Testing the River Continuum Concept with geostatistical stream-network models

Stefano Larsen\textsuperscript{a,b}\textsuperscript{*}; Maria Cristina Bruno\textsuperscript{a}; Ian P. Vaughan\textsuperscript{c}; Guido Zolezzi\textsuperscript{a}

\textsuperscript{a} Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

\textsuperscript{b} Department of Sustainable Agro-ecosystems and Bioresources, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy

\textsuperscript{c} Cardiff School of Biosciences. Cardiff University, Wales, UK

\textsuperscript{*}Corresponding author. email: larsen.stefano@gmail.com
Abstract
The River Continuum Concept (RCC) provided one of the first unifying frameworks in fluvial ecosystem theory. While the RCC predictions held in many empirical tests, other research highlighted how the model overlooked sources of heterogeneity at different scales e.g. the effects of tributaries. Disentangling these effects requires an assessment of variation in key ecosystem variables over the longitudinal and lateral dimension of river networks. However, so far, no empirical tests have employed a spatially explicit statistical approach to this assessment. Here, we show how recently-developed spatially-explicit models for river networks can be used to test predictions of the RCC whilst taking into account cross-scale sources of heterogeneity. We used macroinvertebrate data from 195 monitoring sites from 1st to 4th order streams spread across the Adige River network (NE Italy). We compared theoretical expectations with empirical semivariograms that incorporated network topology to assess the continuity and patchiness in the proportion of invertebrates functional feeding groups (FFG) over Euclidean and in-stream distances. Geostatistical stream-network models were then used to quantify the influence of the longitudinal gradient relative to local-scale water quality and land-use drivers, while accounting for network spatial autocorrelation.

Patterns in the semivariograms based on flow-connected relationships were characterised by a nested structure associated with heterogeneity at multiple scales. Therefore, the longitudinal variation in FFG was better described by a patchy discontinuum rather than a gradient, implying that both in-stream processes and landscape factors influenced stream ecosystem function. The overall shift in FFG along the longitudinal profile was generally consistent with the RCC predictions, although the best models often included water quality and local land-use predictors. Stream-network models further indicated that up to 90% of residual variation (mean=50%) was accounted for by spatial autocorrelation, especially among flow-connected communities. Accounting for such autocorrelation not only improved model performance relative to non-spatial approaches, but indicated that most flow-connected communities were spatially correlated to some extent. This has clear implications for the assessment of the RCC tenets. This is the first test of the river continuum model that explicitly accounted for stream network topology and autocorrelation. Results indicated that in the Adige River, macroinvertebrates feeding groups exhibited heterogeneity along the longitudinal gradient, which appeared punctuated by local habitat transitions. Such transitions could be associated with artificial impoundments that alter the natural continuity of river processes, and we advocate the use of spatially explicit network models to test the RCC in more natural contexts.

Keywords: Functional Feeding Groups; Macroinvertebrates; Longitudinal gradient; Semivariograms; Autocorrelation; Adige River
Introduction

The distribution and diversity of aquatic organisms in river networks is predominantly influenced by the downstream direction of water flow and by the physical changes occurring along the longitudinal gradient (Townsend, 1996; Ward, 1989). This concept is at the heart of early conceptual models aimed at idealizing the structure and function of communities along river systems such as the river zonation (Hawkes, 1975; Illies, 1961) and the River Continuum concepts (Vannote et al., 1980). In particular, the River Continuum Concept (RCC) provides a useful conceptualisation of river networks as open ecosystems characterised by a continuum of physical changes and associated ecological responses, in which the type and availability of organic matter, the structure of invertebrate communities and the partitioning of resources shift gradually along the longitudinal gradient. Although the RCC is a simplification that overlooks the patchy nature of river systems associated with local geology, tributary effects and lateral floodplain inputs (e.g. Thorp et al., 2006), it remains one of the most influential concepts in river science (4,949 citations as of Match 2019; ISI Web of Science database). One of the key strengths of the RCC is that it proposes testable hypotheses regarding changes in stream metabolism (P/R), the main trophic basis of production (carbon sources) and the consequent adjustment of consumer communities (functional feeding groups; Cummins, 2016).

