Biol Invasions (2010) 12:1427–1445
DOI 10.1007/s10530-009-9558-5

ORIGINAL PAPER

Modelling physico-chemical factors affecting occurrences of a non-indigenous planktonic copepod in northeast Pacific estuaries

Jeffery R. Cordell · Lucinda M. Tear · Stephen M. Bollens

Received: 8 October 2008 / Accepted: 28 July 2009 / Published online: 12 August 2009

Abstract Estuarine ecosystems along the Pacific coast of North America are vulnerable to invasions by non-indigenous planktonic copepods, with documented invasions by at least nine species introduced via ship's ballast. One of these, the calanoid copepod *Pseudodiaptomus inopinus*, now occurs in a relatively wide geographical area in coastal estuaries of Washington and Oregon States. Although it appears to be well established in the region, plankton surveys conducted in 1992, 1996, 2000, and 2004 in estuaries from southern Vancouver Island in British Columbia, Canada, to northern California, United States indicate that it has not expanded its range. This static distribution suggests that *P. inopinus* has reached a distributional limit, and it may thus be a good organism for applying models for predicting planktonic invasions, by characterizing estuaries with and without populations of the copepod. In this study, we applied both parametric, linear (discriminant function analysis, logistic regression) and nonparametric, nonlinear (classification trees) techniques to develop models for occurrence of *P. inopinus*, to identify parameters that may lead to successful invasions and to identify specific estuaries or regions that might be at risk of invasion by this species. Both model types had similar results, identifying relatively simple salinity- and stratification-based models as good predictors of *P. inopinus*. While different models selected slightly different sets of variables and thresholds, all models identified relatively low salinity and stratification of water column temperature and salinity as important predictors of *P. inopinus* presence. The models also identified several "false positives" that mainly occurred in more inland waters of Puget Sound—estuaries that did not have *P. inopinus*, but had the conditions that support it, and which may be at risk for future invasions by this species.

Keywords Invasive species · Zooplankton · Predictive modelling · Classification and regression trees · Discriminant function analysis

J. R. Cordell (✉) · L. M. Tear
School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA
e-mail: jcordell@u.washington.edu

S. M. Bollens
School of Earth and Environmental Sciences and School of Biological Sciences, Washington State University Vancouver, Vancouver, WA 98686-9600, USA

Introduction

Invasions of coastal and freshwater ecosystems by non-indigenous planktonic organisms can have large-scale ecological effects. The best known and most profound of these effects followed invasions by the comb jelly *Mnemiopsis leidyi*, which was introduced from the northwest Atlantic to the Black and Azov Seas in the

late 1980s, and subsequently to the Caspian and Baltic Seas. Appearance of this predator in the Black Sea was followed by dramatic decreases in, and in some cases disappearances of, its prey species (Kideys 2002). The resulting reduction in plankton prey resources was accompanied by large declines in pelagic fisheries, and recent research has suggested that similar fishery declines associated with *M. leidyi* may also follow in the Caspian and Baltic Seas (Finenko et al. 2007; Haslob et al. 2007). Other types of invasive zooplankton can also have large effects on invaded systems: for example, several regions in the world have seen invasions by freshwater planktonic cladocerans, with ecosystem-level consequences. *Cercopagis pengoi* is a pelagic predatory cladoceran native to the Ponto-Caspian region that has established itself in the Baltic Sea and the North American Great Lakes (Gorokhova et al. 2000; Therriault et al. 2002), and apparently caused population shifts in other plankton species and food web changes (Benoit et al. 2002; Ojaveer et al. 2004; Gorokhova and Lehtiniemi 2007; Lehtiniemi and Gorokhova 2008). Similarly, when *Bythotrephes longimanus*, another predatory cladoceran native to Eurasia, invaded North American lakes in the early 1980s, it changed native zooplankton communities by selective predation (Yan and Pawson 1997; Yan et al. 2002; Boudreau and Yan 2003) and caused indirect food web effects (Hovius et al. 2007).

Though less studied than freshwater systems, estuarine ecosystems are also known to be vulnerable to invasions by non-indigenous planktonic species. One of the most notable cases has occurred along the Pacific coast of the United States where coastal estuaries have been invaded by at least nine species of planktonic copepods native to Asia, all introduced via ship's ballast (Orsi and Ohtsuka 1999; Bollens et al. 2002). Most of these species appeared first in the San Francisco Estuary in California State and several of them have recently appeared in other estuaries in the region (Cordell et al. 2008a). Although impacts of these invaders on zooplankton communities are less well documented than for invasive freshwater cladocerans, in one case a laboratory and field study of the carnivorous calanoid *Tortanus dextrilobatus* indicated that it can consume remarkably high proportions of its copepod prey and could thus affect zooplankton population dynamics in the San Francisco Estuary (Hooff and Bollens 2004).

The most geographically widespread of the invasive estuarine planktonic copepods in the northeast Pacific, the calanoid *Pseudodiaptomus inopinus*, was found in the Columbia River estuary in 1990. It has since appeared in many other estuaries in Washington and Oregon states, where it can form a virtual monoculture during the late summer/early autumn (Cordell et al. 1992; Cordell and Morrison 1996; Bollens et al. 2002). *Pseudodiaptomus inopinus* is well established in the region, having remained abundant since the early 1990s: however, plankton surveys conducted in 1992, 1996, 2000, and 2004 in estuaries from southern Vancouver Island in British Columbia, Canada, to northern California, United States indicate that it has not expanded its range (Cordell and Morrison 1996; this study). This apparently static distribution suggests that *P. inopinus* has reached a distributional limit in invaded habitats, and it may thus be a good organism for applying models for predicting planktonic invasions, by characterizing estuaries that do not have populations of the copepod and comparing them to those that do.

Predicting invasions in aquatic habitats is particularly desirable, because once established, aquatic invasive species are nearly impossible to eradicate, and information that could prevent the introduction of new species and expansion of those already introduced is important. Both prevention and management can benefit from accurate predictions of potential distributions of invasive species, and a number of techniques have been applied for this purpose. Most of these fall into the category of ecological niche modelling, in which occurrence data on the distribution of a species in its native range is applied to the invaded region (reviewed in Guisan and Zimmerman 2000; Peterson 2003; Elith et al. 2006). Until recently, linear models such as logistic regression, multiple regression, and discriminant function analysis (DFA) have been most commonly used for modelling aquatic invasive species distributions. For example, Allen and Ramcharan (2001) used logistic regression to identify a set of environmental factors that restrict the establishment of invasive zebra mussels (*Dreissena polymorpha*) in commercial waterways of the United States. Distributions of zebra and quagga mussels (*Dreissena bugensis*) have also been predicted using physico-chemical variables in stepwise multiple regression models (Jones and Ricciardi 2005).

Similarly, Buchan and Padilla (2000) used logistic regression to predict the likelihood of invasion of lakes by Eurasian watermilfoil (*Myriophyllum spicatum*) using a suite of variables obtained from public databases and recommended particular models that could be used for different management needs and constraints. Logistic and multiple regression models have also been used to predict establishment, spread, and abundance of invasive freshwater fishes within North America and globally (Marchetti et al. 2004; Ruesink 2005). MacIsaac et al. (2004) used DFA and gravity models to backcast and forecast invasion of US lakes by the spiny water flea (*Bythotrephes longimanus*).

The success of multivariate linear approaches is sometimes limited if data do not meet model assumptions related to the independence of variables, distribution of input variables and model residuals, and linearity or additivity of effects (Olden and Jackson 2002). As a result, non-linear and nonparametric ecological niche models have seen increased use in predicting aquatic invasive species distributions. For example, potential distributions of zebra mussels, New Zealand mud snails, Chinese mitten crabs, and invasive fishes have been predicted using genetic algorithms (Drake and Bossenbroek 2004; Drake and Lodge 2006; Loo et al. 2007; Herborg et al. 2007a, b). Classification trees have also proven useful for relating species abundances to spatial and physical environmental variables (De'ath and Fabricius 2000) and were used by Mercado-Silva et al. (2006) to identify North American lakes vulnerable to invasion by rainbow smelt (*Osmerus mordax*) using physico-chemical lake attributes. Olden and Jackson (2002) compared logistic regression and DFA with neural networks and classification trees to predict the presence of 27 fish species as a function of habitat conditions in 286 temperate lakes in Ontario, Canada. They found that neural networks outperformed the other modelling approaches, although all approaches predicted species presence with moderate to excellent success.

