

# Supporting Information Appendix S3: Additional tables summarizing observation data and model outputs

S. Derville, L. Torres, C. Iovan, C. Garrigue

May 22<sup>nd</sup>, 2018

*Supporting information S3 to "Finding the right fit: Comparative cetacean distribution models using multiple data sources and statistical approaches", 2018, Diversity and Distributions. Contact information: solene.derville@ird.fr*

Table 1: Sample sizes by year used in the crowdsourced and research surveys models of humpback whale distribution in New Caledonia. Estimates of survey effort are provided for the later as number of days and distance covered.

Year	Crowdsourced models	Research surveys models			
	Presence	Presence	Survey effort (days)	Distance covered on-effort (km)	Control points
2003	27	71	39	2,720	1,142
2004	26	38	46	3,714	1,413
2005	33	83	47	3,744	1,479
2006	47	117	44	3,309	1,445
2007	29	123	47	3,307	1,438
2008	40	44	61	5,928	1,962
2009	26	102	45	4,053	1,450
2010	122	140	51 <sup>a</sup>	3,465	1,630
2011	33	187	45	2,805	1,529
2012	45	117	44	3,487	1,183
2013	36	97	49	3,672	1,524
2014	17	47	48	3,875	1,404
2015	54	99	31	1,939	961
2016	90	95	41	3,825	1,710
<b>Total</b>	<b>625</b>	<b>1,360</b>	<b>638</b>	<b>49,843</b>	<b>20,270</b>
<b>Mean</b>	<b>45</b>	<b>97</b>	<b>46</b>	<b>3,560</b>	<b>1,448</b>
<b>SD</b>	<b>28</b>	<b>40</b>	<b>7</b>	<b>883</b>	<b>246</b>

<sup>a</sup> including three days over which two surveys were taking place simultaneously in two separate locations

Table 2: Mean contribution of environmental variables\* to habitat preference models for humpback whales in New Caledonian waters with 9 predictors. Coefficients of variation of the mean contribution calculated over 50 cross-validation runs indicated. Variables in grey were removed in the models presented in the main manuscript. For BRTs, SVMs, and MAXENT models results are reported for the ‘stable model’ tuning with best diff.AUC, and for the ‘predictive model’ tuning with best ext.AUC (except SVM for which contributions were only evaluated in the linear kernel case).

variables	BRT		GAM		MAXENT		GLM		SVM					
	best diff.AUC	best ext.AUC	mean	CV%	mean	CV %	best diff.AUC	best ext.AUC	mean	CV %	best diff.AUC			
A.AVG	2.9	14.0	5.4	8.9	2.9	20.7	0.0	156.3	1.1	28.9	8.5	11.6	0.1	85.9
DEPTH	34.4	4.4	22.7	4.7	24.4	9.8	21.1	13.6	45.9	4.6	12.9	9.4	7.2	21.8
CPRO	0.4	41.0	3.9	11.8	1.1	36.4	0.3	83.3	0.0	76.3	3.5	17.7	0.4	50.7
DISSURF	2.5	13.3	4.6	7.8	26.6	10.5	28.7	10.2	20.1	7.1	18.6	13.7	11.3	8.3
JULIAN	2.2	14.3	4.5	8.1	10.6	10.4	0.2	135.3	4.2	17.3	9.4	10.1	0.2	104.4
K490	20.2	5.5	14.7	5.3	10.5	23.8	0.9	27.0	2.4	20.7	17.7	9.0	0.6	35.0
S.AVG	5.9	8.5	6.2	7.5	1.1	45.5	7.7	34.3	1.1	27.7	3.6	20.1	3.5	28.4
S.COV	5.2	14.8	16.6	4.6	2.2	22.7	0.4	55.0	1.4	65.5	9.2	9.6	0.3	47.4
SST	26.3	6.8	21.5	6.6	20.7	10.1	40.7	7.5	23.6	9.9	16.5	9.0	76.5	1.5

\* average slope (S.AVG), profile curvature (CPRO), average aspect (A.AVG), julian date (JULIAN), coefficient of variation of the slope (S.COV), sea surface temperature (SST), distance to closest reef or land (DISSURF), depth (DEPTH) and diffuse attenuation (K490).

