Biol Invasions
DOI 10.1007/s10530-017-1492-3

ORIGINAL PAPER

Geostatistical distribution modelling of two invasive crayfish across dendritic stream networks

Ana Filipa Filipe · Lorenzo Quaglietta · Mário Ferreira ·
Maria Filomena Magalhães · Pedro Beja

Received: 9 February 2017 / Accepted: 26 June 2017

Abstract Species distribution models combining environmental and spatial components are increasingly used to understand and forecast species invasions. However, modelling distributions of invasive species inhabiting stream networks requires due consideration of their dendritic spatial structure, which may strongly constrain dispersal and colonization pathways. Here we evaluate the application of novel geostatistical tools to species distribution modelling in dendritic networks, using as case study two invasive crayfish (*Procambarus clarkii* and *Pacifastacus leniusculus*) in a Mediterranean watershed. Specifically, we used logistic mixed models to relate the probability of occurrence of each crayfish to environmental variables, while specifying three spatial autocorrelation components in random errors. These components described spatial dependencies between sites as a function of (1) straight-line distances (Euclidean model) between sites, (2) hydrologic (along the waterlines) distances between flow-connected sites (tail-up model), and (3) hydrologic distances irrespective of flow connection (tail-down model). We found a positive effect of stream order on *P. clarkii*, indicating an association with the lower and mid reaches of larger streams, while *P. leniusculus* was affected by an interaction between stream order and elevation, indicating an association with larger streams at higher altitude. For both species, models including environmental and spatial components far outperformed the pure environmental models, with the tail-up and the Euclidean components being the most important for *P. clarkii* and *P. leniusculus*, respectively. Overall, our study highlighted the value of geostatistical tools to model the distribution of riverine and aquatic invasive species, and stress the need to specify spatial dependencies representing the dendritic network structure of stream ecosystems.

Keywords Dendritic ecological networks · Freshwater invasions · Geostatistics · Spatial autocorrelation · Species distribution models · Stream ecology

---

Electronic supplementary material The online version of this article (doi:10.1007/s10530-017-1492-3) contains supplementary material, which is available to authorized users.

Ana Filipa Filipe and Lorenzo Quaglietta have contributed equally to this work.

A. F. Filipe · L. Quaglietta · M. Ferreira · P. Beja (✉)
CIBIO/InBio, Centro de Investigação em Biodiversidade
e Recursos Genéticos, Universidade do Porto, Campus
Agrário de Vairão, Rua Padre Armando Quintas,
4485–601 Vairão, Portugal
e-mail: pbeja@cibio.up.pt

A. F. Filipe · L. Quaglietta · M. Ferreira · P. Beja
CEABN/InBio, Centro de Ecologia Aplicada "Professor
Baeta Neves", Instituto Superior de Agronomia,
Universidade de Lisboa, Tapada da Ajuda,
1349-017 Lisbon, Portugal

M. F. Magalhães
cE3c, Centro de Ecologia, Evolução e Alterações
Ambientais, Faculdade de Ciências, Universidade de
Lisboa, 1749-016 Lisbon, Portugal

Introduction

The invasion of stream ecosystems by exotic species is a cause of concern worldwide due to their negative environmental and economic impacts (Strayer 2010; Walsh et al. 2016). Therefore, understanding how invasive freshwater species spread into novel areas is essential to prevent further expansion, promote eradication, or adapt to their continued presence. This problem has often been addressed using distribution modelling tools, aiming to understand invasion drivers, predict species range expansions or contractions in relation to natural and anthropogenic factors, and to guide early detections (Capinha and Anastácio 2011; Siesa et al. 2011; Larson and Olden 2012; Václavík et al. 2012; Capinha et al. 2013). However, studies often ignore the specificities of aquatic organisms, particularly those living along dendritic stream networks, which may bias model results and ultimately mislead management prescriptions (Peterson et al. 2013).

While terrestrial species live and move in a two-dimensional space, the movements of stream species are constrained by the topology of the dendritic stream network, which strongly affects their distribution, persistence and diversity (Carrara et al. 2012; Altermatt 2013). This applies to strictly aquatic species such as freshwater molluscs and fish, but also to species that move preferentially along waterlines but that can also travel overland such as amphibians and semi-aquatic mammals (e.g., Grant et al. 2010; Quaglietta et al. 2014). In invasive species, the stream spatial structure may constrain the patterns of expansion from initial founder populations, with individuals dispersing up- and downstream and progressively colonizing favourable habitats across the stream network (Bernardo et al. 2011; Bronnenhuber et al. 2011; Hein et al. 2011). Stream reaches that are connected and close to founder populations may thus become occupied first, while farther or unconnected reaches may take more time to be colonized, even if there are favourable habitat conditions. Early during expansion, the distribution is unlikely to be in equilibrium with the environment, as a species may be absent from potentially suitable sites because of colonization time lag and dispersal limitations (Václavík and Meentemeyer 2009, 2012). Over time, the species may progressively spread across the river network and eventually colonize all suitable habitats, thereby converging to an equilibrium with prevalent environmental conditions. At this stage, the stream network topology may still be important because, for instance, reaches sharing the same headwaters are likely to have similarities in terms of flow regime and water chemistry, thus providing similar habitat conditions (Carrara et al. 2012; Peterson et al. 2013; McGuire et al. 2014). Failure to incorporate these spatial processes may introduce errors and biases in distribution modelling, such as over-estimating the importance of environmental factors (Diniz-Filho et al. 2003; Václavík et al. 2012), under-estimating potential distribution ranges (Václavík and Meentemeyer 2012), or over-estimating actual distribution ranges (De Marco et al. 2008; Václavík and Meentemeyer 2009; Václavík et al. 2012).

The use of geostatistical modelling to account for spatial dependencies in physical and ecological processes across stream networks was first introduced by Ver Hoef et al. (2006), providing a valuable tool to improve distribution modelling of aquatic invasive species. Geostatistical models are similar to conventional linear mixed models, with spatial autocorrelation specified in the random errors. In ecological applications, the deterministic mean of the dependent variable is modelled as a linear function of explanatory variables, and local deviations from the mean are modelled using the spatial autocorrelation between nearby sites (Ver Hoef et al. 2006; Ver Hoef and Peterson 2010; Peterson and Ver Hoef 2010). This is specified using covariance functions, which represent the strength of the influence between sites as a function of the distance separating such sites. Distances can be straight-line (Euclidean) distances measured overland, or hydrologic distances measured along the flow lines (i.e. longitudinal connection). Hydrologic distances can represent flow-connected relations, assuming that a point downstream may be influenced by a point upstream, but not the reverse (tail-up models), or both flow-connected and flow-unconnected relations, assuming that influences are not limited by flow direction (tail-down models). The stream network models of Ver Hoef and Peterson (2010) account for these multiple spatial relationships based on a mixture of covariances, each of which may be specified using a moving-average function (e.g. exponential, spherical, linear-with-sill). Covariance functions differ on how they specify the distance-decay of spatial influences, but typically they have three parameters: the nugget effect, representing the variation between sites when their separation distance approaches zero; the sill, representing the variance found among spatially independent sites; and the range, representing how fast the covariance decays with distance (Peterson et al. 2007). Application of these models to aquatic invasive species might provide information on the relative role of environmental versus spatial processes on current distributions, which can help clarifying whether the species is in equilibrium with the environment. Also, they might provide a basis to infer likely mechanisms of expansion (e.g. Siesa et al. 2011), by comparing for instance the relative importance of different spatial components. Finally, they might allow more robust inferences on species occurrences at unobserved sites, using observed values at nearby sites.