Empirical tests of the RCC provided support for its predictions, especially in temperate North American rivers (Curfis et al., 2018; Hawkins and Sedell, 1981; Minshall et al., 1985, 1983; Rosi-Marshall and Wallace, 2002; Webster, 2007), and more recently in different biomes and climatic zones (Greathouse and Pringle, 2006; Jiang et al., 2011; Tomanova et al., 2007). Studies that have criticised the RCC generally emphasise the local heterogeneity of river systems (Perry and Schaeffer, 1987; Statzner and Higler, 1985; Townsend, 1989). For instance, Pool (2002) argued that local factors such as reach geomorphology or bedrock geology could override longitudinal gradients, so that stream communities in a given segment may be just as similar to communities far up or downstream as they are to those in neighbouring stretches. The issue of quantifying the relative contributions of ‘global’ river gradients and local heterogeneity is currently acknowledged (e.g. Thorp, 2014), and may in part stem from the methodological challenges of describing patterns and testing alternative hypotheses in dendritic networks. Standard statistical methods are unable to handle the complexities resulting from linear river reaches arranged into complex branching networks and the influence of directional water movement (Peterson et al., 2013). Spatial autocorrelations are, in fact, particularly complex in river systems as their intensity varies with the connectivity and directionality within the network (Isaak et al., 2014). In this case, models based on Euclidean distances, for instance, might be insufficient to represent the unique spatial relationships found in river systems (Peterson et al., 2013). These considerations are particularly relevant for the RCC where one fundamental aspect of the continuum is that ecological processes in downstream reaches are linked to those occurring upstream (e.g. Minshall et al., 1985). Moreover, critics to the continuum model emphasised how river systems display heterogeneity at multiple spatial scales besides the longitudinal dimension (Perry and Schaeffer, 1987; Poole, 2002). It is therefore surprising that none of the previous empirical tests of the RCC model employed any spatially explicit approach that could either account for autocorrelation or utilise the spatial variance as part of the study.
Fortunately, recent advances in the field of geospatial statistics adapted to dendritic networks provide the tools to quantify the main scales of spatial variation within river networks and allow for more rigorous tests of hypotheses such as the RCC (Peterson and Hoef, 2010; Ver Hoef and Peterson, 2010). Two developments in particular are valuable for assessing RCC. The first is the generalisation of the standard geostatistical tool, the semivariogram, to river networks (called Torgegrams; Peterson et al., 2013). Variograms quantify spatial structure and can reveal the dominant scales of environmental processes (Cressie, 1993). Specifically, semivariograms depict the autocorrelation of a given variable calculating the semivariance between pairs of observations for a range of watercourse distance lags (h) as:

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} [z(s_i) - z(s_i + h)]^2
\]

where \( N(h) \) is the number of observation pairs separated by distance lag \( h \), \( z(s_i) \) is the value of the variable in location \( s_i \), and \( z(s_i + h) \) is the value at distance \( h \) from \( s_i \).

The shapes of the semivariograms can be compared with theoretical expectations reflecting hypothesised spatial structures and dependency (Fig. 1).

In river networks, semivariograms can be calculated based on three spatial distances among sampling locations: flow-connected (watercourse distance between locations connected by water flow), flow-unconnected (watercourse distance between any locations in the network) and Euclidean. Specifically, semivariograms of flow-connected relationship describe the effects of hydrologic transport and upstream dependence and can thus indicate whether the longitudinal gradient represents the dominant scale of variability in the distribution of carbon sources and consumers (a key tenet of the RCC). Conversely, patterns from the semivariograms based on flow-unconnected and Euclidean relationships can inform on the influence of adjacent tributaries and wider landscape properties independent of network position, respectively (McGuire et al., 2014). Thus, when used deductively, empirical spatial patterns can help formulate hypotheses regarding the main processes influencing the distribution of the variables of interest (McIntire and Fajardo, 2009).

The second major development is stream-network models that extend conventional linear models to account for the branching structure of the network and the directionality of water flow, as well as the 2-D terrestrial matrix in which the network is embedded. Therefore, they can simultaneously model both along-channel and across catchment, Euclidean patterns of autocorrelation, thus accounting for different sources of spatial autocorrelation (Ver Hoef and Peterson, 2010). Notably, equivalent models that assume different types of spatial correlation (e.g. flow-connected, flow-unconnected and Euclidean, the latter ignoring the network structure) can be compared to assess alternative hypotheses regarding spatial dependency.

In this study, we first used semivariograms based on Euclidean and stream-network distances (Torgegrams) to visualise patterns and scales of variability in the distribution of macroinvertebrate functional feeding groups (FFG) along the longitudinal river gradients.
(consistent with RCC), and across the whole catchment in a large Alpine river network. We focussed on invertebrate FFG because their variation over the continuum represents a central prediction of the RCC, linking basal carbon sources to consumer communities. Moreover, macroinvertebrate data are routinely collected for bio-monitoring purposes and their feeding habits are well known (Schmidt-Kloiber and Hering, 2015). Subsequently, we used geostatistical stream-network models to quantify the importance of the longitudinal gradient relative to local scale habitat variables, while accounting for spatial autocorrelation within the network. Whilst previous studies have considered the shift of invertebrate FFG along the river gradient (Greathouse and Pringle, 2006; Grubaugh et al., 1996; Jiang et al., 2011; Minshall et al., 1985), this is the first time a geostatistical approach has been employed that specifically accounts for spatial autocorrelation in dendritic networks.

The RCC was originally based on forested temperate high-relief basins in North America. While the Adige River network conforms to these aspects, its longitudinal continuity is altered by numerous hydropower dams distributed over the basin (Chiogna et al., 2016a; Larsen et al., 2019), which are expected to alter the natural continuity of river processes (Hoenighaus et al., 2007; Humphries et al., 2014). Therefore, the present study should not be considered as a formal test of the validity of the RCC model. Rather, we propose a statistically robust approach to test its predictions to a case study that well represents many Alpine river catchments across Europe and North America.
Figure 1 – Hypothetical distribution of a variable mapped along a river network (e.g. proportion of one feeding group in the community) and the associated semivariograms. In A), there is no spatial structure in the variance at the sampled scale. The intercept of the semivariance function (called ‘nugget’) represents the variance due to sampling error or variation at scales finer than the shortest separation distance. In B), spatial dependency at large-scale reflects a single dominant gradient of variation from upstream to downstream. The values where the semivariance function reaches a plateau (called ‘range’), indicates the distance at which values are considered independent from each other. In C), small-scale heterogeneity reflects patchiness and discontinuity where, for instance, factors influencing the variable operate at fine scales, and the range is reached at shorter distances. In D), nested heterogeneity reflects a combination of small-scale patchiness embedded in a larger-scale gradient (with multiple inflection points). In this case, patterns are influenced by factors operating at multiple scales. Figure is re-drawn after McGuire et al. 2014.
Study area and dataset