Table 3: Mean contribution of environmental variables\* to habitat preference models for humpback whales in New Caledonian waters with 7 predictors. Coefficients of variation of the mean contribution calculated over 50 cross-validation runs indicated. For BRTs, SVMs, and MAXENT models results are reported for the ‘stable model’ tuning with best diff.AUC, and for the ‘predictive model’ tuning with best ext.AUC (except SVM for which contributions were only evaluated in the linear kernel case).

variables	BRT		GAM		MAXENT		GLM		SVM					
	best diff.AUC	best ext.AUC	mean	CV%	mean	CV %	best diff.AUC	best ext.AUC	mean	CV %	best diff.AUC			
DEPTH	35.5	4.3	25.6	4.5	23.7	9.7	21.2	13.6	46.6	4.4	13.6	13.0	8.2	22.3
DISSURF	2.6	12.8	5.2	6.3	28.4	10.2	28.9	10.0	20.4	6.6	21.6	13.5	12.9	6.4
JULIAN	2.4	13.4	4.6	7.0	9.8	11.2	0.2	127.2	4.1	17.4	9.9	11.1	0.3	89.5
K490	20.9	5.5	16.5	5.5	10.7	23.4	0.9	28.6	2.4	20.5	19.4	9.7	0.6	36.8
S.AVG	6.1	8.6	6.9	6.5	2.2	27.3	7.7	34.7	1.2	30.0	5.5	17.7	2.5	32.4
S.COV	5.6	13.6	17.4	5.1	2.3	21.7	0.4	54.6	1.4	60.9	11.0	8.8	0.4	38.4
SST	27.0	6.8	23.9	6.5	22.9	8.7	40.8	7.4	23.8	9.6	19.0	8.8	75.1	1.7

\* average slope (S.AVG), profile curvature (CPRO), average aspect (A.AVG), julian date (JULIAN), coefficient of variation of the slope (S.COV), sea surface temperature (SST), distance to closest reef or land (DISSURF), depth (DEPTH) and diffuse attenuation (K490).

Table 4: Model tuning of MAXENT models with 9 predictors of humpback whale habitat preferences in New Caledonian waters. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argsos.

features	beta	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argsos	value
<b>hinge</b>	<b>1</b>	<b>0.748</b>	<b>0.737</b>	<b>0.011</b>	<b>0.42</b>	<b>0.368</b>	<b>44.716</b>	<b>mean</b>
		0.004	0.031	0.034	0.05	0.057	5.029	sd
hinge	4	0.738	0.733	0.005	0.431	0.361	46.937	mean
		0.004	0.031	0.035	0.048	0.056	5.757	sd
hinge	7	0.733	0.729	0.003	0.442	0.355	45.785	mean
		0.003	0.032	0.035	0.049	0.057	6.23	sd
hinge	10	0.73	0.726	0.004	0.448	0.352	46.117	mean
		0.003	0.033	0.036	0.044	0.058	5.99	sd
quadratic	1	0.707	0.702	0.004	0.45	0.312	53.579	mean
		0.004	0.037	0.04	0.039	0.057	7.943	sd
quadratic	4	0.707	0.702	0.004	0.45	0.312	53.587	mean
		0.004	0.037	0.04	0.039	0.058	7.95	sd
quadratic	7	0.706	0.702	0.004	0.45	0.312	53.663	mean
		0.004	0.037	0.041	0.038	0.057	8.165	sd
quadratic	10	0.706	0.701	0.004	0.455	0.311	52.735	mean
		0.004	0.037	0.041	0.037	0.057	8.14	sd
<b>linear</b>	<b>1</b>	<b>0.676</b>	<b>0.676</b>	<b>0</b>	<b>0.427</b>	<b>0.276</b>	<b>53.84</b>	<b>mean</b>
		0.005	0.041	0.046	0.046	0.061	9.961	sd
linear	4	0.676	0.676	0	0.427	0.276	53.837	mean
		0.005	0.041	0.046	0.046	0.061	9.955	sd
linear	7	0.676	0.676	0	0.426	0.275	54.192	mean
		0.005	0.041	0.046	0.047	0.061	10.174	sd
linear	10	0.676	0.676	0	0.428	0.275	54.158	mean
		0.005	0.04	0.045	0.045	0.061	9.954	sd