This study explores the use of geostatistical distribution modelling to clarify the conditions associated with the establishment of two invasive crayfish and to predict their potential expansion range in dendritic stream networks. We focused on the red swamp crayfish *Procambarus clarkii* and the signal crayfish *Pacifastacus leniusculus*, which are the two most widely distributed invasive crayfish worldwide (Gherardi et al. 2011; Hänfling et al. 2011). Previous studies have shown that these crayfish have different environmental requirements and that their invasion range may still not be in equilibrium with the environment (Capinha and Anastácio 2011; Capinha et al. 2013), but they disregarded the potentially important role of spatial processes occurring across dendritic stream networks. Here we use a case study in a Mediterranean watershed of NE Portugal, aiming to: (1) identify the main environmental factors explaining the current crayfish distributions, (2) quantify the relative importance of environmental and spatial processes influencing the distributions; (3) compare the relative importance of spatial processes described by the Euclidean, tail-up and tail-down models; and (4) predict the potential invasion ranges by combining environmental and spatial predictors. Results are used to discuss the value of geostatistical approaches in relation to conventional species distribution modelling to deal with aquatic invasive species.

Methods

Study area and species

The study was conducted in NE Portugal, in the river Sabor watershed (N41°09′–42°00′, W7°15′–6°15′; Fig. 1), which covers a wide range of environmental conditions in terms of elevation (100–1500 m above sea level), total annual precipitation (443–1163 mm), and mean annual temperature (6.9–15.6 °C). Climate is Mediterranean, with precipitation largely concentrated in October–March, while it is virtually absent in the hot summer months (June–August). Flow regime is highly seasonal, with most headwater streams drying out or being reduced to a series of disconnected pools in summer, though the main watercourse and the largest tributaries are permanent. A large hydroelectric power plant involving two dams located near the mouth of the Sabor river were under construction during the study period (Jackson 2011), but otherwise the river was largely free flowing except for a few small impoundments. See Ferreira et al. (2016) for details of the study area.

The two crayfish species studied were intentionally introduced in Europe in the 1960s (*P. leniusculus*) and in the 1970s (*P. clarkii*), due to their economic value (Clavero 2016), and currently have well-established populations (Souty-Grosset et al. 2006; Capinha et al. 2013). In the Sabor watershed, *P. clarkii* was first recorded in the 1990s (Bernardo et al. 2011), but possibly was present earlier because it has spread rapidly in the Iberian Peninsula since the first introduction in 1973 (Habsburgo-Lorena 1978; Ramos and Pereira 1981). Although *P. leniusculus* was also introduced in the Iberian Peninsula in the 1970s, it probably reached the Sabor only in 1994, following a deliberate introduction in the Spanish sector of the watershed (Bernardo et al. 2011). Recent studies suggest that the two species are still expanding in the Sabor watershed, which may be a consequence of dispersal from source populations along the stream network, in both the upstream and downstream directions (Bernardo et al. 2011; Anastácio et al. 2015). Dispersal overland may also occur in at least *P. clarkii* (Cruz and Rebelo 2007; Ramalho and Anastácio 2015), but its contribution to range expansion in the study area is unknown.

