aquatic insects (except perhaps dragonflies), may show rather limited dispersal distances (Finn et al. 2011).

Also, large-sized actively dispersing organisms, such as some diadromous fish or aquatic birds may disperse (or rather migrate) very long distances.

Another group of organismal-based proxies are provided by advances in molecular biology. These include population genetics (Hughes, 2007), DNA-barcoding (Cristescu 2014) and environmental DNA (Bohmann et al. 2014). However, as these advances have been reviewed recently (Manel et al. 2003; Manel and Holderegger 2013), we only mention briefly that they may also be used as proxies for dispersal (Bohonak 1999; Wilcock et al. 2001; Hughes et al. 2009). These methods have also some drawbacks, such as “detecting” a species when it is not actually present at a site in the environmental DNA approach (Bohmann et al. 2014). This is probably because the ‘signal’ of a species’ assumed presence may be carried long distances from occupied sites to other sites where they will result in false presences.

Population genetic approaches used to infer dispersal are manifold, and they have been available to researchers for decades (see reviews by Manel et al. 2003; Manel and Holderegger 2013). They include approaches that inform about past and/or current connections between local populations (Wilcock et al. 2001; Hughes et al. 2009). For example, phylogeography tries to understand the geographic distribution of the different genealogical lineages and can be used to infer past events (including long-term dispersal) by considering the spatial genetic variation of current populations (e.g. Teacher et al. 2009).

More generally, genetic variation across populations (i.e. genetic structure) has been traditionally used as an indirect measure of the current movement of individuals between populations based on molecular markers and statistical methods (e.g. F_{ST}). There have been
some attempts to relate the genetic structure to the dispersal ability of species, showing that
sets of populations exhibiting high genetic diversity are those with low dispersal ability
(Bohonak 1999). Genetic structure can be, however, a biased proxy of dispersal because it
not only informs about gene flow among populations, but also about mutation, genetic drift,
adaptation by natural selection along environmental gradients and colonization history (i.e.
founder effects). Different theoretical and empirical models are currently being used to detect
these different processes (Orsini et al. 2013). Among them, isolation-by-distance (IBD)
models are commonly used to explain spatial genetic variation by gene flow and gradual
genetic drift. In this case, genetic similarity is reduced when geographical distance between
sites increases (Relethford 2004). However, IBD models are neutral models (Orsini et al.
2013) that do not consider changes in the environmental conditions in space and assume that
populations are in gene-flow-drift equilibrium, which is probably not the case of most natural
populations. In addition, disentangling the relative effects of gene flow from genetic drift is a
challenging task. Most direct methods used to measure gene flow require direct estimates of
dispersal, whereas indirect methods, which do not require dispersal information, still consider
equilibrium conditions. Gene flow is supposed to be more advantageous than traditional
dispersal proxies (e.g. mark-recapture methods) because it integrates multiple generations,
indicates successful establishment in the target population (in contrast to mark-recapture that
only assesses if individuals reached the target site) and can be applied across extensive
geographical areas (Bohonak 1999; Baguette et al. 2013). However, even if unbiased gene
flow estimates are obtained, they may not fully represent dispersal because not all dispersers
survive and reproduce at a site (Bohonak and Jenkins 2003). Finally, recent advances based
on high throughput sequencing may lead to promising methods to measure dispersal at the
community level, as they may allow better quantification of genetic structure and its
underlying causes (e.g. Tesson and Edelaar 2013).
Graph-based proxies

Modelling is a prerequisite to examine the possible effects of using different dispersal proxies in ecological research (Rouquette et al. 2013; Weinstein et al. 2014). One of the most promising approaches is to examine the studied system as a graph, a set of nodes and links, in which nodes represent the elements of the system (e.g. habitat patches, individuals, populations or communities) and links specify the connectivity relationships between the elements (Calabrese and Fagan 2004; Urban et al. 2009). In graph-based analyses, spatially explicit data derived from geographic information systems (GIS) can be combined with information on the dispersal of organisms (Calabrese and Fagan 2004). Different distance classes among the nodes can be set up and depicted by adding different weights to the links as a proxy for indicating habitat suitability for the dispersing organisms (e.g. flow and riverbed characteristics for benthic insects) or barriers (e.g. dams or waterfalls for fish). Directed links can refine the graph model representing the importance of upstream vs downstream or watercourse vs overland dispersal (Galpern et al. 2011; Erős et al. 2012).

Potential connections between habitat patches (nodes) can be further refined by incorporating information on the dispersal ability of the focal species. For instance, if the distance between a given pair of patches is larger than a given threshold (here, dispersal distance for a species), the patches may be considered unconnected.

Overall, graphs are useful for quantifying the physical relationships among the landscape elements (i.e. structural connectivity; e.g. Saura and Rubio 2010) and how this topological structure affects the movement of organisms across the landscape (i.e. potential
functional connectivity; e.g. Vasas et al. 2009). Graphs can thus help understanding the role of dispersal in a diverse array of ecological systems in a flexible, iterative and exploratory manner with relatively little data requirements (Urban and Keitt 2001; Calabrese and Fagan 2004; Dale and Fortin 2010).

