Appendix 2. Supporting tables and figures

Tables:

Table S1. Overview of species by survey data providing a summary of presence records across different monitoring schemes in our analysis.

Table S2. Summary of species distribution model performance across all data for MCC.

Table S3. Summary of species distribution model performance across all data for TSS.

Table S4. Summaries of species shadow distribution properties as an average and standard deviation across all species.

Figures:

Figure S1. Difference in model performance depending on subsets of the complete compiled dataset.

Figure S2. Coverage of environmental data in each data subset used for sensitivity testing.

Figure S3. Boxplots of model performance across 8 model evaluation metrics (see Table S2-3). Colour indicates which metric is optimised under the binary threshold method (either TSS or MCC).

Figure S4. Comparison of environmental properties of Sense and Emme rivers.

Figure S5. Spatial distribution of habitat suitability for the 9 focal species based on presence-absence random forest models.

Figure S6 – Figure S13. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Barbus barbus, Cottus gobio, Gobio gobio, Lampetra planeri, Perca fluviatilis, Oncorhynchus mykiss, Squalius cephalus, Thymallus thymallus*

Figure S14. Shapley based response curves across all species and variables. Colour indicates variable importance estimated as the average absolute Shapley value.

Figure S15. Spatial maps of environmental data used as covariates when fitting species distribution models.

Figure S16. Qualitative shadow distributions for all species to understand the relative influence of anthropic threats within species' expected distributions.

Figure S17-18. Multi-species average shadow distribution properties per sub-catchment with alternative methods.

Table S1. Overview of species by survey data providing a summary of presence records across different monitoring schemes in our analysis. University of Bern 2022 refers to the summer sampling campaign undertaken by the manuscript authors. Projetto Fiumi was a Switzerland wide river biodiversity monitoring project coordinated by Jakob Brodersen and Ole Seehausen at EAWAG (1). Kanton Bern provided monitoring data undertaken by the fisheries authorities. Consultancies refers to data compiled by Dr. Pascal Vonlanthen and Dr. Sebastien Lauper from various datasets across private and federal surveys and monitoring initiatives (2).

	University of Bern 2022	EAWAG Projetto Fiumi[1]	Kanton Bern	Consultancies[2]	Total
Alburnoides bipunctatus	**10**	**7**	**32**	**61**	**110**
Barbus barbus	**12**	**24**	**34**	**99**	**169**
Cottus gobio	**30**	**59**	**558**	**247**	**894**
Gobio gobio	**11**	**11**	**34**	**79**	**135**
Lampetra planeri	**2**	**2**	**23**	**59**	**86**
Oncorhynchus mykiss	**2**	**3**	**66**	**11**	**82**
Perca fluviatilis	**6**	**8**	**33**	**33**	**80**
Squalius cephalus	**22**	**24**	**100**	**170**	**316**
Thymallus thymallus	**4**	**3**	**27**	**27**	**61**

1. Brodersen, J., Hellmann, J., & Seehausen, O. (2023). *Erhebung der Fischbiodiversität in Schweizer Fliessgewässern. Progetto Fiumi Schlussbericht*. Eawag: Swiss Federal Institute of Aquatic Science and Technology. https://doi.org/10.55408/eawag:30020

2. https://modul-stufen-konzept.ch/en/fish/

Table S2. Summary of species distribution model performance across all data when the threshold between predicted presence and absence is defined by optimizing the Matthews Correlation Coefficient.

Metrics	*Alburnoides bipunctatus*	*Barbus barbus*	*Cottus gobio*	*Gobio gobio*	*Lampetra planeri*	*Oncorhynchus mykiss*	*Perca fluviatilis*	*Squalius cephalus*	*Thymallus thymallus*
TN	590 (±51)	580 (±26)	490 (±88)	580 (±94)	570 (±52)	520 (±110)	610 (±41)	520 (±70)	620 (±45)
FN	6.4 (±3.4)	8.6 (±6.8)	72 (±25)	6.6 (±3.8)	3.8 (±4)	7.4 (±5.5)	8.4 (±2.7)	7.4 (±7.1)	2.8 (±1.1)
TP	10 (±6.1)	16 (±3.7)	82 (±27)	15 (±3.5)	8.6 (±3.2)	8 (±7.4)	3.4 (±3.8)	40 (±12)	6.2 (±4.3)
FP	26 (±20)	34 (±42)	140 (±56)	40 (±34)	51 (±58)	100 (±120)	12 (±27)	67 (±30)	10 (±6.7)
MCC	0.4 (±0.068)	0.49 (±0.091)	0.27 (±0.043)	0.44 (±0.12)	0.4 (±0.17)	0.16 (±0.041)	0.35 (±0.068)	0.52 (±0.051)	0.49 (±0.072)
Overprediction rate	0.68 (±0.088)	0.54 (±0.23)	0.63 (±0.04)	0.66 (±0.15)	0.67 (±0.24)	0.86 (±0.11)	0.22 (±0.37)	0.61 (±0.081)	0.51 (±0.3)
Underprediction rate	0.4 (±0.23)	0.33 (±0.22)	0.47 (±0.17)	0.31 (±0.17)	0.28 (±0.29)	0.53 (±0.39)	0.73 (±0.26)	0.16 (±0.17)	0.36 (±0.24)
Sorensen index	0.39 (±0.06)	0.48 (±0.12)	0.42 (±0.061)	0.43 (±0.14)	0.36 (±0.21)	0.14 (±0.023)	0.25 (±0.093)	0.52 (±0.061)	0.45 (±0.11)
Jaccard index	0.24 (±0.046)	0.32 (±0.096)	0.27 (±0.047)	0.28 (±0.11)	0.24 (±0.16)	0.077 (±0.014)	0.15 (±0.064)	0.35 (±0.056)	0.3 (±0.09)
Sensitivity	0.6 (±0.23)	0.67 (±0.22)	0.53 (±0.17)	0.69 (±0.17)	0.72 (±0.29)	0.47 (±0.39)	0.27 (±0.26)	0.84 (±0.17)	0.64 (±0.24)
Specificity	0.96 (±0.033)	0.95 (±0.06)	0.77 (±0.095)	0.93 (±0.051)	0.92 (±0.09)	0.84 (±0.19)	0.98 (±0.042)	0.88 (±0.056)	0.98 (±0.011)
TSS	0.56 (±0.21)	0.62 (±0.18)	0.3 (±0.084)	0.63 (±0.15)	0.64 (±0.22)	0.31 (±0.22)	0.25 (±0.22)	0.72 (±0.12)	0.62 (±0.24)
AUC	0.92 (±0.029)	0.93 (±0.036)	0.7 (±0.035)	0.91 (±0.043)	0.9 (±0.078)	0.66 (±0.14)	0.86 (±0.07)	0.92 (±0.026)	0.92 (±0.049)

Table S3. Summary of species distribution model performance across all data when the threshold between predicted presence and absence is defined by optimizing the True Skill Statistic.