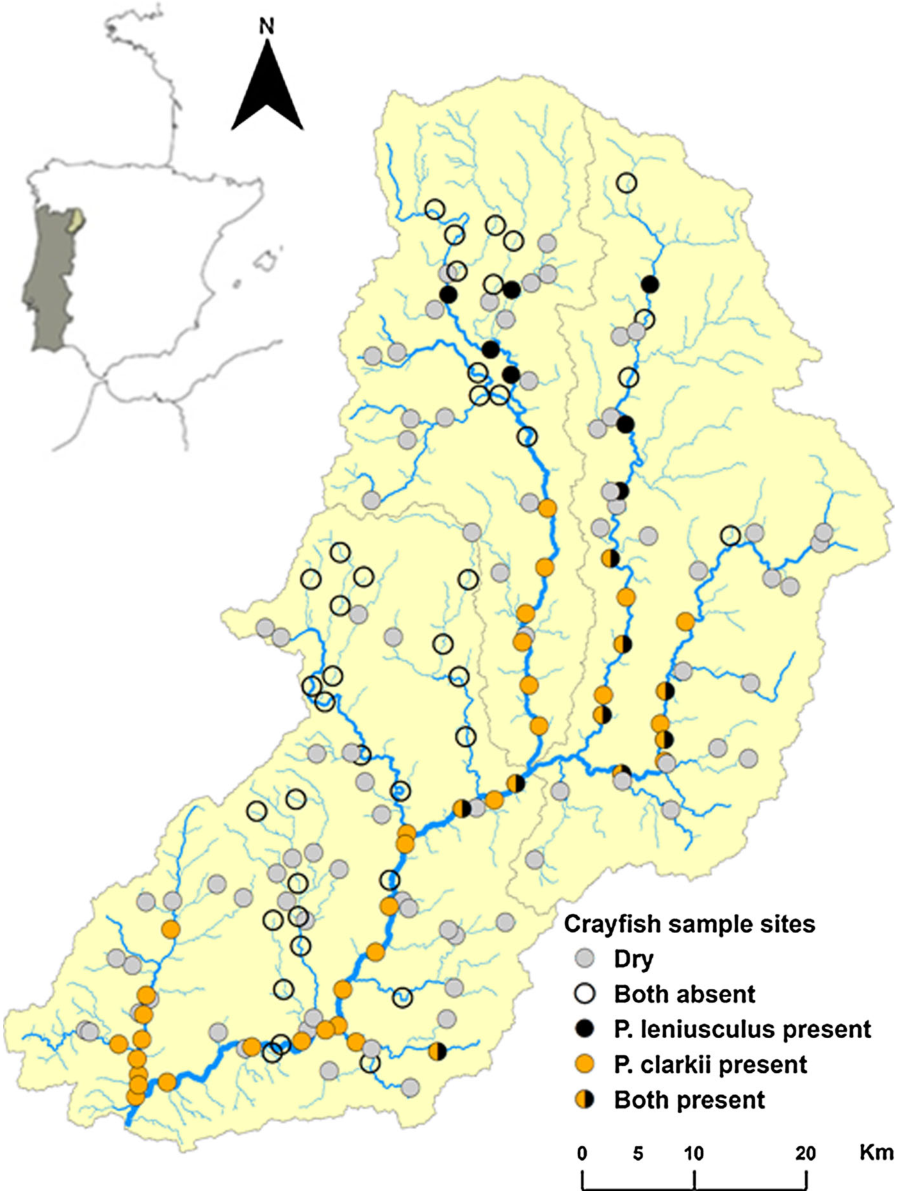

Fig. 1 Map of the Sabor watershed (NE Portugal) and its location in the Iberian Peninsula, showing the spatial distribution of the sites visited, the sites sampled using electrofishing, and the sites where each crayfish species was detected

Crayfish survey

We visited 167 50-m stream reaches (hereafter sites) providing a homogeneous coverage of the Sabor catchment during the summer of 2012 (June 13 to July 15, and August 28 to September 15; Fig. 1). At sites where there was flowing water or isolated pools, we estimated the presence/absence of crayfish using a single anode electrofishing gear, always operated by the same person (MF), and following standard procedures outlined in Ferreira et al. (2016). One additional researcher with a net was always present to pick up stunned crayfishes. After identification, all individuals were eliminated following national regulations for invasive species. The presence of crayfish in dry stream stretches was not assessed, though they may be able to persist therein by burrowing in soft sediments (Guan 1994; Gherardi et al. 2011). Therefore, distribution modelling focused on crayfish that are active in the water during the dry summer season. Although this could potentially underestimate the true crayfish distribution, we believe this was not a serious issue, because incidental observations throughout the year suggest that crayfish are actually absent from headwater streams drying out in summer (Mário Ferreira and Lorenzo Quaglietta, Unpublished Data).

Environmental and spatial data

To model crayfish distribution, we used three potentially influential environmental variables (Capinha and Anastácio 2011, 2013; Moreira et al. 2015) that can be easily extracted from topographic online maps, and could thus be used to extrapolate the species distribution models to the entire watershed. Strahler's stream order was used as a proxy for habitat size and heterogeneity (Hughes et al. 2011; Ferreira et al. 2016), and it was extracted from CCM2 (Catchment Characterization and Modelling database), which is based on a 100-m resolution digital elevation model (DEM) (Vogt et al. 2007). Elevation and slope at each site were also estimated from information available in CCM2. We did not use climate variables, because they were often highly correlated with the topographic variables, and the range of conditions within the study area was well within the much wider climate niches of both species (Capinha and Anastácio 2011, 2013). All variables were screened for potential outliers and influential points.

Spatial data necessary for geostatistical modelling was obtained in a geographic information system (GIS) using the Sabor watershed network extracted from CCM2 and the layer of sampling locations. Estimates included the Euclidean and hydrologic distances (total and downstream hydrologic distances) between every pair of sampling sites (Peterson and Ver Hoef 2010). We also estimated the watershed area draining to each site to be used in tail-up models (see below). Estimates were made using the Spatial Tools for the Analysis of River Systems (STARS) toolbox version 2.0.0 (Peterson and Ver Hoef 2014) for ArcGIS 10.2 (ESRI 2011).

Geostatistical modelling

To visualise spatial dependencies along the stream network, we built empirical semivariograms depicting how semivariance in species presence/absence between pairs of sampling sites changed in relation to the hydrologic distances separating them (Torgegrams), considering either flow-connected or flow-unconnected sites (Peterson et al. 2013). The distribution (presence/absence) of each crayfish was then modelled in relation to environmental variables using a logistic function, with spatial autocorrelation specified through a full covariance mixture-model in the random component (Ver Hoef and Peterson 2010). Therefore, the model incorporated spatial dependencies represented by tail-up (TU), tail-down (TD) and Euclidean (EUC) models. To deal with confluences in tail-up models, the spatial weights to allocate the moving-average function between upstream segments were based on watershed areas (e.g. Ver Hoef et al. 2006; Ver Hoef and Peterson 2010).

For each species, model building followed a two-step procedure (Peterson and Ver Hoef 2010). First, we selected the environmental component while maintaining constant the spatial component, which included exponential tail-up, linear-with-sill tail-down, and Gaussian Euclidean spatial covariance functions. We tested all combinations of environmental variables and their interaction terms, and selected in each case the model with the lowest root-mean-squared-prediction error (RMSPE) estimated through leave-one-out cross-validation (Ver Hoef et al. 2006; Ver Hoef and Peterson 2010; Peterson and Ver Hoef 2010; Frieden et al. 2014). Interactions were only specified when the main effects were also included in the model. We then built the spatial component while maintaining the best environmental model selected in the previous step. We tested all combinations of spatial components, using exponential, spherical, Gaussian and Cauchy functions as alternatives for the Euclidean model, and exponential, linear-with-sill, spherical and Mariah functions as alternatives for the tail-up and tail-down models (Ver Hoef and Peterson 2010). We selected the best function for each spatial component based on the minimization of RMSPE criterion, and kept the three components in the model to allow estimates of the proportion of variation explained by each one. Besides this full "spatial/environmental model", we also built a pure "environmental model" based on a simple logistic regression, and a pure "spatial model", including in each case the variables and functions selected in model building. The discrimination ability of each model was estimated using the area under the receiver operating characteristic curve (AUC) (e.g., Václavík and Meentemeyer 2009), and Cohen's kappa using species prevalence as the threshold for predicted presences (Titus et al. 1984).

We performed all analyses in the R version 3.1.2 (R Core Team 2014) and the spatial stream networks (SSN) packages (Ver Hoef et al. 2014), and used ArcGIS 10.2 for mapping (ESRI 2011). AUC was computed using the package pROC (Robin et al. 2011), and Cohen's kappa was computed using the package irr (Gramer et al. 2012).

Species distribution mapping

The Sabor stream network was divided in 1716 segments, which were used as a basis to predict the distribution of each crayfish using the species distribution models (Ferreira et al. 2016). For each segment, we extracted the environmental variables, the area of the watershed discharging into the segment, and the in-stream and Euclidean distances to every other segment. Prediction of the probability of each species being present in each segment was then computed using universal kriging (Cressie 1993). Maps of species potential distribution across the watershed were then produced, using the prevalence of each species as the threshold for separating segments with predicted presence or absence. Maps were built using either the "spatial/environmental model" or the "environmental model", to assess whether considering the spatial network structure improved the species distribution mapping.