As explained above, the construction of a graph model requires the determination of links (connections) and their weights. In ecological research, many different conceptualizations of physical distance can be used for this purpose, such as Euclidean, network, flow and topographical distances (Olden et al. 2001; Beisner et al. 2006; Jacobson and Peres-Neto 2010; Landeiro et al. 2011; 2012; Maloney and Munguia 2011; Liu et al. 2013; Silva and Hernández 2015; Cañedo-Argüelles et al. 2015; Kärnä et al. 2015; Datry et al. 2016a). Euclidean distance is simply the shortest distance between two sites (Fig. 1). In contrast, network distance takes into account riverine or other ecological corridors and thus measures the shortest route from one site to another via corridors. However, according to Peterson, Theobald and Ver Hoef (2007), “the physical characteristics of streams, such as network configuration, connectivity, flow direction, and position within the network, demand more functional, process-based measures”. These authors made a useful distinction between symmetrical distance (i.e. Euclidean and watercourse distance) and asymmetric distance classes, which include upstream and downstream asymmetric flow distance (Peterson et al. 2007). This is because upstream dispersal is more difficult than downstream dispersal from one site to another, at least for obligatory aquatic organisms. Finally, topographical distance is built on the notion that altitudinal variation and slope may direct the dispersal of terrestrial organisms, whereby they may choose optimal routes by avoiding steep upward slopes (Fig. 1).
Besides the traditional measures of between-site physical distances, cost distance is an alternative family of distance metrics. Cost distance is calculated over a cost surface, representing the resistance to an organism's movement. It can be metaphorically called “as the fox runs” (Kärnä et al. 2015), as a wise animal like fox may choose a path of least resistance in the landscape. Cost distance can be measured either as a least-cost (optimal) path, or as a range of cumulative costs of landscape resistance between sites. Environmental variables used to produce cost surfaces typically include land use, human constructions and topography (Zeller et al. 2012). This technique has been mostly used to model the movement and dispersal of large land mammal species of conservation concern (Larkin et al. 2004; LaRue and Nilsen 2008), but it may also be relevant for the organisms living in freshwater ecosystems (Kärnä et al. 2015).

Previous studies using cost distances have mainly employed categorical variables and have not always taken into account variation in topography. In addition, various other physical structures can be used as costs. For example, the directional effect caused by prevailing wind or flow conditions could be incorporated as part of cost distances (Fig. 1). Additional cost can also consist of waterfalls, dams and other physical barriers for fish (Olden et al. 2001; Pelicice and Agostinho 2008; Filipe et al. 2013) or inhospitable routes through the matrix preventing or reducing dispersal, including pools, ponds and lakes for riffle-dwelling species (Erős and Campbell Grant 2015). The same applies for deforested riparian areas for terrestrial adults of freshwater species (Smith et al. 2009; Erős and Campbell Grant 2015).

Although cost distances, least-cost path modelling and other approaches related to graph-based modelling have been widely applied in ecology (e.g. Pinto and Keitt 2009), the studies to date have mostly considered one species at a time (see review by Sawyer et al.)
In the traditional transport geography, researchers have tried to explain complex human travel patterns by using spatial and spatio-temporal models (Black 2003). The modelling of human travel patterns relies, to a large extent, on the notion of accessibility (Table 2, Fig. 2). Accessibility can be defined as “the potential for reaching spatially distributed opportunities”, and its quantification typically includes the physical distance or cost of travel, as well as the quality and quantity of opportunities that humans want to reach (Páez et al. 2012). In the ecological context, the quality and quantity of opportunities might translate into habitat quality in terms of water chemistry (e.g. pH or nutrients) and quantity of resources (e.g. abundance of prey for predators). These qualities and quantities should be contrasted with the ease to access them, i.e., ecologically meaningful distances between source and destination localities in the landscape.
A number of measures have been devised for describing transport accessibility. These can be broadly divided into connectivity, accessibility of nearest object, cumulated opportunities, gravity and utility measures (Kwan 1998; Rietveld and Bruinsma 1998; Páez et al. 2012). Connectivity measures describe the number or rate of connections for a specific site, such as interconnectivity of a location to other locations within varying topology of a road network (Xie and Levinson 2007). Accessibility of nearest object is measured as least-cost path, for example, by applying street network travel distances to measuring the reach of service facilities (Smoyer-Tomic et al. 2006). Cumulated opportunities measure the number of opportunities (e.g. “available” sites for a species in ecological terms) reached within a certain travel cost, which can be applied to indicate amount of reachable services in an urban environment (Páez et al. 2012). While these measures mostly deal with the presence of a connection between any two sites or the distance separating them, the purpose of gravity measures is to express spatial interactions between sites. Drawing directly on the principles of the law of gravity in physics, gravity measures assume that the attraction of a site increases with size (or any other attribute) and declines with distance, travel time or cost. This is easily translated into dispersal of species between localities in a metacommunity, whereby some sites attract more individuals and species than others given the same dispersal distances, time or cost. Also, for example, potential of human social interaction can be estimated within urban and regional structures by applying daily time and travel constraints of people in relation to residential, work and other activities (Farber et al. 2013). In freshwater systems, this approach can include evaluation of species dispersal with different dispersal abilities within a metacommunity and can be incorporated into the gravity models. Utility measures are similar to gravity measures, but they are based on individual-related choices aiming to maximize utility in the selection of the destination (Geurs and van Wee 2004). This can be
seen as a kind of habitat selection by individual organisms (e.g., oviposition by female insects and nest-site selection by birds), which in turn affects local community structure.

While transport geography is an interesting source of proxies to be conflated with ecological approaches, there is some overlap in the graph-based proxies used in transport geography and metacommunity research. Such overlap is not always easy to detect since vocabulary is not fully consistent across disciplines. Nevertheless, although some of the proxies and terms have been used in metacommunity ecology before, transport geography provides explicit formulas for further ecological applications and defines complex issues in general terms.

There is one potential limitation with the use of physical and transport geography proxies: the lack of suitable landscape-level environmental data in some regions. However, our premise is that when environmental data are needed, they could be acquired from existing databases or using modern geospatial data compilation techniques. These include land use and land cover information using vast sets of airborne or spaceborne remote sensing sensors and topographic information (including delineation of stream networks) from high-resolution digital elevation models. Naturally, micro-scale explorations would require spatially more accurate data than available in most of the global data banks. However, similar remote sensing-based acquisition techniques (e.g., terrestrial hyperspectral and LiDAR imaging) could be applied in fine-scale investigations using the physical and transport geography proxies.
Use of different proxies for dispersal in the literature

In order to roughly estimate the frequency of usage of different proxies for dispersal, we conducted a literature search using the Web of Science database (from 2004 to August 26, 2016) and the terms (Dispers* AND metacommunity*), in the field TOPIC. These terms were combined, also in field TOPIC and using the Boolean operator “AND”, with keywords related to the different proxies evaluated in this review (Table 3). Thus far, terms related to organismal-based proxies were the most frequent, followed by physical distance-based proxies. However, we did not find articles using terms that would indicate the use of transport geography proxies in metacommunity ecology.

