Predicting the distributions of freshwater fish species for all New Zealand's rivers and streams

NIWA Client Report: HAM2008-005
January 2008

NIWA Project: DOC06212

Predicting the distributions of freshwater fish species for all New Zealand's rivers and streams

John Leathwick
Kathryn Julian
Jane Elith
Dave Rowe

NIWA contact/Corresponding author

John Leathwick

Prepared for

Department of Conservation under the Terrestrial and Freshwater Biodiversity Information System (TFBIS) programme

NIWA Client Report: HAM2008-005
January 2008

NIWA Project: DOC06212

National Institute of Water & Atmospheric Research Ltd
Gate 10, Silverdale Road, Hamilton
P O Box 11115, Hamilton, New Zealand
Phone +64-7-856 7026, Fax +64-7-856 0151
www.niwa.co.nz

Contents

Reviewed by:

E Graynoth

Eric Graynoth

Approved for release by:

Elliott

Sandy Elliott

Formatting checked

A. Bartley

1. Introduction

Knowledge of the distributions of species and communities provides a crucial platform for many aspects of conservation management. Superficial consideration would suggest that reasonable knowledge is available for the distributions of some components of New Zealand's freshwater biota. For example, records of fish capture for more than 20,000 sampling events are held in the New Zealand Freshwater Fish Database (NZFFD). However, despite this relatively large number of records, less than 5% of our rivers and streams have been sampled in this manner. The resulting lack of information on the biological communities occurring in the majority of rivers and streams often results in difficult management decision making, requiring either collection of new samples, or estimation of the expected composition of un-sampled streams based on knowledge from sampled sites that have similar physical characteristics.

Here we describe the results of a numerical approach to solving this difficulty. Working with distributional data for the 30 most commonly occurring species in the Freshwater Fish Database, we first use statistical models to describe the average relationship between the probability of capture of each species and environmental conditions at the fishing sites. We then use these statistical models to predict probabilities of capture for each species at all rivers and streams throughout New Zealand, based on the physical environment of discrete stream segments as recorded in a GIS database. Used in this way, our statistical models can be seen as performing two important functions. First, they smooth out the noise that is inherent in most ecological data, providing a robust description of the *average* relationship between species occurrence and the environment. Second, they allow the environment-based *interpolation* of these average probabilities of capture to the large numbers of rivers and streams that are not physically sampled.

This publication provides a working manual for these predictions, which are available as a spatial database built around the river network developed originally as the River Environment Classification (Snelder & Biggs 2002). In terms of scope, these predictions are derived only for rivers and streams – samples from static water bodies such as lakes, wetlands and ponds were specifically excluded, as were those from waters with tidal influence, in which the major method used for sampling fresh-water fish (electric fishing) is ineffective. Finally, no consideration is given of the distributions of introduced fish species in this publication, although the methods described here could be used for analysis of their distributions. Other publications based on these analyses include an exploration of the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species (Leathwick et al. in press), and an exploration of the use of these predictive databases for guiding river restoration (Leathwick et al. in review).

2. Material and methods

The overall approach used in this project was to fit statistical models relating the distributions of 30 fish species, 15 diadromous and 15 non-diadromous species (listed in Tables 1 & 2) to a set of environmental variables chosen for their functional relevance to fish (see Leathwick et al. 2005, Leathwick et al. in press). These statistical models were then combined with equivalent environmental data for all New Zealand rivers and streams to predict the likely probability of capture for each individual fish species.

2.1 Data

2.1.1 Fish distributions

Fish distribution data were drawn from the NZFFD, a database that records the occurrence of freshwater fish from approximately 22,500 sites throughout New Zealand (McDowall & Richardson 1983, http://www.niwa.co.nz/services/free/nzffd). A subset of 13,369 records were selected for this analysis by including sites sampled after January 1980 and for which all species were identified (Figure 1). All sites from tidal rivers, lakes and other still waters were excluded, as were sites sampled using methods that limited selection to some species or life stages (e.g., whitebait nets, plankton nets or diving). Repeated samples from the same site on separate visits were given a reduced weight so that the total weights for all sites were equal. Most sites selected for analysis were sampled using electric fishing (76%), but nets of various construction (7%), traps (6%), and spotlighting (5%) were also used at some sites. Two or more techniques were combined at 6% of sites. Distributional data were extracted for all selected sites for 15 diadromous and 15 non-diadromous species, all of which occurred at 30 or more sites. All records were converted to presence-absence form for this analysis due to difficulties in correcting for the differing catch rates of the different capture methods and/or variation in the area fished.

Users should be aware of the presence of biases in the distributional data used, as this can affect model predictions. In particular, most NZFFD sites represent samples from smaller rivers and streams, and are biased away from larger rivers and/or those with saline influence, as noted above. This reflects the difficulty in sampling large or saline waters using electric fishing, the dominant sampling method used. As a consequence, our predictions of the distributions of species occurring in larger and/or saline waters will be less reliable than for the many smaller rivers and streams that predominate over most of New Zealand.

2.1.2 Environmental predictors

Environmental variables used as predictors in the statistical models were derived from a GIS database representing New Zealand's rivers and streams as a network topology, with each river or stream section between adjacent confluences represented by a unique segment. Individual environmental attributes were available for each of these segments, and each sampling site, although sampling only a small section of river (a reach) was associated with the segment in which it was located. A set of 23 environmental variables were chosen for their relevance to freshwater organisms for use as predictors in the statistical models. These variables were calculated at four spatial scales to describe: local variation in reaches sampled during fishing, the river segments in which fishing occurred, the catchments upstream of fished segments, and access between the fished segments and the coast (Leathwick et al. in press).

Table 1: Six letter codes, scientific and common names of the fifteen diadromous species included in the analysis, along with their number of presences. Taxonomic authorities follow McDowall (1990).

Code	Common name	Scientific name	Presences
Angaus	Shortfin eel	*Anguilla australis*	2670
Angdie	Longfin eel	*Anguilla dieffenbachii*	6650
Chefos	Torrentfish	*Cheimarrichthys fosteri*	1250
Galarg	Giant kokopu	*Galaxias argenteus*	387
Galbre	Koaro	*Galaxias brevipinnis*	1453
Galfas	Banded kokopu	*Galaxias fasciatus*	1649
Galmac	Inanga	*Galaxias maculates*	1372
Galpos	Shortjaw kokopu	*Galaxias postvectis*	348
Geoaus	Lamprey	*Geotria australis*	325
Gobcot	Common bully	*Gobiomorphus cotidianus*	2182
Gobgob	Giant bully	*Gobiomorphus gobioides*	159
Gobhub	Bluegill bully	*Gobiomorphus hubbsi*	661
Gobhut	Redfin bully	*Gobiomorphus huttoni*	2313
Retret	Common smelt	*Retropinna retropinna*	509
Rhoret	Black flounder	*Rhombosolea retiaria*	78

Figure 1: Distribution of fish sample sites used in this analysis. Only rivers with mean annual flows greater than 10 m^3sec^{-1} are shown.

Table 2: Six letter codes, scientific and common names of fifteen non-diadromous species included in the analysis, along with their number of presences. Taxonomic authorities follow McDowall (1990) for all species except *Galaxias* 'species D' and *Galaxias* 'species N', which are undescribed entities formerly included in *Galaxias vulgaris* (McDowall 2006).

Code	Common name	Scientific name	Presences
Galano	Roundhead galaxias	*Galaxias anomalus*	63
Galcob	Lowland longjaw galaxias	*Galaxias cobitinis*	34
Galdep	Taieri flathead galaxias	*Galaxias depressiceps*	170
Galdiv	Dwarf galaxias	*Galaxias divergens*	348
Galeld	Eldon's galaxias	*Galaxias eldoni*	59
Galgol	Gollum galaxias	*Galaxias gollumoides*	78
Galmar	Big-nose galaxias	*Galaxias macronasus*	30
Galpau	Alpine galaxias	*Galaxias paucispondylus*	167
Galpro	Longjaw galaxias	*Galaxias prognathus*	62
Galpul	Dusky galaxias	*Galaxias pullus*	51
Galspd	Clutha flathead galaxias	*Galaxias 'species D'*	110
Galspn	Northern flathead galaxias	*Galaxias. 'species N'*	83
Galvul	Canterbury galaxias	*Galaxias vulgaris*	624
Gobbas	Cran's bully	*Gobiomorphus basalis*	723
Gobbre	Upland bully	*Gobiomorphus breviceps*	1760

Two variables were used to describe effects of local (reach-scale) habitat variation on fish occurrence. *ReachHabitat* describes the weighted average of the relative proportion of seven habitat types (e.g., pool, run, riffle, etc.) and *ReachSubstrate* describes the weighted average of the relative cover on the stream bed of different sized sediments (see Table 3 and Leathwick et al. 2005 for details).

Nine variables were used to describe variation in the environmental character at the segment scale (Table 3). River flow and its seasonal variability were described using two variables derived from hydrological models – *SegLowFlow* describes the 7-day average mean annual low flow (Pearson 1995), with a 4th root transformation giving values that are approximately linearly related to water velocity (Jowett 1998). *SegFlowStability* describes the ratio of the mean annual low flow to the mean average flow, with high values indicating minimal within-year variation in flow and low values indicating marked variation in flow. Segment slope (*SegSlope*) was used as an indicator of local, topographically driven variations in river velocity, derived from GIS calculations for each segment using its length and the difference in elevation between its upstream and downstream ends. A square root transformation was used to compensate for varying rates of change in habitat conditions in relation to slope, i.e., on steep slopes, a given change in slope results in smaller change in habitat conditions than on sites with low slope. Average daily air temperature in summer (January –

SegSumT) and winter (July – *SegTSeas*) were used to describe seasonal variation in river temperatures. As the summer and winter temperatures are highly correlated, winter temperature estimates were transformed so as to describe the deviation in winter temperature from that expected given the summer temperature. The resulting variable describes the seasonal temperature range, with negative values indicating continental or seasonal climates, while positive values indicate maritime climates. The amount of riparian shading (*SegShade*) was estimated from a stream shade model that used as inputs the stream size and the riparian vegetation cover (Leathwick et al. 2005) with the latter estimated from a national satellite image-based vegetation classification. Estimates of stream nitrogen load (*SegN*) were obtained from CLUES, a nutrient leaching model combined with a regionally-based regression model, implemented within a catchment framework (Woods et al. 2006).

Five variables were used to describe downstream conditions likely to influence biological composition, and in particular the distributions of diadromous fish species (Table 3). The first of these (*DSDist)* described the downstream distance from the mid-point of each river segment to the coast. The average downstream slope (*DSAvgSlope*) was calculated from the elevation at the mid-point of the segment and the distance to the coast. *DSMaxSlope* records the maximum value encountered when calculating local slopes at 100 m intervals between the mid-point of each river segment and the coast. *DSDam* was used to indicate river segments in which known downstream obstructions such as dams, culverts or waterfalls were likely to impede fish passage. *DSDist2Lake* was calculated for segments occurring upstream of a lake located on any of the river networks, allowing for the testing of the importance of lakes as breeding habitat for fish. The variable describes the distance from the mid-point of each segment to the nearest downstream lake.

Aspects of the river environment that are affected by conditions in the upstream catchment were described using ten variables (Table 4). Estimates of the average temperature in the upstream catchment were highly correlated with *SegSumT*, so a derived variable (*USAvgT*) was calculated that describes the degree to which temperatures in the upstream catchment depart from those expected given the segment summer air temperature. Negative values indicate river and stream segments for which the upstream catchments have colder temperatures (higher elevations) than expected, in turn, implying steeper upstream gradients and an abundance of energy for sediment transport. Positive values indicate rivers and streams with warmer, lower elevation catchments than expected, resulting in lower elevation catchments, less energy for sediment transport and greater deposition of fine sediments. *USDaysRain* describes the frequency of days of significant rainfall (<25 mm) in the upstream catchment, in turn indicating the likely frequency of elevated flows. The upstream average slope (*USAvgSlope*) and the proportion of land with native vegetation cover (*USNative*) describe factors that are known for their ability to influence the stability of river flows. For example, steep catchments generally show more rapid changes in flow in response to rainfall than catchments with gentle slopes. Similarly, high cover of native vegetation is likely to be positively correlated with stable river flows, while catchments in which such cover has been removed will show more marked variation in river flow. Variation in geological substrates, which affects both flow variability and water chemistry, was quantified using estimates of rock hardness (*USHardness*) and the availability in surface rocks of phosphorus (*USPhosphorus*) and calcium (*USCalcium*) (Leathwick et al. 2005). Two variables were used to describe the local buffering of river flows in the upstream catchment by lakes (*USLake*) and wetlands (*USPeat*). *USGlacier* was used to indicate the proportional cover of glaciers in the upstream catchment, these altering both water temperature and the nature of sediments, in turn having strong effects on biological communities (Castella et al. 2001).

Three variables were used to describe features of the fishing methods (Table 4). *Method* identifies between the four major methods of capture, i.e., electric fishing, nets, spotlights and traps, with a fifth level used to identify fishing using some combination of these methods. *TSin* and *TCos* were used to assess the effects of variation in the time of year at which fishing occurred, following the method described by Flury & Levri (1999).

Table 3: Segment-scale, reach-scale and downstream predictors used to predict the distributions of freshwater fish species.

Segment-scale predictors	**Mean and range**
SegLowFlow – segment mean annual 7-day low flow (m^3/sec), fourth root transformed, i.e., $(\text{low flow} + 1)^{0.25}$	1.092, 1.0 to 4.09
SegFlowStability – annual low flow/annual mean flow (ratio)	0.18, 0 to 0.58
SegSumT – summer air temperature (°C)	16.3, 8.9 to 19.8
SegTSeas – winter air temperature (°C), normalised with respect to *SegSumT*, i.e., $SegTempSeas = \left(\left(\frac{W-\overline{W}}{\sigma_w}\right)-\left(\frac{S-\overline{S}}{\sigma_s}\right)\right)*\sigma_w$ where W is the winter temperature for a segment, $\overline{W}$ is the average winter temperature for all segments, σ_w is the standard deviation of winter temperature, S is the summer temperature, and so on.	0.36, -4.2 to 4.1
SegShade – riparian shade (proportion)	0.41, 0 to 0.8
SegSlope – segment slope (°), square-root transformed (slope+1)	1.59, 1 to 5.5
SegN – log10 of nitrogen concentration (ppm)	-0.19, -2.04 to 1.88
Reach-scale predictors	
ReachSubstrate – weighted average of proportional cover of bed sediment using categories of 1 – mud, 2 – sand; 3 – fine gravel; 4 – coarse gravel; 5 – cobble; 6 – boulder; 7 – bedrock.	3.77, 1 to 7, 2347 missing values
ReachHabitat – weighted average of proportional cover of local habitat using categories of 1– still; 2 – backwater; 3 – pool; 4 – run; 5 – riffle; 6 – rapid; 7 – cascade.	4.04, 1 to 7, 2284 missing values
Downstream predictors	
DSDist – distance to coast (km)	73.8, 0.01 to 432.8
DSAvgSlope – average slope (°), square-root transformed (slope+1)	1.19, 1 to 7.24
DSMaxSlope – maximum downstream slope (°)	7.9, 0 to 40.8
DSDam – presence of known downstream obstruction, mostly dams	absent – 10,978; present – 2391.
DSDist2Lake – distance to a lake (km), where no lake present, set to 500, i.e., greater than the maximum river length.	28.6, 0.04 to 129 (for lakes present)

Table 4: Upstream/catchment-scale environmental and methods predictors used to predict freshwater fish distributions.

Upstream/catchment-scale predictors	
USAvgT – average air temperature (°C), normalised with respect to *SegJanT*	-0.38, -7.7 to 2.2
USRainDays – days/month with rainfall greater than 25 mm	18.0, 1.2 to 103.4
USSlope – average slope in the catchment (°)	14.3, 0 to 41.0
USNative – area with indigenous vegetation (proportion)	0.57, 0 to 1
USPhosphorus – average phosphorous concentration of underlying rocks, 1 = very low to 5 = very high	2.44, 1 to 5
USCalcium – average calcium concentration of underlying rocks, 1 = very low to 4 = very high	1.48, 1 to 4
USHardness – average hardness of underlying rocks, 1 = very low to 5 = very high	3.11, 1 to 5
USPeat – area of peat in catchment (proportion)	0.006, 0 to 1
USLake – area of lake in catchment (proportion)	0.002, 0 to 1
USGlacier – area of glacier in catchment (proportion)	0.003, 0 to 0.53
Fishing method	
Method – fishing method in five broad classes	electric – 10,155 net – 986 spot – 634 trap – 763 mixture – 831
TCos, TSin – month of year fitted as sine and cosine transforms of the Julian day for the middle of the month in which fishing occurred	*TCos* - 0.26, -1 to 1 *TSin* - 0.18, -1 to 1

2.2 Statistical modelling

Relationships between fish occurrence and environment were analyzed using boosted regression trees (BRT), an advanced type of regression model. Only relevant analysis settings are described below – more detailed descriptions are contained in Elith et al. (in press) and Leathwick et al. (in press). All analyses were carried out in R (version 2.0.1 R Development Core Team 2004) using the ‘gbm’ library of Ridgeway (2004), supplemented with purpose-written functions (Elith et al. in press). All models were fitted to allow interactions, using a tree complexity of 5 and with the learning rate generally set to 0.01 – a learning rate of 0.005 was used for three species of very low prevalence (Leathwick et al. in press). Ten-fold cross validation was used to determine the optimal number of trees for each model, i.e., that which gives maximum predictive performance when predicting to new sites. Because of the tendency of BRT models to over-fit the training data, the performance of all models was assessed on predictions to sites that were with-held during cross-validation. Two values were calculated for each model: the predictive deviance and the discrimination as measured by the area under the receiver operator characteristic curve (AUC) (Hanley & McNeil 1982). Values for the predictive deviance provides a measure of the goodness of fit between predicted and raw values when predicting to independent data, and were expressed as a percentage of the null deviance for each species. Values for AUC give a measure of the degree to which fitted values discriminate between observed presences and absences; values can be interpreted as indicating the probability that a presence for species drawn at random from the data will have a higher fitted probability than an absence drawn at random. A value of 0.5 indicates that a model predicts presences and absences no better than tossing a coin, while a value of 1.0 indicates that presences and absences are perfectly distinguished. Values greater than 0.8 can be interpreted as indicating excellent prediction.

The relative importance of the individual predictors was determined using a script in the 'gdm' library that sums, by predictor, reductions in error across all individual regression tree rules. The influences of interactions between predictor variables was evaluated using purpose-written functions as described in Elith et al. (in press). Environmental optima for each species were determined by plotting the distributions of fitted values in relation to each of the predictors.

Finally, we predicted probability of capture across New Zealand's entire river network by combining the regression models with nation-wide environmental predictors identical to those used for model fitting, storing the results in GIS files. For the patchily distributed non-diadromous species we produced two separate predictions, one showing predicted probabilities of capture throughout New Zealand, and the second showing predicted probabilities of capture only for those catchments or sub-catchments in which the species has been recorded as present.

2.2.1 Defining a classification of diadromous species composition

For diadromous species, we also defined a community classification using predictions for all 15 species from those rivers and streams in which the predicted diadromous species richness per fishing event was greater than 0.5, i.e., at least one species could be expected to be encountered in 50% or more of samples. Predicted probabilities of capture were clustered using the Bray-Curtis similarity measure and the ALOC and flexible UPGMA routines contained in the multivariate statistical software PATN (Belbin 1995). Results were imported back into ArcView, where a 10-group level of classification was identified as providing a useful summary of diadromous fish distribution patterns.

Although we also included non-diadromous species in an additional trial classification, results were problematic due to their much more patchy and generally non-overlapping distributions, and are not shown here.

3. Results

3.1 Predictive performance of statistical models

Statistical models for all thirty species showed very high levels of predictive performance (Table 5) as indicated by statistics calculated during cross-validation, with AUC scores averaging just under 0.95. On average, both the amount of deviance explained and AUC statistics were higher for non-diadromous species than for diadromous species, probably reflecting the manner in which the lower prevalence of the non-diadromous species made the predictions of their large areas of non-occurrence relatively straight forward. However the average AUC for the diadromous species is still very high, indicating that the models show a high degree of accuracy in discriminating between river and stream segments where these species are present, compared to those in which they are absent.

Variables describing the accessibility of river and stream segments (*DSAvgSlope*, *DSDist*, *SegSumT*, *DSMaxSlope*) made the largest contributions to model outcomes for diadromous fish species, with most species caught most frequently in coastal locations (Table 6). Although *SegSumT* might not initially be identified as an access-related variable, it is strongly correlated with elevation, which plays an important role in determining accessibility to migrating juvenile fish, particularly in coastal locations. Predictors describing regional scale climate made the two largest contributions to model outcomes for the non-diadromous species, with most of these species caught most frequently at sites with strong seasonal variation in temperature (continental climates) and a low frequency of days with intense rainfall. *DSDist* and *DSMaxSlope* were also important for these species, and most of them were caught most frequently in inland locations with large downstream obstacles. Contributions of the individual predictors for each of the fish species are shown in Appendix 1, and provide important clues as to the factors playing major roles in determining the distributions of species – comments are included where appropriate in the individual species descriptions. Further discussion of the biogeography and ecology of these species, including a more extended interpretation of the statistical models is contained in Leathwick et al. (in press).

Table 5: Comparison of the predictive performance of statistical models relating species distributions to environment, averaged across (a) diadromous and, (b) non-diadromous species. Table values indicate the mean of the predicted deviance and their standard errors, the mean percentage of the total deviance explained, the mean of the AUC scores and their standard errors and the correlation between raw and fitted values and their standard error. Estimates of predictive performance were calculated using 10-fold cross validation.

	(a) Diadromous species	**(b) Non-diadromous species**
Null deviance	0.500	0.150
Predictive deviance (se)	0.312 (0.005)	0.068 (0.003)
% deviance explained	36.5%	58.2%
AUC (se)	0.916 (0.005)	0.983 (0.003)

Table 6: Average percentage contributions of the twelve most important predictors for BRT models relating the distributions of (a) diadromous and, (b) non-diadromous species for environment.

a) Diadromous species		**(b) Non-diadromous species**	
Predictor	**Contribution (%)**	**Predictor**	**Contribution (%)**
DSAvgSlope	11.9	*SegTSeas*	9.8
DSDist	10.6	*USRainDays*	8.5
SegSumT	7.7	*DSDist*	8.2
DSMaxSlope	7.2	*DSMaxSlope*	6.8
USRainDays	6.5	*SegSumT*	6.1
Method	5.1	*SegFlowStability*	5.3
ReachSubstrate	4.6	*USCalcium*	4.8
SegTSeas	3.9	*USNative*	4.6
SegFlowStability	3.7	*DSAvgSlope*	4.4
SegLowFlow	3.6	*USPhosphorus*	4.2
SegShade	3.5	*SegN*	4.0
SegN	3.4	*SegShade*	3.9

3.1.1 Diadromous community groups

At a ten group level of classification there is strong separation between diadromous species assemblages found in rivers and streams in coastal or moderately coastal locations (Groups A–F – upper part of Figure 2 & 3 and Tables 7 & 8) and those found in more inland waters (Groups G–J). Groups A–C, which occur mostly in coastal hill-country streams, are dominated generally by banded kokopu and longfin eels. Longfin eels are also the most frequently caught species in Groups D–F, which occur in more inland locations than the preceding groups, and generally in low gradient rivers and streams. Groups G–I all have relatively simple composition, each being defined by consistent capture of only one species, shortfin eels in Group G, longfin eels in H, koaro in I, and common bullies in J. Of these four groups, Group G occurs in moderately inland, low gradient streams, H occurs in inland hill-country streams, and I in steep montane streams. Group J is associated with inland lakes, mostly in the central North Island.