Metrics	*Alburnoides bipunctatus*	*Barbus barbus*	*Cottus gobio*	*Gobio gobio*	*Lampetra planeri*	*Oncorhynchus mykiss*	*Perca fluviatilis*	*Squalius cephalus*	*Thymallus thymallus*
TN	520 (±50)	540 (±37)	420 (±80)	530 (±93)	550 (±31)	410 (±120)	510 (±71)	500 (±42)	540 (±74)
FN	1.4 (±1.5)	3 (±3.5)	54 (±14)	2.8 (±3)	1.4 (±2.1)	4.4 (±4.8)	1.8 (±2)	3.6 (±2.4)	0.6 (±0.55)
TP	15 (±5.3)	21 (±1.7)	100 (±17)	18 (±3.2)	11 (±2.1)	11 (±4.3)	10 (±1.9)	44 (±10)	8.4 (±4.4)
FP	97 (±36)	75 (±24)	210 (±74)	89 (±44)	77 (±34)	210 (±120)	110 (±47)	94 (±13)	85 (±59)
MCC	0.32 (±0.033)	0.41 (±0.053)	0.27 (±0.044)	0.36 (±0.087)	0.32 (±0.095)	0.14 (±0.073)	0.24 (±0.056)	0.49 (±0.064)	0.3 (±0.13)
Overprediction rate	0.86 (±0.027)	0.77 (±0.064)	0.67 (±0.04)	0.81 (±0.066)	0.86 (±0.067)	0.94 (±0.032)	0.91 (±0.029)	0.68 (±0.054)	0.87 (±0.09)
Underprediction rate	0.083 (±0.096)	0.11 (±0.11)	0.35 (±0.09)	0.13 (±0.14)	0.1 (±0.15)	0.26 (±0.24)	0.15 (±0.16)	0.082 (±0.063)	0.065 (±0.067)
Sorensen index	0.24 (±0.04)	0.36 (±0.072)	0.44 (±0.03)	0.31 (±0.09)	0.24 (±0.1)	0.11 (±0.051)	0.16 (±0.048)	0.47 (±0.067)	0.22 (±0.14)
Jaccard index	0.14 (±0.025)	0.22 (±0.054)	0.28 (±0.024)	0.18 (±0.066)	0.14 (±0.066)	0.058 (±0.029)	0.086 (±0.028)	0.31 (±0.056)	0.13 (±0.087)
Sensitivity	0.92 (±0.096)	0.89 (±0.11)	0.65 (±0.09)	0.87 (±0.14)	0.9 (±0.15)	0.74 (±0.24)	0.85 (±0.16)	0.92 (±0.063)	0.93 (±0.067)
Specificity	0.84 (±0.057)	0.88 (±0.034)	0.67 (±0.11)	0.86 (±0.071)	0.88 (±0.052)	0.66 (±0.2)	0.81 (±0.08)	0.84 (±0.019)	0.86 (±0.096)
TSS	0.76 (±0.074)	0.76 (±0.088)	0.32 (±0.054)	0.72 (±0.12)	0.78 (±0.14)	0.4 (±0.21)	0.67 (±0.13)	0.76 (±0.064)	0.8 (±0.048)
AUC	0.92 (±0.029)	0.93 (±0.036)	0.7 (±0.035)	0.91 (±0.043)	0.9 (±0.078)	0.66 (±0.14)	0.86 (±0.07)	0.92 (±0.026)	0.92 (±0.049)

Table S4. Summaries of species shadow distribution properties as an average and standard deviation across all species.

Property of Shapley values and/or shadow distribution	Mean	SD
% all subcatchments with a positive contribution of discharge	36	4.36
% all subcatchments with a positive contribution of flow velocity	53.43	7.98
% all subcatchments with a positive contribution of slope	53	
% all subcatchments with a positive contribution of distance to lake	50	7.95
% all subcatchments with a positive contribution of minimum temperature	55.89	12.55
% all subcatchments with a positive contribution of maximum temperature	56.22	8.87
% of subcatchments inside niche with 0 threats negative	12.44	8.53
% of subcatchments inside niche with 1 threats negative	32.44	10.6
% of subcatchments inside niche with 2 threats negative	32	6.34
% of subcatchments inside niche with 3 threats negative	18.62	8.38
% of subcatchments inside niche with 4 threats negative	8.33	6.74
% of subcatchments inside niche with 5 threats negative	4	0
% reduction in suitability in threatened areas within niche	-24.78	6.98
% sub-catchments inside ecological niche	45.44	7.13
% sub-catchments with positive shapley values for all natural niche variables	7	2.29
% subcatchments inside niche with a negative contribution of river morphological modification	45.25	5.2
% subcatchments inside niche with a negative contribution of connectivity	17.62	4
% subcatchments inside niche with a negative contribution of floodplain proportion	70.29	7.41
% subcatchments inside niche with a negative contribution of urbanisation	50.89	5.4
% subcatchments inside niche with a negative contribution of wetland proportion	49.5	23.33
% subcatchments inside niche with a negative net effect of all threats summed	22.78	12.43
% subcatchments inside niche with at least one negative threat	87.56	8.53
% threatened catchments in niche with predicted absence	78.56	15.74
% unthreatened catchments in niche with predicted absence	50.11	26.21
mean suitability in threatened sub-catchments	0.48	0.07
mean suitability in unthreatened sub-catchments	0.63	0.08
number of times more absences in threatened compared to unthreatened base rate	2.04	1.18

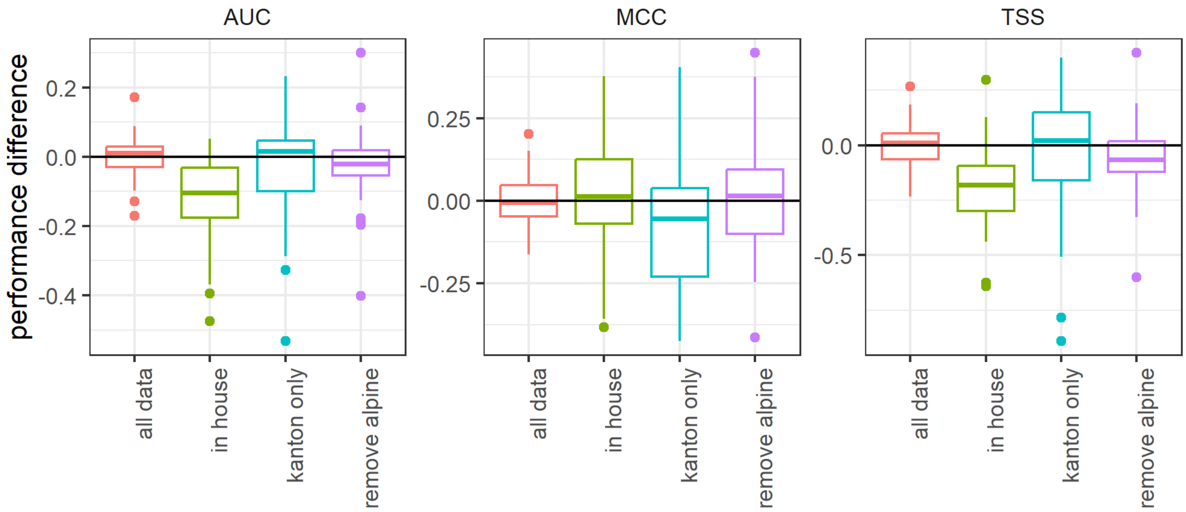

Figure S1. Difference in model performance depending on subsets of the complete compiled dataset. Each panel refers to a difference performance metric as (a) area under receiver operator curve (AUC), (b) Matthews correlation coefficient (MCC) and (c) True Skill Statistic (TSS). We the following subsets i) only "in-house" surveys based on scientific research teams and standardized national monitoring data, ii) Canton of Bern monitoring data, iii) all data but removing up-stream alpine catchments that are generally disconnected, cold and fast flowing rivers. We used analysis of variance to test for differences in the mean performance compared with models using all the data. We designed these subsets to check whether performance was increased by including additional cantonal data sources. The mean of inhouse performance for AUC and for TSS was significantly reduced compared to models including all the data ($p<0.001$). All other comparisons between subsets of the data and the full dataset were not significant. These results indicate that high model performance was not driven by different environments of the upstream alpine catchments. Secondly, these results also indicate that the model performance was driven largely by the high quantity of data provided by cantonal monitoring sources compared to in house monitoring. Note that the model evaluations are performed internally on the subset data sources rather than across data sources. Note also that boxplots contain underlying 5-fold variation in blocked cross validations as well as between species variation in model performance.