In aquatic habitats, both linear and non-linear models have been used mainly for predicting potential distributions and invasion success of freshwater mussels and fishes (e.g., references cited above). A few studies have been conducted with invasive freshwater zooplankton taxa. The vector-based gravity model developed by MacIsaac et al. (2004) successfully backcasted the invasion sequence of inland lakes by *Bythotrephes longimana* based on relative strength of inbound vectors and forecasted new invasions of non-invaded lakes. Their analysis led the authors to propose management intervention through a public education campaign that targets the export of live individuals or resting stages from lakes with the greatest outbound vector flows and those most likely to serve as invasion hubs. Predictive models have also been developed for the invasive tropical cladoceran *Daphnia lumholtzi*, which has rapidly invaded lakes in North America. Based on canonical correspondence and logistic regression analyses, Havel et al. (2005) found that this species was strongly associated with large, eutrophic lakes. They suggested that large lakes may be more susceptible to invasion because they are more likely to have suitable habitats (e.g., larger pelagic zones, greater habitat heterogeneity), may receive higher propagule pressure (e.g., more boat traffic), and can support larger populations less prone to extinction.

For marine and estuarine invaders, there have been relatively few predictive modelling studies and we could find no such work conducted on zooplankton invaders. In this study, we applied both parametric, linear (DFA, logistic regression) and nonparametric, non-linear (classification trees) techniques to develop models for occurrence of the invasive planktonic calanoid copepod *Pseudodiaptomus inopinus* on the west coast of North America. The goals of the analyses were to identify parameters that may lead to successful establishment of *P. inopinus* in Pacific Northwest estuaries and to identify specific estuaries or regions that might be at risk of invasion by this species. Based on previous studies of the copepod in invaded estuaries (Cordell and Morrison 1996; Cordell et al. 2007) and field observations, we hypothesized that water temperature, salinity, and water-column stratification would be important factors governing the presence and absence of this species.

Methods

Sample collection

We sampled the estuaries of 30 rivers from the Campbell River, British Columbia to the Russian River, California (Table 1). Sampling took place in

Table 1 Locations, river discharge, drainage area, sampling years, and sample numbers for river estuaries sampled for zooplankton, N/A, data not available

State	River	Mean annual discharge (m^3 s^{-1})	Drainage area (km^{-3})	Latitude	Number of samples taken			
					1996	2000	2004	Total
California	Russian	66	3,728	38.45	5	5	5	15
	Noyo	66	442	39.43	5	5	5	15
	Eel	210	9,456	40.61	5			5
	Elk	2.4	72	40.75		5		5
	Eureka Slough	N/A	N/A	40.80		5		5
	Klamath	501	40,608	41.54	4	5	5	14
Oregon	Chetco	64	1,130	42.05	5	5	5	15
	Rogue	285	13,348	42.43	5	5	5	15
	Coquille*	51	4,414	43.12	5	5	5	15
	Coos*	33	1,608	43.38	5	5	5	15
	Umpqua*	211	12,133	43.71		5	5	10
	Siuslaw*	56	1,523	43.96	5			5
	Yaquina*	7	296	44.62	5	5	5	15
	Tillamook*	N/a	158	45.48	5	5	5	15
	Youngs*	5.1	65	46.16			4	4
Washington	Naselle*	12	88	46.43	5	5	5	15
	Willapa*	18	209	46.71	5			5
	Chehalis*	115	2,082	47.96			5	5
	Quileute	30	208	47.92		5		5
	Puyallup	94	1,526	47.25	5	5	5	15
	Duwamish	49	708	47.56	5	5	5	15
	Snohomish	269	2,473	47.99	5	5	5	15
	Dungeness	11	251	48.15	5			5
	Samish	6.9	141	48.56		5	5	10
	Skagit	468	4,978	48.36		5	5	10
British Columbia	San Juan	48.3	580	48.56	5			5
	Nanaimo	40	821	49.13	5	5	5	15
	Fraser	3954	234,000	49.12	5	5	5	15
	Squamish	300	2,330	49.69	5	5	5	15
	Campbell	99	1,755	50.05		5	5	10

An asterisk indicates the presence of *Pseudodiaptomus inopinus*

1996, 2000, and 2004 during September and October when *Pseudodiaptomus inopinus* are known to be abundant (Cordell and Morrison 1996; Cordell et al. 2007). We usually sampled five locations to span the salinity transition zone, at a downstream, relatively high salinity station (approximately 9–15 psu) and at salinities of approximately 6, 3, 1.5, and 0 psu. In rivers with short transition zones and sharp haloclines, we sampled at least three stations approximately evenly spaced between the upper and lower stations. At each station surface and bottom salinity and temperature were measured with a YSI Model 30 portable salinity meter. At each station, zooplankton were collected in one vertical haul with a 0.5 m diameter, 253 μm mesh plankton net. The net was lowered to the bottom, the depth was recorded, and after 30 s the net was pulled slowly to the surface. In the laboratory, zooplankton samples were re-filtered through a 153 μm screen and sub-sampled if necessary. For subsamples, the sample was thoroughly stirred with air bubbles and a smaller known volume removed with a Hensen's-stempel pipette, until the

total count for the most numerous species exceeded 100. *Pseudodiaptomus inopinus* were enumerated and abundances were calculated m^{-3}.

Graphic analyses

We visually explored the relationships among physical variables and between physical variables and abundance, presence, and order of magnitude of abundance of *Pseudodiaptomus inopinus* using graphical methods. We compared (1) boxplots of *P. inopinus* abundance vs. different types of categorical variables (e.g., year, presence/absence of shipping, latitude), (2) bivariate scatter plots of relationships among physical variables, (3) relationships between abundance of *P. inopinus* and continuous physical variables (e.g., salinity and temperature) with data points coded by one or two other variables of interest, and (4) sets of scatter plots or boxplots to identify patterns in time or space. Graphs were used to assess the strength and shape of bivariate and multivariate relationships and to identify variables for use in the predictive models.

Predictive models

We initially compared three multivariate techniques for identifying the variables that best explained the presence of *Pseudodiaptomus inopinus* and to determine if they identified similar variables.

Step-wise Discriminant Function Analysis (DFA) was employed to determine if a linear combination of physical variables could be found that would predict the presence or absence of *Pseudodiaptomus inopinus* at a site. DFA is part of a large class of generalized linear models and is most powerful when data meet test assumptions of multivariate normality and residuals having homogeneous variances. Within this class of models, DFA is used to predict categorical dependent variables using continuous and categorical independent variables. In this study, DFA was conducted in SPSS (Version 11.5). DFA models were run using data for individual stations, for annual river means, and for all-year river means.

Classification Trees (CT) (Breiman et al. 1984) were used to try to identify threshold values of one or more physical variables that correctly classify a site as having or not having *Pseudodiaptomus inopinus*. Classification Trees are "non parametric" in that they do not require that data meet any distributional requirements for results to be valid. Classification Tree analyses use iterative "splitting rules" to develop a decision tree of variables and associated splitting (threshold) values that result in the most accurate classification of samples. Different splitting algorithms can be used as well as different methods for weighting data, prioritizing the selection variables used in the model, and quantifying variance. The set of input variables entered into the step-wise DFA was also used in the CT analyses which were conducted using CART software (Version 6.0, Salford Systems). The optimal tree was selected using the 1 SE rule.

We also created models using step-wise logistic regression. Like DFA, logistic regression is a form of generalized linear model that predicts categorical dependent variables; logistic regression predicts binary responses only (DFA can predict more than two levels of response), uses a different error structure than DFA (binomial vs. normal), and involves a different computational algorithm. We tested a variety of combinations of input variables as well as forward- and backward-step-wise selection procedures, but none of the step-wise logistic regression models performed as well as the other methods. Many correlated variables were retained in the final models and classification accuracy was much lower than any of the DFA or CT models. However, when only the variables selected by the DFA were entered simultaneously, the resulting logistic regression model had slightly higher classification accuracy than the DFA model. Given these findings, we do not report results from the step-wise logistic regression results and instead discuss the final logistic regression model as a form of validation of the DFA. Logistic regressions were conducted in SPSS (Version 11.5).