Table 5: Model tuning of BRT models with 9 predictors of humpback whale habitat preferences in New Caledonian waters. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argsos.

learning rate	tree complexity	trees	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argsos	value
<b>0.005</b>	<b>1</b>	<b>5326</b>	<b>0.769</b>	<b>0.738</b>	<b>0.031</b>	<b>0.507</b>	<b>0.365</b>	<b>42.76</b>	<b>mean</b>
		309.417	0.004	0.033	0.036	0.054	0.058	6.447	sd
0.005	2	5542	0.83	0.771	0.059	0.502	0.418	41.035	mean
		321.248	0.004	0.028	0.03	0.056	0.047	5.302	sd
0.005	3	4566	0.852	0.776	0.076	0.498	0.426	39.643	mean
		295.966	0.004	0.027	0.028	0.061	0.045	5.115	sd
0.01	1	3559	0.775	0.74	0.035	0.501	0.372	43.3	mean
		231.827	0.004	0.033	0.036	0.057	0.058	6.614	sd
0.01	2	3646	0.842	0.773	0.069	0.488	0.422	41.55	mean
		267.04	0.005	0.028	0.03	0.062	0.047	5.43	sd
<b>0.01</b>	<b>3</b>	<b>2778</b>	<b>0.863</b>	<b>0.776</b>	<b>0.086</b>	<b>0.47</b>	<b>0.427</b>	<b>41.316</b>	<b>mean</b>
		267.673	0.006	0.027	0.028	0.074	0.046	6.129	sd
0.05	1	1092	0.783	0.74	0.043	0.489	0.369	43.842	mean
		233.946	0.006	0.033	0.037	0.071	0.056	6.923	sd
0.05	2	808	0.846	0.773	0.073	0.483	0.42	41.802	mean
		128.317	0.009	0.027	0.032	0.067	0.047	6.078	sd
0.05	3	537	0.859	0.776	0.083	0.47	0.425	41.354	mean
		74.1	0.009	0.028	0.029	0.075	0.046	6.23	sd

Table 6: Model tuning of SVM models with 9 predictors of humpback whale habitat preferences in New Caledonian waters. “polynomial2” indicate second-order polynomials and “polynomial3” indicate third-order polynomials. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argos.

kernel type	cost	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argos	value
linear	0.00001	0.668	0.67	-0.002	0.002	0	99.184	mean
		0.004	0.039	0.043	0.011	0.001	5.771	sd
<b>linear</b>	<b>0.01</b>	<b>0.67</b>	<b>0.67</b>	<b>0.001</b>	<b>0.066</b>	<b>0.27</b>	<b>70.616</b>	<b>mean</b>
		0.005	0.041	0.046	0.019	0.062	14.472	sd
linear	1	0.672	0.669	0.003	0.066	0.272	61.418	mean
		0.005	0.042	0.046	0.018	0.061	13.951	sd
linear	10	0.672	0.669	0.003	0.066	0.272	61.397	mean
		0.005	0.042	0.046	0.018	0.062	14.163	sd
polynomial2	0.00001	0.644	0.638	0.006	0.003	0	99.544	mean
		0.004	0.036	0.037	0.015	0.003	2.367	sd
polynomial2	0.01	0.69	0.682	0.009	0.079	0.276	45.137	mean
		0.004	0.033	0.036	0.005	0.049	8.681	sd
polynomial2	1	0.738	0.726	0.012	0.085	0.351	41.618	mean
		0.003	0.032	0.034	0.013	0.049	5.071	sd
polynomial2	10	0.746	0.732	0.014	0.088	0.365	40.545	mean
		0.004	0.031	0.034	0.011	0.052	4.504	sd
polynomial3	0.00001	0.663	0.661	0.002	0.003	0.001	99.488	mean
		0.004	0.038	0.042	0.014	0.005	2.552	sd
polynomial3	0.01	0.672	0.665	0.007	0.077	0.265	61.02	mean
		0.004	0.041	0.045	0.004	0.063	6.666	sd
polynomial3	1	0.711	0.684	0.027	0.083	0.301	41.765	mean
		0.004	0.038	0.042	0.005	0.06	5.929	sd
polynomial3	10	0.738	0.699	0.039	0.086	0.324	34.606	mean
		0.004	0.032	0.034	0.005	0.054	4.732	sd
radial	0.00001	0.65	0.643	0.008	0	0	100	mean
		0.004	0.04	0.041	0	0	0	sd
radial	0.01	0.65	0.643	0.008	0.071	0.25	36.88	mean
		0.004	0.04	0.041	0.01	0.059	12.305	sd
radial	1	0.758	0.735	0.023	0.074	0.375	44.529	mean
		0.003	0.029	0.032	0.02	0.051	8.104	sd
<b>radial</b>	<b>10</b>	<b>0.797</b>	<b>0.752</b>	<b>0.044</b>	<b>0.082</b>	<b>0.411</b>	<b>38.052</b>	<b>mean</b>
		0.003	0.028	0.03	0.018	0.047	5.62	sd