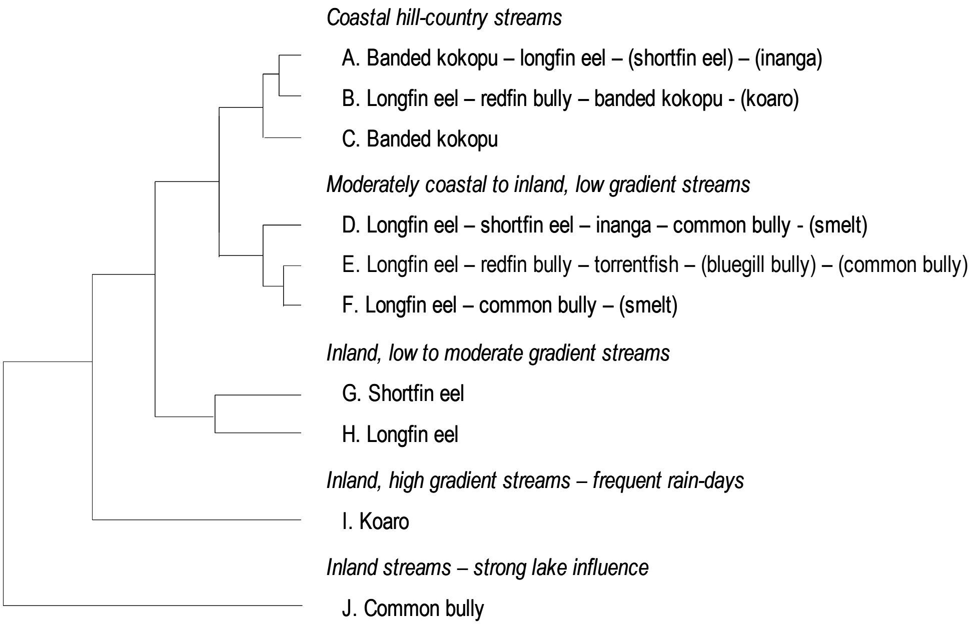

Figure 2: Dendrogram describing similarities between groups defined by numerical classification of predicted probabilities of capture for fifteen diadromous fish species in New Zealand rivers and streams.

Group A: (20,229 km) Banded kokopu (0.78 probability of capture) and longfin eels (0.68) occur most frequently (Table 7), with shortfin eels (0.48) and inanga (0.36) also present. Common bully, redfin bully, smelt, koaro and giant bully are caught only occasionally. This group occurs in coastal locations in warm (18.1 ^{o}C), maritime climates with low frequencies of rain days. It occurs mostly in very small streams (0.012 cumecs) with unstable flows, a predominance of gravelly substrates and high nitrogen concentrations. On average, stream segment and downstream slopes are moderate, and catchment slopes are gentle. Riparian shading is high but cover of native vegetation in the upstream catchment is low. It occurs extensively across the upper North Island, particularly in Northland and Auckland, and along the north and west coast of the South Island.

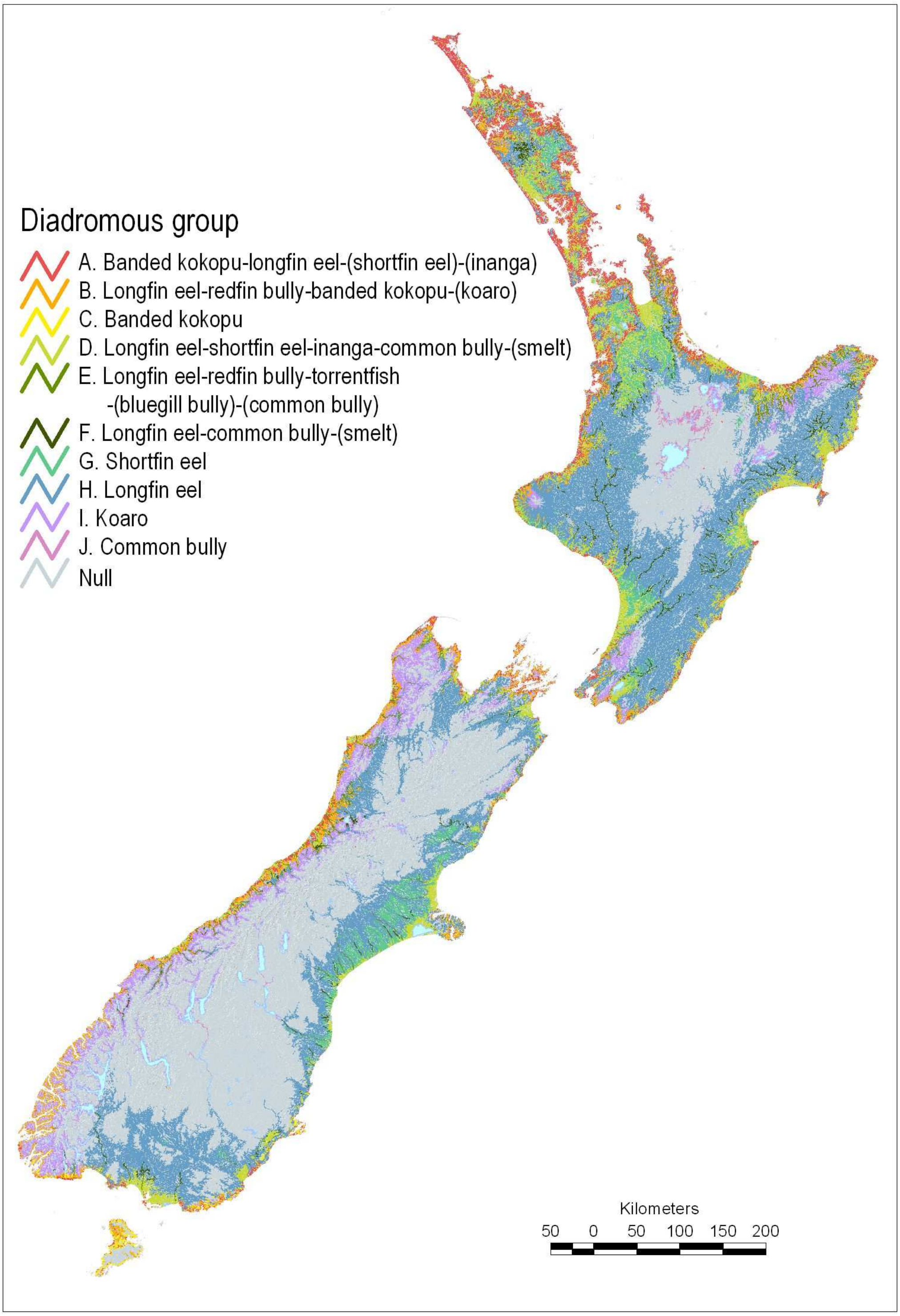

Figure 3: Geographic distribution of 10 groups defined by classification of predicted diadromous fish composition for New Zealand rivers and streams.

Table 7: Species composition of 12 diadromous fish assemblages identified by multivariate classification of predictions for all New Zealand rivers and streams. Numerical values indicate the predicted probability of capture, and are shown only where the average predicted occurrence exceeds a "presence" threshold calculated during cross-validation to give a level of "presence" in the predictions equivalent to the occurrence of the species in the training data. Predictions of occasional capture of species (range: 0.2–threshold) is indicated by a "+". Predictions for two species, Geoaus and Rhoret, failed to reach an average of 0.2 at this level of classification detail and are omitted.

Species Group	Angaus	Galmac	Gobcot	Galfas	Gobhut	Angdie	Chefos	Retret	Galarg	Galbre	Galpos	Gobhub	Gobgob
A	0.48	0.36	+	0.78	+	0.68	-	+	+	+	-	-	+
B	+	+	+	0.51	0.54	0.68	+	-	+	0.44	+	+	-
C	-	+	-	0.67	+	+	-	-	+	+	-	-	-
D	0.69	0.55	0.51	+	+	0.73	+	0.33	+	-	-	-	-
E	+	+	0.34	+	0.64	0.83	0.62	+	-	+	+	0.36	-
F	+	+	0.53	+	+	0.85	+	0.33	-	+	-	-	-
G	0.62	+	+	+	-	+	-	+	-	-	-	-	-
H	+	-	+	+	+	0.72	-	-	-	+	-	-	-
I	-	-	-	+	+	+	-	-	-	0.78	+	-	-
J	-	-	0.62	-	-	+	-	-	-	+	-	-	-
Threshold	0.39	0.32	0.32	0.39	0.40	0.55	0.32	0.23	0.18	0.30	0.21	0.28	0.15

Table 8: Segment scale, catchment and downstream environmental attributes for 10 groups from a multivariate classification of New Zealand’s diadromous fish species.

Classification	Length (km)	SegLowFlow (4^{th} root)	SegLowFlow ($m^3 sec^{-1}$)	SegFlowStability (ratio)	SegSumT (^oC)	SegMinTNorm (normalised ^oC)	SegN (log10 conc)	SegShade (proportion)	SegSlope (^{o})	ReachSubstrate (1=mud to 6=boulders)	USDaysRain (d yr^{-1})	USAvgT (normalised ^oC)	USAvgSlope (^{o})	USNative (proportion)	DSDist (km)	DSAvgSlope (^{o})	DSDam (proportion)
A	20,229	1.003	0.01	0.13	18.13	1.56	0.09	0.66	1.92	2.78	14.40	0.27	10.56	0.39	8.80	1.11	0.02
B	11,215	1.029	0.12	0.19	15.72	1.48	-0.43	0.66	5.46	4.30	35.48	-0.09	20.61	0.87	11.09	2.74	0.02
C	1658	1.002	0.01	0.14	13.12	2.48	-0.23	0.75	5.05	4.03	14.42	1.07	13.20	0.84	5.46	3.26	0.04
D	30,875	1.091	0.42	0.11	17.78	0.95	0.26	0.31	0.38	2.72	12.29	0.11	8.76	0.23	29.10	0.20	0.02
E	5225	1.437	3.26	0.21	17.13	0.85	-0.51	0.33	0.84	4.27	36.37	-1.69	22.76	0.82	20.74	0.28	0.00
F	5412	1.381	2.64	0.18	17.17	0.45	-0.15	0.20	0.51	3.82	19.09	-1.01	15.59	0.44	81.01	0.11	0.08
G	17,203	1.003	0.01	0.08	17.50	0.27	0.70	0.59	0.48	1.99	7.56	0.65	2.89	0.03	66.45	0.14	0.05
H	127,412	1.029	0.12	0.13	16.44	0.42	0.04	0.54	2.47	3.44	12.81	0.29	13.02	0.34	82.42	0.29	0.05
I	24,888	1.079	0.36	0.28	14.41	0.84	-0.65	0.66	8.91	4.66	52.46	-0.79	28.25	0.97	58.37	1.08	0.17
J	12,016	1.596	5.49	0.38	16.53	-0.64	-0.51	0.13	0.44	3.09	19.67	-1.12	13.36	0.41	235.26	0.21	0.99

Group B: (11,215 km) Longfin eels (0.68), redfin bully (0.54) and banded kokopu (0.51) are the most frequently caught species, followed by koaro (0.44). Species caught occasionally include shortfin eels, inanga, common bully, torrentfish, giant kokopu, shortjaw kokopu and bluegill bully. This group occurs in coastal locations in mild (15.7 ^{o}C), maritime climates with moderate frequencies of rain days. It occurs in small streams (0.121 cumecs) with unstable flows, a predominance of coarse gravelly substrates and moderate nitrogen concentrations. Segment and downstream gradients are generally steep, and catchment slopes are moderate. Riparian shading is high and cover of native vegetation in the upstream catchment is very high. It occurs extensively through Fiordland, Westland and the Marlborough Sounds, with smaller areas on Stewart Island, in coastal Southland, on Banks Peninsula, and in scattered locations around the North Island coast.

Group C: (1658 km) Banded kokopu (0.67) is the most frequently caught species in this group, with occasional catches of inanga, redfin bully, longfin eels, giant kokopu and koaro. It occurs in coastal locations in cool (13.1 ^{o}C), strongly maritime climates with low frequencies of rain days, generally in very small streams (0.008 cumecs) with unstable flows, a predominance of coarse gravelly substrates and moderate nitrogen concentrations. Segment and downstream gradients are generally steep, and catchment slopes are gentle. Riparian shading is high and cover of native vegetation in the upstream catchment is very high. It is largely restricted to southern New Zealand, where it occurs extensively on Stewart Island and around the south-eastern coast of the South Island.

Group D: (30,875 km) The most frequently caught species in this group are longfin eels (0.73), shortfin eels (0.69), inanga (0.55) and common bully (0.51), with occasional catches of smelt (0.33). Banded kokopu, redfin bully, torrentfish and giant kokopu are also occasionally caught. This group occurs in moderately coastal locations in warm (17.8 ^{o}C) climates with low frequencies of rain days. It occurs mostly in small streams (0.417 cumecs) with unstable flows, a predominance of sandy to gravelly substrates and high nitrogen concentrations. Segment and downstream gradients are generally very gentle, and catchment slopes are gentle, while riparian shading and cover of native vegetation in the upstream catchment are low. It is widespread throughout the lowland plains of the North Island, occurring extensively in Northland, Auckland, Waikato, Bay of Plenty, Hawke's Bay, Poverty Bay, Manawatu and Wairarapa. In the south, it is largely confined to the coastal plains of Marlborough, Canterbury and Southland.

Group E: (5225 km) This group is dominated by longfin eels (0.83), redfin bully (0.64) and torrentfish (0.62) with occasional bluegill bully (0.36) and common bully (0.34). Shortfin eels, inanga, banded kokopu, smelt, koaro and shortjaw kokopu are also occasionally caught. This group occurs in moderately coastal locations in mild (17.1 °C) climates with moderate frequencies of rain days. It occurs mostly in small rivers (3.264 cumecs) with moderately stable flows and a predominance of coarse gravelly substrates. Stream gradients are generally gentle, with moderate catchment and very gentle downstream slopes. Riparian shading is low and cover of native vegetation in the upstream catchment is very high. This group occurs most frequently on the South Island's west coast and in the eastern Bay of Plenty.

Group F: (5412 km) Longfin eels (0.85) are the most frequently caught species in this group, followed by common bully (0.53) and smelt (0.33). Shortfin eels, inanga, banded kokopu, redfin bully, torrentfish and koaro are also caught occasionally. This group occurs in moderately inland locations in mild (17.2 °C) climates with low frequencies of rain days. It occurs mostly in small rivers (2.637 cumecs) with unstable flows, a predominance of coarse gravelly substrates and moderate nitrogen concentrations. Stream gradients are generally gentle, with moderate catchment and very gentle downstream slopes. Riparian shading is very low and cover of native vegetation in the upstream catchment is moderate. It occurs most widely in the North Island, particularly in Taranaki, Manawatu, Hawke's Bay and Poverty Bay, and along the Canterbury Plains in the South Island.

Group G: (17,203 km) Shortfin eels (0.62) are the most commonly caught species, with occasional catches of inanga, common bully, banded kokopu, longfin eels and common smelt. This group occurs in moderately inland locations in mild (17.5 °C) climates with very low frequencies of rain days, mostly in very small streams (0.012 cumecs) with very unstable flows, a predominance of sandy substrates and very high nitrogen concentrations. Stream gradients are generally very gentle, as are catchment and downstream slopes. Riparian shading is moderate and cover of native vegetation in the upstream catchment is very low. It occurs predominantly in the North Island lowlands, particularly in Waikato, Northland, Manawatu and Hawke's Bay, but also occurs on the Canterbury Plains in the South Island.

Group H: (127,412 km) Longfin eels (0.72) are the only frequently caught species in this, which is the most extensive of all the groups. Shortfin eels, common bully, banded kokopu, redfin bully and koaro are caught only occasionally. It occurs in moderately inland locations in mild (16.4 °C) climates with low frequencies of rain days, mostly in small streams (0.121 cumecs) with unstable flows, a predominance of gravelly substrates and high nitrogen concentrations. Stream gradients are generally moderate, with gentle catchment and very gentle downstream slopes. Riparian shading is moderate and cover of native vegetation in the upstream catchment is low. This group occurs throughout the North and South Islands, with absences only in Auckland and the south-western South Island.

Group I: (24,888 km) Koaro (0.78) is the only frequently caught species, but banded kokopu, redfin bully, longfin eels and shortjaw kokopu are caught on occasion. This group occurs mostly in moderately inland locations in cool (14.4 ^{o}C) climates with high frequencies of rain days. It occurs mostly in small streams (0.355 cumecs) with moderately stable flows and a predominance of cobbly substrates. Stream and catchment gradients are generally steep, and downstream slopes are moderate. Riparian shading is high and cover of native vegetation in the upstream catchment is very high. It occurs extensively throughout Westland and Fiordland, with smaller areas in the North Island on Mount Taranaki and in the Raukumara, Huiarau and Tararua Ranges.

Group J: (12,016 km) Common bullies (0.62) are the only frequently caught species, with only occasional catches of longfin eels and koaro. This group occurs in strongly inland locations in mild (16.5 ^{o}C) climates with low frequencies of rain days. It occurs mostly in small rivers (5.488 cumecs) with stable flows and a predominance of gravelly substrates, and mostly in close proximity to lakes. Stream and downstream gradients are generally very gentle, and catchment slopes are gentle. Riparian shading is very low and cover of native vegetation in the upstream catchment is moderate and upstream dam affect is very high. This group is largely confined to the central North Island, occurring around Lake Taupo, in tributaries of the Waikato River downstream of Taupo, and in rivers and streams flowing into lakes of the Rotorua district. Its presence here probably reflects movement of this species by Maori in pre-European times (see McDowall 1990).

3.2 Maps of predicted distribution for individual species

The remainder of this report focuses on the predicted distributions of the individual fish species, presented as a series of maps, one per species. For each map we show the actual distribution as recorded in the subset of NZFFD sites used for fitting the model, along with the predicted probability of capture for all rivers and streams. For diadromous species, which appear to be well sorted in relation to environment, we show predictions for all rivers and streams, regardless of whether the species is caught in the geographic proximity or not. For non-diadromous species, most of which show marked geographic patchiness in their distributions, we show predicted probabilities of capture only within their broad geographic ranges. However, the digital database containing predictions for these species includes both geographically constrained predictions as shown here, and predictions for all rivers and streams in New Zealand. Broader accounts of the ecology of these species can be found for example in McDowall (1990, 2000).

Although we have not attempted to formally assess the confidence limits about these predictions for all species, analyses for representative species using bootstrap simulation suggest 95% confidence intervals of around +/- 0.1 when predicted probabilities are 0.5, declining to smaller values as predicted values approach either zero or one. Some biases will also be present in the predictions because of the biases inherent in the sampling data from which they were derived. In particular, many more samples are available for small streams than for larger streams and rivers, and electric fishing tends to be less effective in larger rivers. In addition, no attempt has been made to predict the distributions of diadromous species that also occur in lakes, although their enhanced presence in rivers adjacent to lakes is apparent for two species in particular, i.e., koaro and common bullies.

All data are available from the National Institute of Water and Atmospheric Research in ArcView shapefile format. Maps for diadromous species are presented first, followed by those for non-diadromous species.

Shortfin eels are caught most frequently in moderately coastal locations in warm (mean = 18.1°C), maritime climates with low frequencies of high intensity rain days. They are most commonly caught in small streams (0.52 cumecs) with unstable flows and a predominance of sandy substrates. Stream and downstream gradients are generally very gentle, with gentle catchment slopes. Shortfin eels are most commonly caught in streams with low native vegetation cover in the upstream catchment, low riparian shading and high nitrogen concentrations.

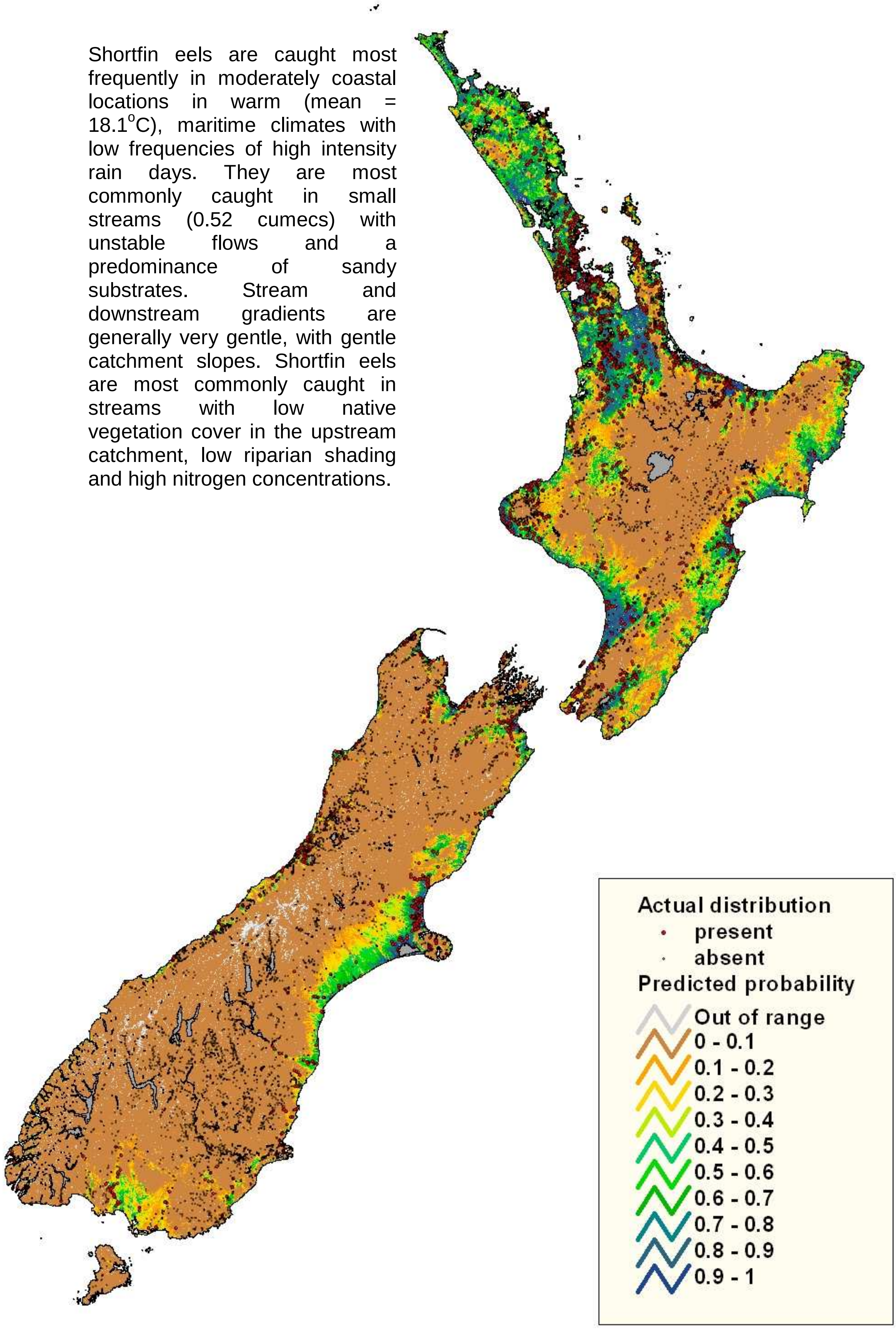

Figure 4: Shortfin eels (*Anguilla australis*).