Results

From the 167 visited sites, 87 had flowing water or isolated pools and were thus sampled for crayfish (Fig. 1). From these, we detected *P. clarkii* at 41 (46.1%) and *P. leniusculus* at 16 (18.0%) sites. Visual analysis of Torgegrams indicated that there were spatial dependencies in species occurrences along the stream network, as suggested by the patterns of change in semivariance between sites in relation to the hydrologic distances separating them (Fig. 2). For *P. clarkii*, there was a rapid and nearly linear increase in semivariance with distance between flow-connected sites, levelling off at about 30 km, while semivariance between flow-unconnected sites was generally much smaller and it increased slowly with distance between sites. For *P. leniusculus*, the Torgegrams showed much less marked patterns than for *P. clarkii*, though semivariance was also smaller between flow-unconnected than flow-connected sites.

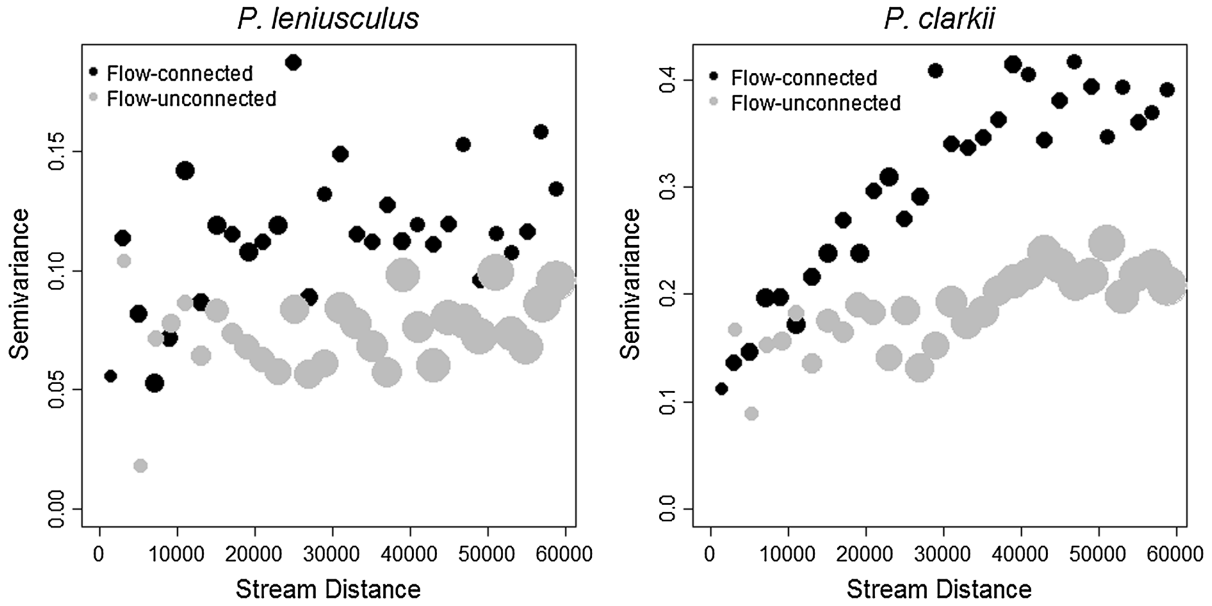

Fig. 2 Torgegrams describing spatial dependencies in the distribution (presence/absence) of *Procambarus clarkii* and *Pacifastacus leniusculus* along the stream network of the Sabor watershed (NE Portugal). The graphs show changes in semivariance between sampling sites in relation to the hydrologic distances separating them, considering either flow-connected (*black circles*) or flow-unconnected (*grey circles*) sites. The size of *circles* is proportional to the number of pairs of sites used to estimate the semivariance

In model building, the environmental component with the lowest RMSPE for *P. clarkii* included only the positive effect of stream order on the probability of species occurrence (Table 1, Table S1). The environmental component for *P. leniusculus* included stream order, elevation, and their interaction term (Table 1, Table S2), indicating that the probability of occurrence increased in higher order streams at higher elevation (Table 1, Fig. 3). Regarding the spatial component, the best covariance structure for *P. clarkii* included a tail-up linear-with-sill function and a Euclidean spherical function (Table 2, Table S3). The model with the three spatial components included also a tail-down linear-with-sill function, and it was very close to the best considering the RMSPE criterion (Table 2, Table S3). The range of the tail-up (31.8 km) and Euclidean (188.2 km) components were similar in both models, and much smaller than the range estimated for the tail-down component (627.4 km). For *P. leniusculus*, the best spatial model included a tail-down exponential function, and a Euclidean Gaussian function (Table 2). The full autocovariance mixture model was also the second best, including a tail-up linear-with-sill-function, a tail-down spherical function, and a Euclidean Cauchy function (Table 2, Table S4). The range estimate for the tail-up component was very small (1.4 km), while those of the tail-

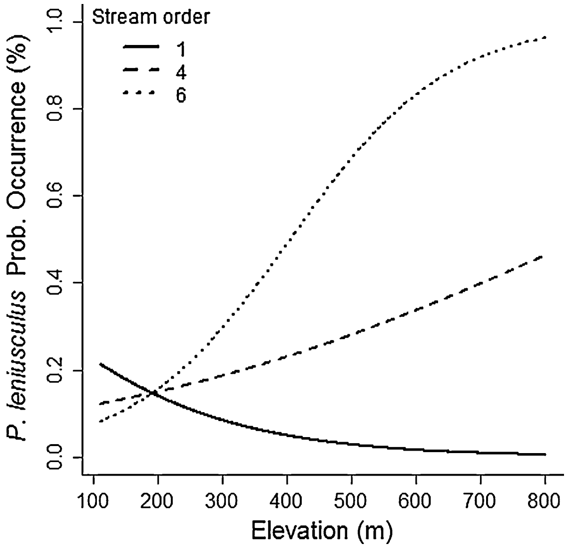

Fig. 3 Response curves estimated from a logistic mixed model relating the probability of occurrence of *Pacifastacus leniusculus* to environmental variables, describing the interaction effect between stream order and elevation

Table 1 Parameter estimates and summary statistics for the environmental component of crayfish distribution models in the Sabor watershed (NE Portugal)