In studies using organismal-based proxies, a possible analytical approach consists in the creation of different matrices comprising taxa with different (yet typically inferred) dispersal abilities. These matrices may then be analyzed using variation partitioning methods (see examples below). The frequency of usage of spatial eigenfunction analysis and simple polynomials of geographic coordinates (i.e. distance-based proxies) was likely underestimated in our search. For example, Soininen (2014; 2016) found a total of 322 data sets, which were analyzed with variation partitioning methods (most of which were from lakes and streams). However, many data points in Soininen’s (2014; 2016) studies originated from one paper (Cottenie 2005), which was also counted as a single paper in our literature searches. We thus believe that our keyword analysis confidently reveals that use of more elaborated proxies for dispersal (considering, for instance, transport geography proxies) are less frequent than simple and possibly too simplistic proxies. In summary, our keyword analysis indicates the need for further comparative studies to better take dispersal into account in metacommunity studies.
There are many spatial statistical approaches to study species distributions and community structure that incorporate physical distance proxies, including Mantel test (Mantel 1967), eigenfunction spatial analysis (Borcard and Legendre 2002) and related methods (for a comprehensive review, see Legendre and Legendre 2012). For example, the flexibility and usefulness of eigenfunction spatial analysis and other similar methods in spatial modelling have been stressed elsewhere (Griffith and Peres-Neto 2006; Dray et al. 2006; Dray et al. 2012), and we briefly emphasize that they deserved their place in community ecologists’ toolbox. Eigenfunction spatial analyses allow one to use different types of distance (e.g. overland, watercourse and flow distance), geographic connectivity matrices and information about directional spatial processes (Blanchet et al. 2008; 2011; Landeiro et al. 2011; Göthe et al. 2013a; Grönroos et al. 2013) as inputs to compute eigenvectors (i.e. spatial predictors for univariate regression or multivariate constrained ordination analyses). This offers important flexibility to model complex spatial phenomena (Griffith and Peres-Neto 2006), such as variation of community structure (Dray et al. 2012). However, it has also been suggested that the explanatory variables derived from spatial eigenfunction analysis may overestimate spatial structure and the potential effects of dispersal on biological communities (Bennett and Gilbert 2010; Smith and Lundholm 2010). Also, spatial patterns in metacommunity structure may have emerged due to the effects of environmental variables, which are themselves spatially patterned and, more importantly considering the scope of this review, due to dispersal processes. In short, after controlling for the effects of environmental variables (e.g. using variance partitioning; see Peres-Neto et al. 2006; Legendre and Legendre 2012), the
spatial variables can be used to infer the relative role of dispersal processes. In studies of metacommunity structure, this inference is valid only if one assumes that no relevant environmental variables have been overlooked and that the effects of biotic interactions on the spatial patterns of community structure are negligible (Peres-Neto and Legendre 2010; Vellend et al. 2014).

Layeghifard et al. (2015) suggested weighting a spatial matrix (be it overland or not) by a dissimilarity matrix derived from a community data matrix. Accordingly, connectivity between a focal site and two other equally-distant sites will not be identical, but are dependent on biological dissimilarity. The more similar the focal site is to one of the sites, the higher is their assumed connectivity (Layeghifard et al. 2015). It is probably possible to modify these methods to accompany more complex relationships between sites in space. For instance, it could be possible to use the suite of distance classes referred to earlier in this review (Table 1). Also, if a gravity model of connectivity is hypothesized to represent dispersal, for instance, from headwaters to mainstreams and the latter accumulates more species, a suitable dissimilarity index may be one that measures species turnover only and not species richness differences (Lennon et al. 2001; Baselga 2010; Legendre 2014).

Combining organismal and physical distance proxies in the same modelling study

A few studies have considered simultaneously organismal and physical distance proxies. For example, Kärnä (2014) and Kärnä et al. (2015) studied a stream insect metacommunity in a subarctic drainage basin in Finland and examined how physical distance proxies affect different groups of insects defined by body size and dispersal mode. As physical distances,
they used (1) overland, (2) watercourse, (3) least-cost path (i.e. optimal routes between sites
in landscape) and (4) cumulative cost (i.e. cumulative landscape resistance between sites
along the optimal route) distances (Kärnä 2014; Kärnä et al. 2015). They calculated Mantel
correlations and partial Mantel correlations between Bray-Curtis biological community
dissimilarities and environmental distances or each of the four types of physical distances. In
these data, environmental and spatial distances were not strongly correlated, and the results of
partial Mantel test were hence very similar to the Mantel tests shown here (Fig. 3). Kärnä et
al. (2015) found that environmental distances between sites were most strongly correlated
with all biological dissimilarity matrices, as has been shown previously for stream
metacommunities (Heino et al. 2015b). However, different types of physical distances were
also often significant for different subsets of stream insect assemblages, even when
environmental effects were controlled for. A similar pattern has also been found in streams of
other climatic zones (Cañedo-Argüelles et al. 2015; Datry et al. 2016b). What is more
important is that the more complex least-cost path distances and cumulative cost distances
were either equally good or sometimes even outperformed the typically-used overland and
watercourse distances in accounting for variation in biological community dissimilarities
between sites, although this varied between different subsets of stream insect assemblages
(Kärnä et al. 2015).
Applications of proxies for dispersal