The Adige River (Fig. 2) is the third largest river basin in Italy, covering more than 12,000 km². Most of the Adige River drains the Alpine region with elevation reaching 3400 m a.s.l. Climate is typically Alpine with dry winters, snow and glacier melt in spring and rather humid summers and autumns (Lutz et al., 2016). Since the beginning of last century, more than 30 dams have been built across the whole network (i.e. dams are distributed over 1st to 4th order streams), mostly for hydropower generation. These dams have altered the natural flow regimes of many reaches (Larsen et al., 2019; Zolezzi et al., 2009), and likely disrupted the natural continuum of sediment and organic matter transport.

Macroinvertebrate data were collected as part of the institutional monitoring programmes of the Environmental Protection Agencies of the Provinces of Trento and Bolzano. Sampling occurred between 2009 and 2014 in 195 reaches from 1st to 4th order streams (between 130 and 1980 m a.s.l.). Sites were sampled multiple times in different seasons. In any given year, between 91 and 161 sites (median = 113) were included. The number of samples per site ranged between 2 and 12 (mean =4.8), but most sites (80%) were sampled 3-9 times, mostly in spring and autumns. Macroinvertebrate densities were averaged to represent the typical community composition of a reach and remove seasonal effects. Sampling followed the multi-habitat approach where 10-replicate Surber samples were distributed over a 20-50 m reach in proportion to the different microhabitat types present (Hering et al., 2004). Macroinvertebrates were identified to genus and family level (Appendix 1).
Methods

Ecological trait and local environmental data

Information about invertebrate feeding traits was gathered from the online database on the ecology of freshwater organisms (www.freshwaterecology.info; Schmidt-Kloiber and Hering, 2015). To test the RCC predictions, we considered the following feeding groups: grazers, shredders, gatherers, filterers and predators. A fuzzy approach was used to assign each taxon an affinity score for each feeding group, thus avoiding restricting taxa to a specific feeding group and effectively taking into account intra-specific variability (Chevenet et al., 1994). Affinities at family and genus levels were obtained by averaging scores over the species known to occur in

Figure 2 - Map of the Adige river network in NE Italy (map inset) with the 195 study sites. Color reflects the sites’ score on the first component of a principal component analysis synthesising the longitudinal gradient, from dark (low order, high altitudes) to light (higher order, lower altitudes).
the region. Affinity scores were then standardised between 0 and 1 and then weighted by each
taxon’s relative abundance (using the ‘functcomp’ command within the FD package in R) to
calculate the community-wide proportion of FFG for each site (Schmera et al., 2014).

To quantify the position along the longitudinal river continuum for each of the 195 study reaches,
a synthetic variable was created using principal component analysis (PCA) of Strahler stream
order, distance from ‘mouth’ (i.e. most downstream reach), altitude and upstream catchment
area. The first principal component (‘Longitudinal PC1’) explained c. 60% of the variation and
was negatively correlated with distance from ‘mouth’ and altitude, and positively correlated with
Strahler order and catchment area (Tab. 1). This PC reflected the longitudinal position and
allowed us to score each sample location over the network according to a continuous longitudinal
gradient that accounted for multiple aspects (Jiang et al., 2011; Tomanova et al., 2007; Vaughan
et al., 2013).

Table 1. Loadings of the variables on the first Principal Component (‘Longitudinal PC1’)
describing the position of the reaches over the longitudinal continuum

<table>
<thead>
<tr>
<th>Variable</th>
<th>Loading on the 1st Principal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance from mouth</td>
<td>-0.66</td>
</tr>
<tr>
<td>Altitude</td>
<td>-0.81</td>
</tr>
<tr>
<td>Upstream catchment area</td>
<td>0.80</td>
</tr>
<tr>
<td>Strahler stream order</td>
<td>0.72</td>
</tr>
</tbody>
</table>

ArcMap 10.5 and the STARS toolset (Peterson and Hoef, 2014) were used to calculate distance
matrices (flow-connected, flow-unconnected and Euclidean), upstream catchment areas and the
spatial weights needed in the network-models (see below). Shapefiles of the river network and
catchment topography were obtained by the Environmental Protection Agencies of the Provinces
of Trento and Bolzano.

Three reach-scale environmental variables were also included in the analyses. The first two were
the proportions of agricultural and forest land cover within a 1-km buffer around each sampling
location as proxies for allochthonous input and shading, i.e., among the main contributing factors
to the longitudinal gradients in the RCC. The third variable was the stream water physico-
chemical status, expressed by the WFD (Water Framework Directive, EU 2000/60) LIMeco index
(Livello di Inquinamento dai Macrodescrittori per lo stato ecologico), which is the official water-
quality indicator in Italy. This is a multimetric index that scores water quality based on threshold
levels for dissolved oxygen, ammonia and nitrate concentrations and total phosphorus (see
Azzellino et al., 2015).

