RF and GAM model summaries

Brian C. Stock[1], Eric J. Ward[2], James T. Thorson[2], Jason E. Jannot[2], Brice X. Semmens[1]

September 4, 2018

Random forest and GAM models

For each (of 15) species we fit the models described below. `fac(x)` is shorthand for treating covariate x as a factor. All random forest models were fit with `ntree=500` and default `mtry`. All GAMs were fit with `family=Tweedie(p=tweedie_p[i]), method="REML"`. All data are at the haul-level.

- Y: response, bycatch/discards (in kg)
- *year*: year (5 levels: 2011-2015)
- *depth_interval*: depth of haul (3 levels: $<$ 125, 126-250, $>$ 250 fathoms)
- *season*: season (2 levels: summer, winter)
- *bimonth*: bimonthly period (6 levels: Jan-Feb, Mar-Apr, ..., Nov-Dec)
- *logret*: log(retained target species catch, in kg)
- *lat*: degrees latitude
- *long*: degrees longitude

For random forest models, we report the percent variance explained (pseudo-$R^2 = \frac{1-mse}{Var(y)}$). Random forest covariate effect plots were created using the `forestFloor` R package and custom code.

For GAM models, we report the output of `mgcv::summary.gam`, including percent deviance explained. GAM covariate effect plots were created using the `visreg` R package.

$\mathbf{RF_{strat}}$

Designed to *mimic the stratified ratio estimator* by treating `year`, `season`, and `depth_interval` as factors. `bimonth` is included as linear and quadratic terms to avoid confounding with `fac(season)`.

randomForest(Y ~ fac(year) + fac(season) + fac(depth_interval) + bimonth + I(bimonthˆ2) + logret + lat + long + I(latˆ2) + I(longˆ2))

$\mathbf{GAM_{strat}}$

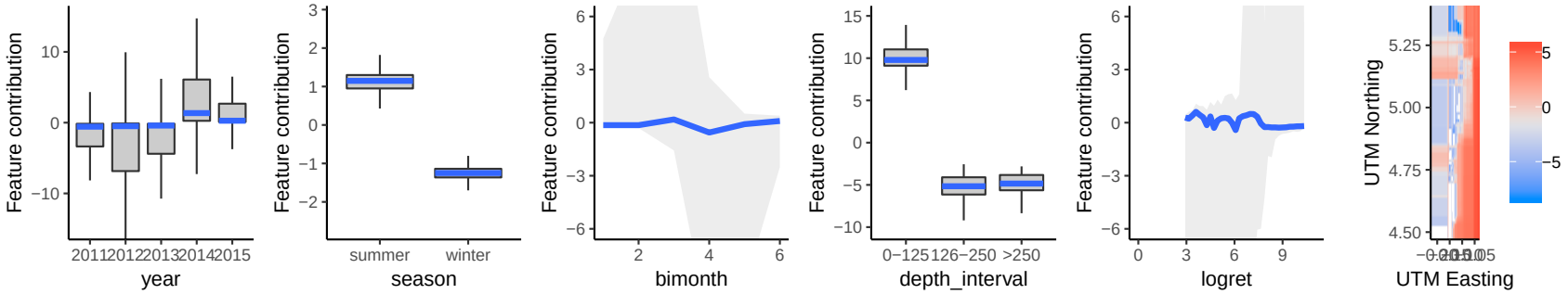
Designed to *mimic the stratified ratio estimator* by treating `year`, `season`, and `depth_interval` as factors. `bimonth` is included as linear and quadratic terms to avoid confounding with `fac(season)`.

gam(Y ~ fac(year) + fac(season) + fac(depth_interval) + bimonth + I(bimonthˆ2) + logret + s(lat, long, k=50))