Applied research befitting from use of dispersal proxies

While the importance of dispersal is well appreciated in fundamental ecology, applied research has lagged behind in integrating dispersal effects on biological communities (Bengtsson 2010; Heino 2013a). For example, current bioassessment approaches infer effects of environmental changes using the responses of bioindicators to environmental factors (Hawkins et al. 2000a; Friberg et al. 2011). However, sole reliance on local environmental control (i.e. species sorting) may be misleading (Heino 2013a; Friberg 2014). In species sorting, adequate dispersal guarantees that all species are available at a locale to be filtered by local environmental factors (Leibold et al. 2004; Holyoak et al. 2005). However, high dispersal rates from unpolluted to polluted sites as in source-sink dynamics (Pulliam 1988) may decrease our ability to detect environmental change through the use of bioindicators. This is because some species indicative of pristine conditions may occur at the polluted site owing to high dispersal rates, even if that site is not favourable for them in the long term, thus masking the influence of anthropogenic changes on local biota. In contrast, owing to dispersal limitation, some pristine reference sites may also lack species that would otherwise occur there, thus affecting bioassessment results. Hence, we support the idea derived from simulation analyses (Siqueira et al. 2014) that potential dispersal effects should be directly integrated in aquatic bioassessment studies (Heino 2013a; Alahuhta and Aroviita 2016).

Restoration ecology is another field that might benefit from greater insights about dispersal. This is because restored sites may lack many species simply because potential
donor communities were all impacted by pollution or habitat degradation in a region, and colonization will thus be slow and initially composed mostly of dispersal-prone species (Bond and Lake 2003). Another possibility in this context relates to delayed recolonization of ecosystems that are recovering from anthropogenic stressors due to dispersal limitation (Blakely et al. 2006; Gray and Arnott 2011; 2012). Restoration ecology should thus take into account ecological corridors for dispersal, which might facilitate the recolonization of previously denuded or restored sites (Tonkin et al. 2014). The efficiency of ecological corridors is also dependent on dispersal ability and the spatial configuration of these corridors in the landscape (Joly et al. 2001). Hence, rather than restoring only local sites, restoration of connectivity is also a prerequisite for successful local restoration outcomes (see also McRae et al. 2012).

Conservation planning is a third field of applied research that should take dispersal directly into consideration. This is because dispersal within and between protected areas should be guaranteed (Jaeger et al. 2014; Barton et al. 2015a), and the network of protected areas should be planned such that they can act as stepping-stones to allow organisms to respond to environmental change (Fahrig and Merriam 1994; Margules and Pressey 2000; Lechner et al. 2015). However, conservation planning is also challenged by the vast numbers of species that should be monitored over broad metacommunities (e.g. Heino 2013a) and macrosystems levels (e.g. Heffernan et al. 2014), which is also exacerbated by the difficulties to measure dispersal over broad spatial scales. As a “science of crisis” (Soulé 1985), conservation biology cannot wait for the development and application of sophisticated, time-consuming and expensive methods to measure dispersal directly for hundreds to thousands of species and, at least in the short-term, the best we can do is to rely on proxies for dispersal.
The importance of integrating dispersal in predictive models of global change

Dispersal should be directly considered in predictive models in ecological research. This is because ecology has become increasingly predictive, most likely due to the need to forecast the effects of the ongoing global change (Evans et al. 2012; Petchey et al. 2015). Over the past decades, several models have been designed to predict how populations, communities or ecosystems will respond to ecological changes in time and space. Predictive models have been used to forecast distributions of species based on their climatic niches using Species Distribution Models (SDMs; Guisan and Zimmerman 2000; Chu et al. 2005) and, for example, to assess ecological status by comparing the observed community in a water body with the one expected under reference conditions (Hawkins et al. 2000a; Clarke et al. 2003). However, despite the wide use of both approaches, predictions can be biased if dispersal is not considered. Suitable habitats can be available for a species, but its real occurrence will ultimately depend on its ability to reach the site.

SDMs have been criticized because most of them only consider niche characteristics of species and neglect biotic interactions (Wisz et al. 2013), evolutionary changes (Thuiller et al. 2013) or dispersal processes. Several attempts have been made to incorporate dispersal into SDMs (e.g. Araújo et al. 2006). This is usually done by considering two extreme degrees of dispersal limitation (e.g. no dispersal vs unlimited dispersal) or intermediate situations using probabilistic methods when data on the dispersal abilities of the species are available (Barbet-Massin et al. 2012). Some modelling endeavours have also acknowledged the need to consider barriers to dispersal (e.g. dams) to improve model accuracy (Filipe et al. 2013). Information on current spatial connectivity across populations based on genetic approaches could also be used in SDMs to improve model accuracy (Duckett et al. 2013).
A possibility to construct models encompassing responses of multiple species at the same time include the River InVertebrate Prediction And Classification System (RIVPACS), first applied in riverine ecosystems (Wright et al. 2000; Clarke et al. 2003), but which can also be applied in other freshwater, marine and terrestrial ecosystems. There have been no empirical attempts to include dispersal in the practical applications of RIVPACS-type models, but simulations have shown the potential importance of dispersal for bioassessment (Siqueira et al. 2014). At best, some of these types of models consider spatial coordinates (i.e. latitude and longitude) as model predictors, but are usually based on assumptions about the niche characteristics of species (i.e. environmental filtering; Friberg et al. 2011). The importance of using dispersal proxies as predictor variables in bioassessment models is of particular significance in the context of metacommunities (Heino 2013a). This is because the spatial connectivity of sites and the dispersal abilities of the species may hinder the ability of models to detect an impact (Alahuhta and Aroviita 2016). This is especially relevant in less impacted and highly isolated sites (Siqueira et al. 2014). In addition, these sites (e.g. isolated headwater streams) usually host species with narrow ecological niches and distribution ranges, which can also have limited dispersal abilities (Finn et al. 2011). Incorporating organismal and physical distance proxies for dispersal in the metacommunity-level bioassessment could help to increase the accuracy of these models and thus the management of constituent freshwater ecosystems.
Questions for further freshwater research

The importance of dispersal proxies can be revealed by a number of questions that should be considered in basic and applied freshwater ecology. Although these ideas are somewhat speculative at present, they may provide useful roadmaps for further studies on dispersal proxies in bioassessment, restoration and conservation biology.