Variables

Based on our exploratory graphical analyses, we selected a set of model input variables from field measurements and other readily available information (Table 2). These included presence/absence of international shipping, latitude of sampling location, maximum depth of sample, and surface and bottom salinity and temperature. Since it has been noted that *Pseudodiaptomus inopinus* occurs mostly in brackish regions of well-mixed estuaries (i.e., salinity is similar

Table 2 Variables used in step-wise discriminant function and classification and regression tree analyses

Variable	Definition
ID	
River	
Year	
Station	
Latitude	
Shipping present	Yes/no
Presence/absence of *P. inopinus*	0 (Not present), 1(present)
Station surface temperature (°C)	
Station bottom temperature (°C)	
Station surface salinity (psu)	
Station bottom salinity (psu)	
Station sample depth (m)	
River average surface temperature (°C)	Mean of 5 stations in a sampling year
River average bottom temperature (°C)	Mean of 5 stations in a sampling year
River average surface salinity (psu)	Mean of 5 stations in a sampling year
River average bottom salinity (psu)	Mean of 5 stations in a sampling year
River average sample depth (m)	Mean of 5 stations in a sampling year
Station salinity stratification index	(Surface salinity + 1)/(bottom salinity + 1)
Station temperature stratification index	(Surface temperature + 1)/(bottom temperature + 1)
River salinity stratification index (stations 1–4)	Average salinity stratification index for stations 1–4
River temperature stratification index (stations 1–4)	Average temperature stratification index for stations 1–4
River salinity stratification index (stations 1–5)	Average salinity stratification index for stations 1–5
River temperature stratification index (stations 1–5)	Average temperature stratification index for stations 1–5
Station % difference between surface and bottom temperature	(Surface temperature − bottom temperature)/surface temperature
Station % difference between surface and bottom salinity	(Surface salinity − bottom salinity)/surface salinity
River % difference between surface and bottom temperature	Mean of 5 stations in a sampling year
River % difference between surface and bottom salinity	Mean of 5 stations in a sampling year
Station surface − bottom temperature	Surface temperature − bottom temperature
Station surface − bottom salinity	Surface salinity − bottom salinity
Station salinity rate-of-change	(Surface salinity − bottom salinity)/station depth
Station temperature rate-of-change	(Surface temperature − bottom temperature/station depth
River average salinity rate-of-change	Mean of 5 stations in a sampling year
River average temperature rate-of-change	Mean of 5 stations in a sampling year

throughout the water column—Cordell and Morrison 1996), we created several temperature- and salinity-based variables to describe mixing or stratification of the water column. The stratification indices were similar, but not always linearly related, and therefore we initially used all of them in the models to determine if one was more informative than others (Table 2).

For each variable, we also calculated a river-wide mean for each sampling year (average of all 5 station values), and a river-wide mean across sampling years (average of all river-wide annual means—Table 2). For salinity and temperature stratification indices, we calculated an additional river-wide mean using only stations 1–4, because at station 5, the 0 psu station, there was always little or no difference between surface and bottom conditions.

Model classification accuracy

The classification accuracy of each model was assessed by creating a simple cross-tabulation (contingency table) of empirical and predicted presence-

absence classes. Each sample was classified in terms of its true, empirical class (*Pseudodiaptomus inopinus* present or *P. inopinus* absent, as rows) and its model-predicted class (as columns) and placed in one of four classes. The predictive accuracy of each model was assessed by computing the true and false positive and true and false negative rates.

Model sensitivity and validation

Assessing model sensitivity to choices made in creating the model is important in interpreting model outcomes. Selection and ordering of input variables, weights applied to data, the magnitude of the input variables relative to one another, computational algorithms used, and the way error is quantified can all affect model performance. Model validation helps to determine if a model performs better than chance in predicting outcomes and is general enough to make correct predictions from other, similar data sets. For example, Olden and Jackson (2002) demonstrated that the prediction accuracy of discriminant function and classification tree analyses were affected by the proportion of samples in each class and that it was possible to develop models that appeared to be accurate but that were performing no better than random assignment of samples to classes.

We conducted sensitivity analyses on the station-level models to assess the sensitivity of model outcomes to model inputs. We compared the final variables selected by each modelling procedure, the classification accuracy, and any coefficients or threshold values derived after varying the way that the model was run. For these sensitivity analyses, we varied (1) the order that input variables were added to DFA, CT, and logistic regression models, because some of the independent variables were correlated; (2) the choice of dependent variable, i.e., developing models that would predict both *Pseudodiaptomus inopinus* absolute abundance (a continuous dependent variable) and those that would predict order-of magnitude abundances (4-level ordinal response vs. 2-level presence/absence) for DFA and CT analyses; (3) the type of error structures, distance measures and model algorithms used, including use of Wilks and Mahalanobis distance measures to control the entry and removal of variables from the step-wise DFA model, comparing logistic regression and DFA, and changing splitting rules in CT analyses; (4) weighting factors used for predicting outcomes (equal and proportional weighting of prior probabilities were used for DFA and CT) and (5) weighting of costs associated with false negative predictions (increasing the relative costs of false negatives vs. false positives from 1:1 to 4:1 in classification trees).

Predictive accuracy of a model can be evaluated by applying the model to data not used to develop it. For all of our models, we divided the data set into "learn" and "test" subsets ("twofold cross validation"). The "learn" data used to develop or "train" the model were from rivers that had been sampled more than once, and the "test" data used to validate the model were from rivers that had been sampled only once. Models were developed using the "learn" data and tested on the "test" data, and classification accuracies for each data set were compared.

All DFA models were also validated using "leave-one-out" cross validation (also known as "*N*-fold cross validation"). Each case was classified into a group according to the classification functions computed from all the data except the case being classified and the classification accuracy of the model was assessed using these predictions.

Classification trees were also validated using "ten-fold cross-validation". In this type of validation, a model is first developed for the entire sample. The size and configuration of the optimal tree is then determined using error rates developed by model testing conducted on 10 subsets of the data. In the testing phase, the data are randomly divided into 10 subsets of approximately equal size and each subset is used as the test sample for models developed using the nine other data subsets.

To determine if the station-level discriminant function model performed better than chance, 1,000 iterations of the following procedure were conducted: presence-absence classes were randomly assigned to stations with $P = 0.3$ (the overall probability of *Pseudodiaptomus inopinus* presence in the dataset), a DFA was conducted and the false positive, false negative, true positive, true negative, and total correct classification rates were calculated. The classification rates for each iteration were sorted in ascending order of the total correct classification rate and the classification accuracy of the discriminant function model developed on the actual data was compared to the accuracy rates associated with the 5th and 95th percentiles of the total correct classification rate of the models developed with randomized responses.

Results

Graphical

Graphical analyses revealed a strong geographic pattern in the presence of *Pseudodiaptomus inopinus* (Fig. 1). It occurred only between latitudes of approximately 43° and 47.5° in Oregon and Washington States. Bivariate scatter plots showing relationships between latitude, salinity, and temperature indicated that this latitude range was associated with estuaries that had low average bottom salinities and were relatively un-stratified (i.e., high stratification index values) (Figs. 2, 3). Other results indicated that *P. inopinus* occurred (1) in almost all stations in a river if it was found at any station; (2) only at temperatures higher than 15°C (Fig. 4; Table 3); (3) mainly in the 0–8 psu salinity range (Fig. 4; Table 3); and (4) mainly in rivers that had relatively well-mixed water columns (i.e., salinity and temperature relatively similar between bottom and surface) (Fig. 4).

The graphical observations were corroborated by *t*-tests of differences between physical variables at stations with and without *Pseudodiaptomus inopinus*. These tests indicated that rivers with *P. inopinus* were warmer (average bottom temperature with *P. inopinus* 17.6°C; without *P. inopinus* 14.5°C, $P < 0.001$) and had lower bottom salinity (average salinity with *P. inopinus* 5.1 psu; without *P. inopinus* 14.9 psu, $P < 0.001$) and less stratification, especially when stratification was defined in terms of salinity (e. g., average percent difference between bottom and surface salinity with *P. inopinus* −0.20; without *P. inopinus* −0.63, $P < 0.001$).

Discriminant function analysis

The stepwise discriminant function model using station-level data with learn ($n = 279$) and test

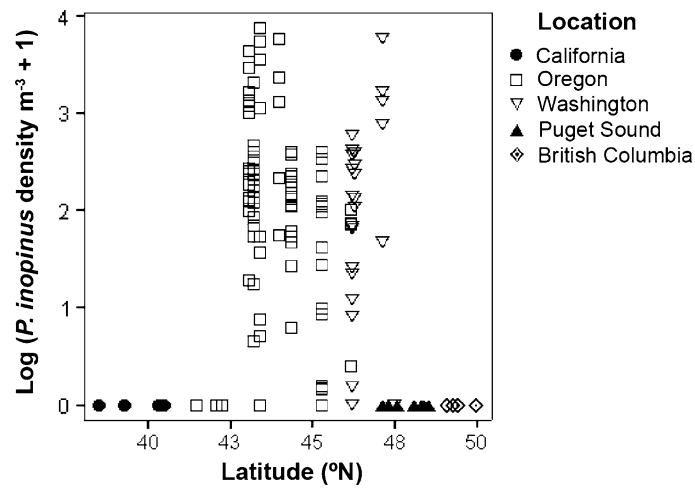

Fig. 1 Log of *Pseudodiaptomus inopinus* densities (numbers m^{-3}) in vertical plankton tows taken in 30 estuaries in 1996, 2000, and 2004