Semivariograms
Empirical variograms and Torgegrams were compared with theoretical expectations (Fig. 1) and used to examine the dominant spatial scale of variation in the proportion of feeding groups along and across the river network. For instance, a continuous single scale of variation in the proportion of a feeding group along the river continuum would produce variograms like those in Fig. 1B. Conversely, a complete dominance of local-scale drivers would produce variograms like Fig. 1C, while a patchy continuum would generate variograms such as Fig. 1D. Using Euclidean and in-stream distances, the continuity and patchiness of the spatial patterns can be separately examined over the 2D and longitudinal dimensions of river networks. Torgegrams were also used as preliminary exploratory analysis to visually examine the patterns in semivariance between flow-connected and flow-unconnected site pairs in order to guide the selection of autocovariance function in the subsequent stream-network modelling (Ver Hoef et al., 2014).

**Stream network models**

In a subsequent analysis, we used stream-network models (Ver Hoef et al., 2014; Ver Hoef and Peterson, 2010) to quantify the importance of the longitudinal continuum (expressed as the Longitudinal PC1) relative to local-scale environmental factors (i.e. water quality and land-use) as predictors of FFG proportions. All the spatial data necessary to analyse stream-network models were generated in ArcMap 10.5 using the STARS toolset. We accounted for the complex autocorrelation structure of dendritic networks using Euclidean as well as in-stream flow-connected and flow-unconnected autocovariance functions. Stream-network models are variance components models that take the general form:

\[ y = X\beta + z_{TU} + z_{TD} + z_E + \varepsilon \]

where \( y \) is the response variable vector (here: logit transformed proportion of feeding groups), \( X \) is the matrix of covariates (here: longitudinal PC1, LIMeco, land-use), \( z_{TU} + z_{TD} + z_E \) are vectors of zero-mean random variables with autocorrelation structure based on tail-up, tail-down and Euclidean functions, and \( \varepsilon \) is the vector of random independent errors. The tail-up and tail-down autocovariance structures are moving-average functions that quantify autocorrelation among flow-connected and flow-unconnected locations, respectively (Isaak et al., 2014; Peterson et al., 2013). The autocovariance functions can take different forms, including linear-with-sill, spherical, Mariah and exponential models (Garreta et al., 2009), but spatial models are generally robust against their mis-specification (Garreta et al., 2009; Isaak et al., 2014). Specifically, the tail-up function permits correlation exclusively between sites that are flow-connected (sites not connected are assumed independent) and uses a weighting approach to up- and down-weight samples that occur upstream of a given location. Here, we used upstream catchment area as surrogate of discharge for the weighting procedure. In this case, the moving-average autocorrelation function is split at confluences so that locations on larger streams have a stronger influence on downstream communities than locations in smaller streams (Peterson et al.,
The tail-down function allows correlation among both flow-connected and unconnected samples and therefore a spatial weighting measure is not necessary (Isaak et al., 2014). Finally, the Euclidean functions is based on Euclidean distances as in the traditional spatial statistics methods. Therefore, stream-network models are flexible tools that can incorporate multiple information into a single model (Ver Hoef and Peterson, 2010). Moreover, by allowing errors to be differently autocorrelated over the longitudinal and lateral network dimension they can indirectly account for the effects of unmeasured variables that have a spatial pattern (e.g. soil, underlying bedrock geology).

The values of FFG in the communities were logit-transformed as recommended for proportional data (Warton and Hui, 2011) and parameter estimation was based on maximum-likelihood. Covariate selection was then based on Akaike Information Criterion (AIC; Burnham and Anderson, 2002), but to describe overall model performance we also report root-mean-square prediction errors (RMSPE), which specifically focus on model predictive power. Overall model development was a step-wise process. We first included all predictors (i.e. longitudinal PC1, LIMeco, and local land-use) and the full mixture of autocovariance functions, which included an exponential tail-up, tail-down and Euclidean models (Ver Hoef et al., 2014). Then we used a manual step-wise approach to remove non-significant predictors from the model. We then refined the spatial component comparing (or removing) different functions for the tail-up, tail-down and Euclidean autocovariance structure to select the final model with the lowest AIC. The spatial components were investigated after the selection of covariates since the model accounts for spatial autocorrelation in the errors after the effects of the covariates have been removed. Therefore, patterns of spatial dependency are data and model specific and can change if the covariates change (e.g. Frieden et al., 2014). The efficacy of the selected spatial model relative to a non-spatial model (ignoring any spatial autocorrelation) was also evaluated (Isaak et al., 2014). Spatial stream-network models were run in R (R Core Team, 2017) using the SSN package (Ver Hoef et al., 2014). For each site, raw data about FFG proportions, taxon richness, Longitudinal PC1, and geographic coordinates are given as Supplementary Material.