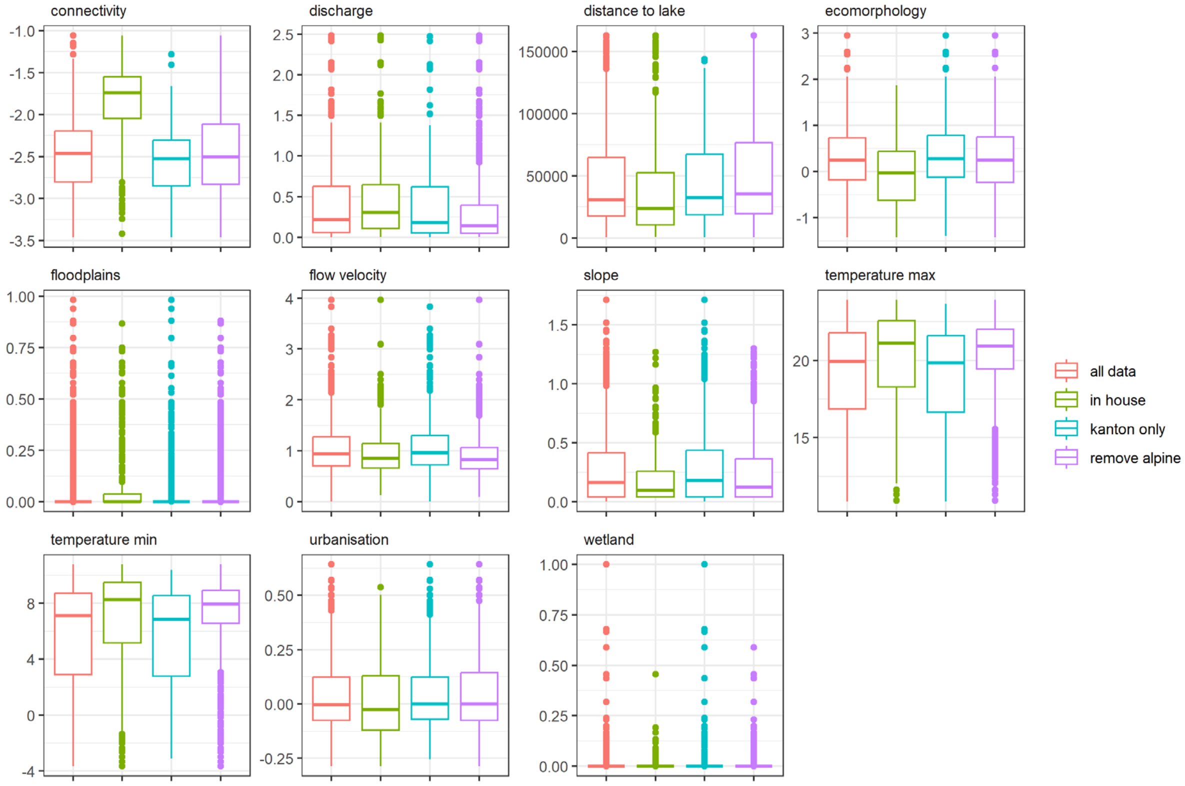

Figure S2. Coverage of environmental data in each data subset used for sensitivity testing. The differential coverage of environmental data supports our decision to combine data sources.

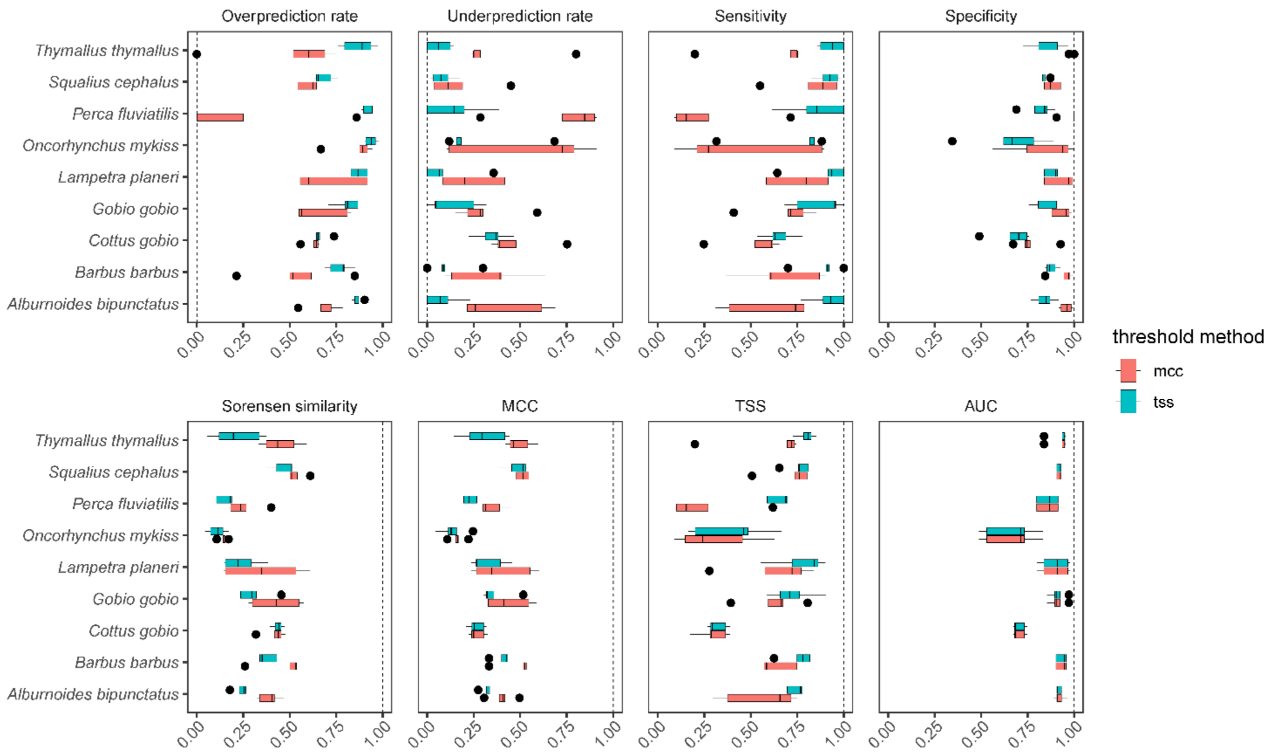

Figure S3. Boxplots of model performance across 8 model evaluation metrics (see Table S2-S3). Colour indicates which metric is optimised under the binary threshold method (either TSS or MCC). Note that AUC is a threshold independent performance metric.