Longfin eels are caught most frequently in moderately coastal locations, but they also extend considerable distance inland, particularly in lower gradient rivers. They occur in mild (17.0^{o}C), maritime climates with moderate frequencies of high intensity rain days. They are most commonly caught in very small streams (0.46 cumecs) with unstable flows and a predominance of coarse gravelly substrates. Stream and catchment gradients are generally moderate, with very gentle downstream slopes, but they also occur in very steep gradient streams. They are generally caught in locations with moderate native vegetation cover in the upstream catchment but tolerate wide variation in riparian shading.

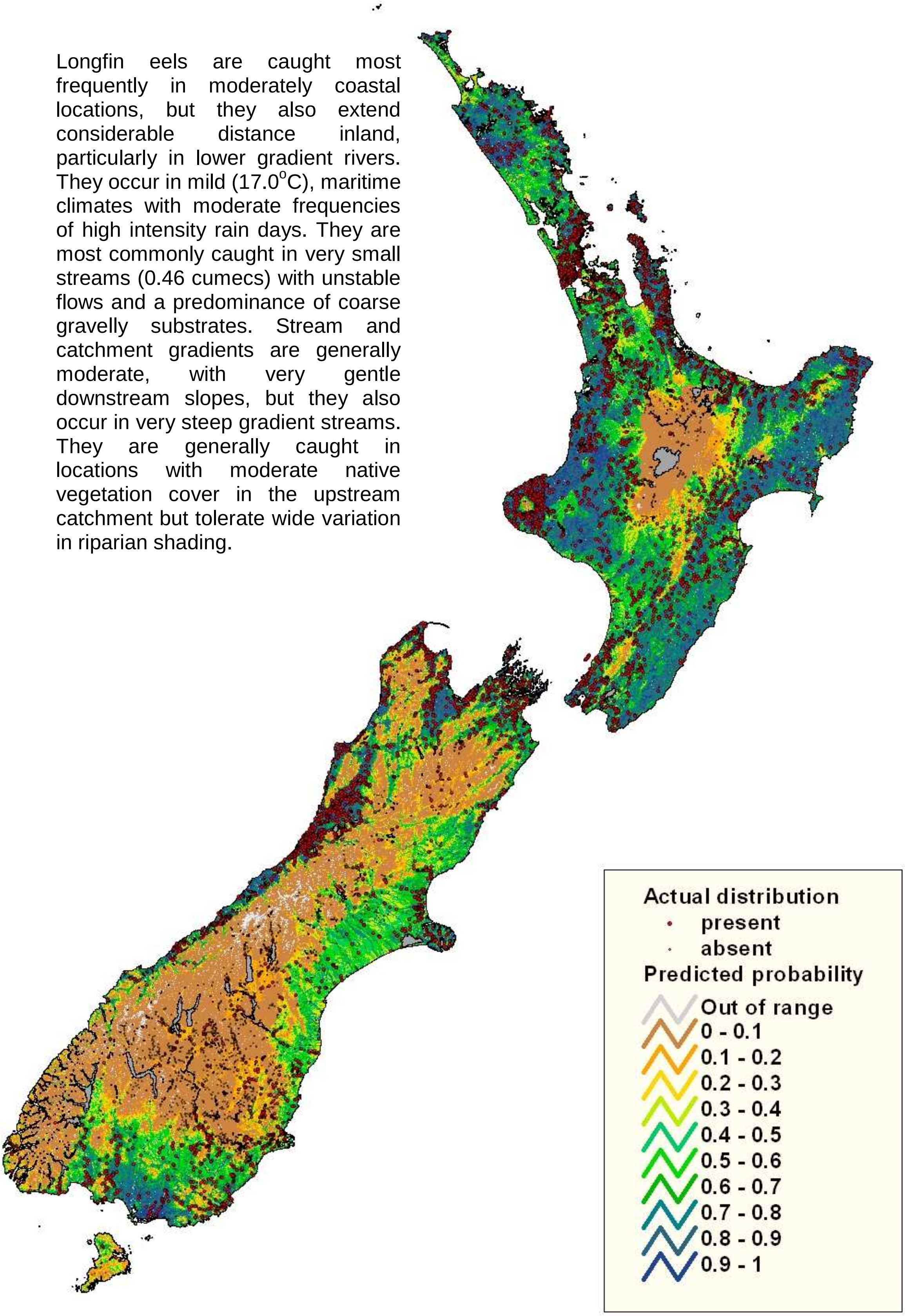

Figure 5: Longfin eels (*Anguilla dieffenbachii*).

Torrentfish are caught most frequently in moderately coastal locations with mild (17.4^{o}C) climates and moderate frequencies of high intensity rain days. They are most commonly caught in moderate-sized streams (2.63 cumecs) with relatively stable flows and a predominance of coarse gravelly substrates. Stream gradients are generally gentle, with moderate catchment and very gentle downstream slopes. Torrentfish are most commonly caught in locations with high levels of native vegetation cover in the upstream catchment but low riparian shading.

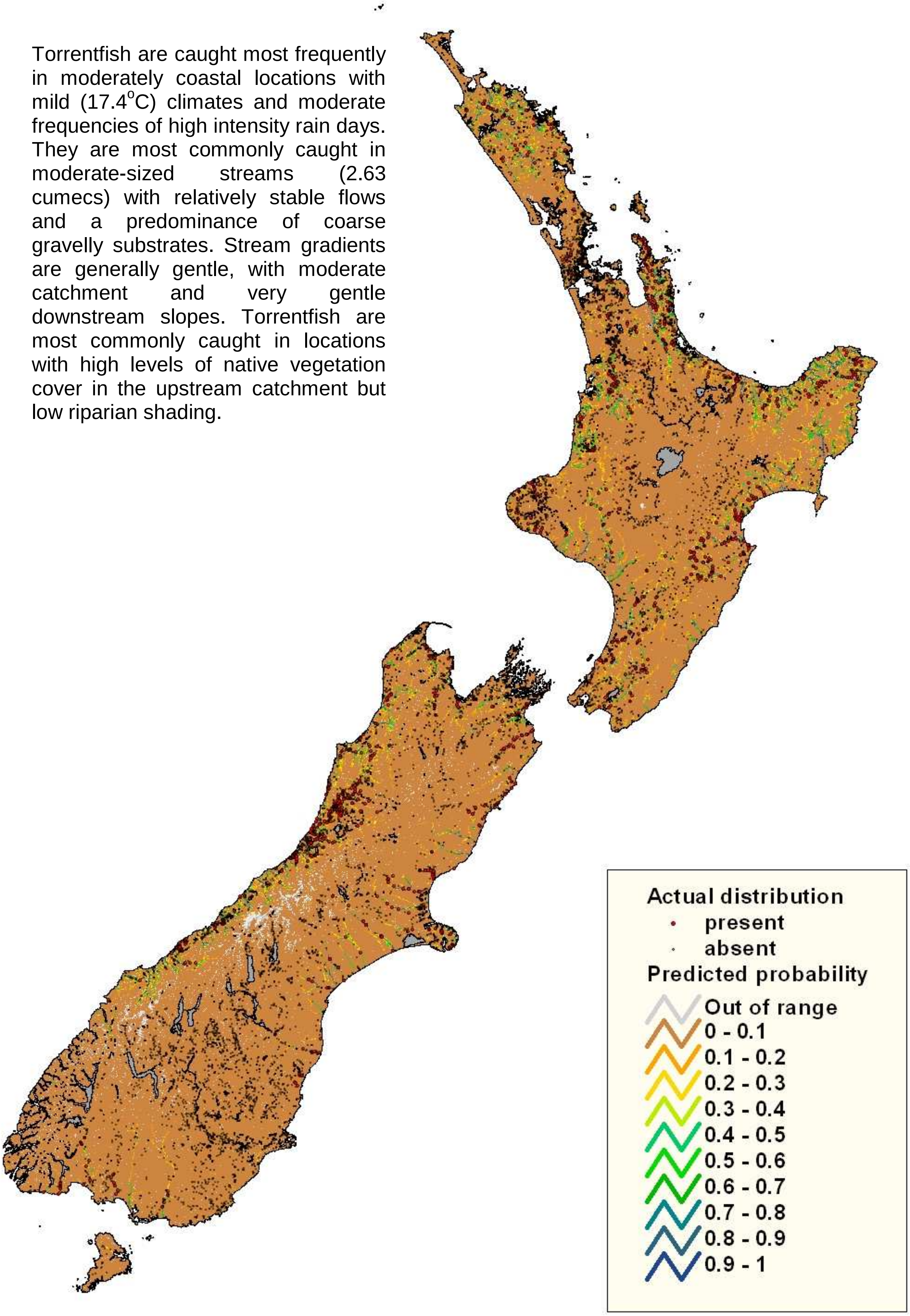

Figure 6: Torrentfish (*Cheimarrichthys fosteri*).

Giant kokopu are caught most frequently in coastal locations with mild (15.6^{o}C), maritime climates with moderate frequencies of high intensity rain days. They are mostly caught in very small streams (0.04 cumecs) with unstable flows and a predominance of gravelly substrates. Stream gradients are generally moderate, with gentle catchment and very gentle downstream slopes. Giant kokopu are caught most frequently in streams with high levels of native vegetation cover in the upstream catchment and moderate riparian shading.

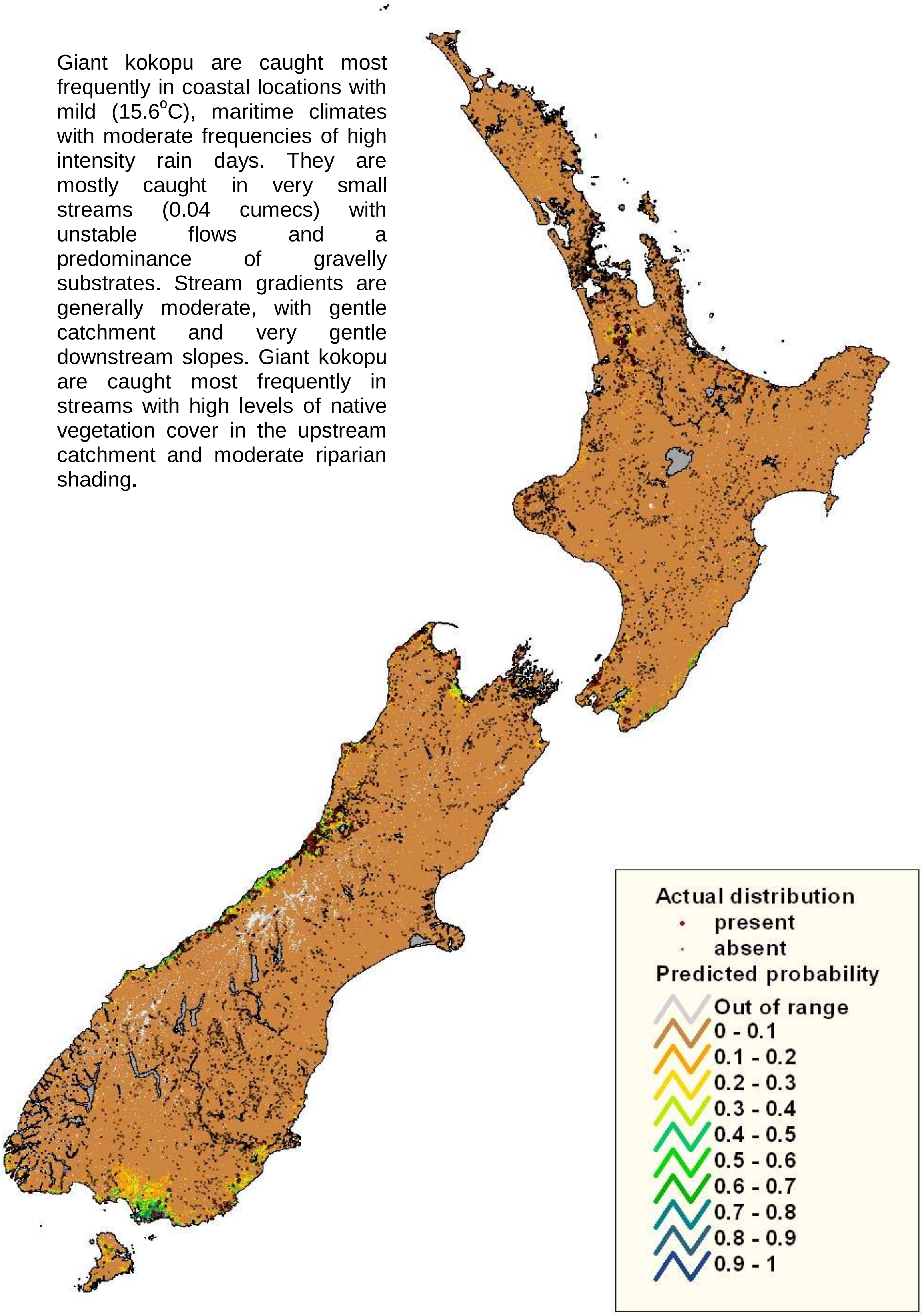

Figure 7: Giant kokopu (*Galaxias argenteus*).

Koaro are caught most frequently in inland locations with mild (15.4°C), maritime climates. They show a preference for sites with high frequencies of high intensity rain days in the upstream catchment. They generally occur in very small streams (0.36 cumecs) with moderately stable flows and cobbly substrates. Stream and down-stream gradients are generally moderate, with steep catchment slopes. Koaro are mostly caught in streams with very high native catchment cover and high riparian shading. They occur in some inland in rivers associated with lakes, presumably because of lacustrine recruitment. Predictions for Fiordland streams are based on a minimal number of field samples.

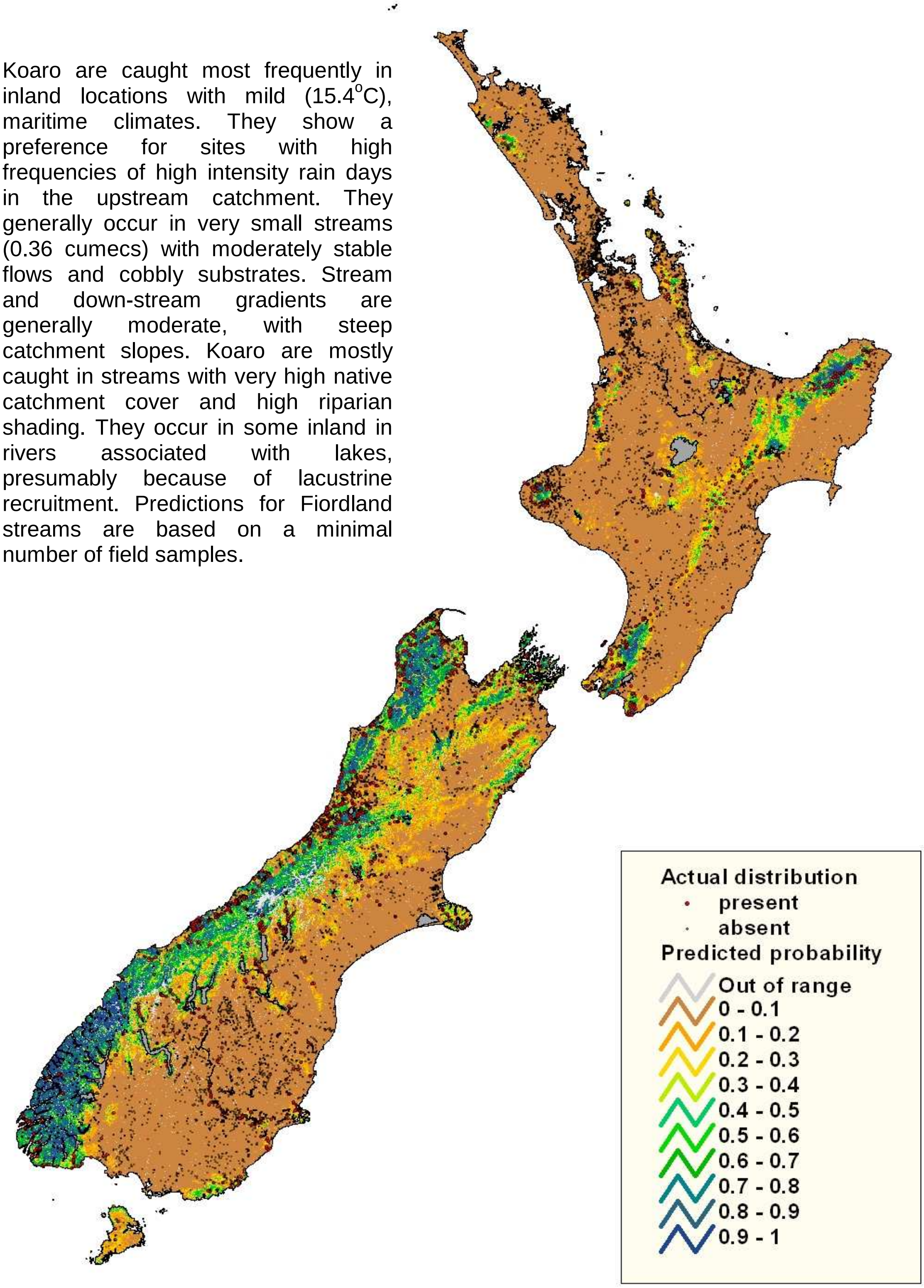

Figure 8: Koaro (*Galaxias brevipinnis*).

Banded kokopu are strongly confined to coastal locations, occurring in sites with warm (17.8°C), maritime climates with low frequencies of high intensity rain days. They are most commonly caught in very small streams (0.04 cumecs) with unstable flows and a predominance of gravelly substrates. Stream gradients are generally moderate, as are downstream and catchment gradients. Banded kokopu are most commonly caught in locations with high native vegetation cover in the upstream catchment and show a strong preference for high levels of riparian shading.

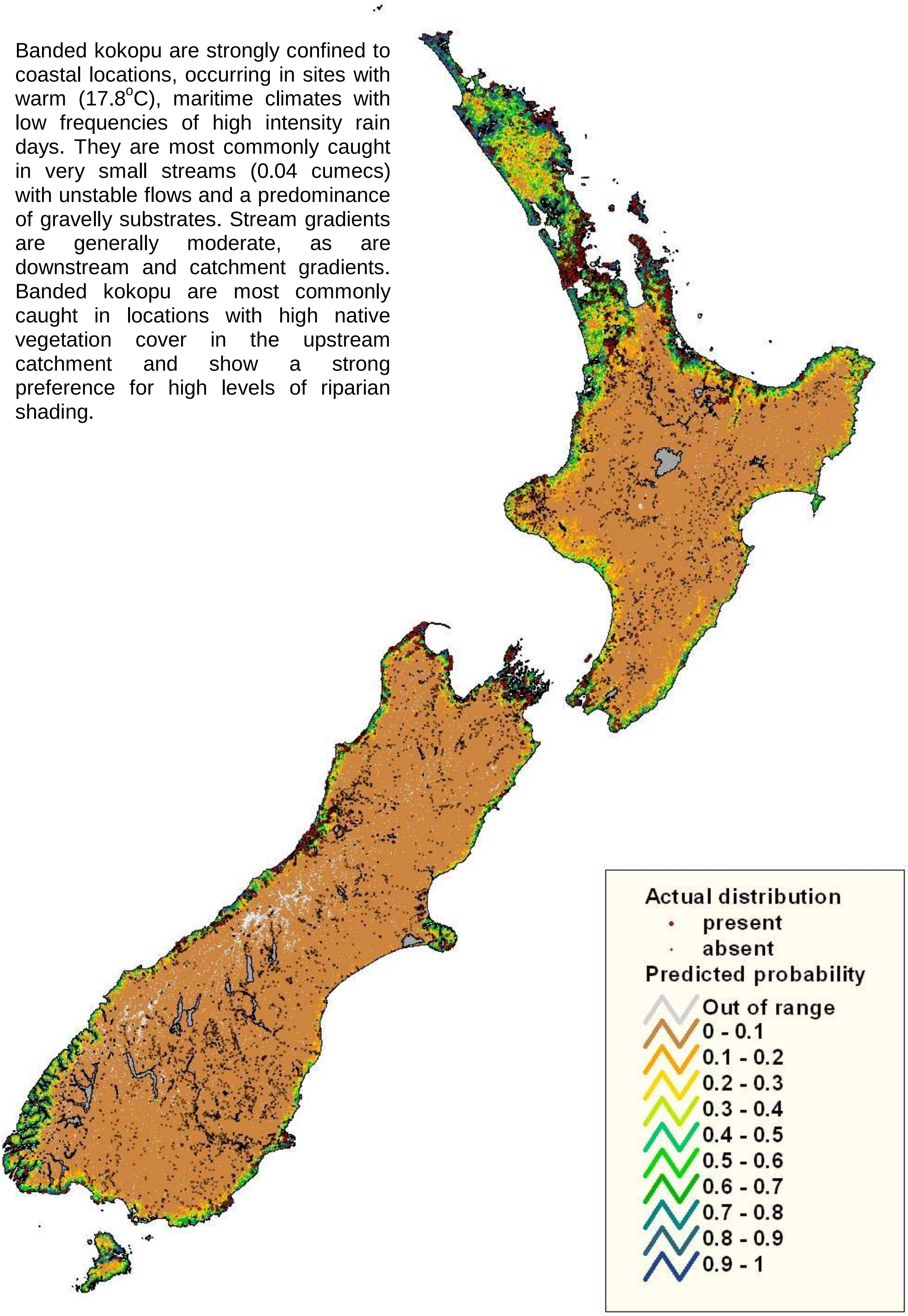

Figure 9: Banded kokopu (*Galaxias fasciatus*).

Inanga are caught most frequently in coastal locations with warm (17.6^{o}C), maritime climates with low frequencies of high intensity rain days. They are mostly caught in small streams (0.46 cumecs) with unstable flows and a predominance of gravelly substrates, but also occur in larger rivers. Inanga show a strong preference for sites with gentle segment and downstream gradients, apparently having little climbing ability. They are most commonly caught in rivers and streams with moderate native vegetation cover in the upstream catchment and low riparian shading.