Results

Empirical semivariograms

The semivariograms for Euclidean and flow-unconnected relationships (Fig. 3) were consistent with the presence of a single dominant spatial structure, with variance progressively increasing with distance for all FFGs. These patterns resemble Fig.1B and suggest the presence of a catchment-wide gradient. Conversely, when accounting for flow-connections (limiting the modelled spatial autocorrelation to occur only among sites connected by water flow), nested spatial structures emerged that are associated with heterogeneity at multiple scales (i.e. multiple inflection points at different distances, resembling Fig. 1D). The flow-connected semivariogram is
the most relevant to the RCC, and shows that the spatial distribution of FFG does not vary as a continuum along the longitudinal gradient, but is highly heterogeneous.

The semivariogram for taxonomic richness was mostly indicative of a single scale of variation, especially across the lateral dimension of the stream network (Euclidean and flow-unconnected relationships). Patterns from the flow-connected relationships exhibited some heterogeneity, which was less marked than that characterising FFG, and with an inflection point evident at larger distances.

**Figure 3** – Empirical semivariograms of the proportion of feeding traits and taxon richness based on Euclidean, flow-connected and flow-unconnected distances. Note the change in the y-axis. The number of observation pairs from which semivariances are calculated differs among distance lags, with fewer pairs contributing to measures at larger distances. The semivariance values are expressed as x1000, except for taxon richness.
Stream-network models

Stream network models were in broad agreement with the RCC predictions (Tab. 2; Fig.4), with an increase in grazers and gatherers along the longitudinal gradient, and a decrease in the shredders. Filterers were not related to the longitudinal gradient while, contrary to the original RCC predictions, predators declined. Taxon richness was unrelated to the longitudinal profile. However, the relative importance of local-scale drivers and the influence of spatial autocorrelation differed substantially among FFG (Tab. 2).

Figure 4. Plot of the proportion of feeding traits and taxonomic richness along the river longitudinal dimension as expressed by the first PC (Adige river network; 195 reaches). A linear regression line is shown when the relationship was significant. Note: the regression lines are corrected for spatial autocorrelation employing the full mixture of autocovariance functions.
Table 2 – Variance components for the most supported non-spatial (in grey) and spatial models for FFG and taxon richness. The sign after the covariates indicates the direction of the effect. The spatial autocovariance functions from stream-network models and their estimated range are also shown. RMSPE = root mean squared prediction error; AIC = Akaike Information Criterion

<table>
<thead>
<tr>
<th>FFG</th>
<th>Covariates</th>
<th>Spatial autocovariance function (range)</th>
<th>Proportion of variance</th>
<th>RMSPE</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grazers</td>
<td>Longitudinal PC1 (+)</td>
<td>None</td>
<td>0.08</td>
<td>0.31</td>
<td>99.05</td>
</tr>
<tr>
<td></td>
<td>Longitudinal PC1 (+)</td>
<td>Mariah tail-up (62 km)</td>
<td>0.05</td>
<td>0.28</td>
<td>89.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential Euclidean (91 km)</td>
<td>0.77</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Shredders</td>
<td>Limeco (+) Agric. land-use (-)</td>
<td>None</td>
<td>0.45</td>
<td>0.54</td>
<td>272.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear-with-sill tail-up (101 km)</td>
<td>0.42</td>
<td>0.51</td>
<td>272.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential Euclidean (1.5 km)</td>
<td>0.17</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Gatherers</td>
<td>Longitudinal PC1 (+) LIMeco (-) Agric. land-use (+)</td>
<td>None</td>
<td>0.24</td>
<td>0.29</td>
<td>67.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spherical tail-up (8.4 km)</td>
<td>0.22</td>
<td>0.29</td>
<td>70.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filterers</td>
<td>LIMeco (-) Wood. land-use</td>
<td>None</td>
<td>0.037</td>
<td>0.11</td>
<td>514</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential tail-down (16 km)</td>
<td>0.003</td>
<td>0.99</td>
<td>501</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential Euclidean (77 km)</td>
<td>0.29</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>
Grazers
Grazers were primarily represented by Ephemeroptera (*Baetis, Ecdyonurus*) and Chironomidae. Longitudinal PC1 explained 8% of the variance in the proportion of grazers in the non-spatial model (i.e. ignoring any spatial autocorrelation), which declined to 5% when autocorrelation was accounted for. There was no evidence for an effect of land-use or water quality parameters. Overall, the model with the lowest AIC and RMSPE was the spatial model with a tail-up and Euclidean autocovariance functions. Most of the residual variation (77%) was captured by the tail-up autocovariance function at a relatively large scale (estimated range of c.60 km). This indicates that grazer communities in flow-connected sites exhibited spatial autocorrelation that accounted for most of their variation along the network.

Shredders
The proportion of shredders (mostly represented by Limnephilidae, *Leuctra* and Gammaridae) declined along the longitudinal gradient, which alone explained 17% of their variation in the non-spatial model and about 16% in the spatial-model. However, according to the AIC, the best models describing shredders variation did not include the longitudinal gradient, but only the LiMeco index and the proportion of agricultural land-use for 1-km buffer around the site. The non-spatial and the spatial models were equally supported according to the AIC (delta AIC =0.2), although the spatial model had smaller prediction errors.
the spatial model, most of the residual variation was accounted for by a small-scale Euclidean autocovariance function (estimated range = 1.5 km).
These results indicated that, overall, the variation in the proportion of shredders over the network was mostly influenced by local scale factors rather than by the longitudinal gradient.