Models	Variables	Coef.	SE	*t* values	*P* values	AUC	*k*
Procambarus clarkii							
Environmental	Intersect	−6.712	0.88	−7.596	<0.001	0.861	0.447
	Stream order	1.354	0.20	6.778	<0.001		
Spatial	Intersect	−1.123	0.18	−6.226	<0.001	0.975	0.632
Spatial/environmental	Intersect	−5.520	1.11	−4.968	<0.001	0.963	0.836
	Stream order	1.150	0.20	5.831	<0.001		
Pacifastacus leniusculus							
Environmental	Intersect	−0.968	2.61	−0.372	0.711	0.782	0.324
	Stream order	−0.761	0.52	−1.457	0.147		
	Elevation	−0.013	0.00	−2.318	0.022		
	Stream order × Elevation	0.004	0.01	3.439	<0.001		
Spatial	Intersect	−2.245	0.26	−8.512	<0.001	0.904	0.216
Spatial/environmental	Intersect	−0.138	1.25	−0.111	0.912	0.823	0.379
	Stream order	−0.531	0.26	−2.050	0.042		
	Elevation	−0.008	0.01	−2.896	0.004		
	Stream order × Elevation	0.003	0.00	3.667	<0.001		

For each species we present the parameter estimates for the pure environmental, the pure spatial, and the spatial/environmental models. For each variable included in the best model, we provide the regression coefficient (Coef.), the standard error of the estimate (SE), the corresponding *t* and *P* values. For each model, we also provide the area under the receiver operating characteristic curve (AUC), and the Cohen's kappa (k)

Table 2 Comparison of mixture models relating the probability of occurrence of two crayfish species to environmental and spatial components

Mixture	Model 1	Model 2	Model 3	RMSPE
Procambarus clarkii				
Nonspatial				3.0813
Spatial				
TU	Linear-sill			3.0039
TD	Linear-sill			2.7982
EUC	Cauchy			2.3776
TU/TD	Linear-sill	Linear-sill		2.3559
TU/EUC	Linear-sill	Spherical		2.2351
TD/EUC	Mariah	Spherical		2.3360
TU/TD/EUC	Linear-sill	Linear-sill	Spherical	2.2371
Pacifastacus leniusculus				
Nonspatial				3.0054
Spatial				
TU	Mariah			3.0785
TD	Mariah			2.3243
EUC	Exponential			1.6234
TU/TD	Mariah	Linear-sill		1.8258
TU/EUC	Mariah	Gaussian		1.5947
TD/EUC	Exponential	Gaussian		1.4116
TU/TD/EUC	Linear-sill	Spherical	Cauchy	1.4238

The models shown represent the best model fit for each mixture type based on the root mean square prediction error (RMSPE). Models are Euclidean (EUC), tail-up (TU) and tail-down (TD), and they were tested using exponential, spherical, Gaussian and Cauchy functions as alternatives for EUC, and exponential, linear-with-sill, spherical and Mariah functions as alternatives for TU and TD

down (634.8–645.93 km) and Euclidean (371.1 km) components were very large.

For both crayfish species, the model combining environmental and spatial components outperformed both the pure environmental and the pure spatial model, in terms of variance explained, AUC and Cohen's kappa (Tables 1, 3). In the spatial/environmental model, the environmental component always explained less variance than the spatial component (Table 3). The full covariance mixture model for *P. clarkii* accounted for about half the variance in species presence/absence, of which about 41% corresponded to tail-up spatial dependencies, another 25% to Euclidean spatial dependencies, and 34% to environmental variables (Table 3). The predictive accuracy of the model was very satisfactory, as measured using either AUC or Cohen's kappa (Table 1). For *P. leniusculus*, the full mixture model accounted for almost all variance, most of which (64%) corresponding to the tail-up component, and the rest by the Euclidean (25%) and environmental (11%) components. However, the very small range estimated for the tail-up component (1.4 km) indicates that autocorrelation was essentially zero between all sampling sites, so this component is acting like a nugget effect. Predictive accuracy was reasonable when measured through AUC, but low when measured using Cohen's kappa (Table 1).

The distribution models based solely on environmental variables produced poor results for the potential distribution of both species, while improving markedly when considering the spatial component (Fig. 4). For *P. clarkii*, the environmental model correctly predicted a continuous distribution along the mid and lower reaches of the Sabor river, but it produced many false absences along the main tributaries. In contrast, the environmental/spatial model predicted a distribution much closer to that observed, highlighting a continuous occurrence in the main river and large tributaries. The environmental model for *P. leniusculus* also predicted a distribution much restricted than that observed, while the environmental/spatial model produced a larger distribution that was closer to that observed. However, this model missed sections of river where the species was detected, while predicting a far more continuous distribution than that observed.

Table 3 Percentage of variance accounted for by models relating the probability of two crayfish species to environmental and spatial components

	Environmental	Spatial			Nugget
		TU	TD	EUC	
Procambarus clarkii					
Nonspatial	21.8	–	–	–	78.2
Best model	17.1	20.8	–	13.0	49.2
Full covariance mixture	17.1	20.8	≈0.0	12.8	49.3
Pacifastacus leniusculus					
Nonspatial	16.2	–	–	–	83.8
Best model	10.8	–	≈0.0	26.0	63.2
Full covariance mixture	11.0	63.0[a]	0.2	24.5	1.3

For each species we present the values for the model including only environmental variables (nonspatial), for the spatial/environmental model minimising the root mean square prediction error (Best model), and for the full mixture spatial/environmental model (Full covariance mixture). Separate percentages are provided for the Euclidean (EUC), tail-up (TU) and tail-down (TD) components of the spatial model. The nugget is the variation unexplained in models

[a] Given the very small range estimated (1.4 km), this component was acting like a nugget effect

Discussion

Our study found that the occurrence of *P. clarkii* and *P. leniusculus* in a Mediterranean stream network was strongly related to environmental variables such as stream order and elevation, but also that species distributions appeared strongly shaped by spatial processes. Actually, we found that spatial dependencies accounted for a larger proportion of variation in species occurrences than environmental variables, and that disregarding spatial effects seemed to strongly underestimate potential species distributions. These general patterns were observed for both species, though model performance was much lower for the species with a more restricted and patchy distribution (*P. leniusculus*), than for the species with a larger a more continuous distribution (*P. clarkii*). Overall, our results point out the importance of considering both environmental and spatial effects when modelling the distribution of invasive aquatic organisms, and stress the need to specify spatial dependencies representing the dendritic network structure of stream ecosystems.

Environmental effects

The only variable included in the environmental model for *P. clarkii* was stream order, indicating that probability of occurrence increased from the headwaters to the large streams and the main river. This is in line with previous studies suggesting that this species often occurs in lowland rivers, with lower slopes, larger width, slower flowing waters, abundant aquatic vegetation and finer sediments (e.g., Cruz and Rebelo 2007; Anastácio et al. 2015; Moreira et al. 2015), which often are associated with high-order streams (Allan and Castillo 2007). On the other hand, absence from low order streams (i.e., small tributaries) was probably a consequence of these drying out in summer (Ferreira et al. 2016), and so they were not occupied by active crayfish at the time of sampling. The pure environmental model correctly predicted the extent of species distribution in the main Sabor river, but it underestimated considerably the extent of occurrence in its main tributaries. This could be a consequence of this study using only a limited set of environmental variables, thereby failing to fully identify the habitat conditions required by the species. It is also possible, however, that the distribution in the tributaries was strongly influenced by spatial processes associated with expansion from the core distribution in the main river (Bernardo et al. 2011), thereby confounding the operation of environmental drivers (see below).