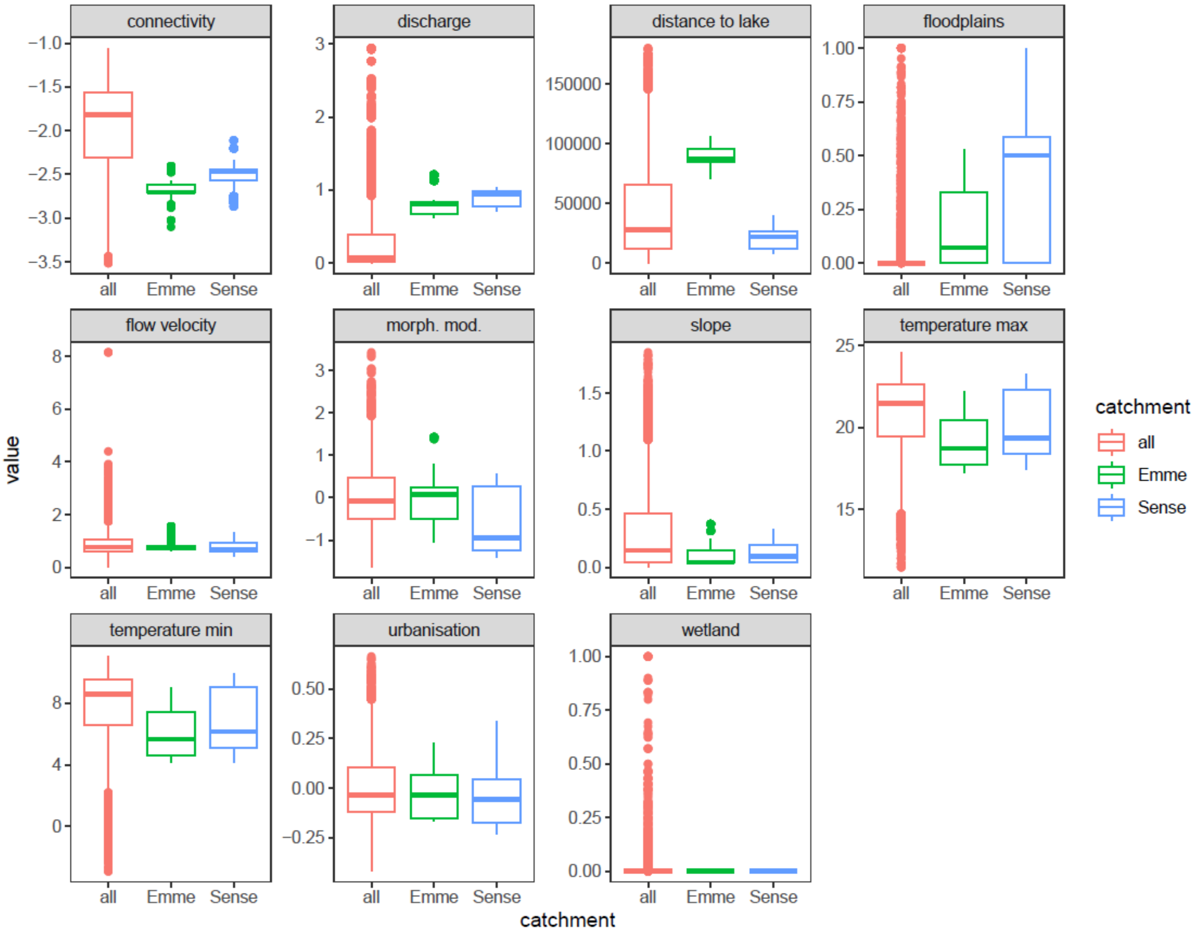

Figure S4. Comparison of environmental properties of Sense and Emme rivers.

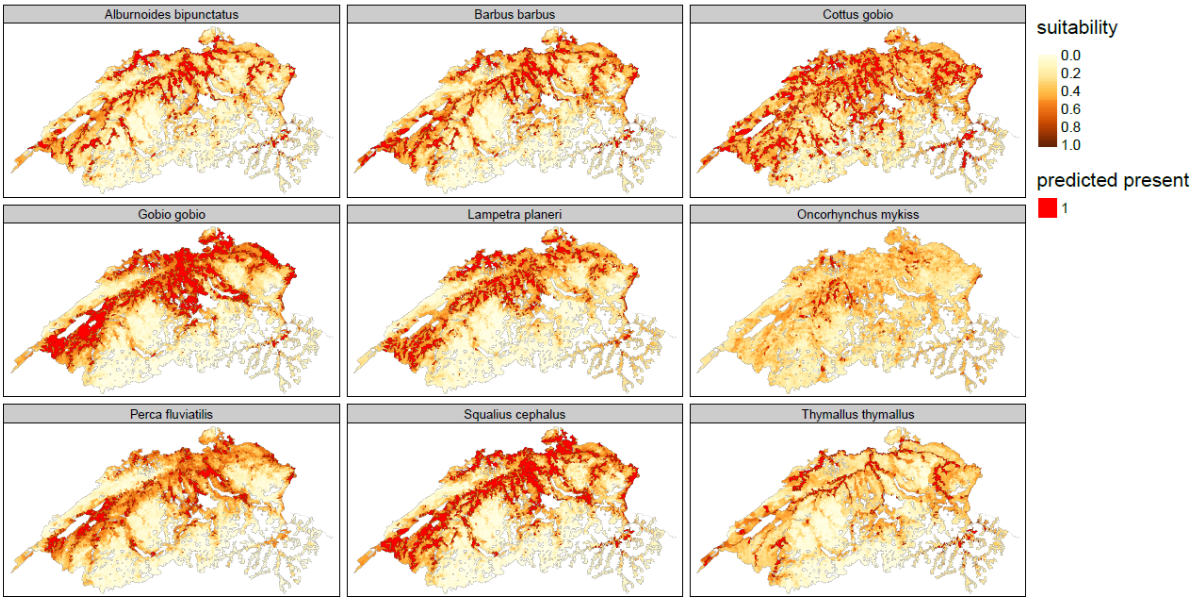

Figure S5. Spatial distribution of habitat suitability for the 9 focal species based on presence-absence random forest models. Red indicates areas where species are predicted as present based on a binary threshold optimising for the True Skill Statistic.