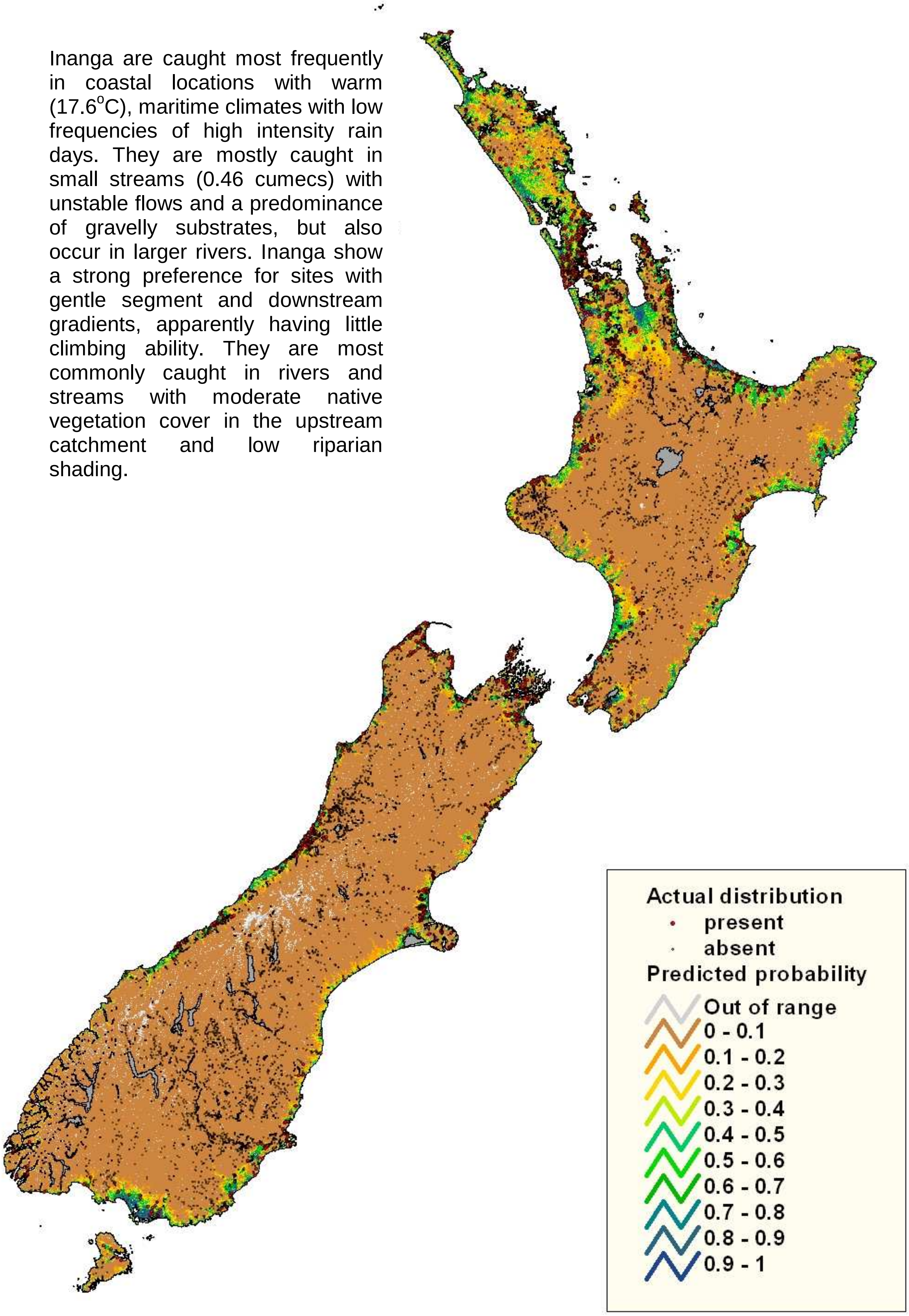

Figure 10: Inanga (*Galaxias maculatus*).

Shortjaw kokopu are caught most frequently in coastal locations with mild (16.1°C), maritime climates. They exhibit a strong preference for sites with moderately high frequencies of intense rainfall. They are most commonly caught in very small streams (0.13 cumecs) with unstable flows and a predominance of bouldery substrates. Stream and catchment gradients are generally moderate, with gentle downstream slopes. Shortjaw kokopu show a strong preference for locations with very high native vegetation cover in the upstream catchment and high riparian shading. They are most commonly recorded by spotlight observation.

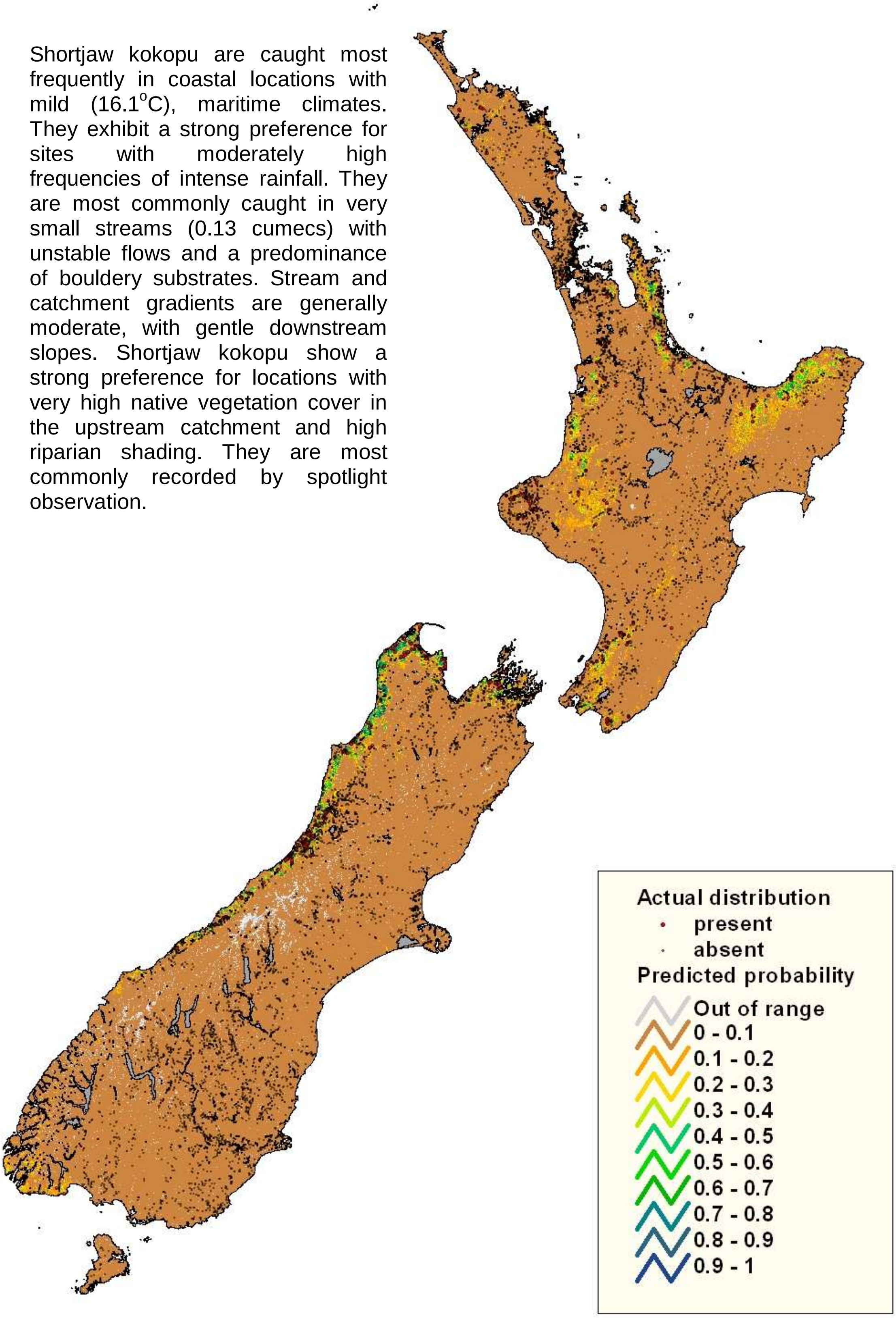

Figure 11: Shortjaw kokopu (Galaxias postvectis).

Common bullies are caught most frequently in moderately coastal sites with mild (17.3°C), maritime climates with low frequencies of high intensity rain days. They are mostly caught in small streams (1.44 cumecs) with unstable flows and gravelly substrates. Stream and downstream gradients are generally very gentle, with moderate catchment slopes. They are generally absent above obstacles such as waterfalls. Common bullies are mostly caught in locations with moderate native vegetation cover in the upstream catchment and low riparian shading. They occur more frequently close to lakes, presumably because of their ability to use them for breeding.

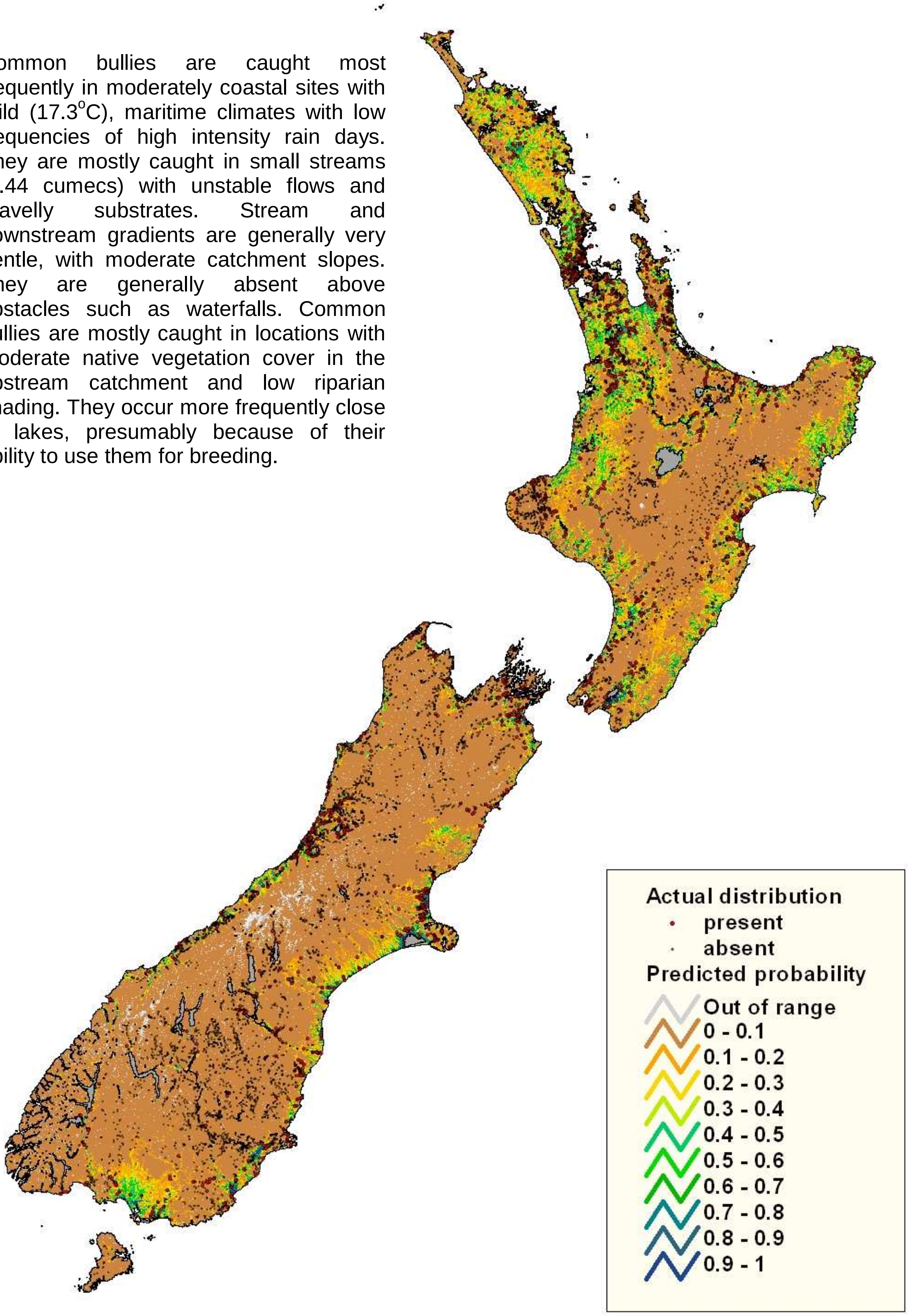

Figure 12: Common bully (*Gobiomorphus cotidianus*).

Giant bullies are caught most frequently in coastal locations with warm (18.4°C), maritime climates with low frequencies of high intensity rain days. They are most commonly caught in very small streams (0.22 cumecs) with unstable flows and a predominance of coarse gravelly substrates. Stream gradients are generally gentle, with moderate catchment and downstream slopes. Native vegetation cover in the upstream catchment is generally high with moderate riparian shading. Records of this species in the database are biased by the difficulty of using electric fishing in waters with tidal influence – anecdotal evidence suggests that they are common in these sites, but our analysis cannot be used to predict their occurrence in there.

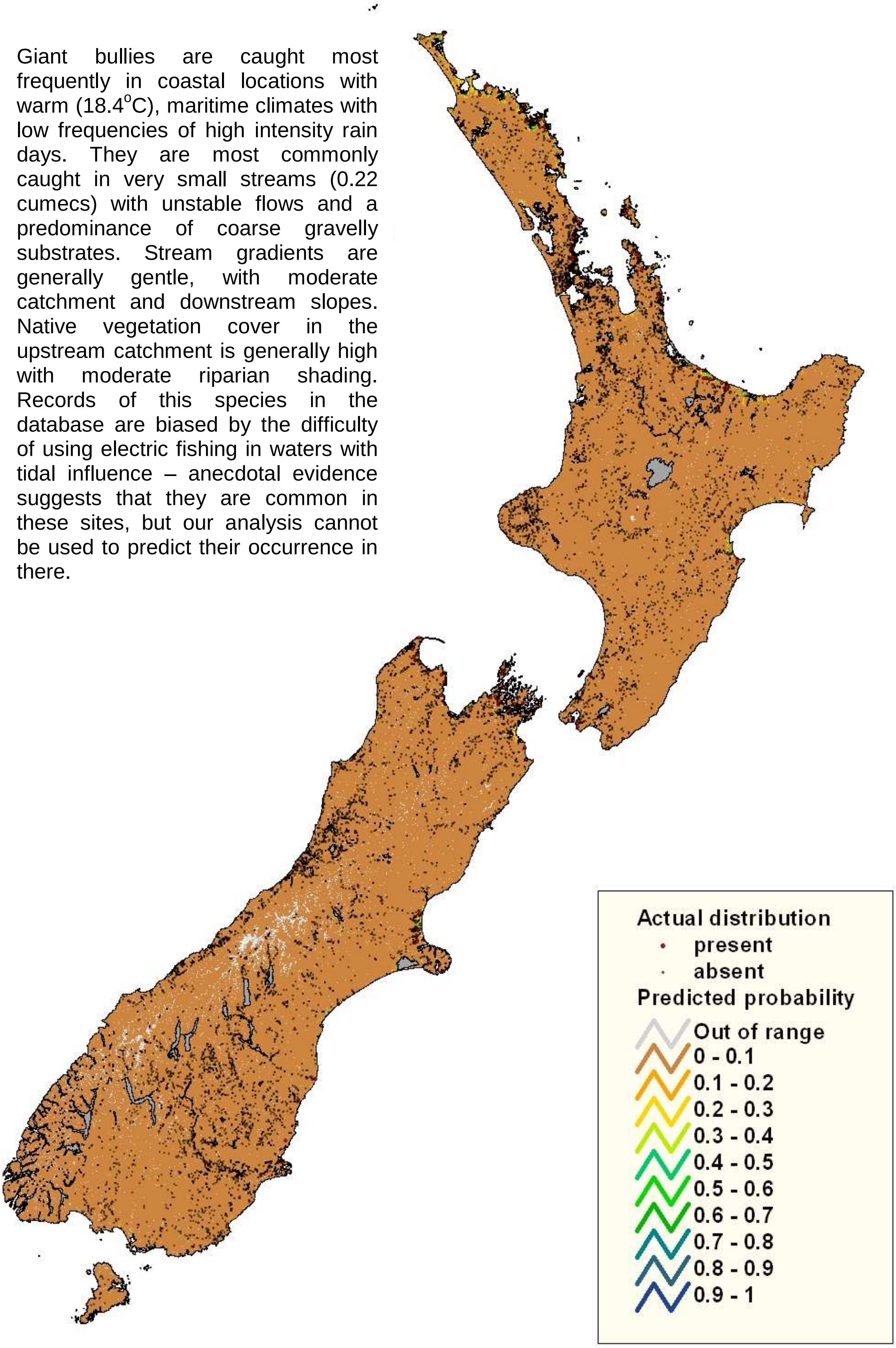

Figure 13: Giant bully (*Gobiomorphus gobiodes*).

Bluegill bullies are caught most frequently in coastal locations with mild (16.4^{o}C), maritime climates with moderate frequencies of high intensity rain days. They are most common in streams (2.63 cumecs) with moderately stable flows and a predominance of cobbly substrates. Stream and downstream gradients are generally gentle, with steep catchment slopes. Bluegill bullies are most commonly caught in locations with high native vegetation cover in the upstream catchment and low riparian shading.

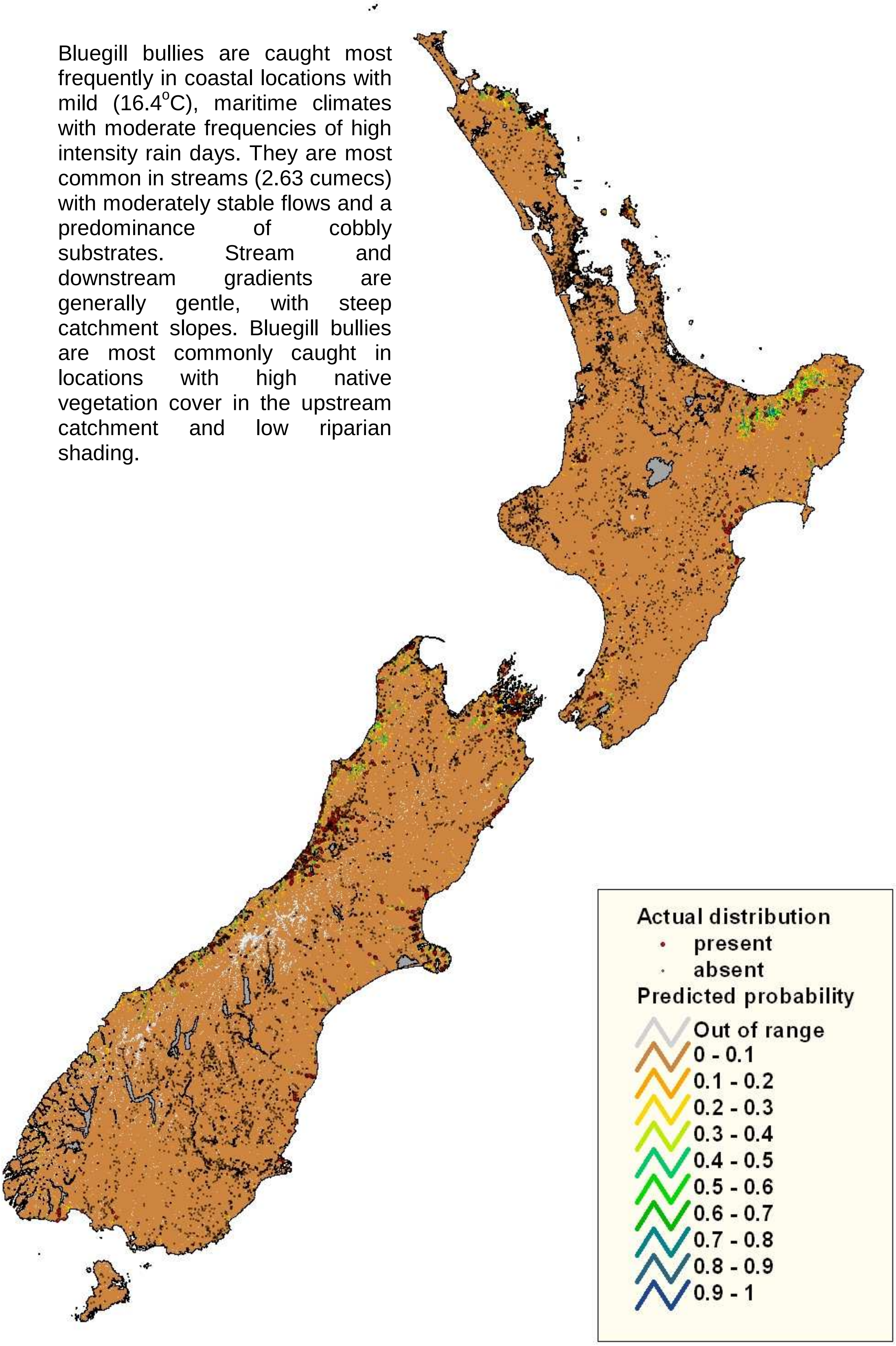

Figure 14: Bluegill bully (*Gobiomorphus hubbsi*).

Redfin bullies show a strong preference for coastal locations, occurring in mild (17.2^{o}C), maritime climates with moderate to high frequencies of high intensity rain days. They are most commonly caught in very small streams (0.31 cumecs) with unstable flows and a predominance of cobbly substrates. Stream and catchment gradients are generally moderate, with gentle downstream slopes. Redfin bullies are most commonly caught in locations with high native vegetation cover in the upstream catchment and moderate riparian shading.

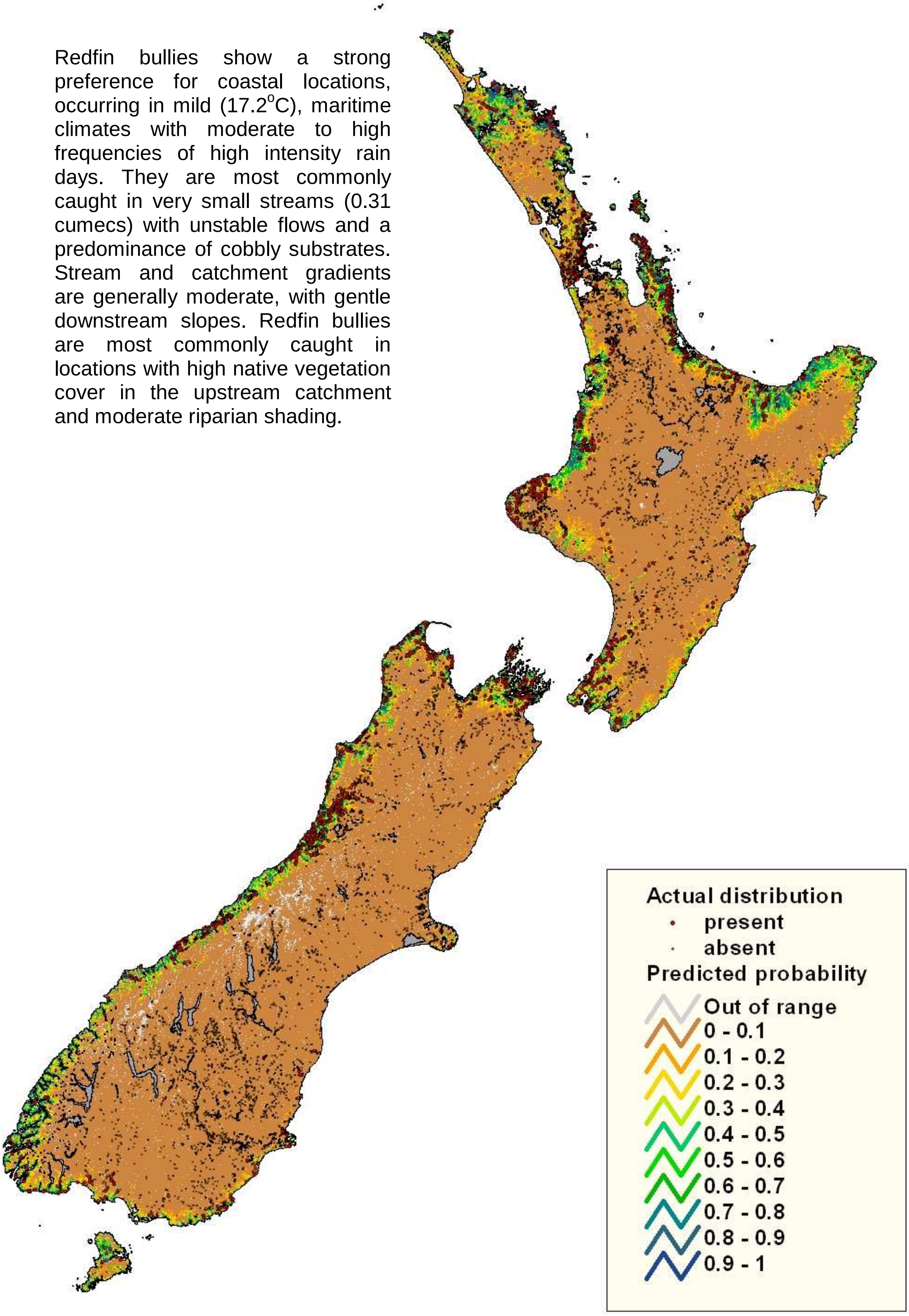

Figure 15: Redfin bully (*Gobiomorphus huttoni*).