Species 1: Big Skate

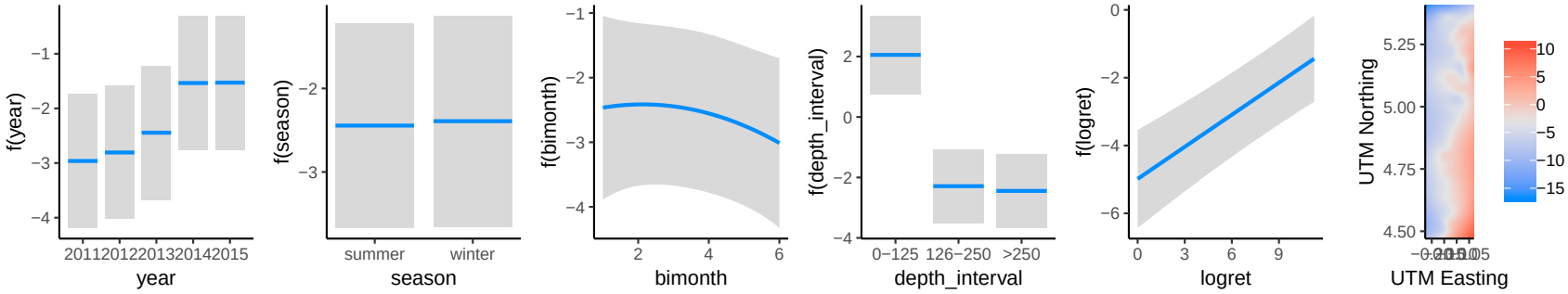
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 1989.101
          % Var explained: 47.8
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.233 Deviance explained = 65.8% -REML = 32567 Scale est. = 102.63 n = 35440



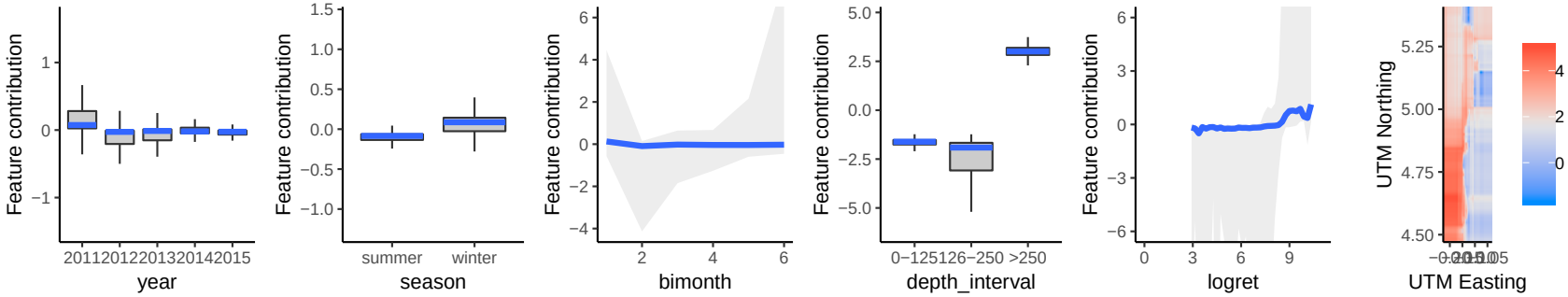
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	46.2	48.5	15.1	8.66e-121	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-6.02	0.806	-7.47	8.29e-14	***
seasonwinter	0.0525	0.248	0.212	0.832	
bimonth	0.169	0.37	0.457	0.648	
I(bimonth^2)	-0.0398	0.0486	-0.819	0.413	
year2012	0.156	0.158	0.987	0.324	
year2013	0.518	0.164	3.15	0.00163	**
year2014	1.43	0.156	9.17	4.96e-20	***
year2015	1.43	0.16	8.97	3.2e-19	***
depth_interval0-125	4.5	0.348	12.9	3.13e-38	***
depth_interval126-250	0.157	0.262	0.598	0.55	
logret	0.315	0.05	6.32	2.73e-10	***

Species 2: Black Skate

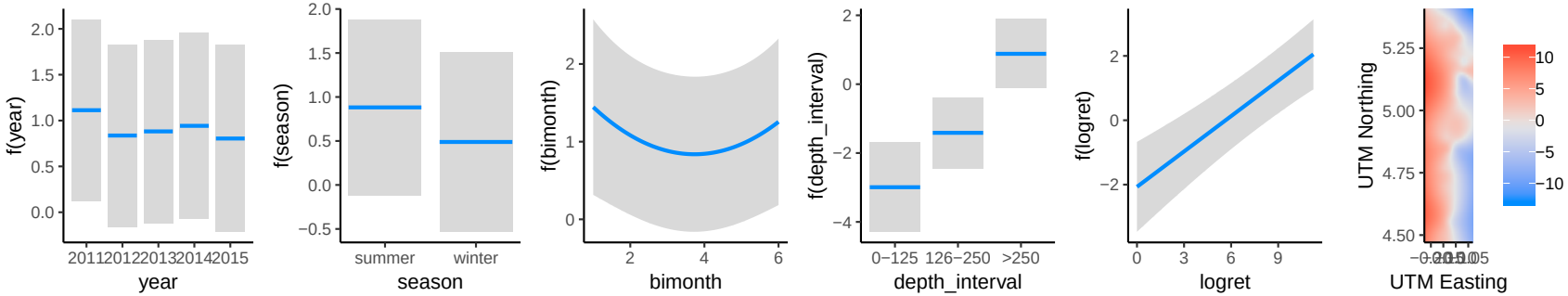
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 253.1007
          % Var explained: 48.13
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = -1.54 Deviance explained = 56% -REML = 33123 Scale est. = 139.75 n = 35440