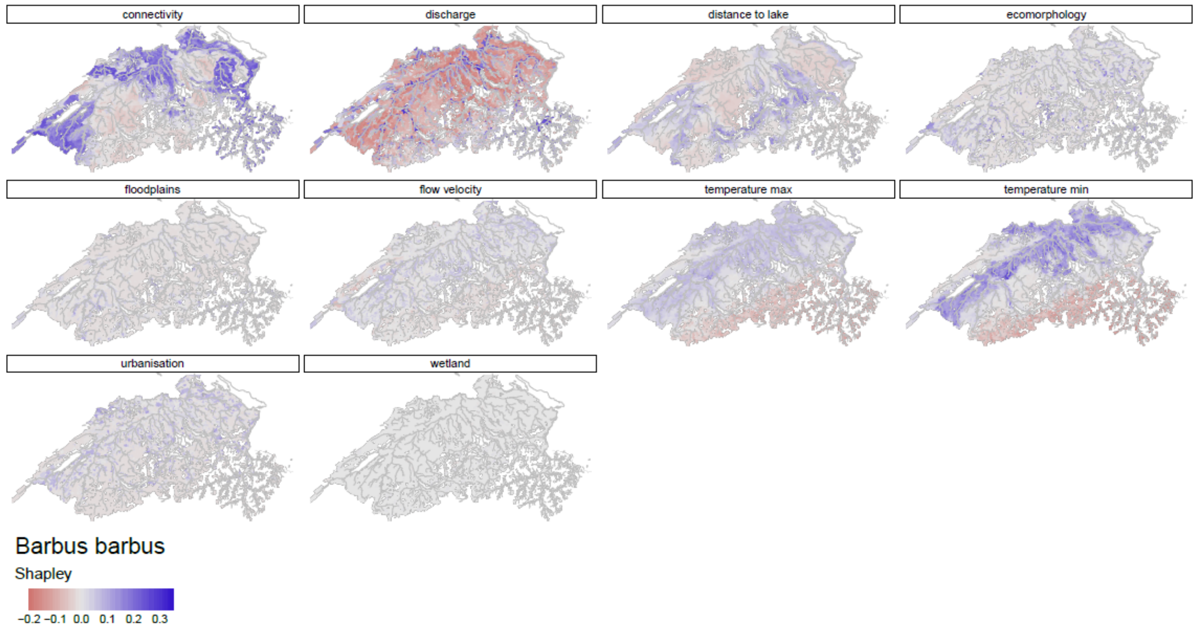

Figure S6. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Barbus barbus.*

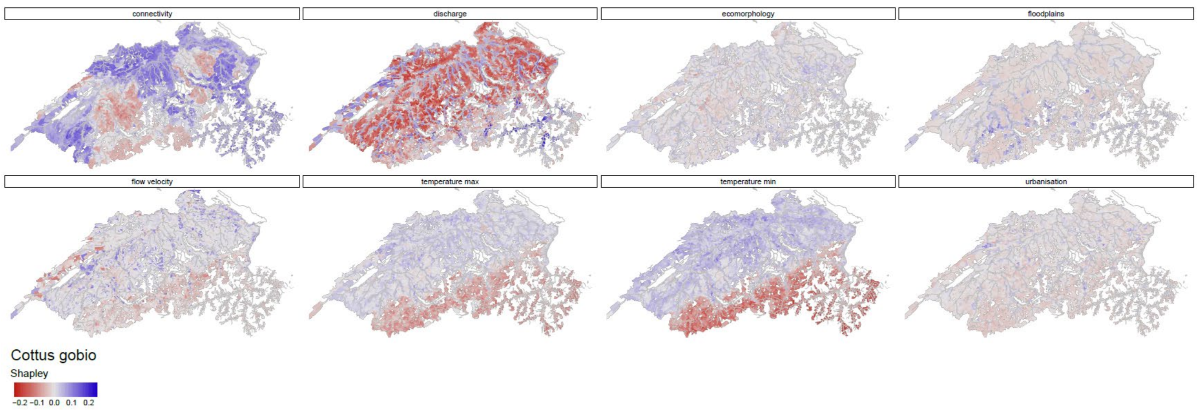

Figure S7. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Cottus gobio*.

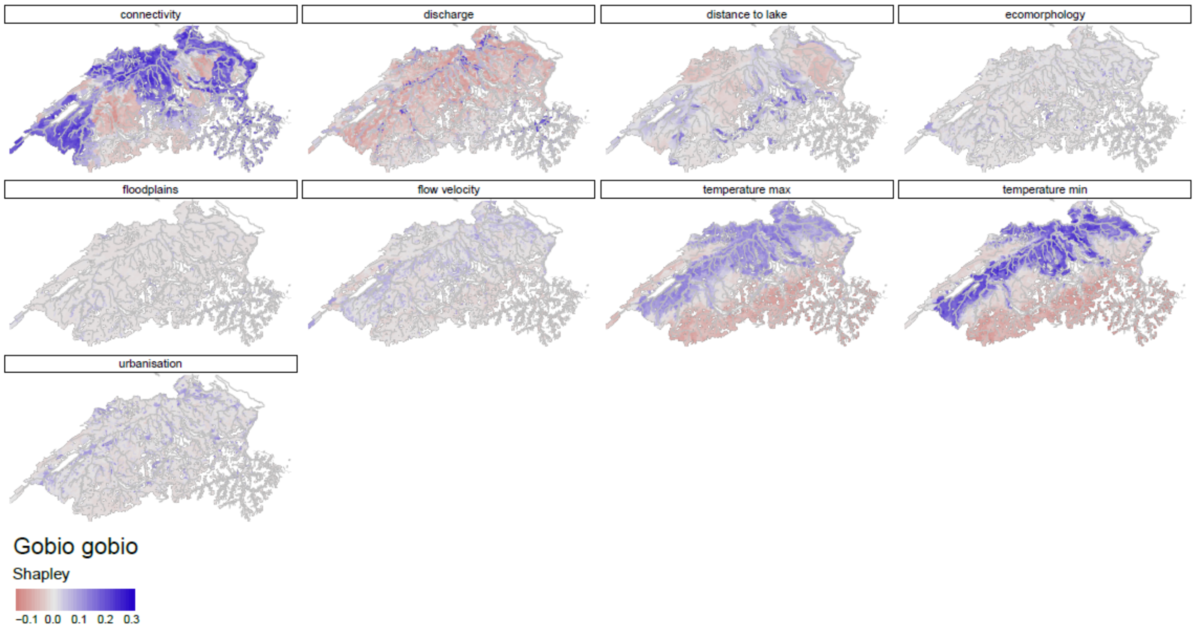

Figure S8. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Gobio gobio*.

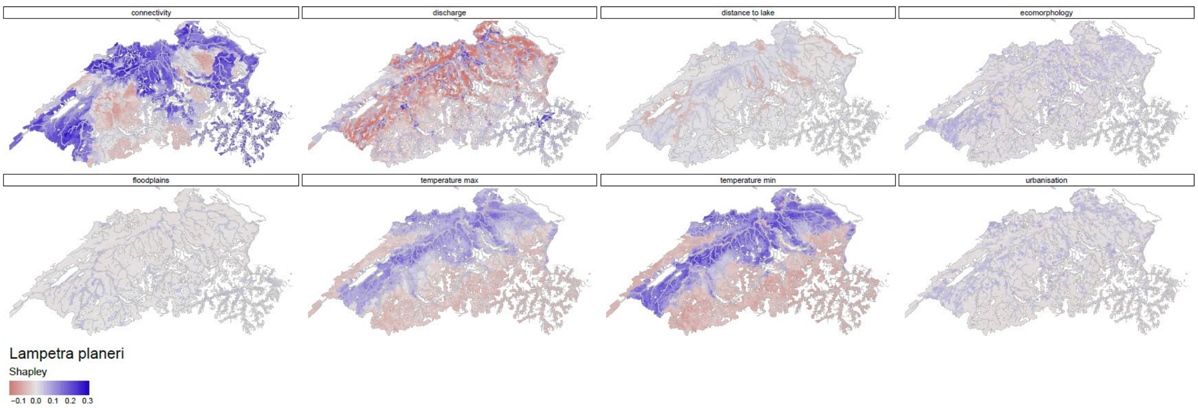

Figure S9. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Lampetra planeri*.

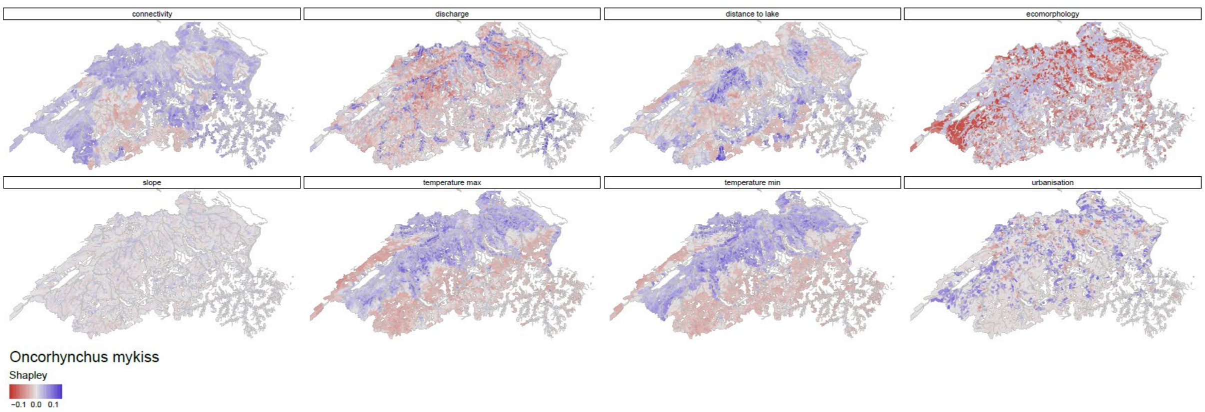

Figure S10. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Oncorhynchus mykiss*.