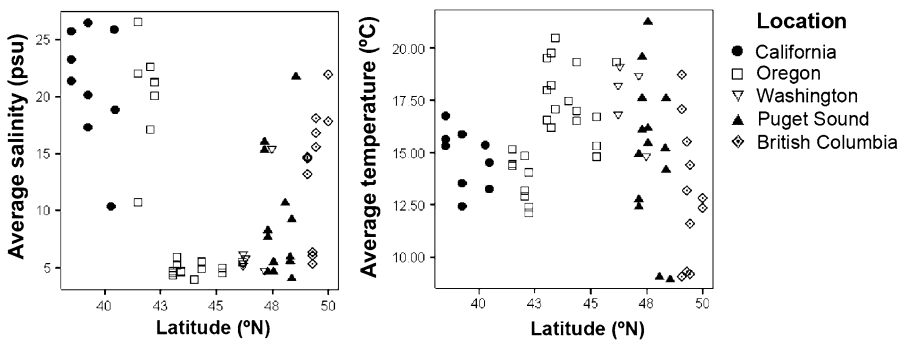

Fig. 2 Mean bottom salinities (psu) and bottom temperatures (°C) along a latitude gradient recorded concurrently with zooplankton samples collected from 30 estuaries in 1996, 2000, and 2004

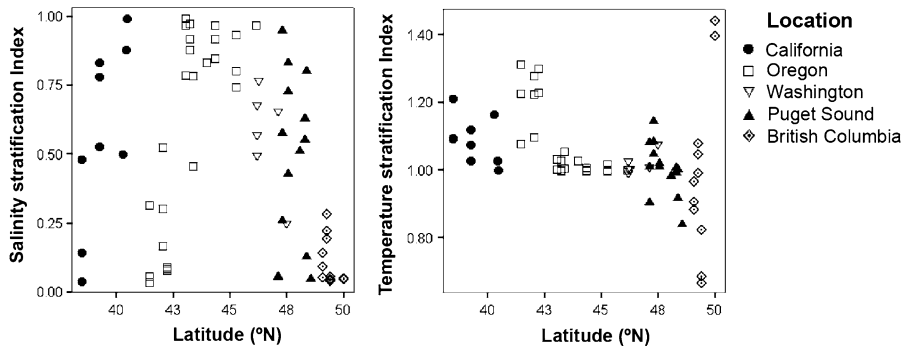

Fig. 3 Mean salinity and temperature stratification indices (based on stations 1–4) along a latitude gradient from zooplankton samples collected from 30 estuaries in 1996, 2000, and 2004

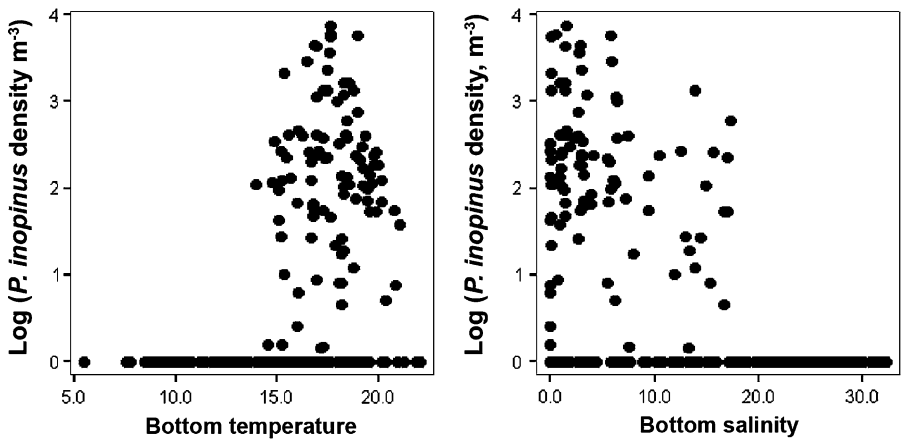

Fig. 4 Relationship between densities of *Pseudodiaptomus inopinus* (numbers $m^{-3}+1$) and bottom temperature (°C) and salinity (psu) from individual zooplankton samples in 30 estuaries in 1996, 2000, and 2004

Table 3 Percent of total *Pseudodiaptomus inopinus* occurring in each bottom temperature*salinity combination from zooplankton samples in 30 estuaries in 1996, 2000, and 2004

		Temperature (°C)									Total
		10	14	15	16	17	18	19	20	21	
Salinity (psu)	<3	0.0	**2.0**	**6.9**	**6.9**	**13.7**	**10.8**	**6.9**	**3.9**	**1.0**	52.0
	3–8	0.0	**2.0**	**1.0**	**4.9**	**3.9**	**6.9**	**5.9**	**2.0**	0.0	26.5
	8–20	0.0	0.0	**2.9**	**2.0**	**3.9**	**6.9**	**2.9**	0.0	0.0	18.6
	>20	**1.0**	**1.0**	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	2.9
Total		1.0	4.9	10.8	13.7	21.6	25.5	15.7	5.9	1.0	100.0

Bold type indicates combinations where *P. inopinus* were present

($n = 49$) validation selected average surface salinity and percent difference between surface and bottom salinity as the only two variables in the discriminant function (Table 4). For combined samples ($n = 328$, no validation), average surface salinity was selected again along with average bottom temperature and average salinity stratification index based on sampling stations 1–4.

Because (1) no station-level variables were selected in either model, (2) *Pseudodiaptomus inopinus* was usually present at all stations in rivers where they were present at any station, and (3) we were interested in predicting *P. inopinus* presence in rivers rather than at stations, we next conducted DFA first using annual river-wide means for physical explanatory variables (hereafter called river*year models) with learn

Table 4 Standardized discriminant function coefficients for variables selected by each DFA

	Stations Learn/test	Stations All cases	Stations All cases (Log.Reg)*	River*year Learn/test	River*year All cases	River All cases
River mean surf salinity (psu)	−0.683	−0.873	−0.536	−0.714	−0.869	
River % diff surf/bottom salinity	1.274			1.290		
River mean bottom temp (°C)		0.220	0.279			0.654
River salinity stratification index (stations 1–4)		1.255	9.423			
River salinity stratification index (stations 1–5)					1.340	
River mean bottom salinity (psu)						−0.561
R^2	0.731	0.751	0.715*	0.743	0.745	0.759

Asterisks indicate logistic regression run using variables selected in all station-level, all cases DFA model; R^2 = Nagelkerke's R^2

($n = 56$), test ($n = 10$), and total ($n = 66$) data sets, and then using river-wide, all-year means. The sample size for the river-wide, all-year data set was too small ($n = 30$) to allow conducting learn-test cross validation, so leave-one-out cross validation was used.

All of the final stepwise DFA models for station, river*year, and river-wide, all-year level data had r^2 values of >0.73 and included both a salinity-based stratification index and a measure of surface and/or bottom salinity (Table 4).

All DFA models were generally better able to predict the presence of *Pseudodiaptomus inopinus* (measured by the true positive rate or model sensitivity which ranged from 83 to 100%) than the absence of *P. inopinus* (measured by the true negative rate or model specificity which ranged from 80 to 83%) (Table 5). Both the true positive and the true negative rates of the DFA models were higher than the 95th percentile of those in randomization tests (73 and 75%, respectively). The 95th percentile of overall model reliability in the randomization tests was 64% (maximum 68%), below the 82–90% reliability of the DFA models. Out of 1000 randomizations of the data, only 403 valid models could be developed using a step-wise discriminant function method: randomly distributed data did not result in many valid models.

DFA misclassification errors

Errors from models that used station-level data were very similar to those that used river*year means. In most sampling years, *Pseudodiaptomus inopinus* was predicted to occur in five Puget Sound rivers—the Duwamish, Dungeness, Samish, Skagit, and Snohomish—where it did not occur (false positive errors). These rivers where *P. inopinus* has not yet been found have surface salinity and percent difference between surface and bottom salinity similar to rivers that currently support *P. inopinus* (Fig. 5). Using station-level data, false positive errors were also made for the Umpqua River (1/2 years), station 3 in the Tillamook River (1/3 years) and stations 2 and 4 in the Naselle River (1/3 years). However, all the single-station errors occurred because the station was the only or one of only a few stations in the river that did not have *P. inopinus*. These stations are likely indications of sampling limitations and that *P. inopinus* was present, though patchy or in low densities.

False negative errors (predicting no *Pseudodiaptomus inopinus* when it did occur) were made for the Umpqua and Naselle rivers, rivers that also had false positive errors. False negative errors disappeared when the river-wide, all-year mean data were used. However, in this case, some false positive errors persisted; the Duwamish, Samish, Skagit, and Snohomish rivers were still predicted to have *P. inopinus* though they did not, but the Dungeness River was correctly classified at this scale. In terms of stratification and mean bottom salinity, the all-year means for these rivers fell well within the range of those that currently have *P. inopinus* populations (Fig. 6).

DFA sensitivity analyses

None of our sensitivity analyses changed the variables selected for the best discriminant function. Model results were extremely robust to changes in the order of input of the variable, the distance measure, and prior weightings.