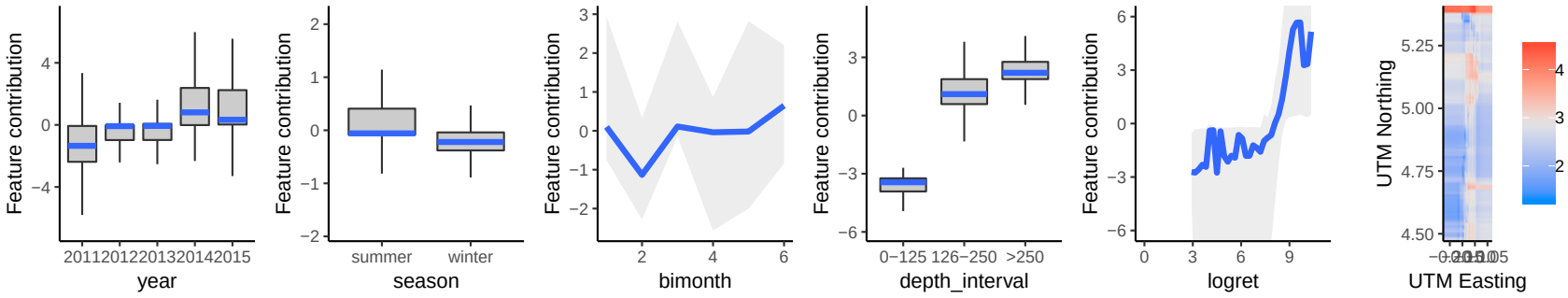
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	47	48.7	8.53	7.16e-59	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-0.848	0.71	-1.19	0.232	
seasonwinter	-0.393	0.203	-1.94	0.0529	.
bimonth	-0.606	0.272	-2.23	0.0257	*
I(bimonth^2)	0.0812	0.0368	2.21	0.0274	*
year2012	-0.276	0.172	-1.61	0.108	
year2013	-0.231	0.169	-1.37	0.171	
year2014	-0.171	0.18	-0.947	0.344	
year2015	-0.309	0.191	-1.61	0.107	
depth_interval0-125	-3.87	0.461	-8.4	4.8e-17	***
depth_interval126-250	-2.29	0.193	-11.9	2.2e-32	***
logret	0.366	0.0634	5.77	7.85e-09	***

Species 3: Brown Cat Shark

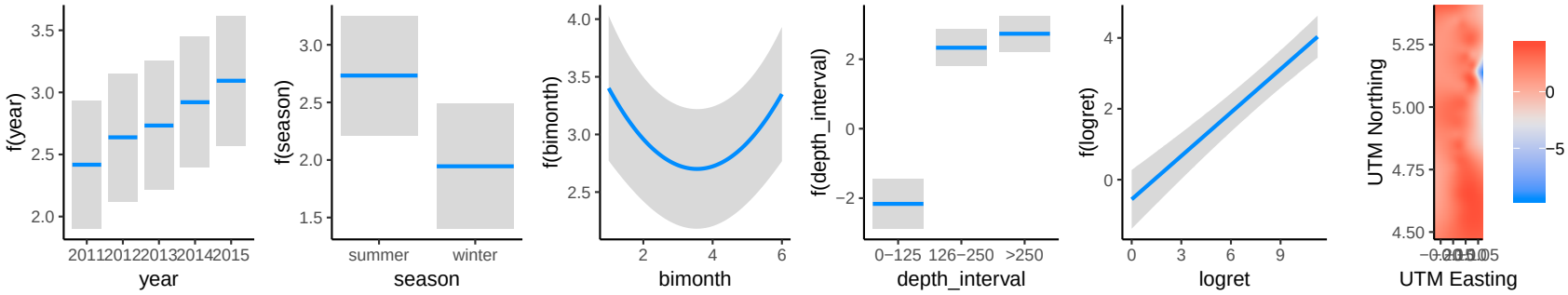
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 356.8372
          % Var explained: 20.8
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.121 Deviance explained = 45.4% -REML = 74672 Scale est. = 88.94 n = 35440