Table 7: Model tuning of MAXENT models with 7 predictors of humpback whale habitat preferences in New Caledonian waters. The selected model is shown in bold. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argsos.

features	beta	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argsos	value
<b>hinge</b>	<b>1</b>	<b>0.747</b>	<b>0.736</b>	<b>0.011</b>	<b>0.413</b>	<b>0.364</b>	<b>46.083</b>	<b>mean</b>
		0.004	0.031	0.034	0.052	0.055	6.198	sd
hinge	4	0.737	0.731	0.006	0.434	0.355	47.575	mean
		0.004	0.032	0.035	0.044	0.056	5.875	sd
hinge	7	0.731	0.728	0.003	0.445	0.351	46.659	mean
		0.003	0.032	0.036	0.049	0.054	6.201	sd
hinge	10	0.729	0.725	0.004	0.45	0.348	46.824	mean
		0.004	0.033	0.037	0.05	0.055	6.919	sd
quadratic	1	0.703	0.699	0.004	0.455	0.304	52.445	mean
		0.004	0.038	0.042	0.042	0.06	10.784	sd
quadratic	4	0.703	0.699	0.004	0.455	0.304	52.417	mean
		0.004	0.038	0.042	0.042	0.06	10.789	sd
quadratic	7	0.703	0.698	0.004	0.457	0.304	52.117	mean
		0.004	0.038	0.042	0.041	0.059	10.592	sd
quadratic	10	0.702	0.698	0.004	0.456	0.303	52.564	mean
		0.004	0.038	0.042	0.041	0.059	10.379	sd
<b>linear</b>	<b>1</b>	<b>0.675</b>	<b>0.675</b>	<b>0</b>	<b>0.426</b>	<b>0.274</b>	<b>53.259</b>	<b>mean</b>
		0.005	0.041	0.046	0.046	0.063	9.771	sd
linear	4	0.675	0.675	0	0.426	0.274	53.269	mean
		0.005	0.041	0.046	0.046	0.063	9.785	sd
linear	7	0.675	0.675	0	0.426	0.274	53.32	mean
		0.005	0.041	0.046	0.044	0.063	9.687	sd
linear	10	0.675	0.675	0	0.427	0.274	53.374	mean
		0.005	0.041	0.046	0.043	0.063	9.455	sd