Table 5 Classification accuracy for each DFA model

Data set	Validation set	True presence/ absence *P. inopinus*	Predicted presence/ absence			True negative rate (specificity)	False negative rate	True positive rate (sensitivity)	False positive rate	Total correct (reliability)
			Absent	Present	Total					
Station*year data learn/ test subsets	Learn	Absent	161	38	199	0.81	0.16	0.84	0.19	0.82
		Present	13	67	80					
	Cross-validated	Absent	161	38	199	0.81	0.16	0.84	0.19	0.82
		Present	13	67	80					
	Test	Absent	25	5	30	0.83	0.00	1.00	0.17	0.90
		Present	0	19	19					
Station*year data all data	Learn	Absent	190	39	229	0.83	0.04	0.96	0.17	0.87
		Present	4	95	99					
	Cross-validated	Absent	190	39	229	0.83	0.04	0.96	0.17	0.87
		Present	4	95	99					
River*year means learn/ test subsets	Learn	Absent	32	7	39	0.82	0.12	0.88	0.18	0.84
		Present	2	15	17					
	Cross-validated	Absent	32	7	39	0.82	0.18	0.82	0.18	0.82
		Present	3	14	17					
	Test	Absent	5	1	6	0.83	0.00	1.00	0.17	0.90
		Present	0	4	4					
River*year means all cases	Original	Absent	37	8	45	0.82	0.10	0.90	0.18	0.85
		Present	2	19	21					
	Cross-Validated	Absent	37	8	45	0.82	0.10	0.90	0.18	0.85
		Present	2	19	21					
River, all year means all cases	Original	Absent	16	4	20	0.80	0.00	1.00	0.20	0.87
		Present	0	10	10					
	Cross-validated	Absent	16	4	20	0.80	0.00	1.00	0.20	0.87
		Present	0	10	10					
Station all cases, logistic regression	Original	Absent	206	23	229	0.90	0.13	0.87	0.10	0.89
		Present	13	86	99					
	Cross-validated	Absent	16	4	20	0.80	0.00	1.00	0.20	0.87
		Present	0	10	10					

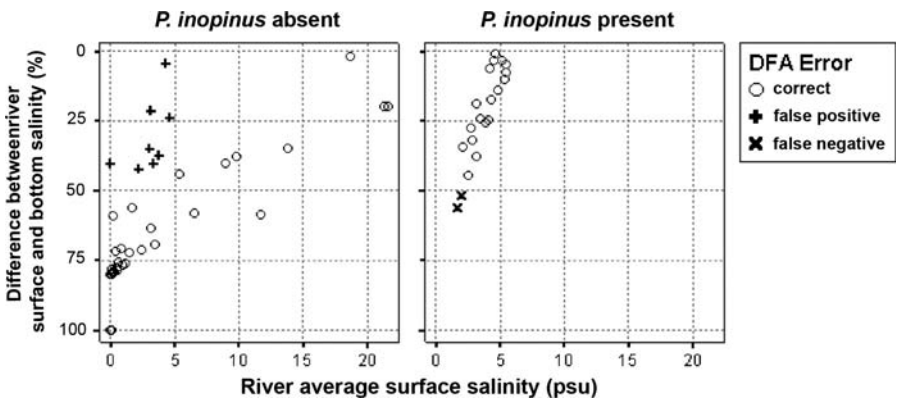

Fig. 5 Scatter plot of relationship between the two main variables selected by river*year mean step-wise discriminant function analysis to predict presence of *Pseudodiaptomus inopinus* for stations with and without *P. inopinus*. See Table 2 for definitions of variables. Data coded by error type in the model

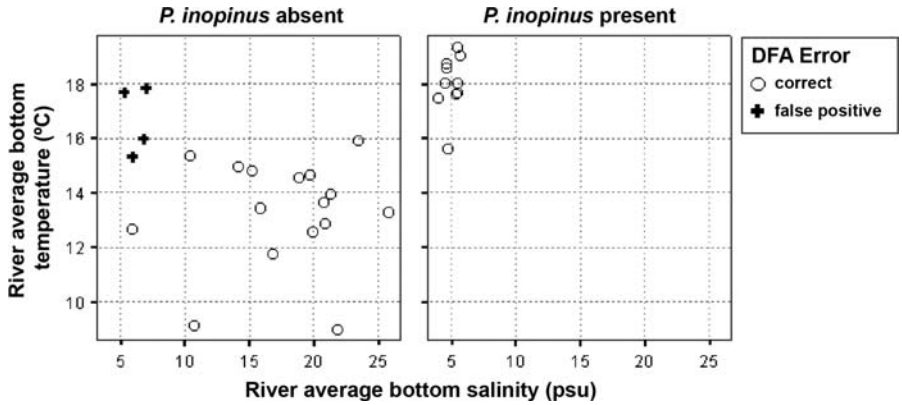

Fig. 6 Scatter plot of relationship between the two main variables selected by river-mean, step-wise discriminant function analysis procedure to predict *P. inopinus* presence in rivers. See Table 2 for definitions of variables. Data coded by error type in the model

Although the logistic regression step-wise model selection process did not produce useful models (many correlated variables were retained even when some culling of correlated variables was conducted and classification errors were higher than DFA), when explanatory variables pre-selected by the DFA were entered simultaneously, the logistic regression model performed very similarly to the DFA. It had a similar r^2 (Table 2), fewer false positives, and more false negatives (Table 7).

Classification trees

Using station-level data, CT analysis identified clear cut-off values that separated sites with and without *Pseudodiaptomus inopinus*. The most efficient model (optimum tree) was a two-node model using only river bottom salinity, but a three-node model that was only marginally less efficient (Fig. 7) increased the true negative rate (model specificity) and the total percent correct by 8 and 5%, respectively (Table 6) and identified the ranges of two variables that describe the domain of *Pseudodiaptomus inopinus* occurrences. The model developed using station-level learn/test subsets identified river mean bottom salinity and river mean salinity stratification index (using sampling stations 1–4) as critical splitting variables. All occurrences of *P. inopinus* were found at stations with river bottom salinity less than 6.1 psu and mean stratification index > 0.4.

Model sensitivity (true positive rate) of the three-node model was 100% in both the learn and test samples for station-level and river*year models,

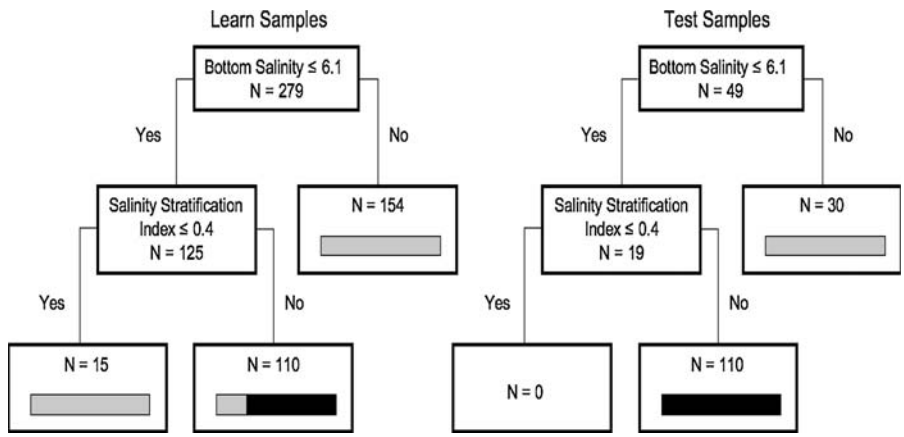

Fig. 7 Classification trees developed using learn and test subsets of station level data. The optimum model consists of the first two levels of the trees; the selected models are the entire trees. Gray pattern represents proportion of samples without *Pseudodiaptomus inopinus*, black pattern represents proportion of samples with *P. inopinus*. Salinity stratification index is based on sampling stations 1–4 (see text and Table 2)

Table 6 Prediction Errors by selected CT models

Data set	Validation set	True presence/ absence *P. inopinus*	Predicted presence			True negative rate (specificity)	False negative rate	True positive rate (sensitivity)	False positive rate	Total correct (reliability)
			Absent	Present	Total					
Station level optimum model—two nodes	Learn	0	154	45	199	0.77	0.00	1.00	0.23	0.84
		1	0	80	80					
	Test	0	30	0	30	1.00	0.00	1.00	0.00	1.00
		1	0	19	19					
Station level preferred model—three nodes	Learn	0	169	30	199	0.85	0.00	1.00	0.15	0.89
		1	0	80	80					
	Test	0	30	0	30	1.00	0.00	1.00	0.00	1.00
		1	0	19	19					
River*year level preferred model—three nodes	Learn	0	33	6	39	0.85	0.00	1.00	0.15	0.89
		1	0	17	17					
	Test	0	6	0	6	1.00	0.00	1.00	0.00	1.00
		1	0	4	4					
Station level priors proportional to learn set—optimum model	Learn	0	174	25	199	0.87	0.13	0.88	0.13	0.87
		1	10	70	80					
	Test	0	30	0	30	1.00	0.26	0.74	0.00	0.90
		1	5	14	19					
Station level priors proportional to learn set—selected model	Learn	0	189	10	199	0.95	0.13	0.88	0.05	0.93
		1	10	70	80					
	Test	0	30	0	30	1.00	0.53	0.47	0.00	0.80
		1	10	9	19					
Station level linear combinations allowed	Learn	0	178	21	199	0.89	0.00	1.00	0.11	0.92
		1	0	80	80					
	Test	0	30	0	30	1.00	0.00	1.00	0.00	1.00
		1	0	19	19					

respectively (Table 6). Although *Pseudodiaptomus inopinus* was not found in all rivers having the range of conditions identified by the CT, all rivers that had *P. inopinus* had those conditions (Table 6, Figs. 7, 8). River*year and river-wide, all-year models were not considered necessary given the accuracy rates of the station-level model.