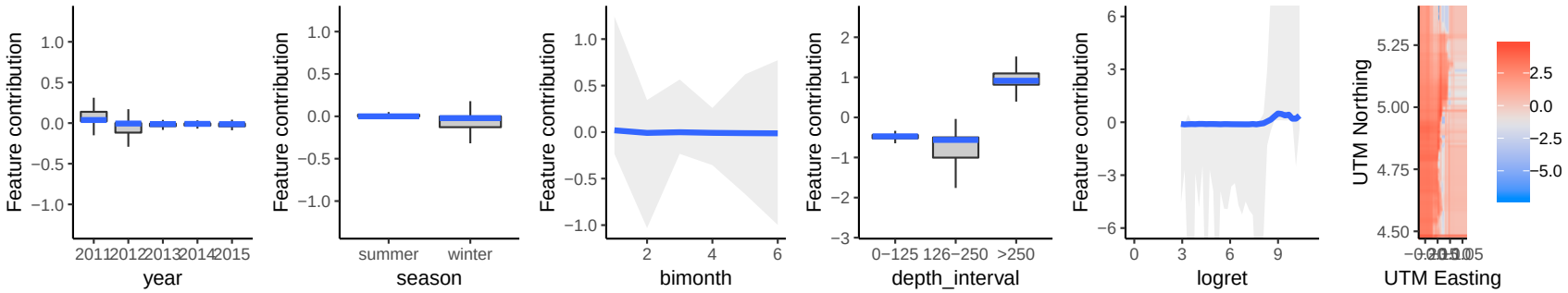
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48.3	49	3.7	6.32e-17	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-0.18	0.479	-0.376	0.707	
seasonwinter	-0.788	0.135	-5.82	5.99e-09	***
bimonth	-0.766	0.184	-4.16	3.13e-05	***
I(bimonth^2)	0.108	0.0249	4.34	1.46e-05	***
year2012	0.221	0.119	1.86	0.0631	.
year2013	0.316	0.115	2.75	0.00591	**
year2014	0.503	0.121	4.14	3.44e-05	***
year2015	0.676	0.122	5.55	2.9e-08	***
depth_interval0-125	-4.9	0.271	-18.1	1.6e-72	***
depth_interval126-250	-0.402	0.0936	-4.3	1.73e-05	***
logret	0.408	0.0423	9.65	5.37e-22	***

Species 4: California Slickhead

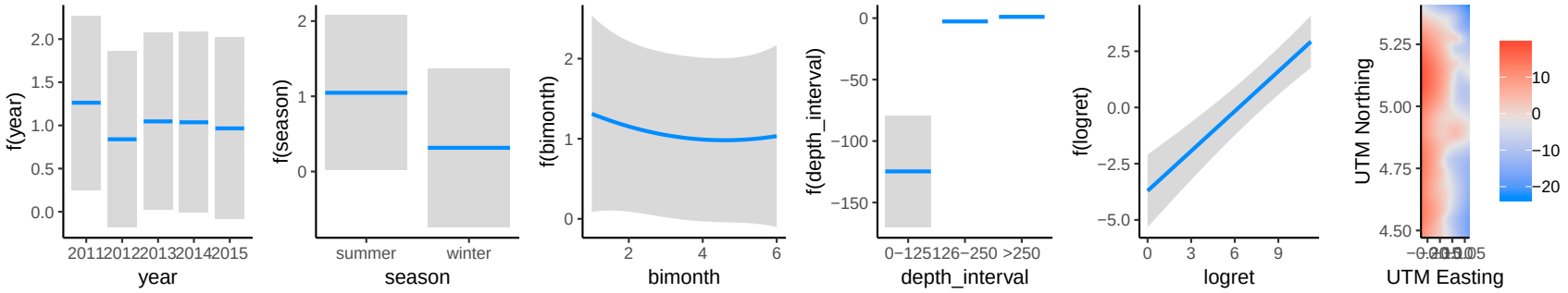
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 145.0971
          % Var explained: 42.62
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = -9.19e+03 Deviance explained = 65.7% -REML = 18895 Scale est. = 97.632 n = 35440