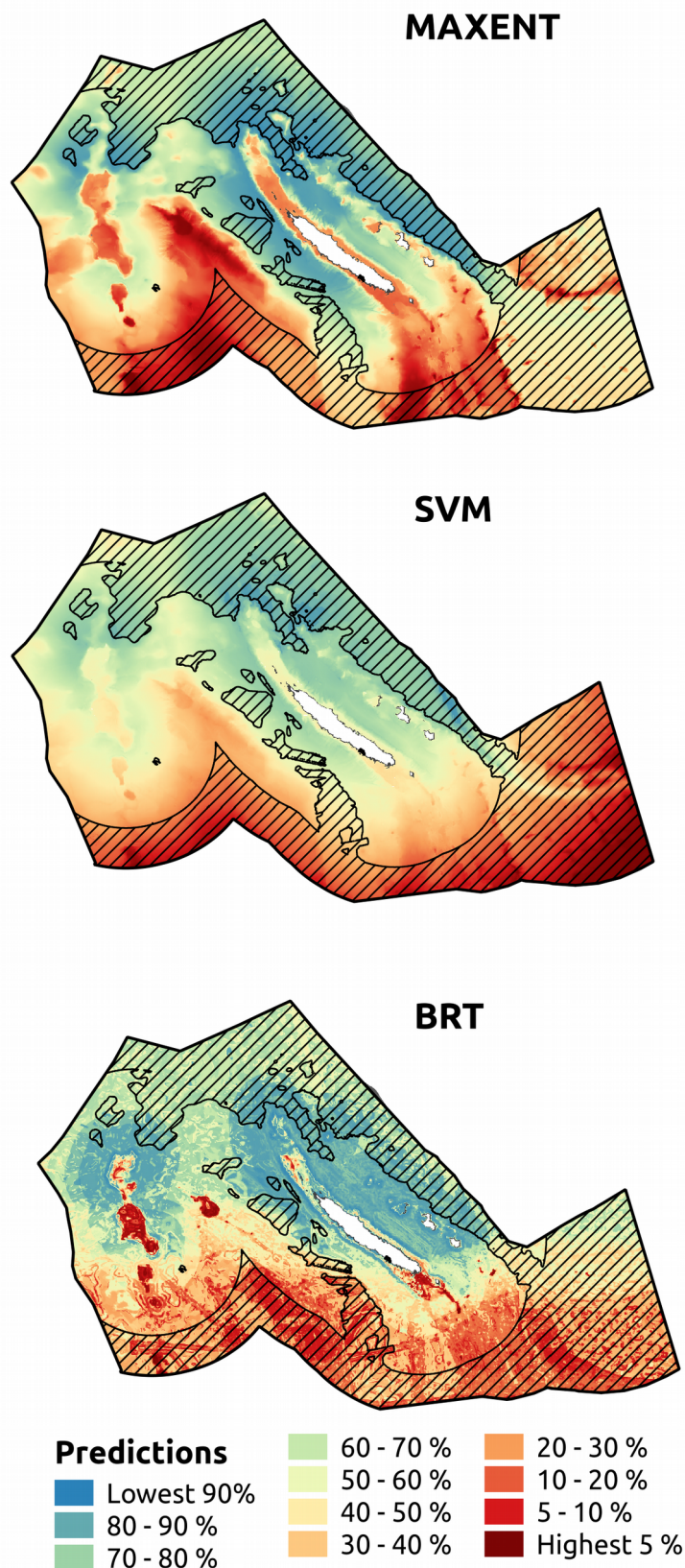
Table 8: Model tuning of BRT models with 7 predictors of humpback whale habitat preferences in New Caledonian waters. The selected model is shown in bold. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argsos.

learning rate	tree complexity	trees	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argsos	value
<b>0.005</b>	<b>1</b>	<b>5206</b>	<b>0.767</b>	<b>0.738</b>	<b>0.029</b>	<b>0.507</b>	<b>0.364</b>	<b>43.895</b>	<b>mean</b>
		255.479	0.004	0.033	0.036	0.056	0.056	7.785	sd
0.005	2	5245	0.823	0.77	0.053	0.506	0.416	41.762	mean
		374.063	0.004	0.029	0.031	0.062	0.048	5.839	sd
<b>0.005</b>	<b>3</b>	<b>4361</b>	<b>0.843</b>	<b>0.775</b>	<b>0.069</b>	<b>0.496</b>	<b>0.425</b>	<b>40.832</b>	<b>mean</b>
		301.745	0.004	0.027	0.029	0.066	0.045	5.934	sd
0.01	1	3378	0.771	0.739	0.032	0.501	0.369	44.239	mean
		199.274	0.004	0.033	0.037	0.063	0.056	7.917	sd
0.01	2	3356	0.833	0.772	0.061	0.499	0.42	41.742	mean
		286.364	0.005	0.028	0.031	0.063	0.048	5.363	sd
0.01	3	2668	0.854	0.775	0.079	0.485	0.425	41.244	mean
		235.554	0.006	0.027	0.03	0.072	0.045	5.974	sd
0.05	1	895	0.776	0.739	0.036	0.495	0.369	44.667	mean
		181.617	0.005	0.034	0.038	0.058	0.056	6.773	sd
0.05	2	745	0.837	0.772	0.065	0.484	0.417	43.058	mean
		118.343	0.008	0.028	0.031	0.078	0.047	7.35	sd
0.05	3	542	0.853	0.775	0.079	0.48	0.424	41.576	mean
		76.505	0.009	0.027	0.031	0.078	0.047	6.177	sd