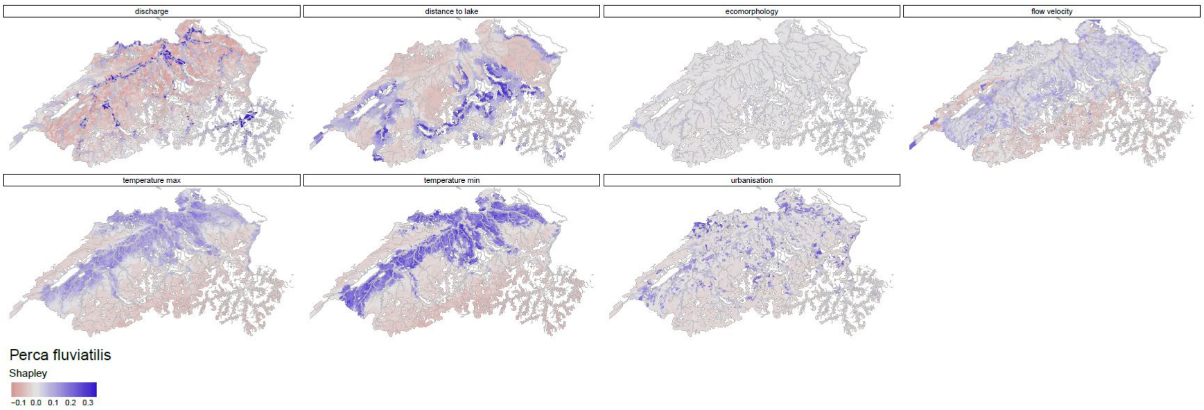

Figure S11. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Perca fluviatilis*.

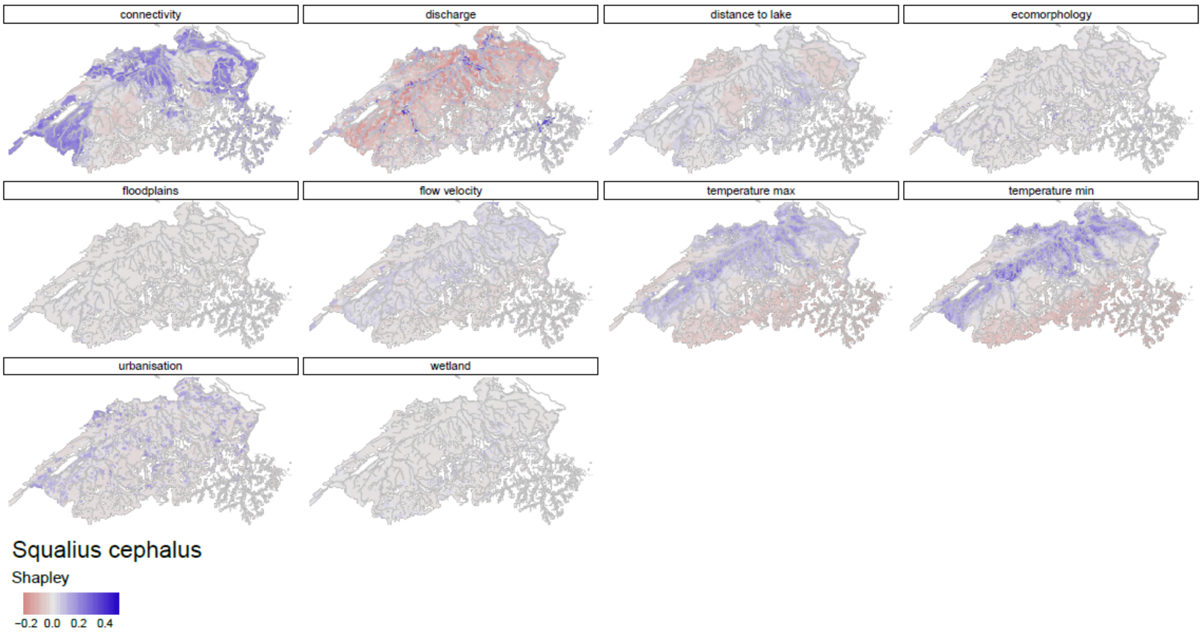

Figure S12. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Squalius cephalus*.

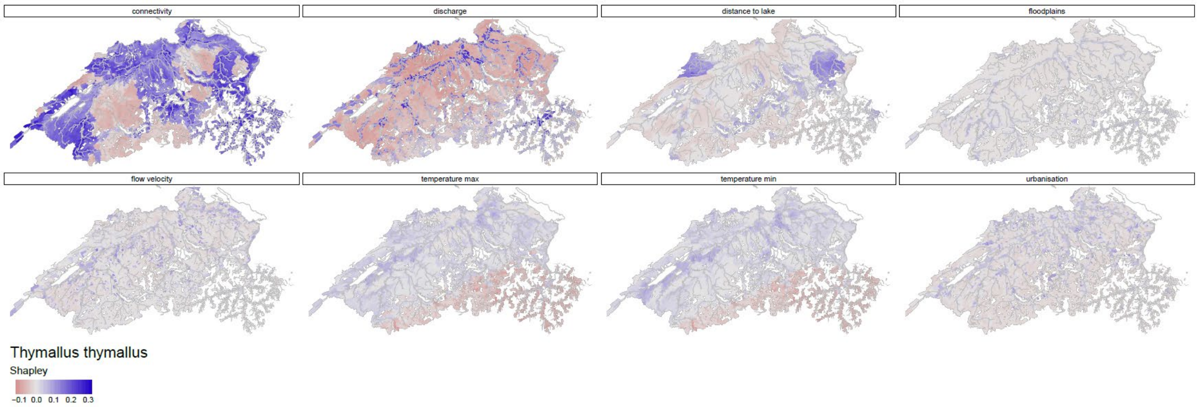

Figure S13. Comparison the spatial distribution of Shapley value distributions for all relevant variables for *Thymallus thymallus*.

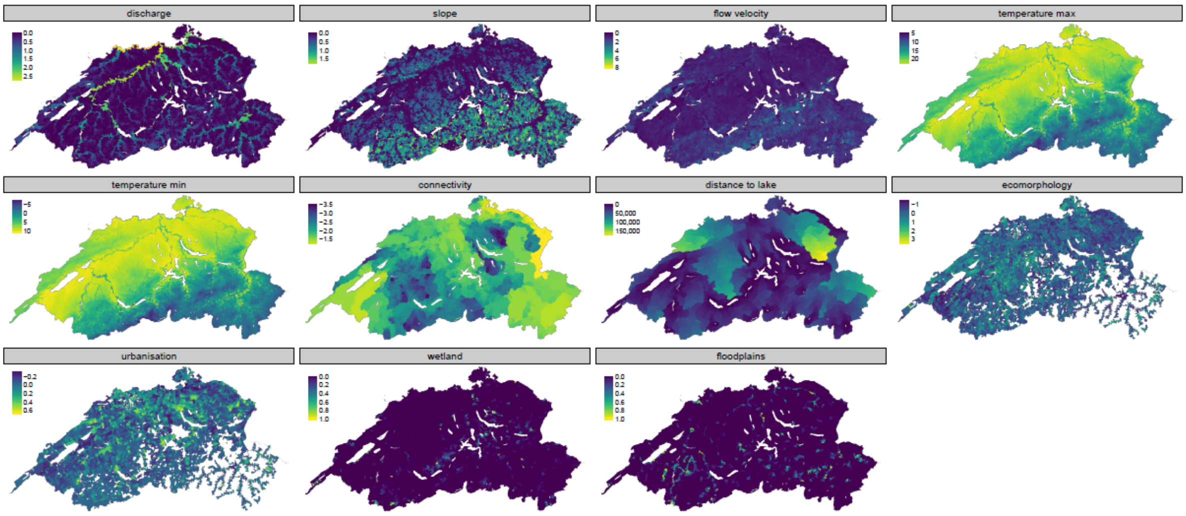

Figure S14. Spatial maps of environmental data used as covariates when fitting species distribution models. Full variable descriptions are provided in our ODMAP protocol (Appendix 1).