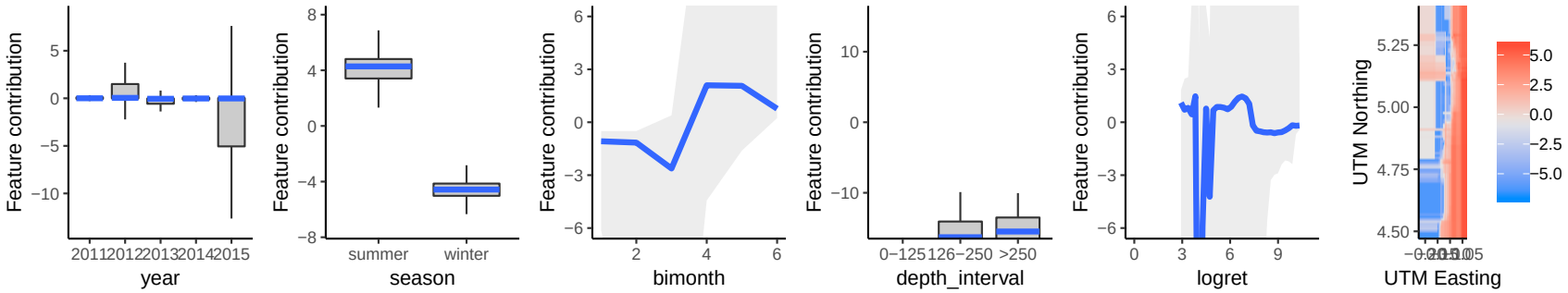
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	43.7	46.4	19.1	1.78e-153	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-5.31	0.979	-5.42	5.87e-08	***
seasonwinter	-0.731	0.252	-2.9	0.00375	**
bimonth	-0.235	0.343	-0.686	0.492	
I(bimonth^2)	0.0256	0.0463	0.553	0.58	
year2012	-0.423	0.214	-1.98	0.048	*
year2013	-0.216	0.211	-1.03	0.304	
year2014	-0.227	0.23	-0.986	0.324	
year2015	-0.297	0.242	-1.23	0.219	
depth_interval0-125	-126	23.1	-5.43	5.6e-08	***
depth_interval126-250	-3.74	0.251	-14.9	2.68e-50	***
logret	0.59	0.0808	7.31	2.83e-13	***

Species 5: Dungeness Crab

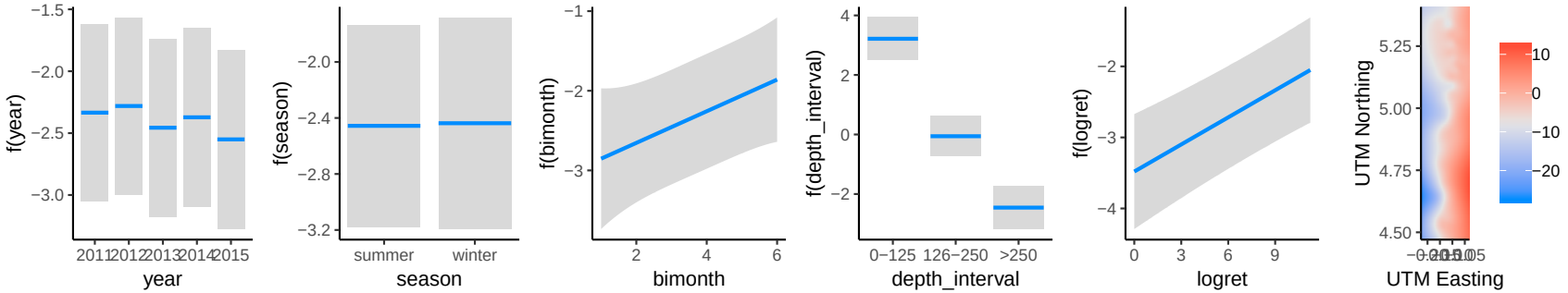
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 11727.1
          % Var explained: 28.81
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.213 Deviance explained = 75.9% -REML = 65965 Scale est. = 55.075 n = 35440