Table 9: Model tuning of SVM models with 7 predictors of humpback whale habitat preferences in New Caledonian waters. “polynomial2” indicate second-order polynomials and “polynomial3” indicate third-order polynomials. The selected model is shown in bold. ‘thresh’ is the threshold used to calculate TSS and sensitivity.argsos.

kernel type	cost	int.AUC	ext.AUC	diff.AUC	thresh	TSS	sensitivity. argos	value
linear	0.00001	0.669	0.669	-0.001	0	0	100	mean
		0.004	0.04	0.044	0	0	0	sd
<b>linear</b>	<b>0.01</b>	<b>0.669</b>	<b>0.669</b>	<b>0</b>	<b>0.066</b>	<b>0.27</b>	<b>70.908</b>	<b>mean</b>
		0.005	0.041	0.046	0.019	0.062	14.521	sd
linear	1	0.671	0.668	0.003	0.066	0.27	60.728	mean
		0.005	0.042	0.047	0.018	0.062	13.318	sd
linear	10	0.671	0.668	0.003	0.066	0.271	60.384	mean
		0.005	0.042	0.047	0.018	0.062	13.257	sd
polynomial2	0.00001	0.638	0.631	0.006	0.001	0	99.724	mean
		0.003	0.037	0.04	0.01	0.003	1.949	sd
polynomial2	0.01	0.674	0.666	0.008	0.073	0.249	53.07	mean
		0.004	0.035	0.038	0.003	0.05	6.875	sd
polynomial2	1	0.722	0.712	0.01	0.093	0.33	41.161	mean
		0.004	0.034	0.037	0.016	0.053	5.992	sd
polynomial2	10	0.732	0.719	0.012	0.087	0.347	40.434	mean
		0.004	0.032	0.035	0.014	0.053	4.835	sd
polynomial3	0.00001	0.662	0.66	0.002	0.001	0.001	99.783	mean
		0.004	0.039	0.043	0.009	0.004	1.534	sd
polynomial3	0.01	0.669	0.665	0.005	0.073	0.254	65.208	mean
		0.004	0.04	0.044	0.004	0.061	7.459	sd
polynomial3	1	0.688	0.674	0.014	0.076	0.288	54.5	mean
		0.004	0.04	0.043	0.005	0.058	6.692	sd
polynomial3	10	0.711	0.687	0.024	0.081	0.3	46.517	mean
		0.005	0.037	0.04	0.006	0.058	7.196	sd
radial	0.00001	0.667	0.658	0.008	0	0	100	mean
		0.006	0.039	0.041	0	0	0	sd
radial	0.01	0.667	0.658	0.008	0.074	0.276	43.471	mean
		0.006	0.039	0.041	0.011	0.055	13.977	sd
radial	1	0.742	0.725	0.017	0.073	0.357	48.948	mean
		0.003	0.031	0.034	0.02	0.053	7.972	sd
<b>radial</b>	<b>10</b>	<b>0.772</b>	<b>0.744</b>	<b>0.028</b>	<b>0.08</b>	<b>0.39</b>	<b>42.779</b>	<b>mean</b>
		0.003	0.029	0.032	0.02	0.047	7.042	sd

Figure 1: Maps of mean predicted humpback whale habitat suitability from research survey models: SVM.stable, BRT.stable and MAXENT.stable (models selected with best diff.AUC). Habitat suitability was averaged over 50 cross validation runs for each statistical algorithm and a colored log-scale was applied to values ranging from 0 to 1. Colors represent fixed percentages of probability distributions of the suitability predicted values (e.g., the highest 10% corresponds to the decile with highest values over each map). Areas of extrapolation where at least one environmental variable expanded outside the range observed in the training dataset are dashed.