Results of the CT analysis can be displayed in a bivariate plot of river mean bottom salinity and river stratification index (Fig. 8). In such a figure, the upper left quadrant (quadrant 1) represents salinity and stratification conditions that support *Pseudodiaptomus inopinus*, while the conditions represented in quadrants 2, 3, and 4 do not. Rivers without *P. inopinus* that fall in the upper left quadrant of the plot are false positives in the CT model. Rivers with *P. inopinus* that fell into quadrants 2, 3, or 4 are false negatives. No false negatives errors were made with the CT model (Fig. 8; Table 6).

Five of the false positives from the DFA were also false positives in the CT model and appear to have conditions that could support *Pseudodiaptomus inopinus* (Fig. 8, "A"). Two DFA false positives were correctly predicted by the CT model (Duwamish 2004; Dungeness 1996) and one DFA false positive was also a false positive in the CT model (Samish 2004) (Fig. 8, "B"). One of the CT false positives was correctly predicted by the DFA (Dungeness 1996) (Fig. 8, "C"). The two cases of false negatives from the DFA river*year model (Umpqua 2000; Naselle 1996) were correctly predicted by the CT model as having *P. inopinus* although they had a stratification index close to the threshold identified by the CT as supporting *P. inopinus* (Fig. 8, "D", right panel, upper left quadrant).

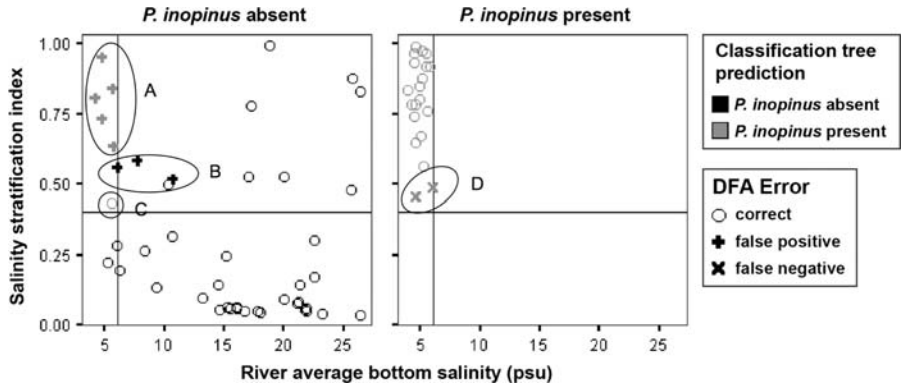

Fig. 8 Relationship between river mean salinity stratification index (based on stations 1–4) and river mean bottom salinity for rivers with and without *Pseudodiaptomus inopinus* showing predictions by classification tree (CT) and discriminant function analysis (DFA) river*year models using learn/test subsets. The upper left quadrant of both panels includes the combination of salinity and stratification conditions in which *P. inopinus* was found and the other three quadrants include combinations where it was not found. Gray symbols in the upper left quadrant of the "*P. inopinus* absent" panel (*A* and *C*) were falsely predicted by the CT model to have *P. inopinus* and cross symbols in any quadrant of that panel (*A* and *B*) were falsely predicted to have *P. inopinus* by the DFA model. In the "*P. inopinus* present" graph, all points in the upper left quadrant were true positives by the CT model (no false negative predictions) and all but two were true positives by the DFA model (*D*)

CT Model validation

Tests conducted to determine the sensitivity of the final CART model to assumptions of the different models used indicated that the final model was very robust to modelling assumptions and that our findings are not an artifact of the modelling algorithms used. For example, changing the algorithms used to split the data or the costs associated with making different types of errors (as described in methods) did not change the optimal model selected. Changing the probabilities used to assign sites to classes or allowing for linear combinations of splitting variables had minor effects on the selection of variables selected at the second split level, but did not improve the total percent correct (Table 6) or alter the general results that areas with low salinity and low stratification support *Pseudodiaptomus inopinus.* Use of tenfold cross classification as a model validation approach resulted in more complex models than the simple three- to five-node models developed using learn-and-test data sets (twofold cross classification), but these models appeared too specific to the data set used to be useful in evaluating other data sets.

The usefulness of the final splitting variables that were selected can be evaluated for each node in the model relative to other potential splitting variables. The software program used provides a summary of the improvement in model prediction accuracy for each splitting variable relative to other potential splitters. At the first node, mean river bottom salinity (6.1 psu), provided an additional 3% improvement in model prediction accuracy compared to the next strongest potential splitter, river salinity rate-of-change from surface to bottom (Table 7) and as much as 6% improvement in prediction accuracy compared to other stratification variables. At the second node, the stratification index that used all 5 stations provided the same percent improvement in model predictions as the 4-station stratification variable that was selected. These stratification variables provided an additional 1.5–2% improvement over the next strongest competitors at the second node, all of which were simple temperature or salinity variables.

Discussion

The fact that the invasive estuarine copepod *Pseudodiaptomus inopinus* has not expanded either north or south from its region of occurrence in Washington and Oregon during 14 years of monitoring indicates that it has likely reached a distributional limit. The results from our models demonstrate that in the northeast Pacific, distribution of *P. inopinus* is strongly related to salinity, temperature, and water

Table 7 Node competitors from CT models in order of improvement showing variable, split value, and percent improvement (0–1) in model predictions

Competitor	Split	Improvement
Root node competitors		
River average bottom salinity (psu)	6.12	0.316
River average salinity rate-of-change	−1.545	0.283
River % difference between surface and bottom temperature	−0.335	0.261
River salinity stratification index (stations 1–4)	0.445	0.254
River salinity stratification index (stations 1–5)	0.56	0.254
River average bottom temperature (°C)	16.19	0.213
River average surface salinity (psu)	1.56	0.201
Node 2 competitors		
River salinity stratification index (stations 1–5)	0.56	0.053
Station bottom temperature (°C)	13.75	0.039
River surface temperature (°C)	14.49	0.035
River bottom temperature (°C)	14.02	0.035
River surface salinity (psu)	1.0	0.035

column stratification. Observations from the graphical analyses, that *P. inopinus* occurs within defined salinity and temperature ranges, were also reflected in the consistency of the results in DFA and CT analyses that created predictive models with salinity and temperature terms. In particular, most occurrences of *P. inopinus* were at stations with bottom salinities less than 6 psu, bottom temperatures greater than 15°C (Fig. 4), and relatively well-mixed water columns (Fig. 8). The various models that we applied identified slightly different sets of variables and slightly different thresholds, but all identified areas with relatively low bottom salinity and relatively unstratified water columns as areas that support *P. inopinus* populations.

These findings correspond to other recent studies that found that *P. inopinus* occurred mainly between salinities of 0 and 12 psu, with highest densities usually occurring around 0–6 psu in relatively short reaches within estuaries (Cordell and Morrison 1996; Cordell et al. 2007). The results from these studies, plus our model results and field observations, lead us to the conclusion that there are two main estuary types among those sampled, one that is suitable for colonization by *P. inopinus* and one that is not. More stratified estuaries with warmer fresh water overlying colder salt water and little mixing of the water column do not have populations of *P. inopinus*. This type of estuary occurred mainly at the northern and southern ends of the region sampled (Fig. 3), and included both small rivers without shipping and large rivers such as the Fraser River in British Columbia with intensive international shipping and presumed high inoculation rates of non-indigenous planktonic organisms (Levings et al. 2004). Estuaries that have been invaded by *P. inopinus* have regions where there is less stratification, and brackish or oligohaline water occurs throughout the water column. These types of estuaries occurred mostly in coastal Oregon and Washington, but also occurred in a few other areas that have not been invaded by *P. inopinus* (Fig. 3). Invaded estuaries included relatively large and small systems (in terms of both drainage area and discharge—Table 1) and those with and without shipping ports. From previous field observations, we noted that invaded estuaries also have low gradients and relatively long regions of tidal excursion (Cordell and Morrison 1996), which we assume contribute to slower net outflow and more opportunity for mixing of salt and fresh water. Thus, regional geomorphology may be an important factor determining the presence of *P. inopinus*: higher gradient rivers that occur mainly in British Columbia and northern California do not have the salinity, temperature, and stratification conditions to support *P. inopinus* while lower gradient rivers in coastal Oregon and Washington do.