Common smelt are caught most frequently in coastal to moderately inland locations, showing a strong preference for more northern, warm (18.2°C) climates with low frequencies of high intensity rain days. They are mostly commonly caught in moderate sized streams (6.41 cumecs) with moderately stable flows and sandy substrates. Other evidence indicates that they also occur in larger rivers, including those with tidal influence, but such sites are difficult to sample with electric fishing, and so are not represented by our data. Stream and downstream gradients are generally very gentle, with moderate catchment slopes. Common smelt are most commonly caught in locations with low native vegetation cover in the upstream catchment and very low riparian shading. They appear tolerant of high nitrogen concentrations. Note that lake populations of this species are not shown.

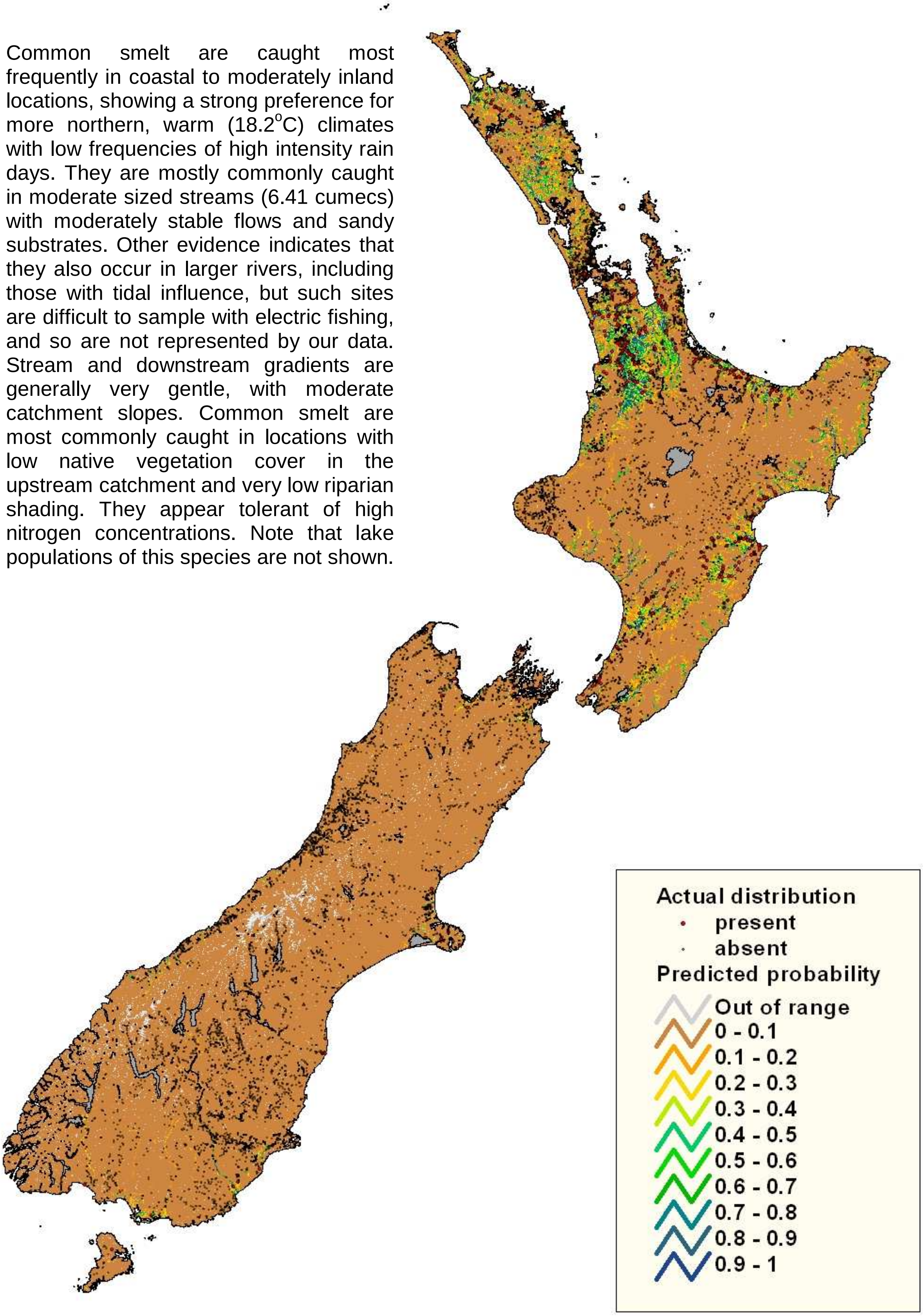

Figure 16: Common smelt (*Retropinna retropinna*).

Black flounder are caught most frequently in coastal locations with mild (16.6°C) climates with low frequencies of high intensity rain days. They are most commonly caught in moderate sized streams or small rivers (4.92 cumecs) with unstable flows and a predominance of sandy substrates. Stream and downstream gradients are generally very gentle, with moderate catchment slopes. Black flounder are most commonly caught in locations with low native vegetation cover in the upstream catchment, very low riparian shading and high nitrogen concentrations. Our data do not contain records from the lower reaches of rivers with strong saline influence, where this species is also likely to be common.

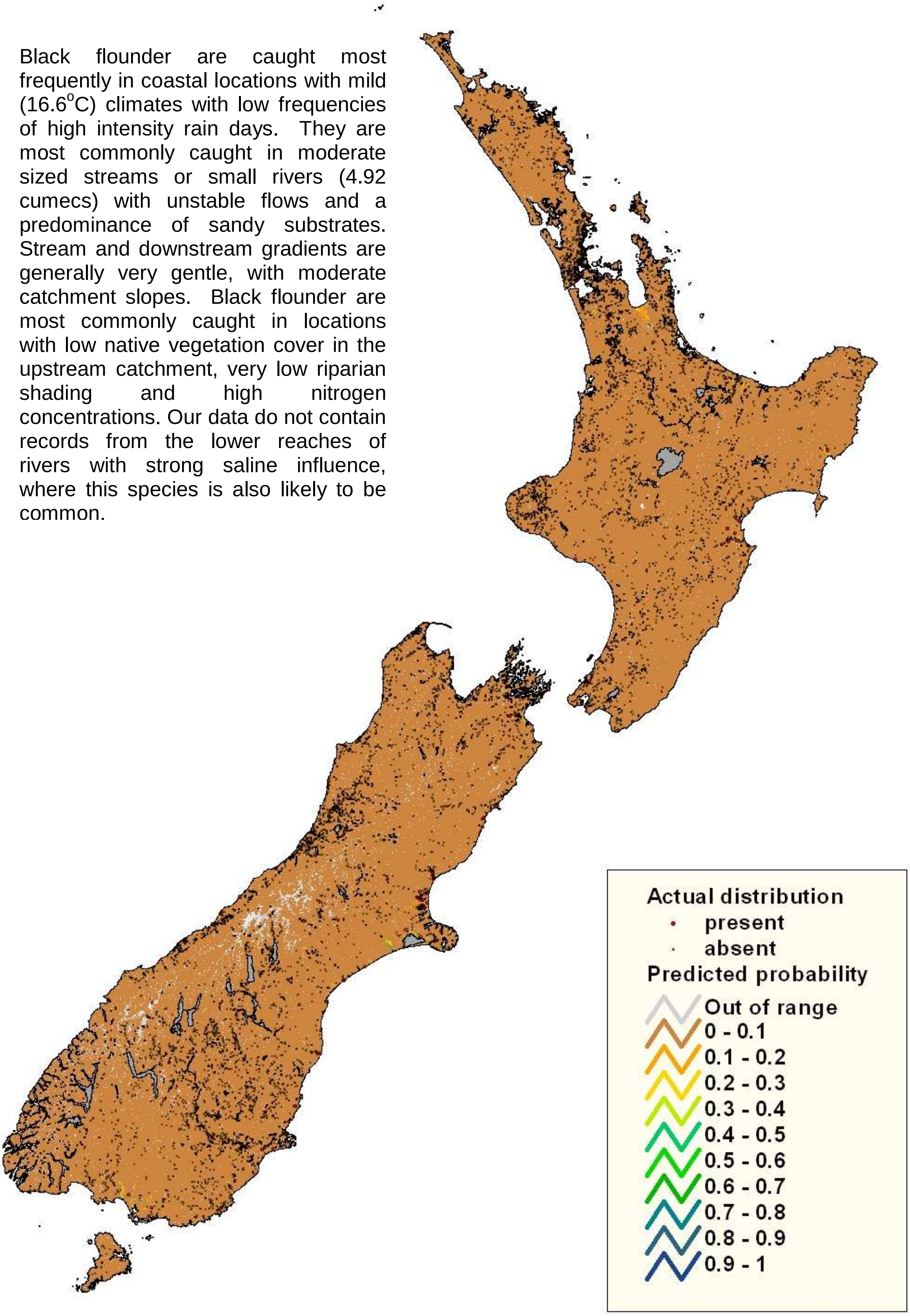

Figure 17: Black flounder (*Rhombosolea retiaria*).

Roundhead galaxias occur only in the middle to upper reaches of the Taieri River. Here, they are caught most frequently at sites with cool (14.9°C), strongly continental climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams (0.04 cumecs) with very unstable flows and a predominance of coarse gravelly substrates. Stream gradients are generally gentle, with moderate catchment and very gentle downstream slopes. These sites generally have low native vegetation cover in the upstream catchment, low riparian shading and high nitrogen concentrations.

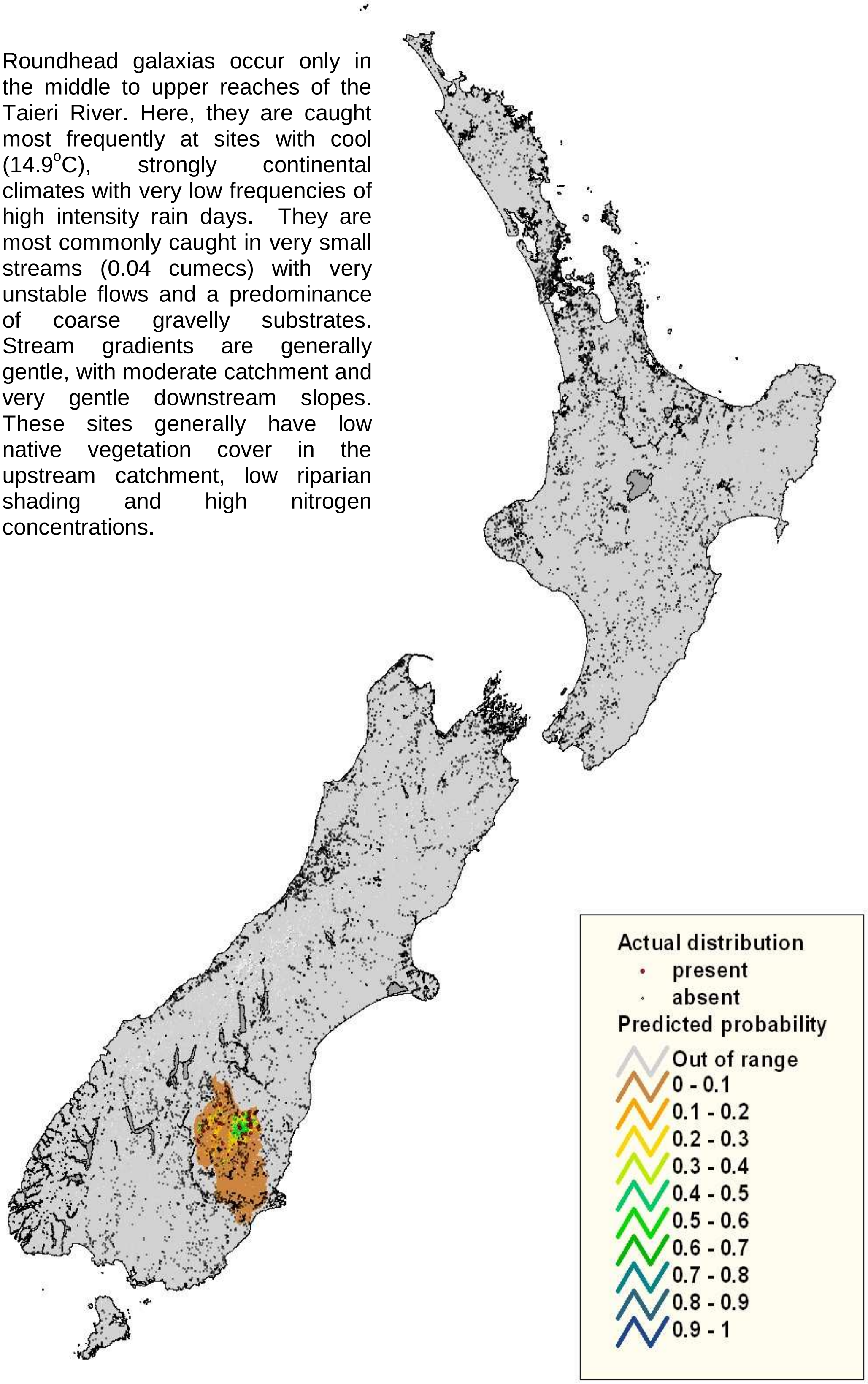

Figure 18: Roundhead galaxias (*Galaxias anomalus*).

Lowland longjaw galaxias are restricted to a few sites in the catchments of the Waitaki and Kakanui River. They are caught most frequently in moderately coastal locations with mild (15.3°C), maritime climates with very low frequencies of high intensity rain days. They are most commonly caught in small streams (0.57 cumecs) with unstable flows and a predominance of coarse gravelly substrates. Stream and downstream gradients are generally very gentle, with steep catchment slopes. These locations generally have low native vegetation cover in the upstream catchment, very low riparian shading and high predicted nitrogen concentrations.

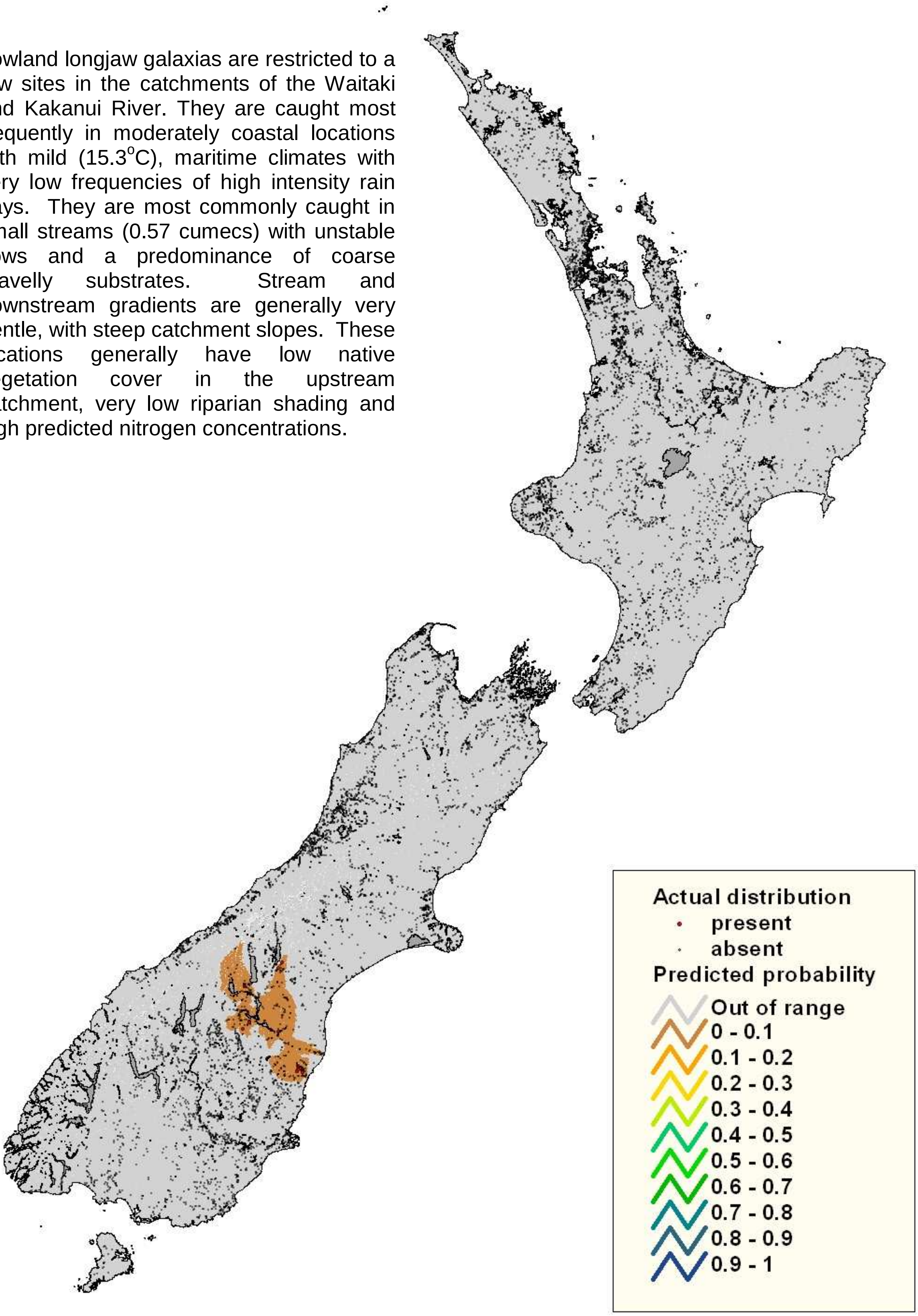

Figure 19: Lowland longjaw galaxias (*Galaxias cobitinis*).

Taieri flathead galaxias are largely confined to the mid-reaches of the Taieri River, with a few occurrences in the nearby Akatore Creek and in an eastern tributary of the Tokomairiro. They occur mostly in inland locations with cool (12.7°C) climates with very low frequencies of high intensity rain days. They occur most commonly in very small streams (0.041 cumecs) with unstable flows and a predominance of bouldery substrates. Stream and catchment gradients are generally moderate, with very gentle downstream slopes. These locations generally have moderate native vegetation cover in the upstream catchment and low riparian shading.

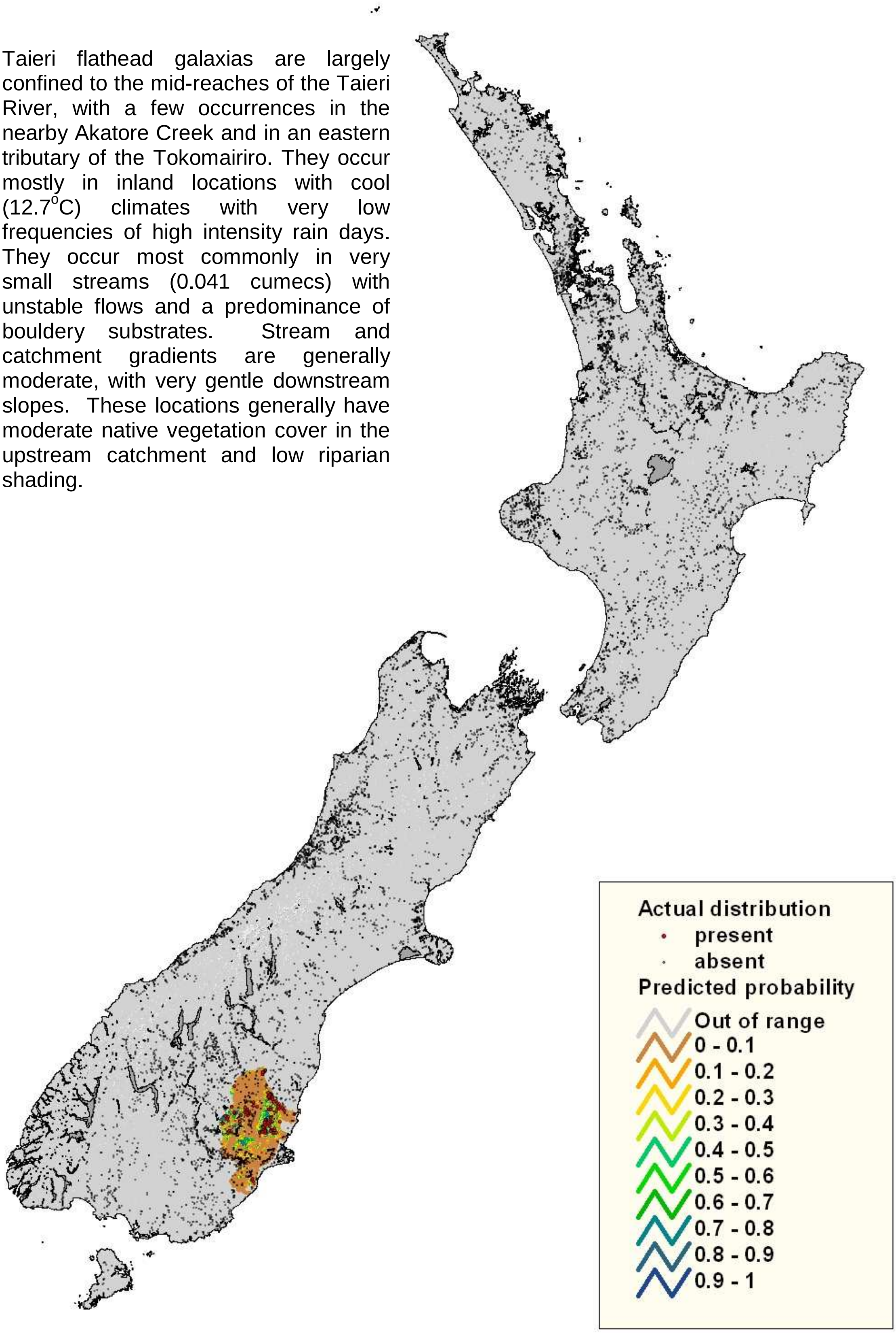

Figure 20: Taieri flathead galaxias (*Galaxias depressiceps*).