In the case of *P. leniusculus*, the environmental model included the interactive effects of stream order and elevation, suggesting that the species prefers high-

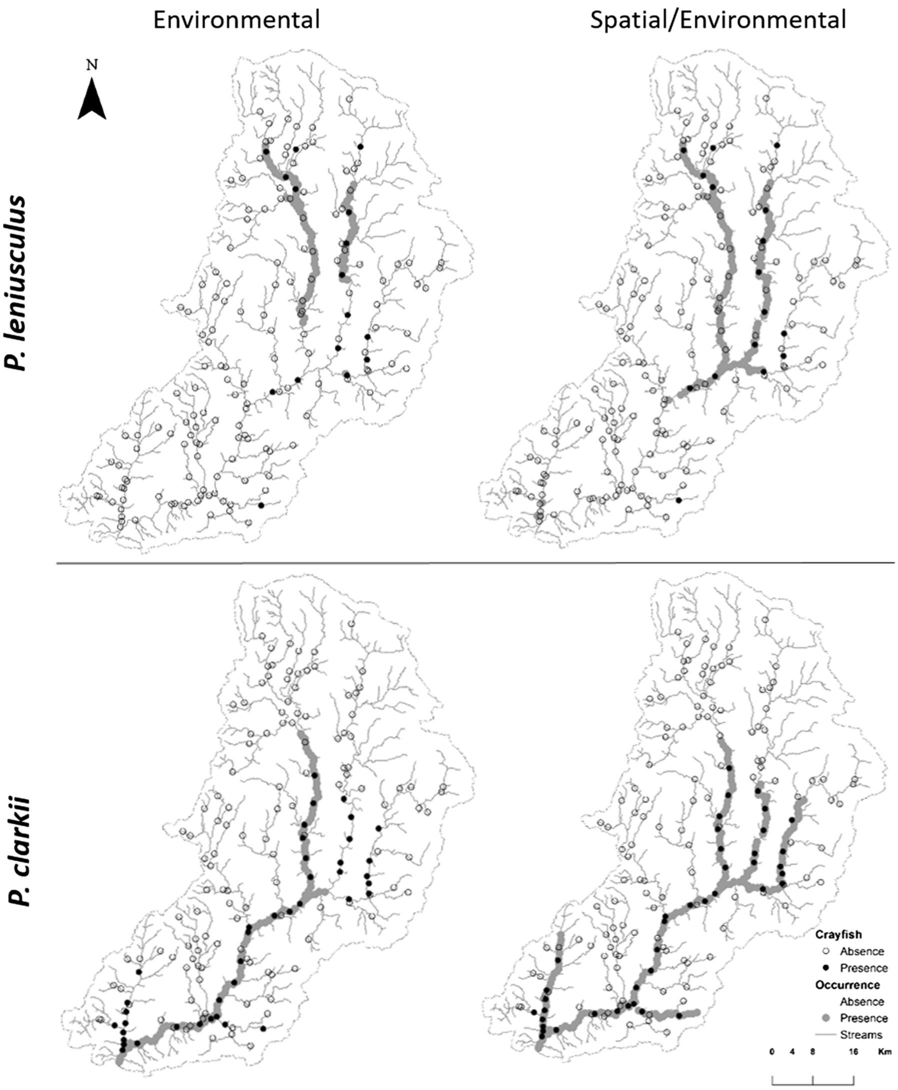

Fig. 4 Maps of potential distribution of *Procambarus clarkii* and *Pacifastacus leniusculus* in the river Sabor watershed (NE Portugal), predicted from either simple environmental models (*left panels*), or models including both environmental and spatial components (*right panels*). The threshold for predicted presences was set equal to the observed prevalence of each species

order streams at higher elevation, while being absent from lowland high-order streams. This is in line with previous studies suggesting that the species is highly adapted to mountain rivers, where it prefers riffle habitats with abundant riparian vegetation (Rallo and García-Arberas 2002; Anastácio et al. 2015). However, the species was absent from the small, lower order streams, possibly because the headwaters of the Sabor watershed generally dry out in summer (Ferreira et al. 2016). The pure environmental model provided a relatively crude picture of the species distribution, producing both false absences and false presences. This may be a consequence, at least partly, of the relatively small number of presences detected in our study (16), which can cause problems in logistic models (Vittinghoff and McCulloch 2007). In alternative, this may be a consequence of the species being still in rapid expansion from several points of introduction (Bernardo et al. 2011; Amilcar Teixeira, personal communication), and thus remaining out of equilibrium with the environmental conditions. Whatever the mechanisms, these results suggest that the distribution of both species in stream networks may be poorly predicted by purely environmental models.

Spatial effects

The models of both species greatly improved when autocovariance functions accounting for spatial processes were considered. For *P. clarkii*, there was a particularly strong tail-up component, suggesting that although presences were spatially autocorrelated along streams, the prevalence of the species can vary widely between stream branches just above a confluence. This result, in combination with the spatial distribution pattern inferred from the environmental/spatial model, is compatible with the idea that the species expands in an invasion front that moves upstream from the lowland sections of the main river. However, expansion may advance primarily through some stream branches, possibly corresponding to higher order streams, as underlined by the environmental model, while colonization may be slower or do not occur in other branches, possibly the smaller tributaries. Previous studies in the Sabor watershed have indeed shown a progressive upstream expansion (Bernardo et al. 2011), and radiotracking revealed that individual crayfish have the ability to move considerably in the upstream direction (up to 250 m in half day; Anastácio et al. 2015). In contrast to our results, previous studies found that *P. clarkii* occurrences were spatially autocorrelated only up to about 2–2.5 km (Cruz and Rebelo 2007; Siesa et al. 2011), which is much smaller than the ≈ 30 km range estimated in our study. This may be due to differences in habitat characteristics, with previous studies largely focusing on naturally disconnected waterbodies such as ponds and small lakes, where dispersal may be more difficult than in a continuous stream network such as that of our study area.