How important are stepping-stones for dispersal and how they can be recognized?

Ecological stepping-stones can be defined as sites or areas that help species to disperse from a site to other suitable sites across inhospitable landscapes. Stepping-stones can be expected to be very important for species dispersal (Saura et al. 2014; Barton et al. 2015a), but their recognition may be difficult. If we can recognize such sites in the landscapes by applying organismal and physical distance proxies in combination or based on transport geography measures, there are better possibilities to plan the conservation of metapopulations and metacommunities. For instance, we should be able to recognize sites having high accessibility for multiple species and subsequently plan a network of such sites across a broader landscape.

Graph-based modelling can also help if field-based measures fail to highlight the importance of stepping-stones for dispersal (Galpern et al. 2011). For example, network analyses can reveal how connectivity relationships change in the landscape if stepping-stones are deleted from the network of habitat patches. The importance of stepping-stones and other patches can be prioritized using different indices (e.g. Rayfield et al. 2011), which quantify
the importance of the focal habitat to maintaining connectivity between the patches (e.g. Pereira et al. 2011). Their more widespread application is warranted, especially for network-like stream systems, where habitat patches and their boundaries may be not so easily recognized (Erő and Campbell Grant 2015).

Are very low or very high dispersal rates affecting bioassessment?

Dispersal limitation may lead to a situation where not all species are available in reference sites (Pärtel et al. 2011; Cornell and Harrison 2014). A traditional approach has been to use a regional stratification to focus on smaller geographical areas, which could ensure that all species are able to reach all sites within a relatively small region (e.g. Hawkins et al. 2000b) and persist on them (e.g. Cornell and Harrison 2014). This should facilitate the detection of species sorting mechanisms and help define reference conditions. However, temporary local extinctions at suitable sites may not always be counterbalanced by immediate colonization if other suitable sites are located far away from the focal site even within a small region (Heino, 2013a) and/or if species have weak dispersal ability. In this case, we may classify sites in the wrong reference site group (or as impacted) if some species that should occur according to environmental conditions are absent from a site. It might be possible to adjust our predictive modelling efforts by using physical distance proxies (see Table 2), which might lead to a better prediction success. Alternatively, we could focus on a subset of good dispersers in our dataset, which should show minor effects of dispersal limitation, or focus on resident species (i.e. those species that do not show strong propensity for migration), which may show stronger associations with environmental gradients than entire assemblages (Bried et al. 2015).
The mass effects perspective in metacommunity ecology (Mouquet and Loreau 2003) suggests that high dispersal between localities may homogenize, at least to some degree, community structure in adjacent sites. On the other hand, some species may be absent from a site owing to not having been able to reach the site yet due to low dispersal rates or small source population size (Leibold et al. 2004). Either way, it may be difficult to assess if anthropogenic stressors have impacted a site, as extra species may be present or some expected species are missing (Siqueira et al. 2014). This limits our bioassessment by not detecting change correctly. Using information about the species composition of nearby sites might help us to decipher if either high or limited dispersal is affecting our bioassessment and restoration endeavours (Tonkin et al. 2014). These could be quantified by taking simultaneously into account a site’s accessibility and relative quality in the landscape, and how it attracts dispersers from the surrounding metacommunity. For instance, the measures from transport geography described above (e.g., gravity or utility measures, Table 2) could be used to show that the lower than expected biological differences between reference and impacted sites are due to their strong spatial connectivity and species exchange in terms of high dispersal.

Will species reach all potential future habitats in the face of global environmental changes?

Even though environmental conditions change, not all species may be able to track those changes (Heino et al. 2009; Poff et al. 2012). Poor dispersers or those with small source populations may not be able to disperse to suitable new habitats in other areas, at least if not assisted by humans. If such poorly-dispersing species can be identified based on their organismal traits, there are more possibilities for success (Bhowmik and Schäfer 2015). Also, if their actual dispersal routes can be approximated using physical distance proxies, the
success of the species for founding self-maintaining metapopulations and metacommunities may be better in the face of global change. For example, global change may lead to increase in temporal fragmentation of river networks, i.e., the degree of intermittency, which should affect the connectivity between stream sites (Datry et al. 2014). Improving our ability to predict changes in stream communities using distance-based proxies accounting for this fragmentation will improve our capacity to assess, estimate and mitigate the effects of global changes on intermittent streams (Datry et al. 2016c).

How can the dispersal of invasive species be predicted using proxies?

Knowing the dispersal ability of an invasive species (i.e. an organismal-based proxy) helps to predict its rate and potential to spread over large areas. Furthermore, knowing how landscape resistance (i.e. a physical distance proxy) may hinder its spread may have obvious benefits for predicting or preventing its dispersal. In this case, applications of the gravity or utility measures originated from the transport geography might also be useful, as the accessibility and attraction of sites for invasive species could be revealed using suitable proxy measures. Hypothetically, some widely recognized man-made structures that impair dispersal of native species such as dams (Winemiller et al. 2016) might, at the same time, boost the spread of invasive species (Havel et al. 2005).

How can we best detect and restore dispersal routes between near-pristine sites?

Local populations and communities at near-pristine or pristine sites need to be connected by gene and organism flows in order to remain viable (Fahrig 2003). Conservation and restoration efforts should also target the maintenance of the most efficient dispersal routes to
and from these pristine sites, although identifying these routes remains a challenge. Dispersal proxies could offer an efficient tool to identify these routes for all types of species, from poor to strong dispersers, and therefore provide insights to ecosystem managers for designing restoration and conservation projects (Tonkin et al. 2014; Cañedo-Arguelles et al. 2015; Kärnä et al. 2015; Datry et al. 2016a).

Can restoration measures fail due to lack of dispersers from neighbouring sites?