Gatherers
The proportion of gatherers (represented by Ephemeroptera and Plecoptera such as *Serratella* and *Amphinemura* among others) increased along the longitudinal river gradient, which alone explained 19 and 18% of their variance in the non-spatial and spatial model, respectively. According to the AIC, the most supported model was the non-spatial model that included the longitudinal gradient as well as the LIMeco and the agricultural land-use. The best spatial model required only a small-scale tail-up function (estimated range = 8 km), which accounted for about 17% of the residual variation. Overall, the proportion of gatherers appeared determined by both local scale factors and the longitudinal position along the network with relatively fine-scale autocorrelation among flow-connected communities.

Filterers
Filterers (mostly Hydropsychidae and Simuliidae) did not show any significant relationship with the longitudinal gradient in either non-spatial or spatial models. The most supported model was a spatial model with only the LIMeco index as covariate and a tail-down (representing flow-unconnected relationships) and Euclidean autocovariance functions, which accounted for a similar proportion of residual variation. The best non-spatial model included the LIMeco index and the proportion of woodland in 1-km buffer. Both models however showed high residual errors (~1) indicating that the proportion of filter feeders was influenced by other factors besides the one considered here. Contrary to what observed in other feeding groups, autocorrelation among flow-connected filterer communities was not included in the best spatial models. Rather, autocorrelation occurred mostly across the lateral dimension of the network as modelled by the tail-down and Euclidean functions (relationship between locations not upstream-downstream of each other).

Predators
Predators primarily included Plecoptera (*Isoperla, Perlodes*) and Trichoptera Rhyacophilidae. In contrast to the RCC prediction, the proportion of predators declined along the longitudinal gradient, which alone explained 29 and 27% of variation in the non-spatial and spatial model respectively. The non-spatial model was the most supported and included the longitudinal gradient and the LIMeco index as covariates, jointly accounting for 41% of predators' variation. The spatial model had, however, the lowest prediction errors and included the same covariates with a large-scale tail-up (range = 100km) and Euclidean autocovariance functions that accounted for 40% of residual variation.
Taxon richness was not related to the longitudinal river gradient. The best non-spatial and spatial models selected partially different covariates. The most supported model included a tail-up and Euclidean autocovariance functions that jointly explained c.40% of the residual variation, while the selected covariates only explained 8%. The estimated range for the flow-connected autocovariance model was much longer than the total length of the river network (1000 km), indicating that measurements of invertebrate richness in all flow-connected communities were correlated to some extent.

Discussion

The River Continuum Concept is one of the most influential theoretical frameworks in river ecology, idealising river network as open ecosystems in which the physical template and the associated ecological processes change predictably along the longitudinal continuum. The RCC immediately stirred a lively debate that stimulated empirical tests as well as conceptual revisions (e.g. Minshall et al., 1985; Statzner and Higler, 1985). Critics to the RCC argued that the model overlooked the importance of lateral floodplain inputs, tributary effects, and fine-scale heterogeneity, as well as human impacts (Perry and Schaeffer, 1987; Poole, 2002). As such, ecological processes and functions along the river network were better described by punctuated discontinuity rather than by a continuum.

Testing the key tenets of the RCC and its caveats rely upon detecting spatial patterns and discriminating between patterns generated by different processes. So far, however, empirical tests of the RCC - either supporting or confuting its predictions – have not used spatially explicit statistical approaches. Ignoring spatial autocorrelation (i.e. non-independence) among field measurements can produce bias in parameter estimates and increase the chance of Type I errors (Legendre, 1993). In addition, patterns of spatial autocorrelation, which are often perceived as data nuisance, can instead be used to appraise the dominant scale of variation in a given variable (Dray et al., 2012). This has clear implications for the assessment of the RCC model, or any spatial patterns in river networks.

Here we combined the analyses of semivariograms with geostatistical stream-network models, to i) assess the continuity and heterogeneity in the proportion of invertebrate feeding groups over the longitudinal and lateral dimensions of the network and ii) quantify the relative importance of the longitudinal gradient vs local drivers, while specifically accounting for the spatial autocorrelation inherent to dendritic networks.

Semivariograms

The semivariograms indicated that in the Adige river network variation in feeding groups along the longitudinal continuum (flow-connected relationships) was characterised by nested spatial
structures with multiple inflection points. This supports the hypothesis that downstream variation in carbon sources and associated consumers were better represented by a patchy discontinuum rather than by a gradient (Rice et al., 2001). This implies that both in-stream factors and local-scale drivers influenced invertebrates’ structure and function along the longitudinal gradient. Part of the observed discontinuity in FFG variation along the Adige river is likely associated with the presence of artificial impoundments. Hydropower dams are expected to alter the relative availability of different carbon sources in complex ways, by changing not only flow patterns, but also temperature regime and water chemistry (Zolezzi et al., 2011, 2009). For instance, consumers in reaches below dams often derive their carbon sources from local production and riparian input rather than upstream transport, thus creating a gap in the downstream transition (Hoenighaus et al., 2007; Poole, 2002). Assessing this, however, would require more detailed local-scale sampling up- and downstream of impoundments.