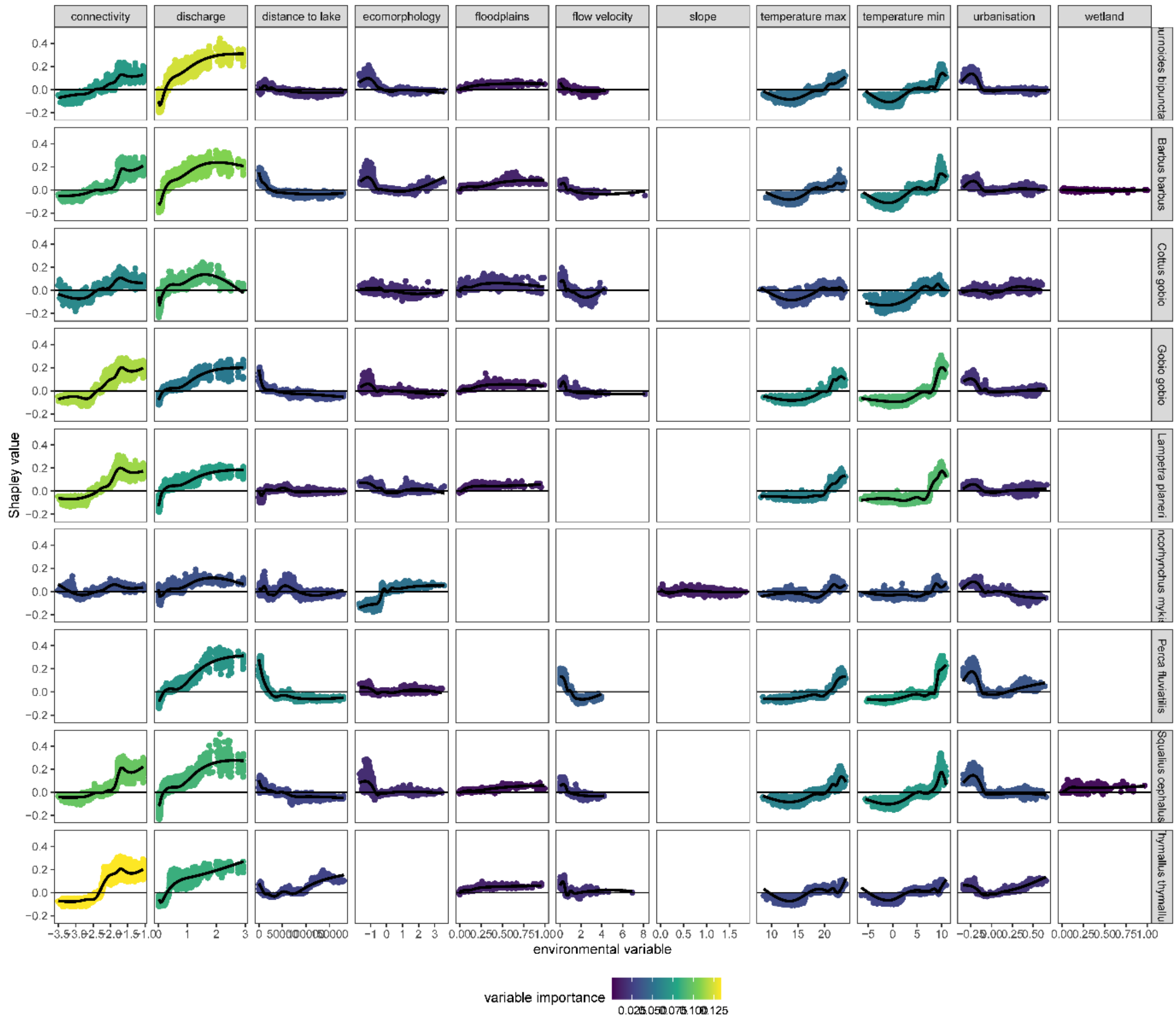

Figure S15. Shapley based response curves across all species and variables. Colour indicates variable importance estimated as the average absolute Shapley value.