Estuaries in Puget Sound that were false positives in both DFA and CT models have the combination of stratification and salinity that support *Pseudodiaptomus inopinus* elsewhere and may therefore be at risk for invasion. The reason that this species does not occur in Puget Sound estuaries is probably not from lack of propagule pressure. The ports of Everett and Seattle are located in estuaries (Snohomish, Duwamish) predicted to have *P. inopinus*, and receive ballast water from a variety of ship types and sources (Cordell et al. 2008b). In fact, recent studies of ballast in ships arriving in Puget Sound have found high

frequencies of occurrence of non-indigenous copepods from both international and previously invaded domestic ballast sources, including *P. inopinus* and several other *Pseudodiaptomus* species (Cordell et al. 2008a, b). One possible explanation for the Puget Sound false positives is that the models lacked one or more variables that determine *P. inopinus* presence. Examples of unmeasured variables could include (1) inoculation pressure: although our use of the shipping/no shipping term was a surrogate for this, estuary-specific inoculation pressure could be important, especially if it was mediated by factors other than shipping (e.g., proximity to and long shore transport from other estuaries that have established populations of *P. inopinus*); (2) range of variability in physical conditions: magnitude and periodicity of river flooding, yearly temperature regimes, and length of time mixed brackish-water conditions persist could affect the ability of *P. inopinus* to establish itself; and (3) biotic factors: our models did not include data on other planktonic organisms, and the presence of competitors or predators could have discouraged invasion by *P. inopinus*. For example, recent mesocosm studies suggest that native zooplankton communities can provide biotic resistance against the establishment of the invasive freshwater cladoceran *Daphnia lumholtzi* (Dzialowski et al. 2007). It may be useful to monitor additional physical and biological variables in rivers with and without *P. inopinus* to determine additional attributes that are associated with new invasions, or with any changes in the status of existing *P. inopinus* populations.

While CT models provided good results in this study, and are known to perform better than traditional linear models with environmental data (Rejwan et al. 1999; Olden and Jackson 2002), in our case both DFA and CT models performed similarly. Both model types had similar accuracy rates and incorrectly classified many of the same rivers. Both types also selected river means over station-level physical data as explanatory variables, and included a salinity variable and some combination of surface and bottom salinity or temperature variables that provided a proxy for stratification. Although logistic regression was not helpful for the step-wise selection of variables, the fact that its results were similar to a DFA using the same variables validates the DFA results and indicates that results were not sensitive to error structure or computational algorithm within the realm of generalized linear models. In addition, error rates for station-level DFA models were much lower than the fifth percentile error rates from randomization tests and indicate that the DFA models performed better than chance. The robust results and the strong performance of the DFA relative to chance indicate that conclusions about the distribution of *Pseudodiaptomus inopinus* are not highly sensitive to model type.

With CT, we found that the use of learn/test subsets of data tended to produce simpler models with apparently greater generality (and management applicability) than models generated using tenfold cross validation. Our station-level data set had 328 samples, around the minimum sample size recommended for learn/test validation. The fact that this validation approach produced straightforward results with a marginal sample size also indicates that the results are robust. In CT, using tenfold cross validation, the most efficient model was one with many levels. As would be expected, with more levels, model sensitivity increased, but the number of levels provided by the most efficient model may not be warranted for practical use by resource managers. Instead, the combination of our DFA and CT results suggests that relatively simple salinity- and stratification-based models may work well for predicting invasion of *Pseudodiaptomus inopinus*. While different models select slightly different sets of variables and thresholds, all of them identify relatively low salinity and low stratification as important in predicting the presence of *P. inopinus.*

DFA and CT have been used in several studies to identify the conditions that support freshwater invasive species and they appear to have worked well in this study with an estuarine zooplankton species. CT is sometimes identified as a preferred approach because it requires fewer assumptions, can identify "nonlinear" effects that do not occur equally across the entire data set, and provides approachable, easy to interpret results. In our study CT had slightly lower error rates and was more useful at identifying thresholds, as compared to DFA. However, we suggest that using more than one model to explore a data set is important. Consistent results across model types increases the confidence in the results and conclusions. When model results differ, there is an opportunity to explore how input variables, assumptions, and other factors affect the outcome of different models. This can be particularly helpful

with biological data when it may be tempting to avoid or overlook complex, indirect interactions when a simple conclusion seems to emerge from a data set.

In predictive modelling for invasive species occurrences, it is desirable to obtain data from the native range. Data taken only from the invaded range may underestimate potential distributions because the invader may not have occupied the full range of conditions that it can tolerate. In this study, we used data only from the invaded range, because zooplankton, temperature, salinity, and water column stratification data are not available from published literature about the native habitats of *Pseudodiaptomus inopinus*: we conducted a thorough literature search and found only a few publications that generally related *P. inopinus* abundances to brackish waters (Suh et al. 1991; Ueda et al. 2004; Islam et al. 2006). However, a case can also be made for using data from the invaded range, because the factors that limit the growth and spread of species in the native region may differ from those in the invaded region, i.e., it occupies different realized niches within its fundamental niche in each region (Drake and Bossenbroek 2004; Loo et al. 2007; Steiner et al. 2008). In the case of *P. inopinus*, it appears for the time being to have reached the limits of its physico-chemical tolerances along the west coast of North America, because its distribution has remained static since 1992, and thus this case may represent a good scenario for predictive modelling using invaded range data.

In this paper we describe the results of linear and non-linear models to describe physico-chemical attributes that coincide with and predict the presence of an invasive estuarine planktonic copepod. The models provide testable predictions about the occurrence of *Pseudodiaptomus inopinus* in certain uninvaded northeast Pacific estuaries, and these kinds of models could also be applied to other regions that have potentially similar conditions but no invasive estuarine zooplankton species. For example, Chesapeake Bay has large estuarine ports and high input of non-indigenous propagules from the ballast of commercial ships, but has not experienced invasions by *Pseudodiaptomus* or other non-indigenous planktonic taxa (Ruiz et al. 2000). Predictive models could identify specific areas to sample for these species, and could help to define those factors contributing to the presence or absence of non-indigenous taxa. Also, because the models (especially those derived from CT) identify thresholds, they may serve as useful predictors for shifting environmental factors. It has been recently demonstrated that, like many other organisms, zooplankton are affected by climate change (e.g., southern species and assemblages moving into northern regions—Beaugrand et al. 2002; Panov et al. 2007; Sullivan et al. 2007). A warming climate is also projected to increase salinity in estuaries and coastal areas, with accompanying changes in the likelihood of non-native species becoming established (Rahel and Olden 2008). This may be particularly important in the case of brackish-water organisms such as *P. inopinus*, which may eventually be able to invade estuaries that currently have salinities too low to allow establishment. In addition to climate change, estuarine organisms are subjected to numerous environmental perturbations related to human activities, including physical changes in hydraulic processes (e.g., dams, channel and harbor widening and deepening), nutrient profiles, and disruptions in land-water linkages (e.g., shoreline armoring, loss of natural uplands), as well as biological changes (e.g., other species introductions). The extent to which any or some of these factors contribute to establishment of non-indigenous copepods or other exotic organisms is largely unexplored, but in theory many combinations of factors could be used to predict their presence. In this study, the data and models suggest that vulnerable uninvaded estuaries (i.e., false positives in the models that occurred mainly in the Puget Sound region), may only have to experience small changes in temperature, salinity, or stratification to become invaded.

Acknowledgments Many students and staff from the Wetland Ecosystems Team, University of Washington helped with sample collections. O. Kalata analyzed the majority of the zooplankton samples. J. Olden provided a thoughtful and comprehensive review that improved the manuscript. This research was sponsored in part by Washington State Sea Grant.