Dwarf galaxias occur spasmodically from the Hauraki Plains to the South Island's West Coast. They generally occur in moderately inland locations with mild (16.3°C) climates with moderate frequencies of high intensity rain days. They are most commonly caught in very small streams (0.46 cumecs) with unstable flows and a predominance of cobbly substrates. Stream gradients are generally moderate, with steep catchment and very gentle downstream slopes. Dwarf galaxias are most commonly caught in locations with high native vegetation cover in the upstream catchment and moderate riparian shade.

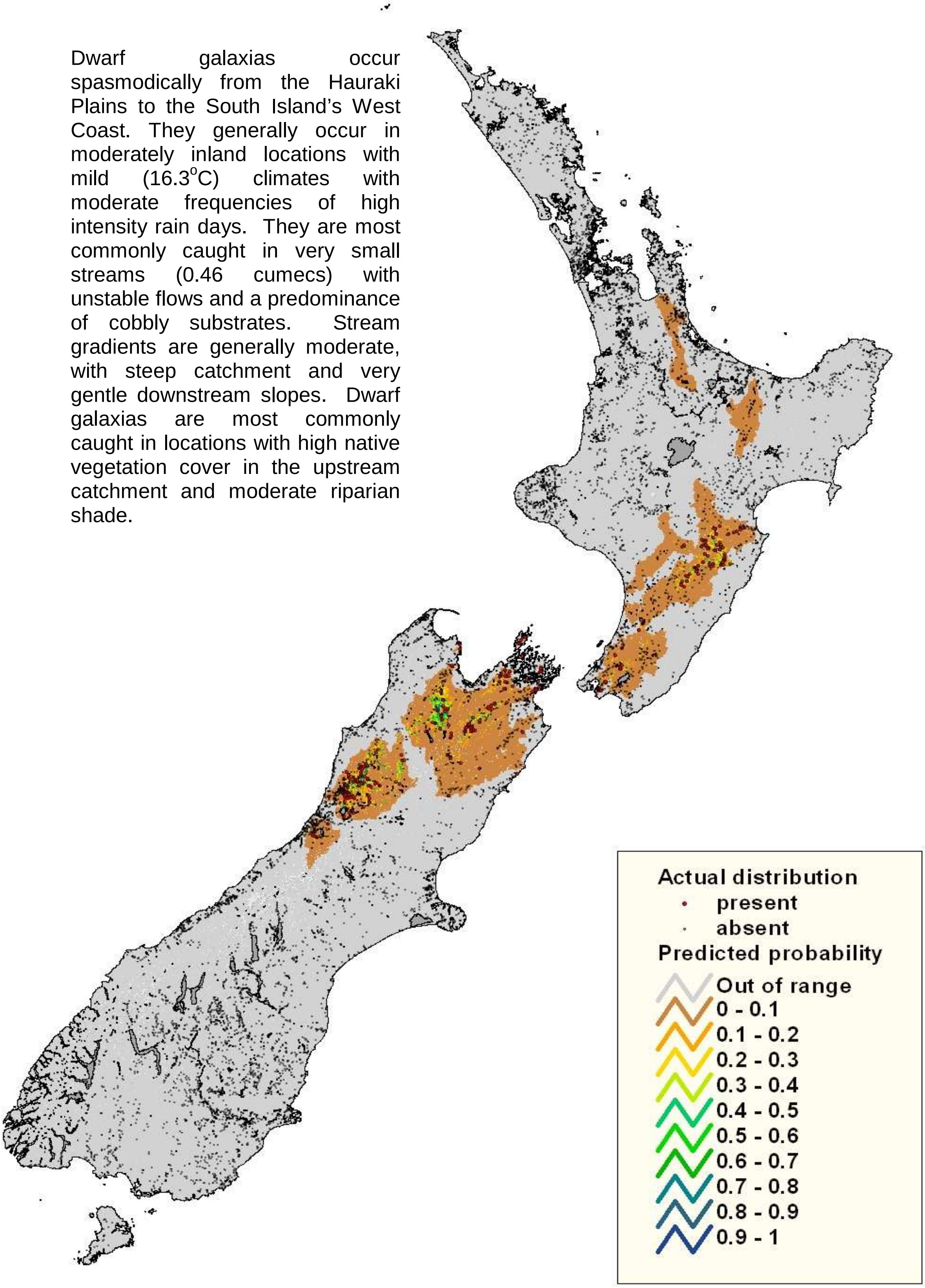

Figure 21: Dwarf galaxias (*Galaxias divergens*).

Eldon's galaxias are found only in eastern Otago, principally in the lower Taieri River. They occur most frequently in moderately inland locations with cold (11.8°C) climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams with unstable flows and a predominance of bouldery substrates. Stream gradients are generally moderate, with gentle catchment and very gentle downstream slopes. Eldon's galaxias are most commonly caught in locations with moderate native vegetation cover in the upstream catchment and moderate riparian shading.

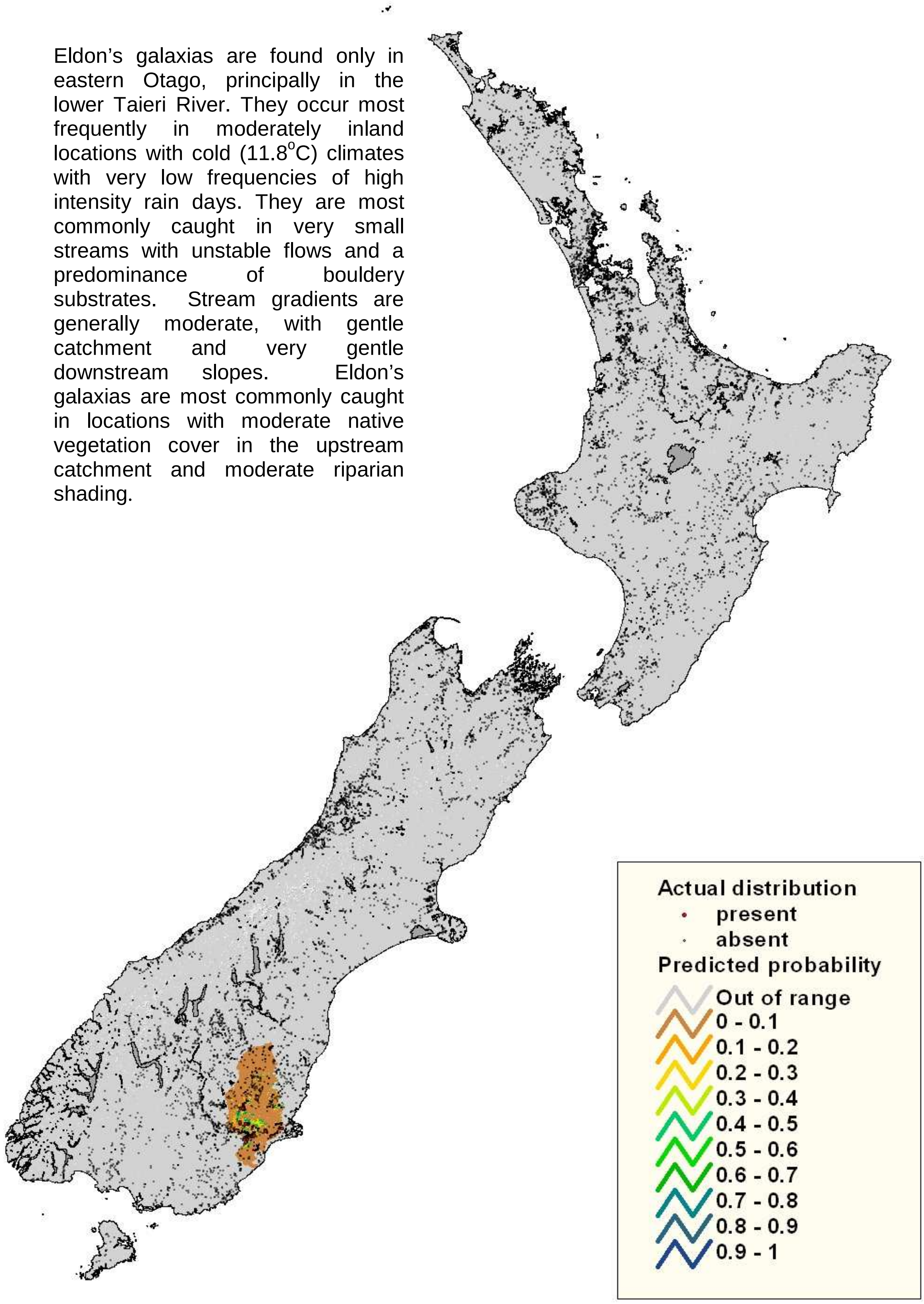

Figure 22: Eldon's galaxias (*Galaxias eldoni*).

Gollum galaxias occurs across Southland from the lower Mataura to the upper Waiau, with small populations in two small, southern tributaries of the Clutha. It is caught most frequently in inland locations with cool (13.7°C), maritime climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams (0.04 cumecs) with unstable flows and a predominance of gravelly substrates. Stream gradients are generally moderate, with gentle catchment and very gentle downstream slopes. This species is most commonly caught in locations with moderate riparian shading and native vegetation cover in the upstream catchment and high nitrogen concentrations.

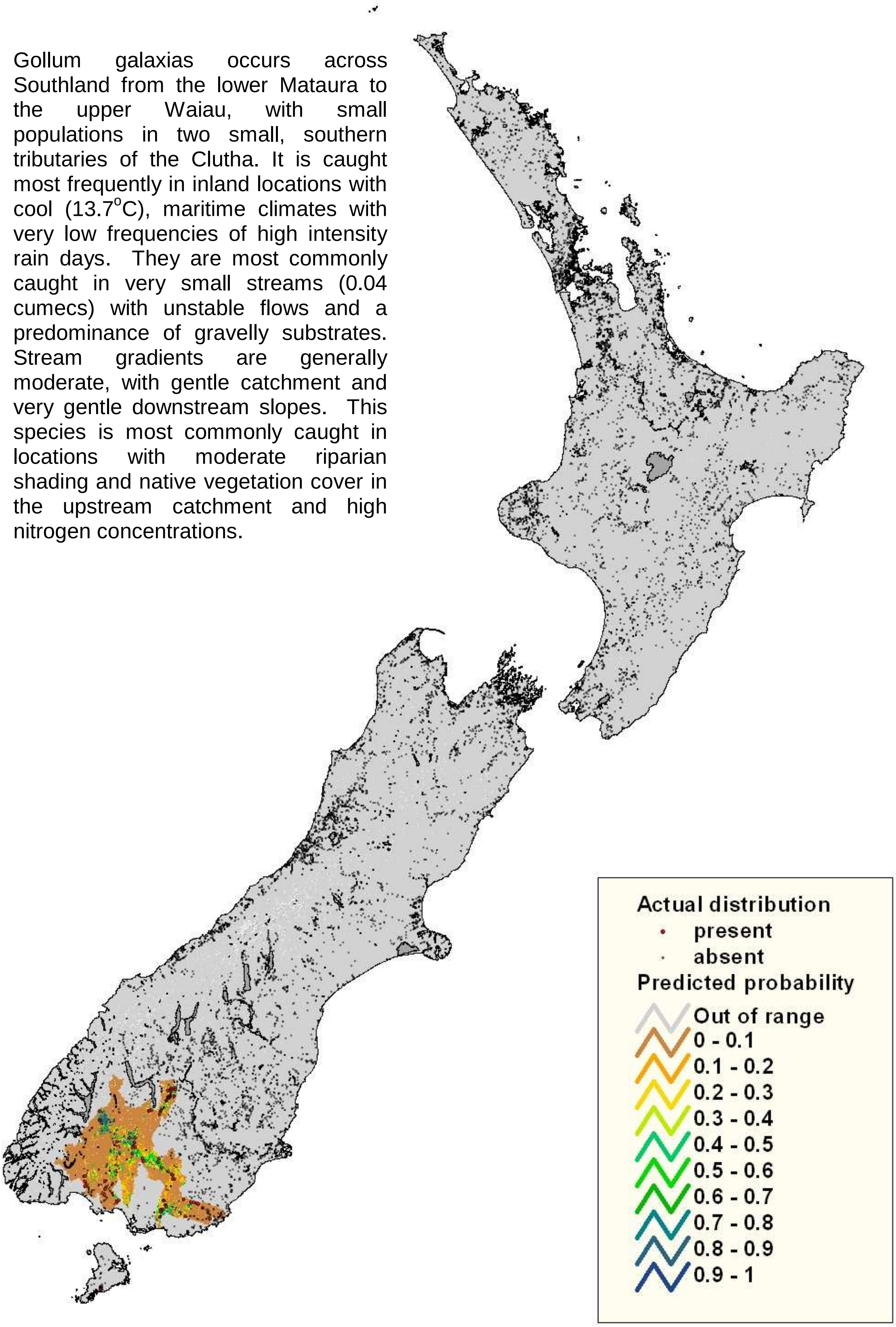

Figure 23: Gollum galaxias (*Galaxias gollumoides*).

Big-nose galaxias are confined to the Waitaki catchment, occurring most frequently in inland locations with mild (15.7^{o}C), strongly continental climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams with stable flows and a predominance of gravelly substrates. Stream and catchment gradients are generally gentle, with very gentle downstream slopes. Big-nose galaxias are most commonly caught in locations with very high native vegetation cover in the upstream catchment, high riparian shading and very high predicted nitrogen concentrations.

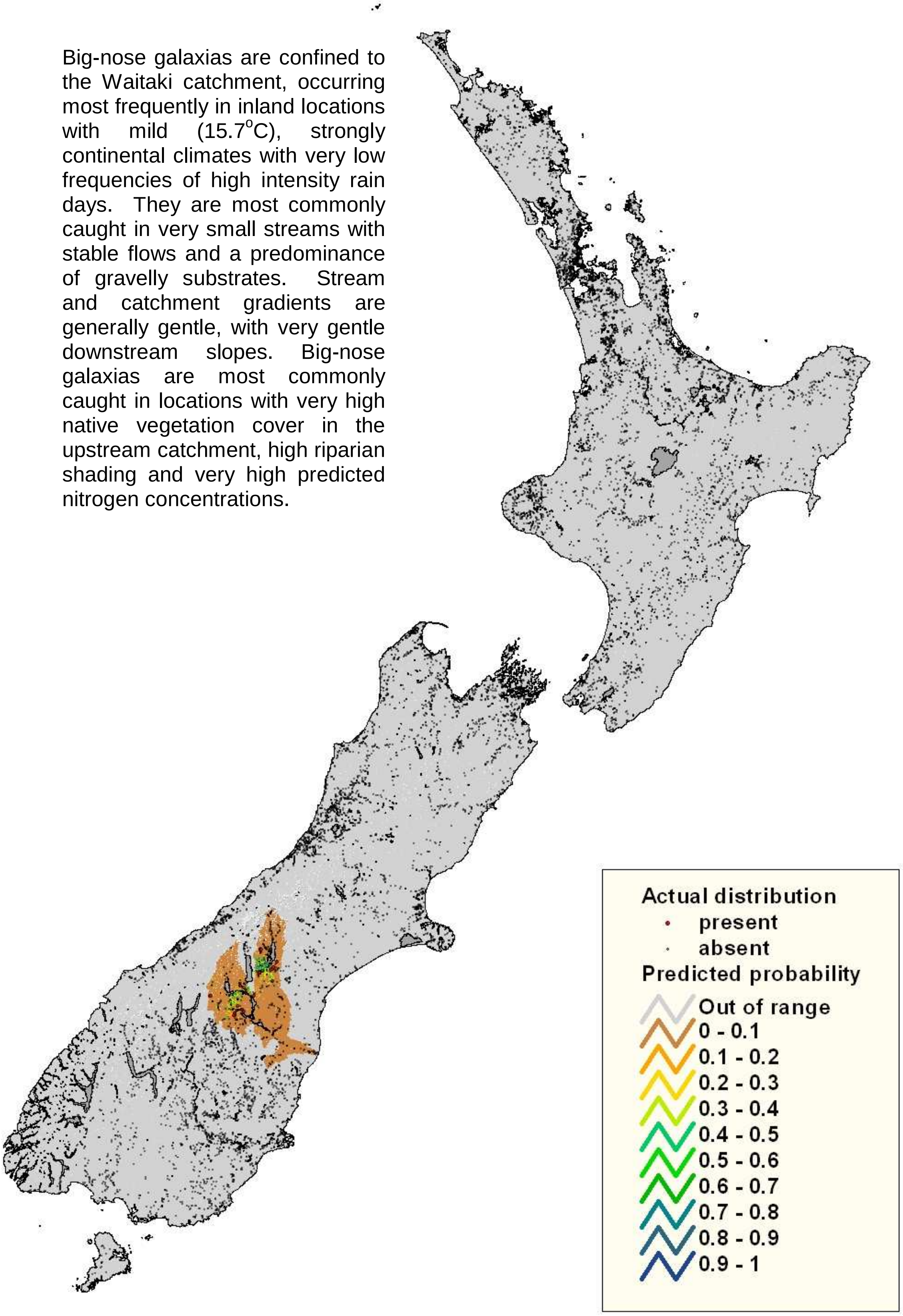

Figure 24: Big-nose galaxias (*Galaxias macronasus*).

Alpine galaxias are widespread in Marlborough and Canterbury, with isolated occurrences including the upper Maruia and Oreti Rivers. They are caught most frequently in inland locations with cool (14.1°C), continental climates with low frequencies of high intensity rain days. They are commonly caught in small upland streams (1.86 cumecs) with stable flows and a predominance of gravelly substrates. Stream gradients are generally moderate, with steep catchment and very gentle downstream slopes. Alpine galaxias are most commonly caught in locations with very high native vegetation cover in the upstream catchment but very low riparian shading.

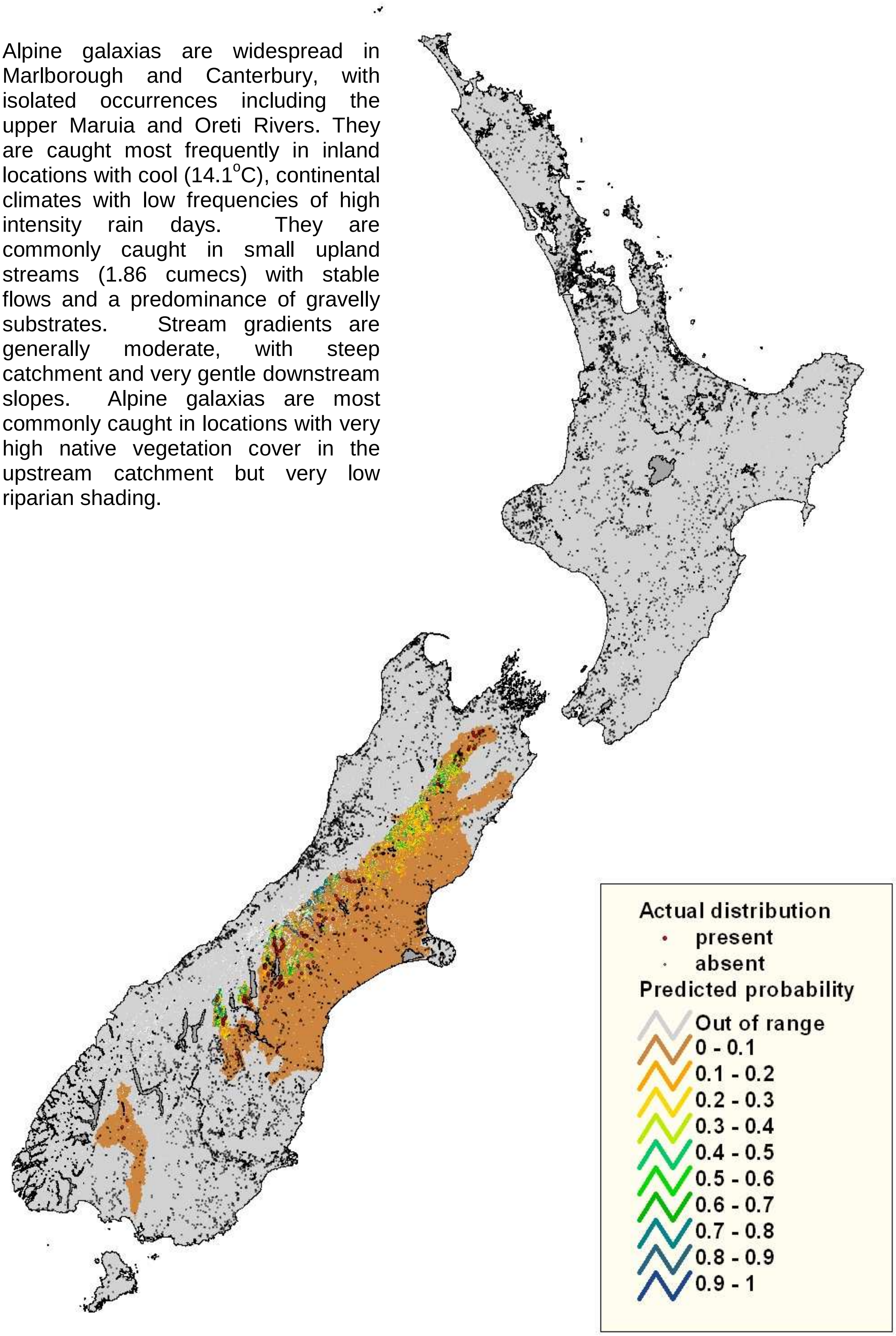

Figure 25: Alpine galaxias (*Galaxias paucispondylus*).

Longjaw galaxias are largely confined to Canterbury, but also occur in the upper Maruia. They are caught most frequently in inland locations with cool (14.3°C), continental climates with moderate frequencies of high intensity rain days. They are most commonly caught in moderate sized streams (9.97 cumecs) with stable flows, and a predominance of gravelly substrates. Stream gradients are generally gentle, with steep catchment and very gentle downstream slopes. Longjaw galaxias are most commonly caught in locations with very high native vegetation cover in the upstream catchment and very low riparian shading.