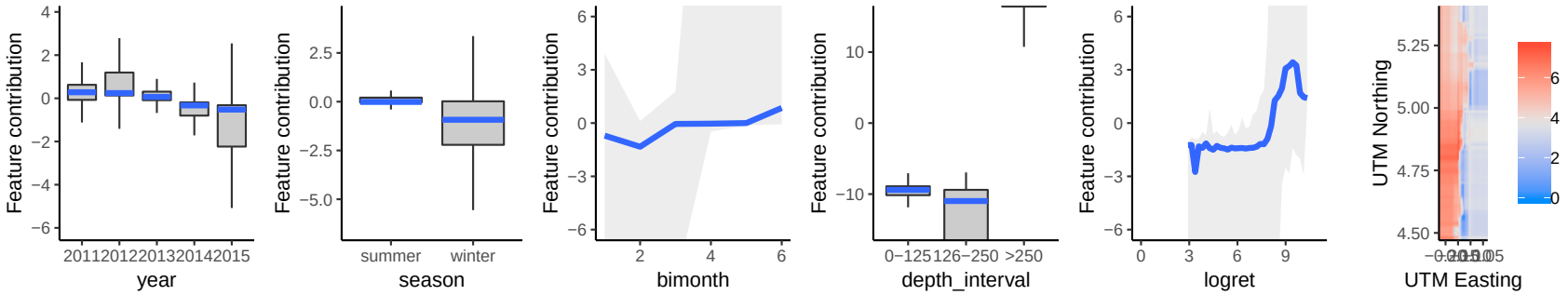
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	47.9	48.9	19.5	9.75e-166	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-5.35	0.542	-9.87	6.04e-23	***
seasonwinter	0.0186	0.162	0.115	0.909	
bimonth	0.198	0.26	0.763	0.445	
I(bimonth^2)	6.62e-05	0.0336	0.00197	0.998	
year2012	0.0538	0.0808	0.666	0.506	
year2013	-0.122	0.0853	-1.42	0.154	
year2014	-0.0376	0.0877	-0.429	0.668	
year2015	-0.216	0.0901	-2.4	0.0165	*
depth_interval0-125	5.67	0.234	24.3	3.16e-129	***
depth_interval126-250	2.4	0.195	12.3	1.03e-34	***
logret	0.127	0.0246	5.16	2.5e-07	***

Species 6: Grenadier

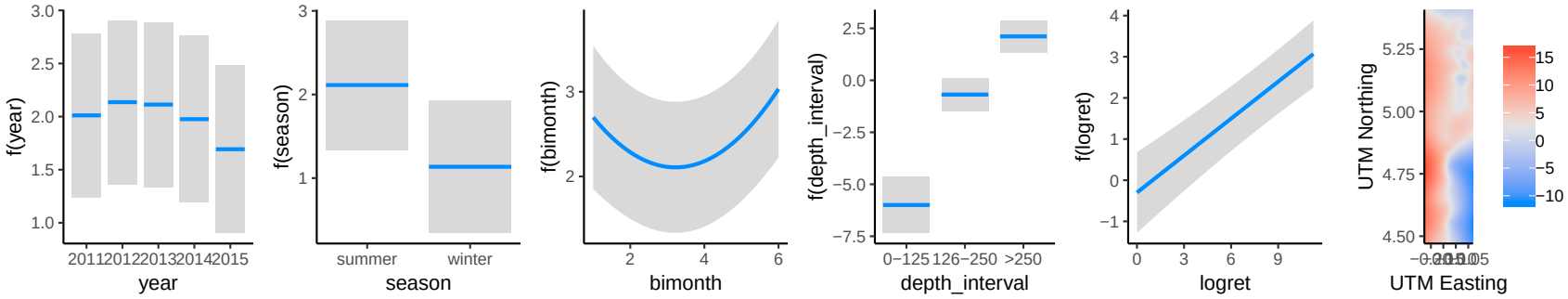
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 18071.56
          % Var explained: 49.32
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = -241 Deviance explained = 66.8% -REML = 63079 Scale est. = 133.84 n = 35440



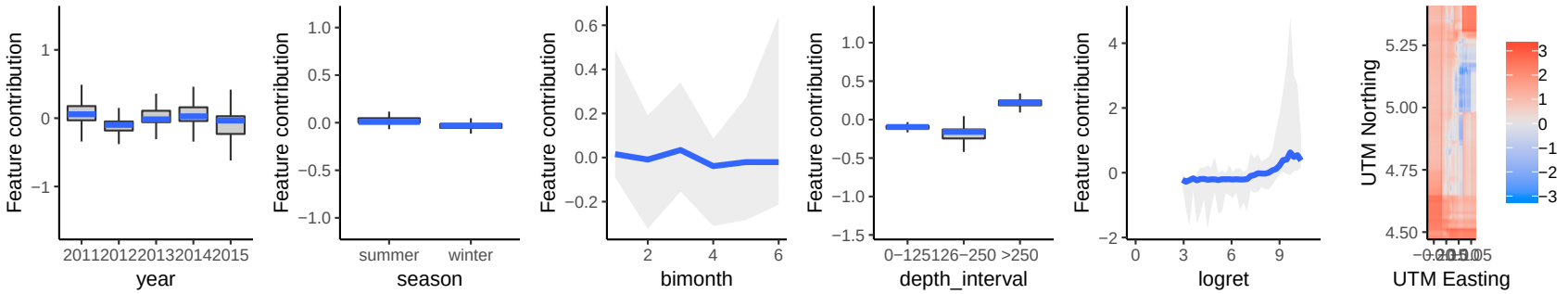
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48	48.9	34.4	8.71000001824745e-316	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	1.6	0.462	3.47	0.000516	***
seasonwinter	-0.977	0.131	-7.48	7.64e-14	***
bimonth	-0.771	0.181	-4.27	1.95e-05	***
I(bimonth^2)	0.12	0.0244	4.91	9.18e-07	***
year2012	0.125	0.105	1.19	0.235	
year2013	0.102	0.106	0.964	0.335	
year2014	-0.0354	0.118	-0.299	0.765	
year2015	-0.318	0.13	-2.46	0.014	*
depth_interval0-125	-8.11	0.58	-14	2.96e-44	***
depth_interval126-250	-2.8	0.137	-20.5	7.95e-93	***
logret	0.299	0.0394	7.6	2.95e-14	***