References

Allen YC, Ramcharan CW (2001) *Dreissena* distribution in commercial waterways of the US: using failed invasions to identify limiting factors. Can J Fish Aquat Sci 58:898–907

Beaugrand G, Reid PC, Ibañez F, Lindley AJ, Edwards M (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. Science 296:1692–1694

Benoit HP, Johannsson OE, Warner DM, Sprules WG, Rudstam LG (2002) Assessing the impact of a recent predatory invader: the population dynamics, vertical distribution, and potential prey of Cercopagis pengoi in Lake Ontario. Limnol Oceanogr 47:626–635

Bollens SM, Cordell JR, Avent S, Hooff R (2002) Occurrences, causes and consequences of zooplankton invasions: a brief review, plus two case studies from the northeast Pacific Ocean. Hydrobiologia 480:87–110

Boudreau SA, Yan ND (2003) The differing crustacean zooplankton communities of Canadian Shield lakes with and without the nonindigenous zooplanktivore *Bythotrephes longimanus*. Can J Fish Aquat Sci 60:1307–1313

Breiman L, Friedman JH, Olshen A, Stone CG (1984) Classification and regression trees. Wadsworth International Group, Belmont

Buchan LAJ, Padilla DK (2000) Predicting the likelihood of Eurasian Watermilfoil presence in lakes, a macrophyte monitoring tool. Ecol Appl 10:1442–1455

Cordell JR, Morrison SM (1996) The invasive Asian copepod *Pseudodiaptomus inopinus* in Oregon, Washington, and British Columbia estuaries. Estuaries 19:629–638

Cordell JR, Morgan CA, Simenstad CA (1992) Occurrence of the Asian calanoid copepod *Pseudodiaptomus inopinus* in the zooplankton of the Columbia River estuary. J Crust Biol 12:260–269

Cordell JR, Rassmussen M, Bollens SM (2007) Biology of the invasive copepod *Pseudodiaptomus inopinus* in a northeast Pacific estuary. Mar Ecol Progr Ser 333:213–227

Cordell JR, Bollens SM, Draheim R, Sytsma M (2008a) Asian copepods on the move: recent invasions in the Columbia–Snake River system, USA. ICES J Mar Sci 65:000

Cordell JR, Lawrence DA, Ferm NC, Tear LM, Smith SS, Herwig RP (2008b) Factors influencing densities of nonindigenous species in the ballast water of ships arriving at ports in Puget Sound, Washington, United States. Aquat Conserv: Mar Freshw Ecosyst 19:322–343

De'ath G, Fabricius KE (2000) Classification trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192

Drake JM, Bossenbroek JM (2004) The potential distribution of zebra mussels in the United States. Bioscience 54:931–941

Drake JM, Lodge DM (2006) Forecasting potential distributions of nonindigenous species with a genetic algorithm. Fisheries 31:9–16

Dzialowski AR, Lennon JT, Smith VH (2007) Food web structure provides biotic resistance against plankton invasion attempts. Biol Invasions 9:257–267

Elith J et al (2006) Novel methods improve prediction of species distributions from occurrence data. Ecography 29:129–151

Finenko GA, Kideys AE, Anninsky BE, Shiganova TA, Roohi A, Tabari MR, Rostami H, Bagheri S (2007) Invasive ctenophore *Mnemiopsis leidyi* in the Caspian Sea: feeding, respiration, reproduction and predatory impact on the zooplankton community. Mar Ecol Prog Ser 314:171–185

Gorokhova E, Lehtiniemi M (2007) A combined approach to understand trophic interactions between *Cercopagis pengoi* (Cladocera: Onychopoda) and mysids in the Gulf of Finland. Limnol Oceanogr 52:685–695

Gorokhova E, Aladin N, Dumont H (2000) Further range extensions of *Cercopagis* genus (Crustacea, Cladocera) with notes on taxonomic composition and ecology. Hydrobiologia 429:207–218

Guisan A, Zimmerman NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Haslob H, Clemmesen C, Schaber M, Hinrichsen H, Schmidt JO, Voss R, Kraus G, Köster FW (2007) Invading *Mnemiopsis leidyi* as a potential threat to Baltic fish. Mar Ecol Prog Ser 349:303–306

Havel JE, Shurin JB, Jones JR (2005) Environmental limits to a rapidly spreading exotic cladoceran. Ecoscience 2:376–385

Herborg L-M, Jerde CL, Lodge DM, Ruiz GM, MacIsaac HJ (2007a) Predicting invasion risk using measures of introduction effort and environmental niche models. Ecol Appl 17:663–674

Herborg L-M, Rudnick DA, Siliang Y, Lodge DM, MacIsaac HJ (2007b) Predicting the range of Chinese mitten crabs in Europe. Conserv Biol 21:1316–1323

Hooff RC, Bollens SM (2004) Functional response and potential predatory impact of *Tortanus dextrilobatus*, a carnivorous copepod recently introduced to the San Francisco Estuary. Mar Ecol Prog Ser 277:167–179

Hovius JT, Beisner BE, McCann KS, Yan ND (2007) Indirect food web effects of *Bythotrephes* invasion: responses by the rotifer *Conochilus* in Harp Lake, Canada. Biol Invasions 9:233–243

Islam MS, Ueda H, Tanaka M (2006) Spatial and seasonal variations in copepod communities related to turbidity maximum along the Chikugo estuarine gradient in the upper Ariake Bay, Japan. Est Coast Shelf Sci 68:113–126

Jones LA, Ricciardi A (2005) Influence of physicochemical factors on the distribution and biomass of invasive mussels (*Dreissena polymorpha* and *Dreissena bugensis*) in the St. Lawrence River. Can J Fish Aquat Sci 62:1953–1962

Kideys AE (2002) Fall and rise of the Black Sea ecosystem. Science 297:1482–1484

Lehtiniemi M, Gorokhova E (2008) Predation of the introduced cladoceran *Cercopagis pengoi* on the native copepod *Eurytemora affinis* in the northern Baltic Sea. Mar Ecol Prog Ser 362:193–200

Levings CD, Cordell JR, Ong S, Piercey GE (2004) The origin and identity of invertebrate organisms being transported to Canada's Pacific coast by ballast water. Can J Fish Aquat Sci 61:1–11

Loo SE, MacNally R, Lake PS (2007) Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges. Ecol Appl 17:181–189

MacIsaac HJ, Borbely JVM, Muirhead JR, Graniero PA (2004) Backasting and forecasting biological invasions of inland lakes. Ecol Appl 14:773–783

Marchetti MP, Moyle PB, Levine R (2004) Invasive species profiling? Exploring the characteristics of non-native fishes across invasion stages in California. Freshw Biol 49:646–661

Mercado-Silva N, Olden JD, Maxted JT, Hrabik TR, Vander Zanden MJ (2006) Forecasting the spread of invasive rainbow smelt in the Laurentian Great Lakes region of North America. Conserv Biol 20:1740–1749

Ojaveer H, Simm M, Lankov A (2004) Population dynamics and ecological impact of the non indigenous *Cercopagis pengoi* in the Gulf of Riga (Baltic Sea). Hydrobiologia 522:261–269

Olden JD, Jackson DA (2002) A comparison of statistical approaches for modelling fish species distributions. Freshw Biol 47:1976–1995

Orsi JJ, Ohtsuka S (1999) Introduction of the Asian copepods *Acartiella sinensis*, *Tortanus dextrilobatus* (Copepoda: Calanoida), and *Limnoithona tetraspina* (Copepoda: Cyclopoida) to the San Francisco Estuary, California, USA. Plankton Biol Ecol 46:128–131

Panov E, Rodionova V, Bolshagin PV, Bychek EA (2007) Invasion biology of Ponto-Caspian cladoceran cladocerans (Crustacea: Cladocera). Hydrobiologia 590:3–14

Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modelling. Quart Rev Biol 78:419–433

Rahel FJ, Olden JD (2008) Assessing the effects of climate change on aquatic invasive species. Conserv Biol 22:521–533

Rejwan C, Collins NC, Brunner J, Shuter BJ, Ridgway MS (1999) Tree regression analysis on the nesting habitat of smallmouth bass. Ecology 80:341–348

Ruesink JL (2005) Global analysis of factors affecting the outcome of freshwater fish introductions. Conserv Biol 19:1883–1893

Ruiz GM, Fofonoff PW, Carlton JT, Wonham MJ, Hines AH (2000) Invasion of coastal marine communities in North America: apparent patterns, processes, and biases. Ann Rev Ecol and Syst 31:481–531

Steiner FM, Schlick-Steiner BC, VanDerWal J, Reuther KD, Christian E, Stauffer C, Suarez AV, Williams SE, Crozier RH (2008) Combined modelling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. Divers Distr 14:538–545

Suh H-L, Soh HY, Cha SS (1991) Salinity and distribution of zooplankton in the estuarine system of Mankyong river and Dongjin river. J Oceanol Soc Korea 3:181–192

Sullivan BK, Costello JH, Van Keuren D (2007) Seasonality of the copepods *Acartia hudsonica* and *Acartia tonsa* in Narragansett Bay, RI, USA during a period of climate change. Est Coast Shelf Sci 73:259–267

Therriault TW, Grigorovich IA, Kane DD, Haas EM, Culver DA, MacIsaac HJ (2002) Range expansion of the exotic zooplankter *Cercopagis pengoi* (Ostroumov) into western Lake Erie and Muskegon Lake. J Gt Lakes Res 28:698–701

Ueda H, Terao A, Tanaka M, Hibino M, Islam MS (2004) How can river-estuarine planktonic copepods survive river floods? Ecol Res 19:625–632

Yan ND, Pawson TW (1997) Changes in the crustacean zooplankton community of Harp L. Canada, following invasion by *Bythotrephes cederstroemi*. Freshw Biol 37:409–425

Yan ND, Girard R, Boudreau S (2002) An introduced invertebrate predator (Bythotrephes) reduces zooplankton species richness. Ecol Lett 5:481–485