Restoration practices may not attain the planned objectives, or only attain them after long periods, if species are not able to colonize restored habitats in a strongly human-impacted landscape (Bond and Lake 2003; Tonkin et al. 2014; Barton et al. 2015b). Accordingly, restoration measures should be initially focused on sites connected to non-impacted source habitats or be planned to encompass entire landscapes or catchments that include some source localities (Bond and Lake 2003). Also, restoration practices should be coupled with the restoration of adjacent ecosystems to enhance suitable habitat corridors for dispersing species (Smith et al. 2009). Identifying such habitat corridors using the physical-based or transport geography proxies might be useful in this context.
Where to go from here?

Barton et al. (2015a) suggested that ecologists have made little effort to validate the use of proxies in ecology. For example, from a bioassessment perspective, the generally assumed conceptual model (e.g. environmental change → local community structure) suggests that a change in the environment (e.g. pollution) causes a change in local community structure (e.g. changes in species composition and relative abundances of species). However, dispersal disrupts this basic model and, to tease apart this effect, one needs a proxy for dispersal, which would function as a covariate (e.g. environmental change → local community structure ← proxy for dispersal). This covariate should, for instance, take mass effects or dispersal limitation into account. As shown in this essay, there are a number of ways to express the level of spatial relationships between sites and the best way may well be case-specific, depending on a study system, regional environmental conditions, between-site connectivity and characteristics of biotas. Thus, we propose that freshwater ecologists should evaluate and quantify the relationship between the biological dataset at hand and different proxies for dispersal (e.g. organismal-based dispersal traits, Euclidean, watercourse, least-cost path distances, and more). However, for the sake of generality, testing the predictability of different proxies in different regions, with different groups of organisms and in different points in time is also warranted (Barton et al. 2015a). In this context, a promising direction for future work would be to utilize the data from previous studies on bioassessment, restoration, conservation biology and community ecology, with the objective of quantifying the relative importance of different proxies for dispersal using a meta-analytical approach. A second objective would be, after knowing which proxy to use, how to integrate a proxy into
practical management of biodiversity. This is an open call for researchers interested in such
proxies for dispersal.

Conclusions

Dispersal proxies include traditional physical distances used in ecological research, such as
Euclidean distances, network distances, and various organismal-based proxies, such as body
size, dispersal mode and dispersal ability. More recent approaches include graph-based
methods, which show considerable promise for freshwater research. Future studies should
also consider applying methods developed in other disciplines, such as transport geography.
Application of these proxies should not be limited to fundamental ecological research, but
they should also be widely considered in applied fields, such as bioassessment, conservation
and restoration ecology. As dispersal is an essential element affecting species distributions, it
should be communicated easily to environmental managers and policy makers responsible for
practical conservation, management and assessment issues (Barton et al. 2015b). Hence,
while dispersal proxies should be efficient enough in capturing dispersal as a phenomenon,
they should also be simple enough to be useful in practical solutions. We propose that
organismal, physical and transport geography proxies for dispersal should be widely
considered as tools guiding environmental management and decision making.
Acknowledgments

The writing of this essay was supported by grants from the Academy of Finland to J. Heino, J. Hjort and J. Soininen. T. Datry and N. Bonada were supported by the French Foundation for Research on Biodiversity and the French National Agency for Water and Aquatic Environments in the context of the CESAB project "Intermittent River Biodiversity Analysis and Synthesis" (IRBAS: http://irbas.cesab.org/). A.S. Melo and L.M. Bini received research fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grants 309412/2014-5 and 304314/2014-5, respectively). T. Erős was supported by the OTKA K104279 grant.

References


Kärnä, O.-M. 2014. Effects of spatial location, environmental factors and species dispersal ability on invertebrate community structure in subarctic streams. MSc thesis, Department of Geography, University of Oulu. [In Finnish]


Saura, S., and Rubio, L. 2010. A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. Ecography 33: 523-537.


Table 1. Comparisons of the pros and cons of different dispersal proxies available to study metacommunities.

<table>
<thead>
<tr>
<th>Dispersal proxy</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organismal-based proxies</strong></td>
<td>More closely related to organisms’ traits and thus dispersal per se than physical distances between sites.</td>
<td>Often very coarse measures, as sufficient autecological information is available only for a few species or a few organismal groups.</td>
</tr>
<tr>
<td>1. Body size</td>
<td>Very easily obtainable for most organismal groups.</td>
<td>Although body size may be related to dispersal mode and capacity, it is also related to many, if not most, other organismal characteristics and functions.</td>
</tr>
<tr>
<td>2. Dispersal mode</td>
<td>Rather easily available information for comparisons of broad organismal groups.</td>
<td>Dispersal mode may not effectively relate to actual dispersal distances or dispersal rates between sites.</td>
</tr>
<tr>
<td>3. Dispersal ability</td>
<td>Potentially, informs greatly about dispersal distances of organisms between sites.</td>
<td>Difficult to obtain information for most organismal groups that cannot be easily tracked.</td>
</tr>
<tr>
<td>4. Population genetic structure</td>
<td>Are more direct measures than other organismal-based proxies, and may reveal complex dispersal routes between sites.</td>
<td>Genetic structure can be a biased proxy of dispersal because it not only informs about gene flow among populations, but also about mutation, genetic drift, adaptation by natural selection along environmental gradients and colonization history (i.e. founder effects). Hardly</td>
</tr>
</tbody>
</table>
feasible for a high number of species at the same time.