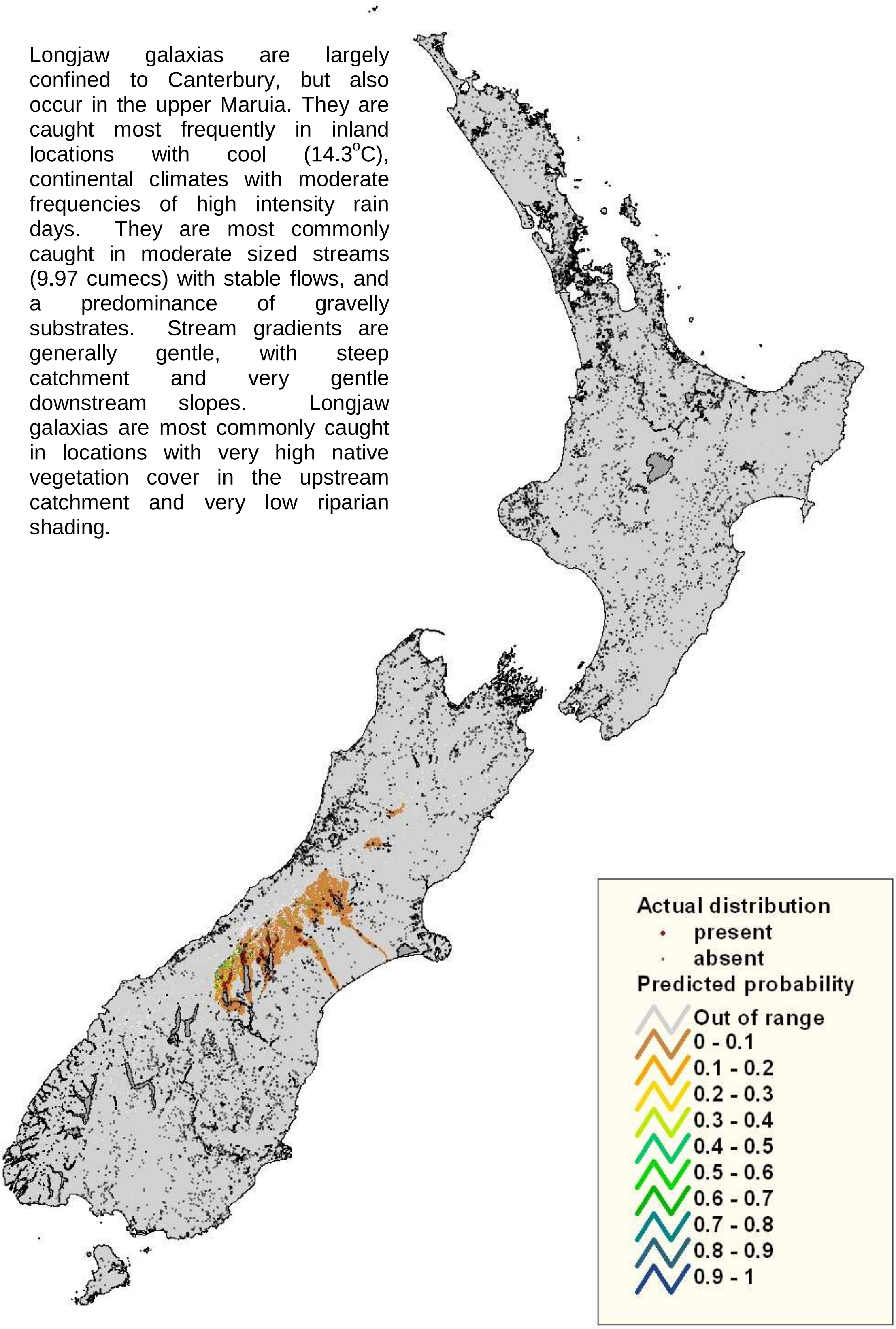

Figure 26: Longjaw galaxias (*Galaxias prognathus*).

Dusky galaxias are largely confined to the Waipori River, although with a few records from tributaries of the Tokomairiro. They are caught most frequently in moderately inland locations with cold (11.8°C), maritime climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams with unstable flows and a predominance of cobbly substrates. Stream gradients are generally moderate, with gentle catchment and very gentle downstream slopes. Dusky galaxias are most commonly caught in locations with very high native vegetation cover in the upstream catchment and moderate riparian shading.

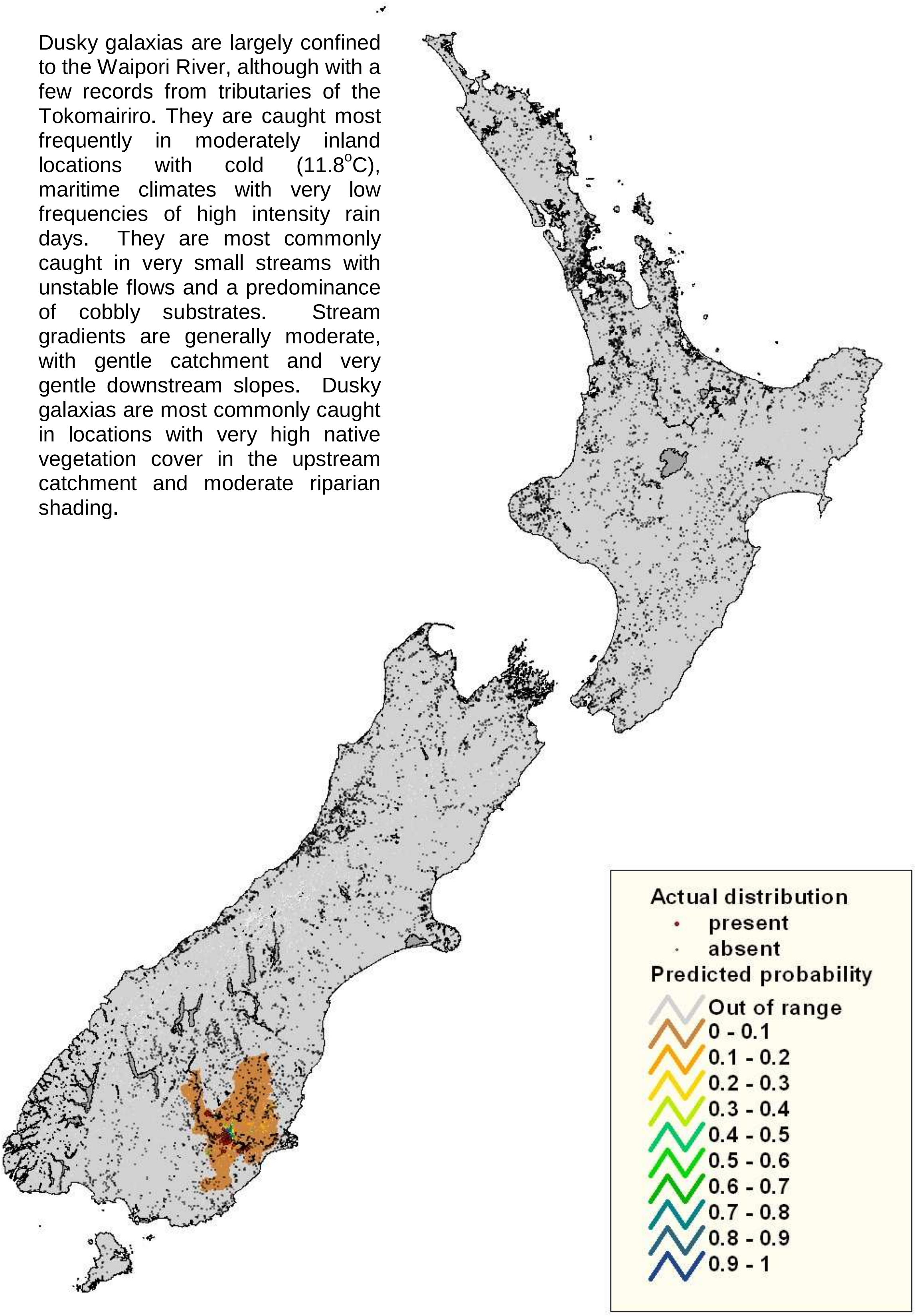

Figure 27: Dusky galaxias (*Galaxias pullus*).

Clutha flathead galaxias are most widespread in the Clutha River, but also occur in smaller coastal catchments to the south. They are caught most frequently in inland locations with cool (14.3^{o}C), continental climates with very low frequencies of high intensity rain days. They are most commonly caught in very small streams with unstable flows and a predominance of gravelly substrates. Stream and catchment gradients are generally moderate, with very gentle downstream slopes. Clutha flathead galaxias are most commonly caught in stream with high native vegetation cover in the upstream catchment, moderate riparian shading and high nitrogen concentrations.

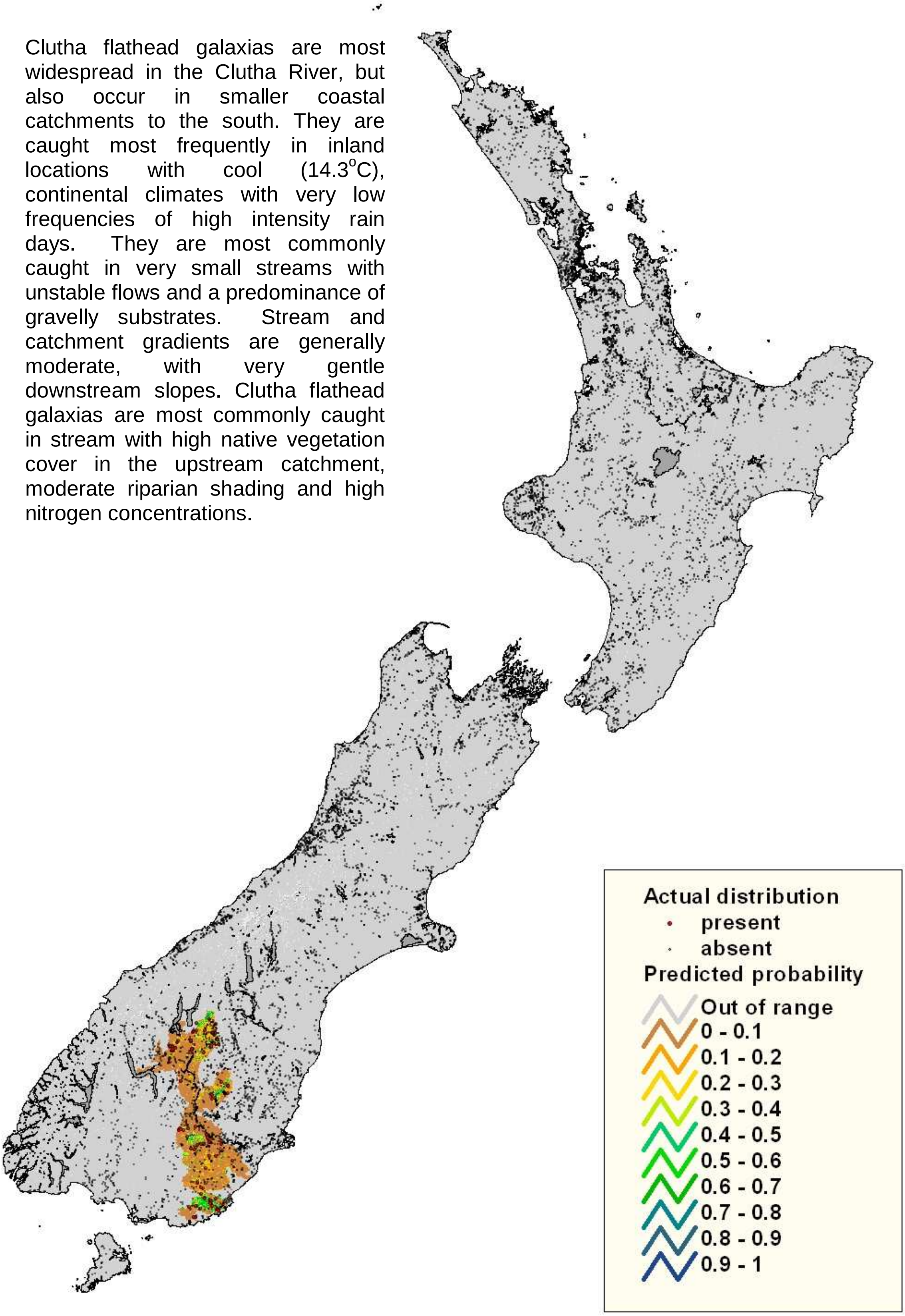

Figure 28: Clutha flathead galaxias (*Galaxias 'species D'*).

Northern flathead galaxias occur in several northern catchments including upper tributaries of the Buller, Motueka, Wairau, Awatere and Clarence Rivers. They are caught most frequently in inland locations with cool (14.9°C), continental climates with low frequencies of high intensity rain days. They are most commonly caught in small streams (0.75 cumecs) with moderately stable flows and a predominance of cobbly substrates. Stream gradients are generally moderate, with steep catchment and very gentle downstream slopes. Northern flathead galaxias are most commonly caught in locations with very high native vegetation cover in the upstream catchment but low riparian shading.

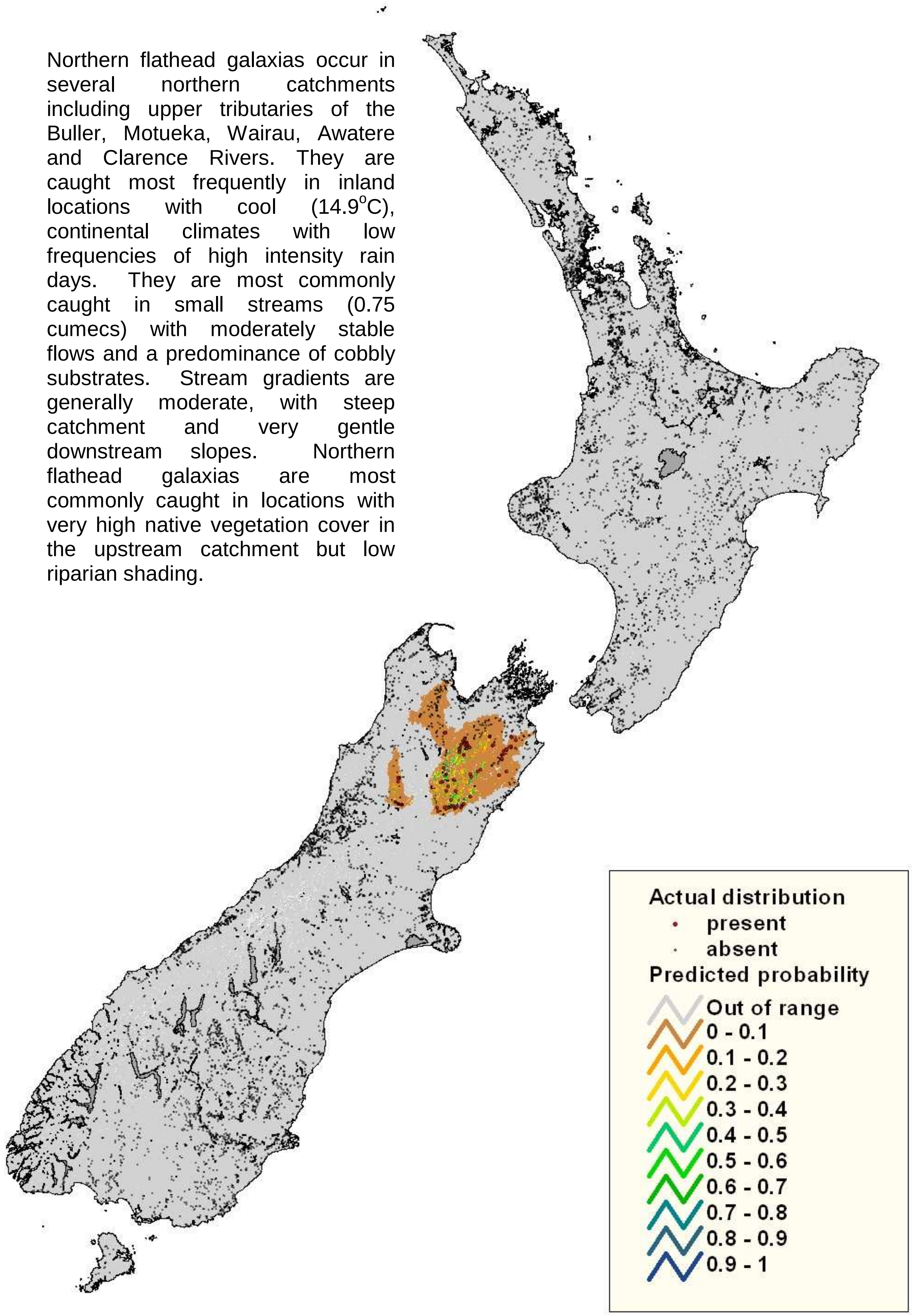

Figure 29: Northern flathead galaxias (*Galaxias 'species N'*).

Canterbury galaxias, as their name suggests, are widespread throughout Canterbury from just north of Kaikoura south to the Kakanui River. They are caught most frequently in moderately inland locations with mild (15.3°C), continental climates with low frequencies of high intensity rain days. They are most commonly caught in small streams (1.68 cumecs) with moderately stable flows and a predominance of coarse gravelly substrates. Stream gradients are generally gentle, with steep catchment and very gentle downstream slopes. Canterbury galaxias are most commonly caught in locations with high native vegetation cover in the upstream catchment but very low riparian shading.

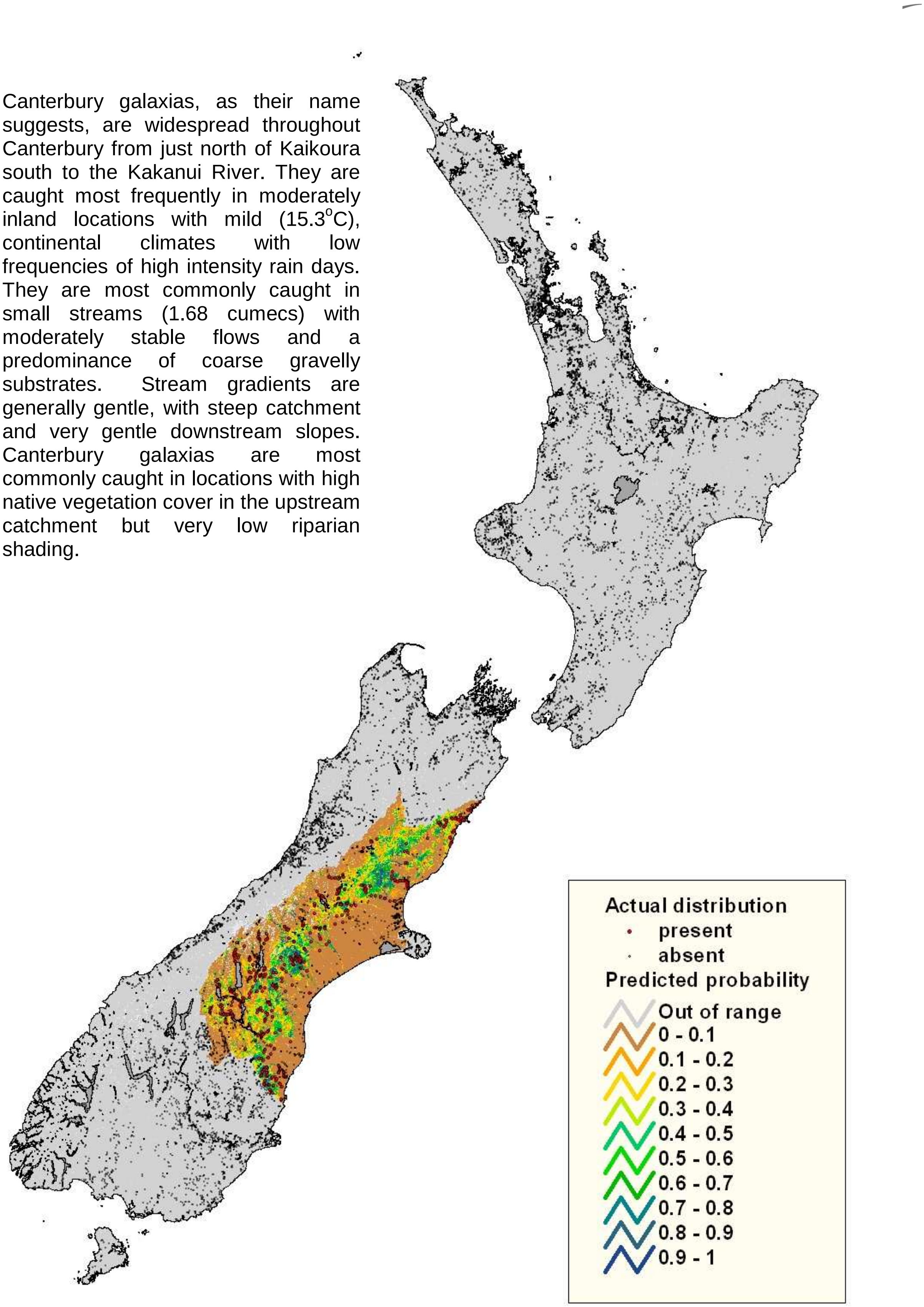

Figure 30: Canterbury galaxias (*Galaxias postvectis*).

Cran's bullies are widely distributed throughout the North Island. They are caught most frequently in moderately inland locations with warm (18.2°C), maritime climates with low frequencies of high intensity rain days. They are most commonly caught in very small streams (0.17 cumecs) with unstable flows and a predominance of gravelly substrates. Stream gradients are generally gentle, with moderate catchment and very gentle downstream slopes. Cran's bullies are most commonly caught in locations with moderate riparian shading and native vegetation cover in the upstream catchment and appear tolerant of high nitrogen concentrations.

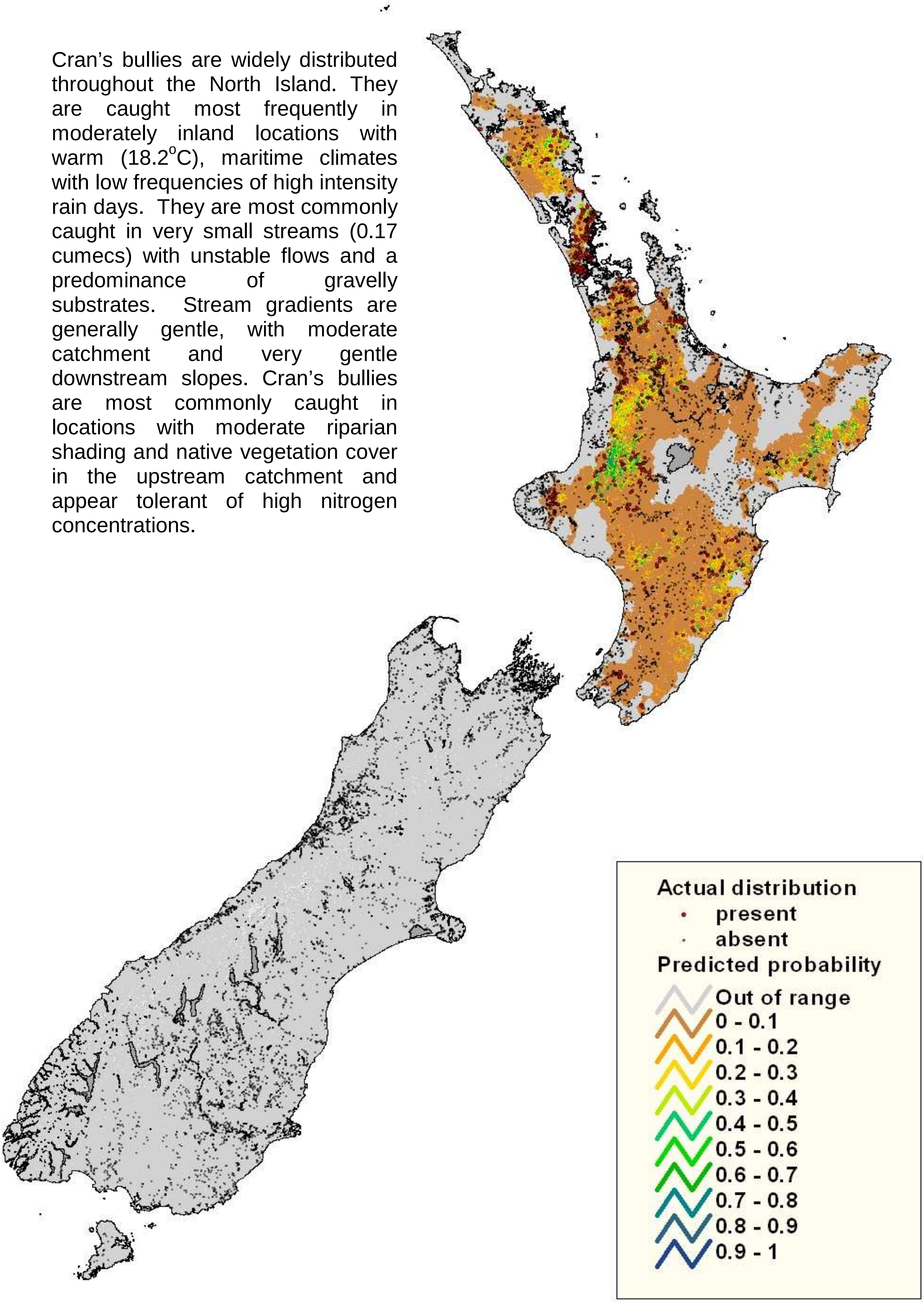

Figure 31: Cran's bully (*Gobiomorphus basalis*).