The model for *P. clarkii* also included a significant Euclidean component, but the underlying mechanism was unclear. However, the large range estimated (≈ 190 km) suggests that spatial autocorrelation was mainly due to broad scale distribution trends, rather than small scale processes. This may be a consequence of the topology of the watershed, as the species was consistently absent in zones occupied by headwater streams, while it was present in three waterlines that run parallel and at short distance to each other. These circumstances probably determined a positive relation between occupancy status and straight-line distances between sites, albeit with no ecological meaning. It could not be ruled out, however, that the Euclidean pattern was at least partly due to the effect of an unmeasured, spatially-structured environmental variable, though it is uncertain what this variable might be as all the area is within the climate niche of the species (Capinha et al. 2013). It is also unlikely that the Euclidean component reflected colonisations occurring by individuals moving through terrestrial habitats. Although this process has been described for *P. clarkii* inhabiting lowland habitats (Cruz and Rebelo 2007; Ramalho and Anastácio 2015), it is unlikely that it contributed much to species expansion in a dry and mountainous area such as ours, where dispersal overland would imply long movements across very dry habitats and steep slopes.

The spatial models developed for *P. leniusculus* differed considerably and were generally poorer than those of *P. clarkii*. Furthermore, there appeared to be some model instability, as shown by the full covariance mixture model, where the tail-up component acted like a nugget effect, with a range estimate (1.4 km) that was smaller than the typical distance between nearest sites (≈ 5 km). Nevertheless, it is noteworthy that there were spatial dependencies described by the Euclidean component, which explained over twice the variation in *P. leniusculus* distribution as the environmental component. As for *P. clarkii*, the large range estimated for the Euclidean component (≈ 370 km) suggests that it described primarily broad scale trends in species distribution, rather than small scale processes. Including the spatial component improved the performance of the pure environmental model, but even so the predictive ability of the best environmental/spatial model was low, with a large number of false presences and absences. These patterns are probably a consequence of the low number of presences detected for this species, as noted for the environmental component, but they may also result from its patchy distribution, which was possibly caused by multiple introductions followed by progressive expansions that are still far from complete. This likely affected both environmental and spatial relationships, because the species was likely absent from sites with adequate environmental conditions, while the spatial dependencies were inconsistent, possibly due to the spatial scattering of the introductions. As this species is still rapidly spreading in the study area (Bernardo et al. 2011; Anastácio et al. 2015; Amilcar Teixeira, personal communication), we expect that the spatial distribution will keep changing for some more time, eventually reaching a stable pattern corresponding to the equilibrium with environmental conditions. In the meantime, we predict that sites that were identified in our study as false absences will soon be occupied, and that the species will progressively occupy most, if not all larger streams at higher elevation.

Distribution modelling of invasive species across stream networks

Our study adds to recent research showing the importance of explicitly accounting for dendritic spatial structures when modelling the distribution of physical and ecological processes across stream networks, and the value of geostatistical tools to undertake such modelling (Peterson et al. 2013; McGuire et al. 2014). Also, our study provides novel insights on the application of this approach to invasive species, whose distribution modelling is affected by a specific set of challenging processes such as non-equilibrium with environmental conditions and highly dynamic distribution patterns (de Marco et al. 2008; Václavík and Meentemeyer 2012). First, our results reinforce the idea that that both environmental and spatial processes need to be incorporated to predict the distribution of invasive species (de Marco et al. 2008; Siesa et al. 2011; Václavík et al. 2012). Second, we confirmed that Euclidean distances are insufficient to incorporate spatial structure in distribution models for invasive stream organisms, as they may fail to account for their strongly constrained dispersal along waterlines and may produce biased and biologically meaningless results (Filipe et al. 2010; Altermatt 2013; Peterson et al. 2013). Third, considering the tail-up autocovariance function may be generally recommended when testing for spatial dependencies, as it accounts for biological meaningful ecological processes such as connectivity along flow-connected waterlines (Carrara et al. 2012). Finally, geospatial models such as those applied here may improve inferences on the invasion processes of aquatic organisms from distribution data (e.g., Siesa et al. 2011; Václavík et al. 2012), by accounting for biologically more meaningful spatial dependencies along waterlines, though careful interpretation of results is required. Care should particularly be taken when dealing with species at the early stages of invasion, as it was the case of *P. leniusculus* in our study, where a small number of occurrences and a very fragmented distribution may produce models that are difficult to interpret. Overall, we suggest that geostatistical modelling across stream networks provides an important addition to the toolbox of researchers interested in biological invasions of aquatic organisms, which may contribute to address this global environmental problem by helping to understand driving mechanisms and to predict future distributions (Strayer 2010).

Acknowledgements This work is part of the Baixo Sabor Long Term Ecological Research (LTER) project, funded by Portuguese Science and Technology Foundation (FCT) through LTER/BIA-BEC/0004/2009, and by EDP Energias de Portugal. AFF was supported by the FRESHING Project funded by FCT and COMPETE (PTDC/AAG-MAA/2261/2014 – POCI-01-0145-FEDER-356 016824), and MF was supported by FCT PhD grant SFRH/BD/95202/2013. We thank Pedro Silva, Rita Severino, Sara Ivone and Sérgio Henriques for their collaboration in field work, and Duarte Prata for preliminary analysis of the data.

References

Allan JD, Castillo MM (2007) Stream ecology: structure and function of running waters, 2nd edn. Springer, Dordrecht

Altermatt F (2013) Diversity in riverine metacommunities: a network perspective. Aquat Ecol 47:365–377

Anastácio PM, Banha F, Capinha C, Bernardo JM, Costa AM, Teixeira A, Bruxelas S (2015) Indicators of movement and space use for two co-occurring invasive crayfish species. Ecol Indic 53:171–181

Bernardo JM, Costa AM, Bruxelas S, Teixeira A (2011) Dispersal and coexistence of two non-native crayfish species (*Pacifastacus leniusculus* and *Procambarus clarkii*) in NE Portugal over a 10-year period. Knowl Manag Aquat Ecosyst 401:28

Bronnenhuber JE, Dufour BA, Higgs DM, Heath DD (2011) Dispersal strategies, secondary range expansion and invasion genetics of the nonindigenous round goby, *Neogobius melanostomus*, in Great Lakes tributaries. Mol Ecol 20:1845–1859

Capinha C, Anastácio P (2011) Assessing the environmental requirements of invaders using ensembles of distribution models. Divers Distrib 17:13–24

Capinha C, Brotons L, Anastácio P (2013) Geographical variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish. J Biogeogr 40:548–558

Carrara F, Altermatt F, Rodriguez-Iturbe I, Rinaldo A (2012) Dendritic connectivity controls biodiversity patterns in experimental metacommunities. Proc Natl Acad Sci USA 109:5761–5766

Clavero M (2016) Species substitutions driven by anthropogenic positive feedbacks: Spanish crayfish species as a case study. Biol Cons 193:80–85

Cressie NA (1993) Statistics for spatial data, Revised edn. Wiley, New York

Cruz MJ, Rebelo R (2007) Colonization of freshwater habitats by an introduced crayfish, *Procambarus clarkii*, in Southwest Iberian Peninsula. Hydrobiologia 575:191–201