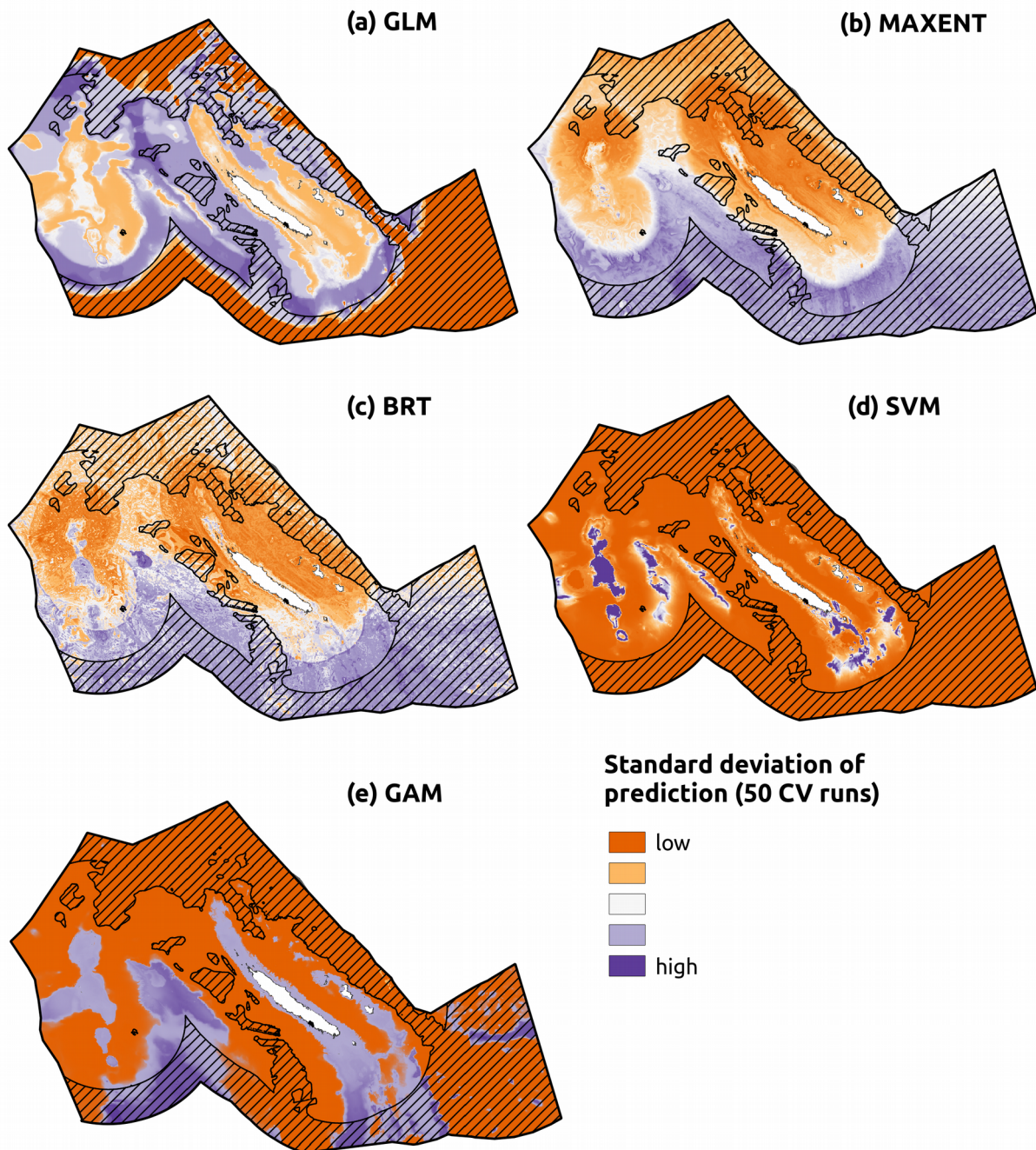


Figure 2: Maps of standard deviation of the predicted humpback whale habitat suitability from research survey models (with “.pred” settings). Standard deviation was calculated over 50 cross validation runs for each statistical algorithm. Color scales are not standardized across maps. Areas of extrapolation where at least one environmental variable expanded outside the range observed in the training dataset are dashed.