Species 7: Octopus Unid

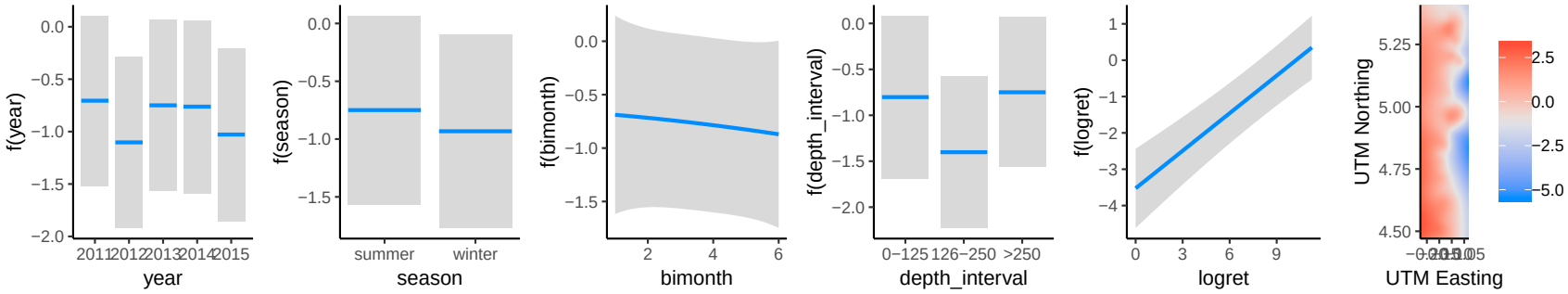
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 36.92363
          % Var explained: 2.19
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.0433 Deviance explained = 20.2% -REML = 27714 Scale est. = 56.927 n = 35440



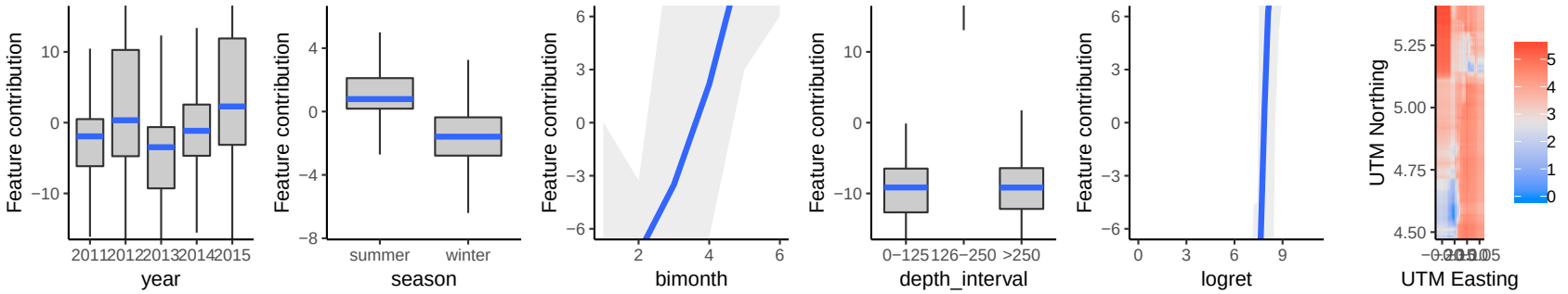
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	43.5	47.7	7.92	1.99e-52	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-2.76	0.559	-4.95	7.61e-07	***
seasonwinter	-0.183	0.164	-1.12	0.264	
bimonth	-0.0225	0.226	-0.0993	0.921	
I(bimonth^2)	-0.002	0.0304	-0.0658	0.948	
year2012	-0.398	0.137	-2.91	0.00363	**
year2013	-0.0443	0.126	-0.35	0.726	
year2014	-0.0568	0.136	-0.418	0.676	
year2015	-0.323	0.143	-2.25	0.0242	*
depth_interval0-125	-0.0528	0.219	-0.241	0.809	
depth_interval126-250	-0.652	0.125	-5.24	1.65e-07	***
logret	0.345	0.0474	7.27	3.76e-13	***