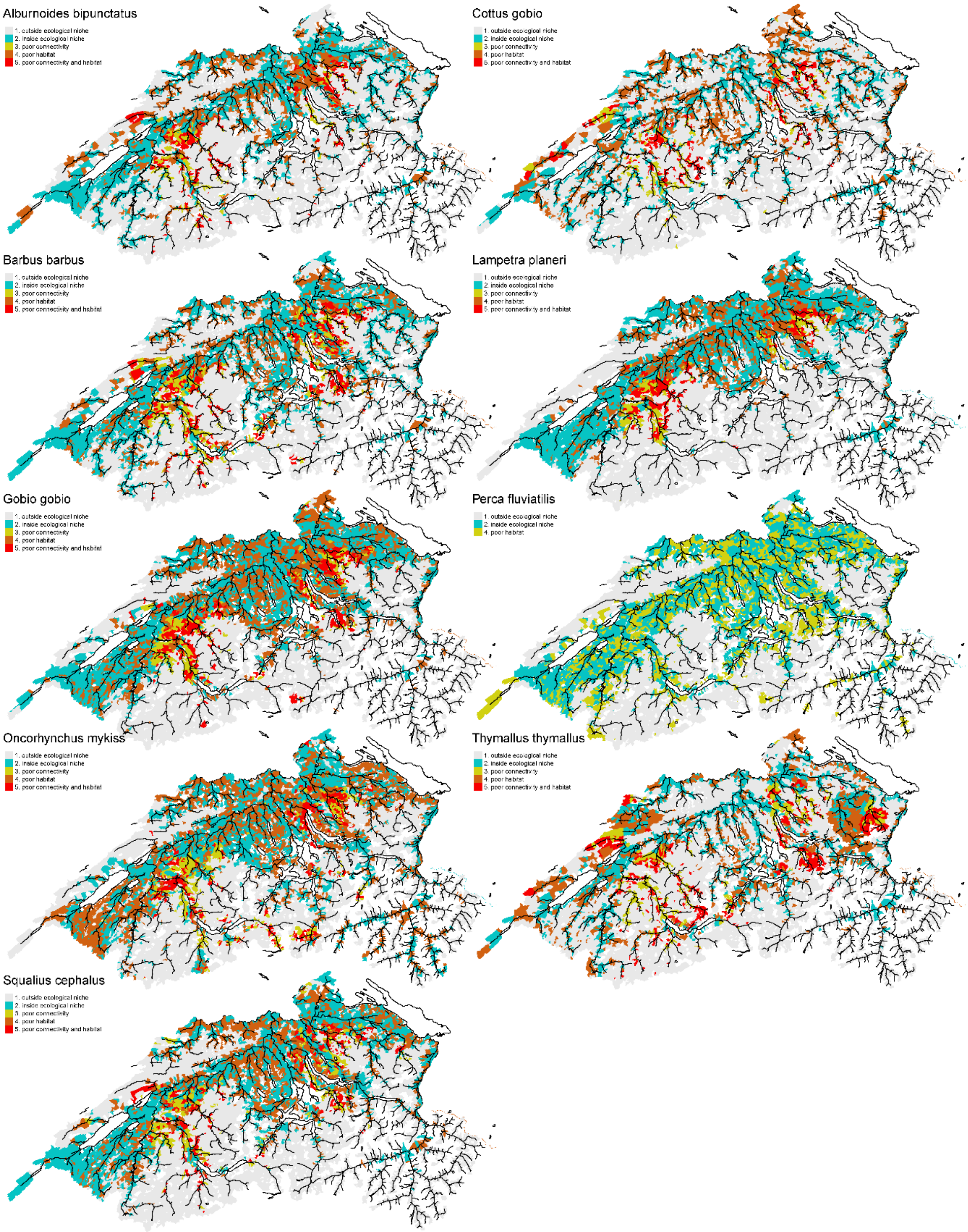

Figure S16. Qualitative shadow distributions for all species to understand the relative influence of anthropic threats within species' expected distributions. For full description see Figure 4 in the main manuscript which provides details for *A. bipunctatus.* Table S3 presents the summary statistics across these distributions.

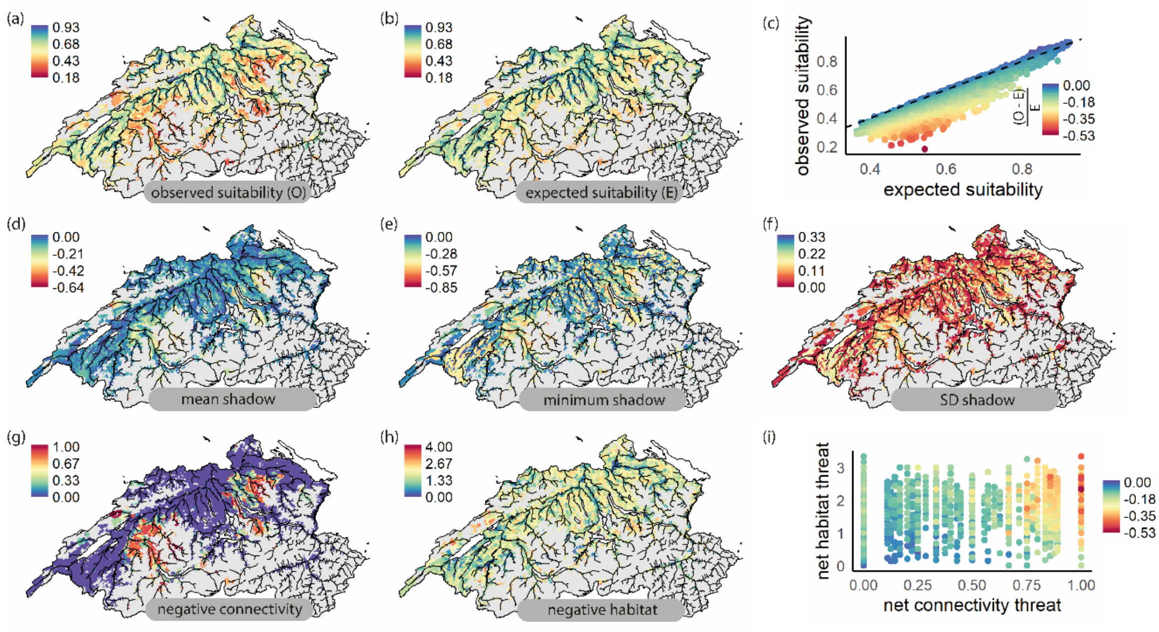

Figure S17. Multi-species average shadow distribution properties per sub-catchment, for full description see Figure 5 of the main manuscript. Here we calculate shadow distributions turning any negative Shapley values for threats to 0 (see methods). To summarise the key results as in the main text for this method Across all sub-catchments, we found environmental suitability in the observed distribution was reduced by 9.4% (averaged across species per sub-catchment = 0.38) compared to the expected distribution (0.47; t = -57; p<0.001). The lowest 10th quantile of catchments had an average suitability reduced by 29%. The most negatively impacted species in each sub-catchment had a habitat suitability reduction of 25% on average across sub-catchments.

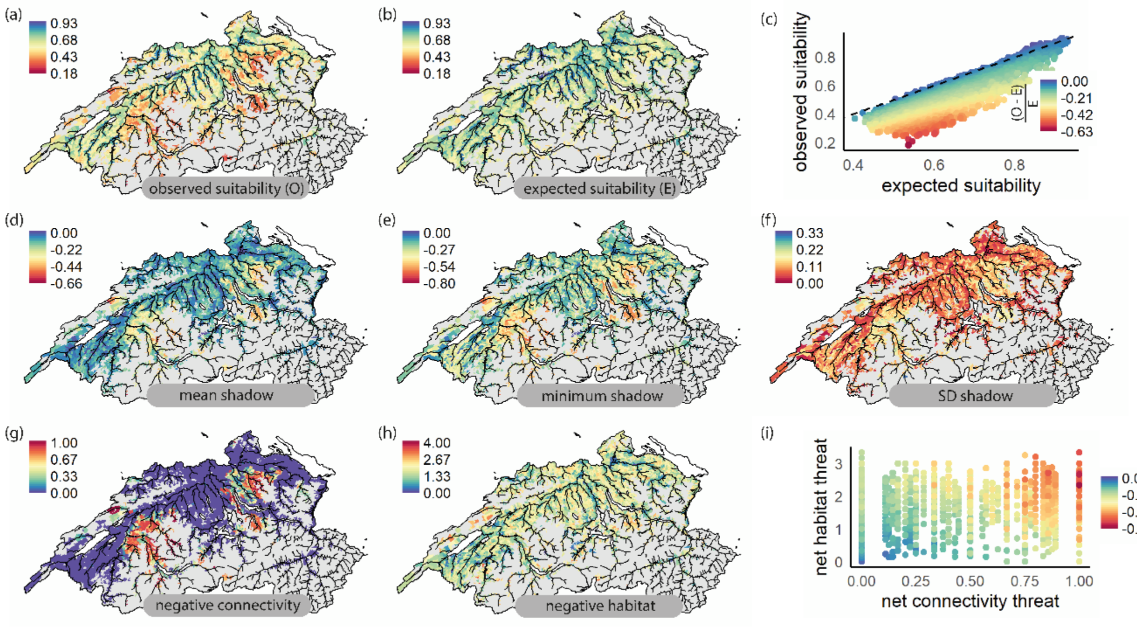

Figure S18. Multi-species average shadow distribution properties per sub-catchment, for full description see Figure 5 of the main manuscript. Here we calculate shadow distributions turning any negative Shapley values for threats to the mean positive contribution across all catchments (see methods). To summarise the key results as in the main text for this method Across all sub-catchments, we found environmental suitability in the observed distribution was reduced by 15% (averaged across species per sub-catchment = 0.38) compared to the expected distribution (0.49; t = -73; p<0.001). The lowest 10th quantile of catchments had an average suitability reduced by 39%. The most negatively impacted species in each sub-catchment had a habitat suitability reduction of 29% on average across sub-catchments.