<table>
<thead>
<tr>
<th>Graph-based proxies</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Physical distance-based proxies</strong></td>
<td>Easily measurable from maps when available.</td>
<td>Are coarse proxies that may not always portray true dispersal routes for many species.</td>
</tr>
<tr>
<td>1. Euclidean distance</td>
<td>Very easily measurable as shortest linear distance between sites.</td>
<td>Not applicable for organisms, such as fish, relying exclusively on riverine corridors for dispersal.</td>
</tr>
<tr>
<td>2. Network distance</td>
<td>Distance between sites in a network may be useful if dispersal is restricted to such networks (e.g. riverine networks for obligatory aquatic organisms).</td>
<td>Some species may show more or less unexpected ‘out-of-network’ dispersal, which cannot be portrayed by network distances between sites.</td>
</tr>
<tr>
<td>3. Flow distance</td>
<td>May well model a) upstream vs downstream dispersal in riverine systems or b) headwind vs. tailwind dispersal in terrestrial systems.</td>
<td>It is not always known for how large a portion of species upstream/headwind dispersal is more costly than downstream/tailwind dispersal.</td>
</tr>
<tr>
<td>4. Topographical distance</td>
<td>May sometimes model well altitudinal features that may either prevent or facilitate dispersal. Rather easy to obtain from maps using geographic information systems (GIS).</td>
<td>Topographic features in a landscape may be important for terrestrial animals, but may be less important for those able to fly and cross higher landscape features.</td>
</tr>
<tr>
<td>5. Cost distances</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
May be used to model more complex landscape features than just topographic characteristics in a landscape. Potentially may be well used to model dispersal routes in heterogeneous landscapes. Sometimes lack of suitable maps may prevent calculating more complex cost distances between sites. Also, what, how and when to consider a landscape feature suitable or not suitable for dispersal may be difficult.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. Transport geography proxies</strong></td>
<td>Network-specific proxies which can be enhanced by route geometry, travel cost attributes, and pulling and pushing factors, when suitable data are available</td>
</tr>
<tr>
<td>1. Access to network</td>
<td>A simple, binary indicator.</td>
</tr>
<tr>
<td>2. Direct network connections or links</td>
<td>A comprehensible indicator expressing the presence of neighbouring localities which can be accessed without passing through other location.</td>
</tr>
<tr>
<td>3. Travel cost to (nearest) destination</td>
<td>A comprehensible indicator expressing the proximity to other locations.</td>
</tr>
<tr>
<td>4. Cumulated opportunities</td>
<td>Represents the quantity of accessible locations within a predefined network distance.</td>
</tr>
<tr>
<td></td>
<td>Needs topologically correct data and careful calibration of routing data or algorithm, when environment or population specific attributes are applied.</td>
</tr>
<tr>
<td></td>
<td>A highly coarse indicator, dependent on how network geometry and connectivity are defined and specified in the first place.</td>
</tr>
<tr>
<td></td>
<td>A coarse indicator which does not indicate the distances that need to be travelled.</td>
</tr>
<tr>
<td></td>
<td>Cannot consider the quality and quantity of accessed locations.</td>
</tr>
<tr>
<td></td>
<td>The indicator is strongly dependent on the threshold value, and does not take gradual distance decay into account.</td>
</tr>
<tr>
<td>5. Potential accessibility, gravity-based measures</td>
<td>Represents the quantity of accessible locations while taking into account the distance decay associated with travelling in the network, and the attraction of the location.</td>
</tr>
<tr>
<td>6. Population attraction and competition between destinations</td>
<td>Allows the determination of the probability for selecting a given destination while taking the distance decay associated with traversal in the network into account.</td>
</tr>
</tbody>
</table>
Table 2. Characteristics of transport geographic accessibility measures (for additional information, see Huff 1963; Kwan 1998; Rietveld and Bruinsma 1998; Páez et al. 2012) and their potential applicability as dispersal proxies in metacommunity ecology.

<table>
<thead>
<tr>
<th>Accessibility measure/index</th>
<th>Description</th>
<th>Formulae* for accessibility</th>
<th>Example case in transport geographic context</th>
<th>Examples of potential applications in metacommunity ecology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to network (A)</td>
<td>Access or connectivity exists or not</td>
<td>$c = \begin{cases} 0 &amp; \text{if not connected} \ 1 &amp; \text{if connected} \end{cases}$</td>
<td>To get value 1, city has to be connected to railway network.</td>
<td>Value 1 indicates that the ecological entity** of a locality is connected to the network.</td>
</tr>
<tr>
<td>Direct network connections or links (B)</td>
<td>Number of direct connections or links to other nodes in the network</td>
<td>$a = \sum_{j=1}^{n} c_{ij}$, $c = \begin{cases} 0 &amp; \text{if } c \text{ is indirect} \ 1 &amp; \text{if } c \text{ is direct} \end{cases}$</td>
<td>Amount of direct railway links that connect city to other cities.</td>
<td>Number of direct links connecting particular ecological entity** to other communities.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E.g. number of species’ direct connections to other populations in the dispersal network, which can, for example, consist of streams or terrestrial paths. Value 0 indicates isolated populations, having no direct connections.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E.g. headwater streams are linked simply to the downstream reach, whereas confluences are linked to three stream reaches (two upstream and one downstream reaches).</td>
</tr>
<tr>
<td>Travel cost to</td>
<td>Least cost path</td>
<td>Travel cost (e.g. time or distance) from the city to the nearest other city.</td>
<td>Travel cost (e.g. time or distance) for fish through riverine corridors from a lake to the nearest other lake.</td>
<td>Travel cost (e.g. time or distance) for a vertebrate through ecological corridors from one protected area to another.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(nearest)</td>
<td>to (most</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>destination (C)</td>
<td>accessible)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>object</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Cumulated opportunities (D)   | Number of       | Number of other cities within certain travel cost.                           | Number of localities within certain travel cost for actively or passively dispersing aquatic, semi-aquatic or terrestrial organisms. | Species opportunities to reach other populations (or communities or metacommunities) through dispersal network depending on species dispersal abilities. |
|-------------------------------| objects within   |                                                                                |                                                                                                |                                                                                                |
|                               | defined travel   |                                                                                |                                                                                                |                                                                                                |
|                               | cost threshold   |                                                                                |                                                                                                |                                                                                                |
|                               | $a = \sum_{j=1}^{n} A_j \times d_{ij},$ |                                                                                     |                                                                                                |                                                                                                |
|                               | $d = \begin{cases} 0 & \text{if } d \geq \text{cost threshold} \\ 1 & \text{if } d < \text{cost threshold} \end{cases}$ |                                                                                     |                                                                                                |                                                                                                |