Upland bullies occur in the southern half of the North Island and through much of the northern and eastern South Island. They are caught most frequently in moderately inland locations with mild (16.1°C), continental climates with very low frequencies of high intensity rain days. They are most commonly caught in small streams (1.14 cumecs) with unstable flows and a predominance of coarse gravelly substrates. Stream and downstream gradients are generally very gentle, with moderate catchment slopes. Upland bullies are most commonly caught in locations with moderate native vegetation cover in the upstream catchment and very low riparian shading.

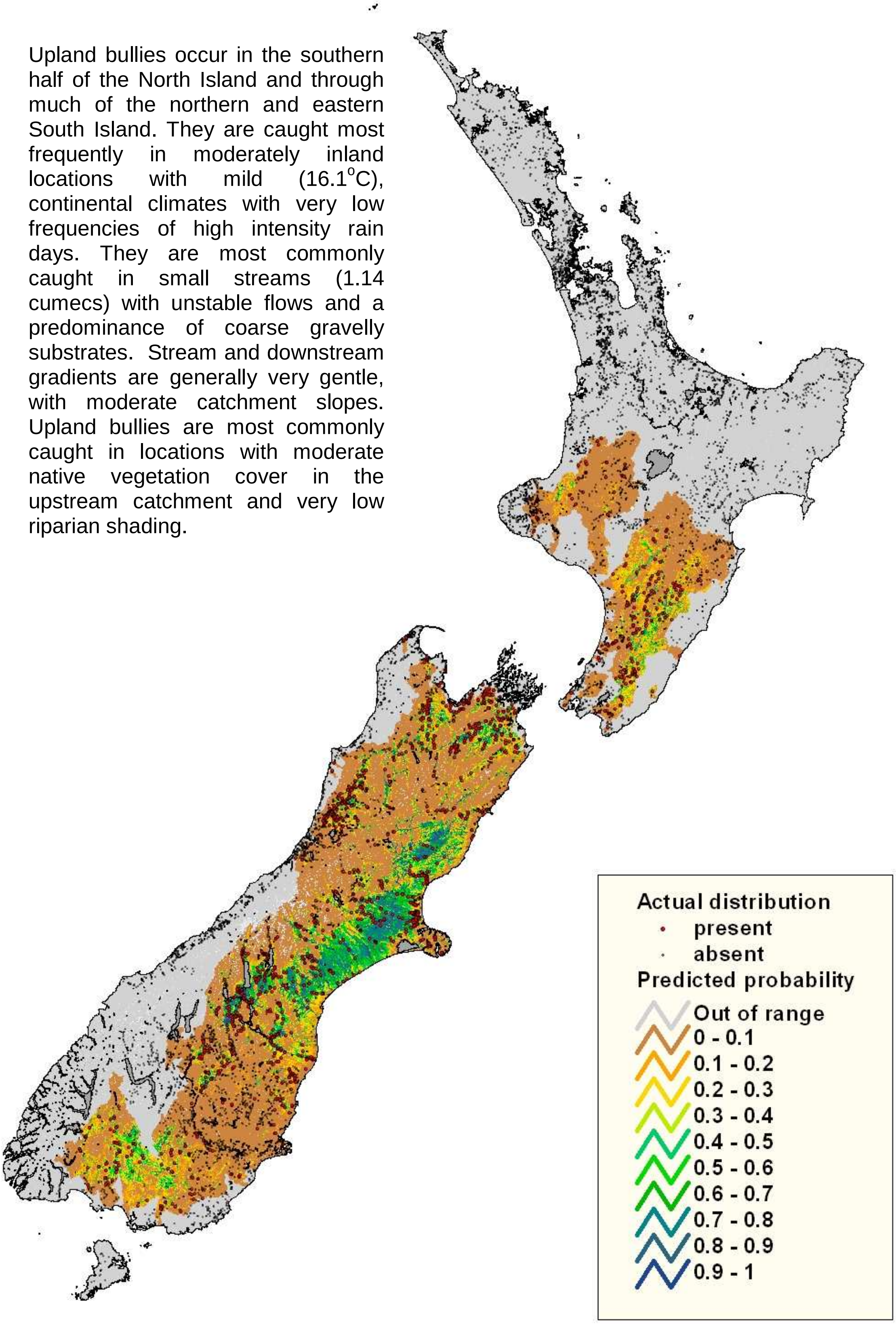

Figure 32: Upland bully (*Gobiomorphus breviceps*).

4. Acknowledgements

This work was funded under the Terrestrial and Freshwater Biodiversity Information System (TFBIS) programme, a programme funded by the Government to help to achieve the goals of the New Zealand Biodiversity Strategy, and administered by the Department of Conservation. A large number of other people also contributed to the outcomes of this study, particularly colleagues from NIWA. Jody Richardson, who has been responsible over a long period for maintenance of the New Zealand Freshwater Fish Database, enabled access to the records used for the analysis, and provided advice on various aspects of the data. Other colleagues provided invaluable input to the development of ecologically sensible predictors, including Ian Jowett, Bob McDowall, John Quinn, David Rowe and Ton Snelder. Helen Roulston provided access to much of the GIS data used in the analyses. Trevor Hastie (Stanford University) provided invaluable advice on the fitting of BRT models, while Lindsay Chadderton and Bruno David provided valuable comment on our predictions of the distributions of individual species.

5. References

Belbin, L. (1995). PATN analysis package. Division of Sustainable Ecosystems, CSIRO, Canberra. 237 p.

Castella, E.; Adalsteinsson, H.; J.E.; Gislason, G.M.; Lehmann, A.; Lencioni, V.; Lods-Crozet, B.; Maiolini, B.; Milner, A.M.; Olafsson, J.S.; Saltveit, S.J. & Snook D.L. (2001). Macrobenthic invertebrate richness and composition along a latitudinal gradient of European glacier-fed streams. *Freshwater Biology 46*: 1811–1831.

Elith, J.; Leathwick, J.R. & Hastie, T. (in review). Boosted regression trees – a new technique for modelling ecological data. Submitted to *Journal of Animal Ecology*.

Flury, B.D. & Levri, E.P. (1999). Periodic logistic regression. *Ecology 80*: 2254-2260.

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology 143*: 29–26.

Jowett, I.G. (1998). Hydraulic geometry of New Zealand rivers and its use as a preliminary method of habitat assessment. *Regulated Rivers: Research and Management 14*: 451–466.

Leathwick, J.R.; Rowe, D.; Richardson, J.; Elith, J. & Hastie, T. (2005). Using multivariate adaptive regression splines to predict the distribution of New Zealand's freshwater diadromous fish. *Freshwater Biology 50*: 2034–2052.

Leathwick, J.R.; Elith, J.; Chadderton, W.L.; Rowe, D.; Hastie, T. (in press). Dispersal, disturbance, and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography*.

McDowall, R.M. (1990). *New Zealand freshwater fishes: a natural history and guide.* Revised edn. Heinemann Reed, Auckland. 553 pp.

McDowall, R.M. (2000). *New Zealand freshwater fishes.* Reed Books, Auckland. 224 pp.

McDowall, R.M. (2006). The taxonomic status, distribution and identification of the *Galaxias vulgaris* species complex in the eastern/southern South Island and Stewart Island. National Institute of Water and Atmospheric Research Client Report CHCDOC2006-081.

McDowall, R.M. & Richardson, J. (1983). The New Zealand freshwater fish survey, a guide to input and output. *New Zealand Ministry of Agriculture and Fisheries, Fisheries Research Division Information Leaflet 12*: 1–15.

Pearson, C.P. (1995). Regional Frequency Analysis of Low Flows in New Zealand's Rivers. *Journal of Hydrology (NZ) 33*: 94–122.

R Development Core Team. (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL http://www.R-project.org.

Ridgeway, G. (2004). gbm: Generalized Boosted Regression Models. R package, version 1.3–5. http://www.i-pensieri.com/grege/gbm.shtml

Snelder, T.H.; Biggs, B.J.F. (2002). Multiscale river environment classification for water resources management. *Journal of the American Water Resources Association 38*: 1225–1239.

Woods, R.; Elliot, S.; Shanker, U.; Bidwell, V.; Harris, S.; Wheeler, D.; Clothier, B.; Green, S.; Hewitt, A.; Gibb, R. & Parfitt, R. (2006). The CLUES Project: Predicting the Effects of Land-use on Water Quality – Stage II. NIWA Client Report MAF05502.

6. Appendix I – Summaries of predictive performance (Tables 1 & 2) and contributions of predictors (Tables 3 & 4) for individual species.

Table 1: Diadromous species predictive performance. Table entries show for each species the final number of trees fitted, the null deviance, and cross-validated estimates and standard errors (se) of the predictive deviance, correlation, and area under the receiver operator characteristic curve (AUC).

Species	No. of trees	Deviance				Correlation		AUC	
		Null	Residual	se	percent	cor	se	auc	se
Angaus	5550	0.8799	0.5518	0.006	37.3%	0.622	0.006	0.895	0.003
Angdie	5000	1.2274	0.817	0.005	33.4%	0.622	0.003	0.857	0.002
Chefos	5000	0.5272	0.3123	0.006	40.8%	0.607	0.010	0.927	0.004
Galarg	2450	0.2357	0.1556	0.005	34.0%	0.467	0.023	0.919	0.005
Galbre	5150	0.6145	0.3919	0.006	36.2%	0.574	0.010	0.901	0.002
Galfas	4700	0.6661	0.3513	0.009	47.3%	0.668	0.010	0.939	0.003
Galmac	4000	0.5854	0.3661	0.005	37.5%	0.549	0.008	0.913	0.003
Galpos	2500	0.2065	0.1291	0.005	37.5%	0.517	0.017	0.933	0.007
Geoaus	2100	0.1961	0.1486	0.004	24.2%	0.350	0.022	0.883	0.010
Gobcot	7650	0.7807	0.5351	0.005	31.5%	0.563	0.006	0.874	0.003
Gobgob	1200	0.1149	0.0797	0.003	30.6%	0.286	0.017	0.941	0.006
Gobhub	4400	0.3133	0.1725	0.006	44.9%	0.641	0.015	0.956	0.003
Gobhut	6650	0.8092	0.4405	0.010	45.6%	0.674	0.008	0.929	0.003
Retret	2750	0.2904	0.1798	0.005	38.1%	0.504	0.019	0.927	0.005
Rhoret	1250	0.0582	0.0419	0.004	28.0%	0.357	0.053	0.941	0.016
Average	4023	0.500	0.312	0.005	36.5	0.533	0.015	0.916	0.005

Table 2: Non-diadromous species predictive performance. Table entries show for each species the final number of trees fitted, the null deviance, and cross-validated estimates and standard errors (se) of the predictive deviance, correlation, and area under the receiver operator characteristic curve (AUC).

Species	No. of trees	Deviance				Correlation		AUC	
		Null	Residual	se	percent	cor	se	auc	se
Galano	4750	0.042	0.018	0.002	56.6%	0.712	0.025	0.991	0.004
Galcob	4050	0.028	0.014	0.002	51.1%	0.584	0.056	0.980	0.009
Galdep	2150	0.094	0.026	0.003	72.3%	0.844	0.016	0.997	0.001
Galdiv	4150	0.212	0.104	0.004	51.2%	0.612	0.014	0.968	0.005
Galeld	1400	0.036	0.010	0.002	71.5%	0.789	0.036	0.997	0.001
Galgol	1500	0.069	0.029	0.003	58.8%	0.639	0.047	0.988	0.005
Galmar	1100	0.031	0.013	0.001	56.6%	0.606	0.038	0.991	0.003
Galpau	1950	0.121	0.051	0.003	58.0%	0.648	0.017	0.984	0.002
Galpro	1000	0.052	0.022	0.003	58.2%	0.630	0.055	0.988	0.004
Galpul	900	0.047	0.016	0.003	65.7%	0.670	0.037	0.995	0.002
Galspd	1100	0.090	0.044	0.002	50.8%	0.487	0.022	0.987	0.002
Galspn	1350	0.070	0.027	0.003	61.1%	0.572	0.044	0.993	0.003
Galvul	3600	0.320	0.119	0.003	63.0%	0.752	0.009	0.985	0.001
Gobbas	4850	0.366	0.188	0.008	48.7%	0.649	0.017	0.958	0.004
Gobbre	6150	0.673	0.346	0.010	48.6%	0.700	0.012	0.947	0.003
Average	2667	0.150	0.068	0.003	58.2%	0.660	0.030	0.983	0.003

Table 3: Diadromous species predictor contributions, ranked in order of decreasing average contribution.

Predictor	**Angaus**	**Angdie**	**Chefos**	**Galarg**	**Galbre**	**Galfas**	**Galmac**	**Galpos**	**Geoaus**	**Gobcot**	**Gobgob**	**Gobhub**	**Gobhut**	**Retret**	**Rhoret**	**Average**
DSAvgSlope	9.5	5.0	3.3	4.6	9.1	3.7	12.7	2.6	31.5	4.9	28.9	3.6	3.9	22.0	33.3	11.9
DSDist	3.7	3.8	6.9	5.4	4.1	26.6	23.7	9.1	8.1	7.9	8.9	13.4	28.4	4.5	4.9	10.6
SegSumT	22.8	13.4	10.9	4.8	4.8	5.8	5.5	4.6	5.3	8.3	8.1	5.7	3.9	10.6	0.6	7.7
DSMaxSlope	13.9	8.0	9.8	5.9	3.2	2.2	12.1	2.6	5.1	14.1	5.7	6.2	5.8	5.9	8.2	7.2
USRainDays	2.2	5.5	5.2	11.7	14.0	4.6	2.7	12.7	4.1	4.4	3.6	6.3	11.8	4.8	3.9	6.5
Method	7.1	11.2	4.3	9.6	1.9	3.5	4.3	10.4	2.0	3.4	6.0	4.2	3.6	4.1	1.3	5.1
ReachSubstrate	3.0	6.4	5.6	3.0	4.7	2.4	3.6	6.9	4.5	5.3	3.9	4.6	5.1	4.4	6.3	4.6
SegTSeas	2.2	6.7	2.3	4.1	4.1	5.4	3.7	4.1	3.8	3.8	2.8	3.9	5.3	2.0	4.5	3.9
SegFlowStability	2.3	2.8	5.2	7.0	3.1	2.6	2.9	5.2	3.1	4.1	3.6	5.1	3.4	2.4	2.3	3.7
SegLowFlow	1.6	1.9	11.9	2.7	2.6	5.2	3.8	2.6	3.3	2.8	1.1	6.7	1.8	4.2	2.5	3.6
SegShade	2.4	1.9	1.5	3.1	3.3	11.1	1.7	1.9	2.3	4.5	1.7	2.1	1.4	6.0	8.1	3.5
SegN	5.6	2.2	3.5	2.9	5.6	1.8	2.5	8.0	2.4	3.4	2.3	5.8	2.0	1.6	1.3	3.4
USSlope	3.0	3.0	2.5	2.7	6.9	2.6	2.1	3.9	3.3	4.1	1.2	7.1	5.3	1.2	1.5	3.4
USAvgT	1.3	1.8	7.9	4.6	2.5	1.9	2.5	3.3	2.8	2.1	1.9	5.6	2.9	5.4	2.1	3.2
SegSlope	2.3	1.9	1.5	5.6	5.8	4.6	2.6	3.1	2.7	4.1	5.7	2.8	1.5	1.4	2.2	3.2
USPhosphorus	2.5	2.2	2.9	2.9	2.7	3.0	2.2	3.3	2.4	3.0	2.4	3.0	3.2	4.6	3.9	2.9
ReachHabitat	2.2	2.0	4.1	4.4	3.4	2.7	2.0	2.6	2.2	2.7	2.6	3.4	2.2	2.1	3.0	2.8
USHardness	1.7	2.7	2.6	3.0	2.2	1.5	1.8	3.2	3.4	2.4	1.8	2.7	2.4	2.7	0.9	2.3
USNative	4.4	1.5	1.1	3.0	4.3	2.1	1.5	3.9	2.1	2.5	2.2	2.3	2.0	1.1	0.6	2.3
USCalcium	2.1	1.1	2.3	1.9	1.8	3.2	1.1	2.9	1.8	2.4	1.4	2.5	1.5	2.8	3.0	2.1
TCos	1.9	2.9	0.8	1.4	1.2	0.8	3.0	0.9	1.7	2.0	1.8	1.6	1.0	4.0	2.1	1.8
TSin	1.3	0.6	0.8	2.0	1.0	0.7	1.2	2.1	1.6	1.4	1.7	1.3	0.7	1.4	2.0	1.3
DSDist2Lake	0.3	0.9	1.1	2.3	7.1	1.5	0.1	0.2	0.2	3.8	0.0	0.2	0.6	0.1	0.9	1.3
DSDam	0.1	9.9	1.3	0.0	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.8
USPeat	0.2	0.3	0.2	1.6	0.1	0.3	0.7	0.0	0.1	0.5	0.8	0.1	0.2	0.4	0.1	0.4
USLake	0.4	0.5	0.5	0.1	0.2	0.1	0.2	0.0	0.1	2.1	0.0	0.1	0.1	0.5	0.0	0.3
USGlacier	0.0	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.6	0.1

Table 4: Non-diadromous species predictor contributions, ranked in order of decreasing average contribution.

Predictor	**Galano**	**Galcob**	**Galdep**	**Galdiv**	**Galeld**	**Galgol**	**Galmar**	**Galpau**	**Galpro**	**Galpul**	**Galspd**	**Galspn**	**Galvul**	**Gobbas**	**Gobbre**	**Average**
SegTSeas	8.4	4.2	7.6	10.0	2.4	19.1	11.2	9.1	10.2	8.1	8.1	9.8	13.7	5.2	20.7	9.8
USRainDays	13.7	4.6	14.6	10.7	5.4	7.0	2.7	7.1	4.3	17.4	12.2	5.5	10.6	5.2	5.9	8.5
DSDist	9.8	4.3	8.9	9.9	2.5	12.2	5.1	7.0	2.3	9.0	13.4	11.2	6.2	15.5	5.9	8.2
DSMaxSlope	10.9	0.3	17.9	9.4	3.8	5.7	0.0	3.4	1.1	7.8	13.4	10.5	3.7	4.8	8.6	6.8
SegSumT	1.7	1.9	10.0	5.9	7.9	6.8	1.0	9.6	9.5	6.3	2.8	2.2	2.2	12.3	11.4	6.1
SegFlowStability	6.4	2.1	3.0	7.6	5.7	6.4	3.9	10.3	3.8	9.4	8.3	2.9	4.1	3.2	2.3	5.3
USCalcium	6.6	4.7	0.9	3.7	0.5	3.4	3.6	7.1	19.0	0.0	0.7	5.3	9.5	3.7	3.2	4.8
USNative	7.5	22.9	1.1	1.1	17.3	3.4	1.4	0.8	1.9	3.3	3.1	0.8	1.0	1.8	1.3	4.6
DSAvgSlope	0.3	1.7	1.3	4.0	2.9	4.6	4.2	9.1	0.8	8.5	3.8	4.9	12.7	4.4	2.5	4.4
USPhosphorus	4.7	5.1	4.4	5.7	1.1	2.4	17.9	2.9	1.5	0.0	0.7	1.8	5.8	5.1	4.4	4.2
USSlope	2.9	8.3	2.9	5.4	2.0	2.7	2.7	4.9	2.8	3.8	2.6	8.3	4.3	2.9	3.0	4.0
SegN	1.4	3.9	3.0	3.7	3.1	1.0	6.0	3.3	2.2	6.2	8.4	5.6	3.1	3.3	4.5	3.9
SegShade	4.0	7.8	0.9	1.3	4.2	0.9	9.6	4.8	2.0	2.3	3.9	3.5	3.4	3.0	4.0	3.7
ReachSediment	1.6	3.9	4.7	3.1	3.9	2.9	6.3	1.5	3.5	1.1	4.5	2.0	1.6	8.1	2.6	3.4
USAvgT	1.4	2.6	2.1	2.1	7.3	4.2	1.4	3.1	6.9	3.6	2.5	2.2	2.5	2.3	1.9	3.1
TCos	1.3	7.5	2.8	1.7	14.7	2.2	0.6	1.4	1.7	1.0	0.9	5.9	0.7	1.7	1.0	3.0
DSDist2Lake	9.7	0.2	1.1	1.5	0.6	0.3	14.0	3.0	1.1	0.9	1.5	5.6	1.0	1.8	2.0	3.0
ReachHabitat	2.7	2.1	0.6	2.6	10.7	1.0	3.6	2.1	3.4	1.8	3.6	1.4	1.6	2.0	1.4	2.7
USHardness	1.4	2.0	0.8	3.8	1.5	3.8	1.1	1.8	3.7	4.5	2.1	1.7	2.0	2.2	1.8	2.3
SegSlope	1.7	2.3	0.7	1.6	1.7	4.1	0.8	2.4	2.1	1.5	2.1	1.9	1.5	2.7	6.0	2.2
SegLowFlow	0.8	6.8	0.6	1.9	0.6	1.8	2.0	1.7	2.3	0.5	0.4	5.9	2.2	2.3	1.8	2.1
USGlacier	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	11.8	0.0	0.0	0.0	4.8	0.0	0.8	1.3
TSin	0.3	0.7	1.2	1.9	0.3	4.1	0.9	1.0	1.4	1.8	0.5	1.1	1.0	1.2	0.7	1.2
DSDam	0.5	0.0	8.9	0.3	0.0	0.1	0.0	0.4	0.0	1.2	0.1	0.0	0.0	0.1	0.1	0.8
Method	0.4	0.0	0.0	1.2	0.1	0.0	0.1	0.3	0.4	0.0	0.6	0.0	0.9	3.8	1.9	0.6
USLake	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.9	0.1	0.1
USPeat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.1	0.0