De Marco P, Diniz-Filho JAF, Bini LM (2008) Spatial analysis improves species distribution modelling during range expansion. Biol Lett 4:577–580

Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. Glob Ecol Biogeogr 12:53–64

ESRI (2011) ArcGIS Desktop: Release 10.2. Environmental Systems Research Institute, Redlands, CA

Ferreira M, Filipe AF, Bardos DC, Magalhães MF, Beja P (2016) Modeling stream fish distributions using interval-censored detection times. Ecol Evol 6:5530–5541

Filipe AF, Magalhães MF, Collares-Pereira MJ (2010) Native and introduced fish species richness in Mediterranean streams: the role of multiple landscape influences. Divers Distrib 16:773–785

Frieden JC, Peterson EE, Webb JA, Negus PM (2014) Improving the predictive power of spatial statistical models of stream macroinvertebrates using weighted autocovariance functions. Environ Model Softw 60:320–330

Gherardi F, Aquiloni L, Diéguez-Uribeondo J, Tricarico E (2011) Managing invasive crayfish: is there a hope? Aquat Sci 73:185–200

Gramer M, Lemon J, Fellows I, Singh P (2012) Various coefficients of interrater reliability and agreement. https://cran.r-project.org/web/packages/irr/irr.pdf

Grant EHC, Nichols JD, Lowe WH, Fagan WF (2010) Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. Proc Natl Acad Sci USA 107:6936–6940

Guan RZ (1994) Burrowing behaviour of signal crayfish, *Pacifastacus leniusculus* (Dana), in the river Great Ouse, England. Freshw Forum 4:155–168

Habsburgo-Lorena AS (1978) Present situation of exotic species of crayfish introduced into Spanish continental waters. Freshw Crayfish 4:175–184

Hänfling B, Edwards F, Gherardi F (2011) Invasive alien Crustacea: dispersal, establishment, impact and control. Biocontrol 56:573–595

Hein CL, Öhlund G, Englund G (2011) Dispersal through stream networks: modelling climate-driven range expansions of fishes. Divers Distrib 17:641–651

Hughes RM, Kaufmann PR, Weber MH (2011) National and regional comparisons between Strahler order and stream size. J N Am Benthol Soc 30:103–121

Jackson AL (2011) Renewable energy vs. biodiversity: policy conflicts and the future of nature conservation. Glob Environ Chang 21:1195–1208

Larson ER, Olden JD (2012) Using avatar species to model the potential distribution of emerging invaders. Glob Ecol Biogeogr 21:1114–1125

McGuire KJ, Torgersen CE, Likens GE, Buso DC, Lowe WH, Bailey SW (2014) Network analysis reveals multiscale controls on streamwater chemistry. Proc Natl Acad Sci USA 111:7030–7035

Moreira FD, Ascensão F, Capinha C, Rodrigues D, Segurado P, Santos-Reis M, Rebelo R (2015) Modelling the risk of invasion by the red-swamp crayfish (*Procambarus clarkii*): incorporating local variables to better inform management decisions. Biol Invasions 17:273–285

Peterson EE, Ver Hoef JM (2010) A mixed-model moving-average approach to geostatistical modeling in stream networks. Ecology 91:644–651

Peterson EE, Ver Hoef JM (2014) STARS: an ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. J Stat Softw 56:1–17

Peterson EE, Theobald DM, Ver Hoef JM (2007) Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. Freshw Biol 52:267–279

Peterson EE, Ver Hoef JM, Isaak DJ, Falke JA, Fortin MJ, Jordan CE, McNyset K, Monestiez P, Ruesch AS, Sengupta A, Som N, Steel EA, Theobald DM, Torgersen CE, Wenger SJ (2013) Modelling dendritic ecological networks in space: an integrated network perspective. Ecol Lett 16:707–719

Quaglietta L, Fonseca VC, Mira A, Boitani L (2014) Sociospatial organization of a solitary carnivore, the Eurasian otter (*Lutra lutra*). J Mammal 95:140–150

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rallo A, García-Arberas L (2002) Differences in abiotic water conditions between fluvial reaches and crayfish fauna in some northern rivers of the Iberian Peninsula. Aquat Living Resour 15:119–128

Ramalho RO, Anastácio PM (2015) Factors inducing overland movement of invasive crayfish (*Procambarus clarkii*) in a ricefield habitat. Hydrobiologia 746:135–146

Ramos MA, Pereira TM (1981) Um novo Astacidae para a fauna portuguesa: *Procambarus clarkii* (Girard, 1852). Boletim do Instituto Nacional de Investigação Pescas, Lisboa 6:37–47

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 12:77

Siesa ME, Manenti R, Padoa-Schioppa E, De Bernardi F, Ficetola GF (2011) Spatial autocorrelation and the analysis of invasion processes from distribution data: a study with the crayfish *Procambarus clarkii*. Biol Invasions 13:2147–2160

Souty-Grosset C, Holdich DM, Noel PY, Reynolds JD, Haffner P (2006) Atlas of crayfish in Europe. Museum National d' Histoire Naturelle, Paris

Strayer DL (2010) Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. Freshw Biol 55:152–174

Titus K, Mosher JA, Williams BK (1984) Chance-corrected classification for use in discriminant analysis: ecological applications. Am Midl Nat 111:1–7

Václavík T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecol Model 220:3248–3258

Václavík T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Divers Distrib 18:73–83

Václavík T, Kupfer JA, Meentemeyer RK (2012) Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). J Biogeogr 39:42–55

Ver Hoef JM, Peterson EE (2010) A moving average approach for spatial statistical models of stream networks. J Am Stat Assoc 105:6–18

Ver Hoef JM, Peterson EE, Theobald D (2006) Spatial statistical models that use flow and stream distance. Environ Ecol Stat 13:449–464

Ver Hoef JM, Peterson EE, Clifford D, Shah R (2014) SSN: an R package for spatial statistical modeling on stream networks. J Stat Softw 56(3):1–43. http://www.jstatsoft.org/v56/i03/. Accessed Jan 2015

Vittinghoff E, McCulloch CE (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 165:710–718

Vogt JV, Soille P, de Jaeger AL, Rimaviciute E, Mehl W (2007) A Pan-European river and catchment database. EC-JRC (Report EUR 22920 EN) Luxembourg. http://desert.jrc.ec.europa.eu/action/documents/CCM2-Report_EUR-22920-EN_2007_STD.pdf. Accessed Jan 2015

Walsh JR, Carpenter SR, Vander Zanden MJ (2016) Invasive species triggers a massive loss of ecosystem services through a trophic cascade. Proc Natl Acad Sci USA 113:4081–4085