Cost-distance attributes and thresholds may be specified in relation to the characteristics of the ecological entity**
Potential accessibility, gravity based measures (E)

<table>
<thead>
<tr>
<th>Potential accessibility, gravity based measures (E)</th>
<th>Potential for interaction with other cities in relation to distance, attraction attributes and interests to move.</th>
<th>An insect female’s potential to reach suitable habitats in relation to travel cost to other populations within its lifespan. Here, lifespan can be understood as a species’ ability or interest to move in relation to travel cost that can vary during a season (term β in formula).</th>
</tr>
</thead>
<tbody>
<tr>
<td>High and/or close opportunities provide better potential for interaction in comparison to low and/or distant opportunities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Potential for interaction with other cities in relation to distance, attraction attributes and interests to move.

\[ a = \sum_{j=1}^{n} A_{ij} \times e^{-\beta d_{ij}} \]

Population attraction and competition between destinations (F)

<table>
<thead>
<tr>
<th>Population attraction and competition between destinations (F)</th>
<th>Amount of interaction with a specific city in relation to other cities, by taking distance, attraction attribute and interests to move into account.</th>
<th>Amount of interaction among habitats with variable environmental quality for female insect or migratory bird individuals from a certain population in relation to travel cost during its lifespan. Here, lifespan can be understood as a species’ ability or interest to move in relation to travel cost that can vary during a season (term β in formula).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability for selecting an attraction amongst all attractions in the space in competitive situation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ P_j = \frac{\sum_{j=1}^{n} A_{ij}^{a} d_{ij}^{-\beta}}{\sum_{j=1}^{n} A_{ij}^{a} d_{ij}^{-\beta}} \]

* Explanation of terms used in formulations: \( a \) is accessibility related for each origin, \( c \) is connecting link between origin and destination nodes, \( d \) is travel cost (e.g. distance, time or other measurable friction) between origin and destination nodes, \( n \) is number of destination nodes, \( A_j \) is
attribute wanted to be accessed in destination(s) (e.g. quantified habitat attraction), $i$ refers to (number of) origin and $j$ to destination and $\beta$ is parameter for interest to move in relation to travel cost.

** May be an organism, a species, a group of species (i.e. a community), a specific habitat or a biome.
Table 3. Number of articles \( (n) \) retrieved according to the Web of Science database (from 01/01/2004 to 26/08/2016) using different combinations of keywords related to the use of dispersal proxies in metacommunity studies.

<table>
<thead>
<tr>
<th>Proxies</th>
<th>keywords</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organismal-based proxies</td>
<td>&quot;Body size*&quot; AND Dispers* AND metacommunit*</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>&quot;Dispersal mode*&quot; AND Dispers* AND metacommunit*</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>&quot;Dispersal capacit*&quot; OR &quot;Dispersal abilit*&quot; AND Dispers* AND metacommunit*</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>genetic* AND Dispers* AND metacommunit*</td>
<td>45</td>
</tr>
<tr>
<td>Physical distance-based proxies</td>
<td>&quot;euclid* distance*&quot; AND Dispers* AND metacommunit*</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>&quot;network* distance*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;watercourse distance*&quot; AND Dispers* AND metacommunit*</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>&quot;flow distance*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;Topographic* distance*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;cost distance*&quot; AND Dispers* AND metacommunit*</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mantel AND Dispers* AND metacommunit*</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>&quot;Spatial eigenfunction*&quot; AND Dispers* AND metacommunit*</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>&quot;Moran* Eigenvector*&quot; AND Dispers* AND metacommunit*</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>&quot;principal coordinates of neighbor matrices&quot; AND Dispers* AND metacommunit*</td>
<td>1</td>
</tr>
<tr>
<td>Transport geography proxies</td>
<td>&quot;Access to network*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;Direct network* connection*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;Travel* cost*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;Cumulat* opportunit*&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&quot;Potential accessibility&quot; AND Dispers* AND metacommunit*</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure captions

Fig. 1. A schematic figure of potential dispersal routes for species in dendritic systems (light blue colour) among three sites (red dots). A describes Euclidean (orange), overland (green) and watercourse (blue) distances; B describes cost distance as related to topography (brown) and stream flow resistance (blue); C describes two species (light green vs dark green) which have different optimal dispersal routes between sites in relation to the cost imposed by land cover or land use; and D describes two optimal dispersal routes for a species in response to the dominant wind direction.

Fig. 2. A schematic figure of transport geographic accessibility measures (Huff 1963; Kwan 1998; Rietveld and Bruinsma 1998; Páez et al. 2012) and their potential applicability as ecological dispersal proxies. The letters (A-F) correspond to the description of the measures of accessibility in Table 2.

Fig. 3. An example of different physical and organismal dispersal proxies in stream insect research (figures redrawn based on Kärnä, 2014; Kärnä et al. 2015). Mantel correlations between Bray-Curtis biological community dissimilarities and environmental distances (based on various local environmental variables) or each of the four types of physical distances are shown. Separate analyses were run for all species, different body size classes and dispersal modes (active or passive). Asterisk indicates a significant correlation. In these data, environmental and physical distances were not strongly correlated, and partial Mantel test were hence very similar to these Mantel tests shown here. See text for further information.
Fig. 1.
Fig. 2.
Fig. 3.

The figure shows a graph of Mantel correlation (r) against the species included in various categories. The categories include:

- All
- 0-0.25 cm
- 0.25-0.5 cm
- 0.50-1 cm
- 1-2 cm
- 2-4 cm
- Active
- Passive

The graph compares the environments of Overland and Watercourse with Optimal and Cumulative. The significance of the correlations is indicated by asterisks. For example, the Mantel correlation for All species shows a statistically significant relationship, as indicated by the asterisk. The same applies to the other categories, with some showing significant correlations and others not.