Conversely, patterns of variation over the whole catchment, ignoring the effects of flow direction, (flow-unconnected and Euclidean relationships), were consistent with the presence of large-scale spatial structure, with little or no fine-scale heterogeneity (cf. Fig. 1B). These semivariograms describe relationships among communities associated with wider landscape properties (e.g. gradient in underlying catchment geology; Chiogna et al., 2016b) and showed that similarity in the proportion of feeding groups generally decreased with spatial separation.

The shape of the semivariograms for taxonomic richness indicated a rather homogeneous distribution, especially over Euclidean and flow-unconnected distances and with large estimated ranges. Indeed, while composition is expected to differ with increasing spatial separation (Soininen et al., 2007), the richness of communities can be similar over large distances and across a range of conditions (Bonada et al., 2012; Larsen et al., 2018). Moreover, the variance from flow-connected relationships was always lower than that based on lateral relationships, indicating that connected communities were generally more similar to each other in terms of taxon richness.

It is important to note, however, that the shape of the variograms is influenced by the overall sampling design and especially by the minimum and maximum distance between samples as well as the dispersal capacity of the organisms involved (Ettema and Wardle, 2002). Therefore, variograms across different studies and/or organisms should be compared with caution.

**Stream-network models**

When significant, the effect of longitudinal gradient explained between 5% and 29% of variation. However, the best models did not always include the longitudinal gradient as covariate, as in the case of shredders and taxon richness. It is well known that many factors across a range of scales influence macroinvertebrate richness and function (e.g. Karaus et al., 2013; Richards et al., 1997), with effect that can be independent from the longitudinal dimension. This was evident here in the inclusion of local land-use and water quality in most of the supported models. Moreover, stream-network models indicated that FFG exhibited predictable spatial patterns that could not be accounted for by local variables with substantial autocorrelation especially among flow-connected communities. Taken together the autocovariance functions explained between 17% and 94% of residual variation with a mean of 50%. That is, half of the variation in the proportion of feeding groups was, on average, accounted for by spatial autocorrelation either
along the stream route or across the network. Accounting for this autocorrelation always improved model performance as measured by the prediction errors (RMSE), as often observed in other studies using stream-network models (Isaak et al., 2014). Autocorrelation among flow-connected locations is the most relevant to the RCC because it represents the relationship along the downstream continuum. Tail-up functions alone explained between 14% and 77% of variation, with an estimated range that varied greatly, from 6 km, to 100 km and up to 1000 km in the case of taxon richness. This means that most flow-connected communities exhibited some degree of spatial correlation. Therefore, the most relevant continuum in the Adige river system appears to be the spatial correlation that exists among flow-connected locations.

The importance of spatial autocorrelation over the lateral dimension was well captured by the Euclidean autocovariance function. This explained up to 39% of variation in feeding groups with effects estimated at both fine (1.6 km range) and large-scale (150 km range). This again indicates that landscape features and local drivers other than the position along the continuum contributed to the observed variation in feeding groups.

Understanding the processes underpinning the observed autocorrelation is, however, not straightforward. In fact, autocorrelation can stem from autogenous factors, such as biotic competition, dispersal and intraspecific aggregation as well as exogenous factors related to spatially structured environment variables. This means that, while our models were able to quantify the spatial dependency in the residual errors, the underlying cause is likely a complex interaction of endogenous and exogenous processes (Frieden et al., 2014).

Specific FFG patterns

If the original RCC predictions applied to our case study, we expected shredders to decrease from headwater to larger streams as the input of terrestrial coarse organic matter declined. Conversely, grazers were expected to increase downstream reflecting changes in aquatic primary production. Although grazers were originally predicted to show a hump-shaped relation with stream size (Vannote et al. 1980), with maximum representation in mid order streams, our sampling was truncated at stream order 4, so that a continuous increase was expected (e.g. Greathouse and Pringle, 2006). Gatherers that rely on fine organic matter should increase along the river continuum reflecting a progressive reduction in detritus particle size. However, in the original RCC, gatherers were grouped with filterers and no clear predictions were given. In their revision of the RCC, Minshall et al. (1985) predicted a relatively constant representation of gatherers and filterers from first to fourth order streams. The proportion of predators was originally expected to remain constant along the continuum (Vannote et al., 1980). However, in our study system we expected predatory invertebrates to decline along the gradient because they were generally dominated by Plecoptera species, which are restricted to cold headwater habitats. Finally, the RCC predicted taxonomic richness to peak in mid-order streams. However, as in the case of grazers, richness should increase along the longitudinal gradient in the Adige River, since communities were sampled between 1st and 4th order reaches.
Although the shapes of the semivariograms provided in our analyses indicate a rather patchy distribution of FFG along the downstream gradient, and the stream-network models highlighted the importance of local-scale drivers, the overall trend in the proportion of grazers, shredders, filterers and, to some extent, gatherers followed the original RCC predictions. Grazers were generally well represented in the benthic communities with an estimated increase of c.20% along the downstream gradient, likely reflecting a parallel increase in aquatic primary production. Surprisingly, however, neither local land-use nor water quality parameters were included as covariates in the best model explaining grazers variation. Rather, most of the residual variation was associated with autocorrelation along the stream route, with a marginal contribution from the longitudinal gradient.

Shredders constituted a minor group in the communities representing less than 5% in the downstream reaches. Although shredders declined by almost 50% along the longitudinal gradient, their variation was better explained by local factors associated with agricultural land-use and water quality. In addition, a large proportion of their residual variation was explained by small-scale autocorrelation over Euclidean distances, further highlighting the influence of local landscape factors. Shredders rely on allochthonous coarse organic matter, i.e., the production and downstream transport of plant litter originating from riparian or catchment vegetation. This explains the decreasing representation of shredders in agricultural-dominated reaches.

The proportion of gatherers increased almost 50% along the longitudinal gradient and was also influenced by water quality and local land-use. Their distribution in the Adige river network likely reflected the downstream increase in fine organic matter, which could also be of anthropogenic origin considering the positive influence of local agricultural land-use. We cannot conclude whether the distribution of gatherers conformed to the RCC model, as no clear predictions were originally given. However, other studies reported a general increase in their proportion with stream size (Greathouse and Pringle, 2006; Jiang et al., 2011). Interestingly, gatherers were the only group for which the non-spatial and spatial models performed equally well in terms of prediction errors. Autocorrelation was in fact only detected at fine-scales and contributed relatively little compared to what estimated for other feeding groups.

As expected, the proportion of predators declined along the longitudinal gradient. This likely reflect the fact that, in the Adige system, most predatory invertebrates belonged to the Plecoptera order, which are restricted to cold and well oxygenated headwater habitats. This is further supported by the positive association of predators with water quality as expressed by the LIMeco index.

Finally, it is important to note that our quantification of FFG proportion was based on the relative densities of the taxa (as. e.g. Hawkins and Sedell, 1981), whereas the original RCC predictions were based on biomass. Estimates of functional composition based on these two measures can differ, especially when taxa with large body-size are included (e.g. crayfish; Lugthart and Wallace, 1992), although this was not the case in the present work. However, information on macroinvertebrate biomass was not available at present.
Conclusion

To our knowledge, this is the first study to use semivariograms to assess the variation in functional feeding groups over multiple spatial dimensions of a river network, and to quantify the importance of the longitudinal gradient while explicitly accounting for spatial autocorrelation. Although we used the RCC to guide our hypotheses, the principal aim of the study was not to formally evaluate the validity of this model, but to provide a novel and robust statistical approach to test its main tenets. This approach was applied to a case-study where continuity was expected to be partially interrupted by hydropower dams distributed over the network. The Adige River system, however, is representative of the majority of Alpine river basins in Europe and elsewhere where the natural hydro-morphological connectivity has long been affected (Maiolini and Bruno, 2007; Marnezy, 2008). Our results revealed a rather heterogeneous distribution of feeding groups along the longitudinal dimension (flow-connected), in support of a patchy discontinuous view. Similarly, stream-network models highlighted how the downstream gradient and local-scale drivers jointly influenced the functional composition of benthic invertebrates. This is expected in river networks where both local-scale factors and large-scale network topography combine to regulate biodiversity patterns (Altermatt et al., 2013; Carrara et al., 2014). Overall, the trends in FFG were mostly in agreement with the RCC model, implying that the key theme of the model hold: changes in stream size determine changes in basal resources and consumers. However, most of the variation in FFG was actually attributed to spatial autocorrelation, with effects that were stronger than those attributed to the covariates and evident at both fine and broad scales. Ignoring such spatial dependence could have implications far beyond the evaluation of river ecosystem models. Benthic macroinvertebrates are routinely sampled and used in bio-assessment programmes and are included in the EU Water Framework Directive as one of the biological element used to evaluate the ecological status of running waters. When spatially explicit issues are investigated, such as those related to land-use, water pollution and habitat change, ignoring autocorrelation could produce bias in parameter estimates and ultimately weaken statistical inference. Nonetheless, field studies specifically accounting for spatial autocorrelation in benthic invertebrates are still rare, but the few available data indicate that spatial dependence can be very strong, especially along the stream route (Bonada et al., 2012; Frieden et al., 2014; Lloyd et al., 2006).

The fact that we observed heterogeneous variations of FFG over the longitudinal dimension, which were accompanied by general trends that followed the original RCC predictions, supports a more holistic and contemporary view of river ecosystems where both patch- and continuum-based processes simultaneously regulate in-stream metabolism and biodiversity (Collins et al., 2018; Humphries et al., 2014). While the Adige river network might not be the most appropriate system for testing the RCC model due to extensive human modification, this is a problem inherent to many RCC tests (e.g. Collins et al., 2018; Greathouse and Pringle, 2006), including the use of relatively impaired systems in North America for the development of the model itself (Statzner and Higler, 1985). We advocate the use of spatially explicit approaches such as the one used here for future evaluations of river ecosystem models in more pristine catchments.
Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 748969, awarded to SL. The authors want to thank the Environmental Agency of the Autonomous Province of Trento (APPA-TN) and the Environmental Agency of the Autonomous Province of Bolzano (APPA-BZ) for providing the macroinvertebrates and LIMeco datasets, and Elisa Stella for the shapefiles used for the GIS-based analyses. Two anonymous reviewers provided valuable suggestions that improved the manuscript.

References