Species 8: Pacific Hake

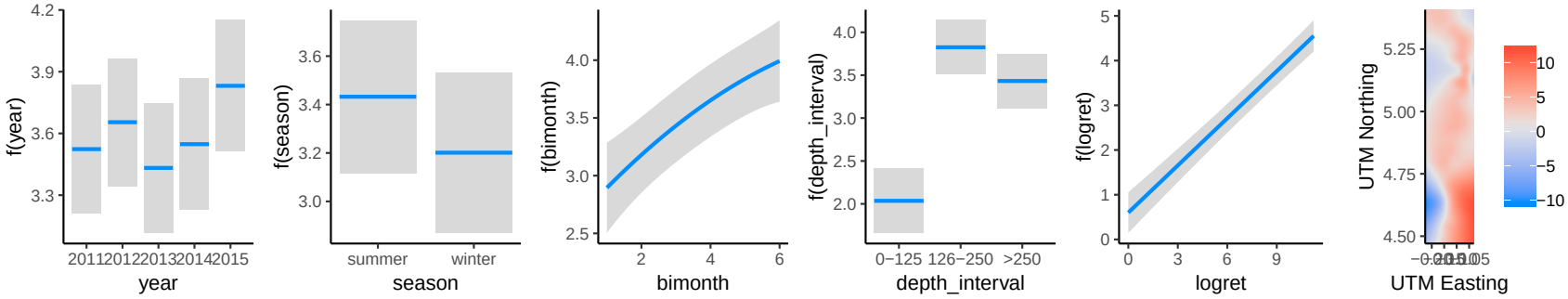
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 18788.2
          % Var explained: 20.61
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.0796 Deviance explained = 22.8% -REML = 1.2473e+05 Scale est. = 49.425 n = 35440



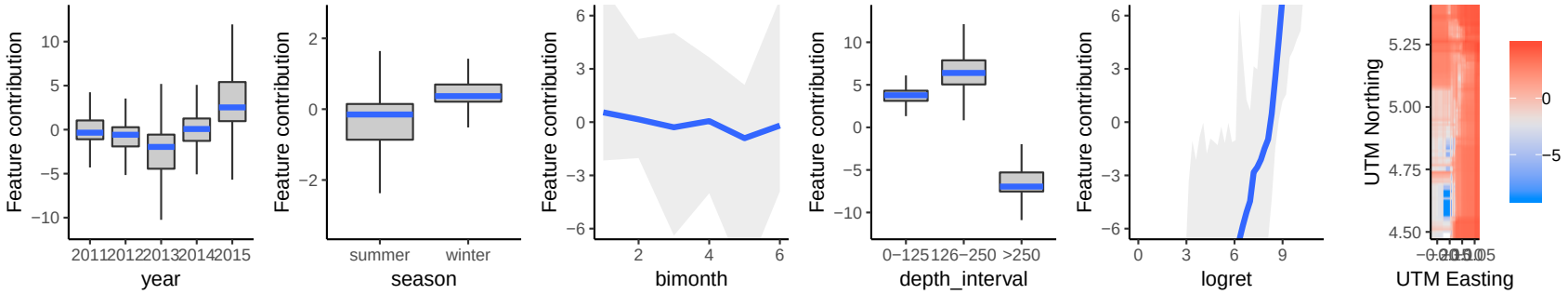
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48.7	49	21.1	2.24e-182	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	0.00278	0.277	0.01	0.992	
seasonwinter	-0.231	0.0825	-2.8	0.00511	**
bimonth	0.334	0.119	2.81	0.00496	**
I(bimonth^2)	-0.0164	0.0159	-1.03	0.301	
year2012	0.131	0.0643	2.03	0.0419	*
year2013	-0.0913	0.0637	-1.43	0.152	
year2014	0.0242	0.0684	0.354	0.723	
year2015	0.308	0.0679	4.54	5.78e-06	***
depth_interval0-125	-1.4	0.117	-12	5.58e-33	***
depth_interval126-250	0.393	0.0567	6.93	4.36e-12	***
logret	0.352	0.0216	16.3	2.56e-59	***

Species 9: Pacific Halibut

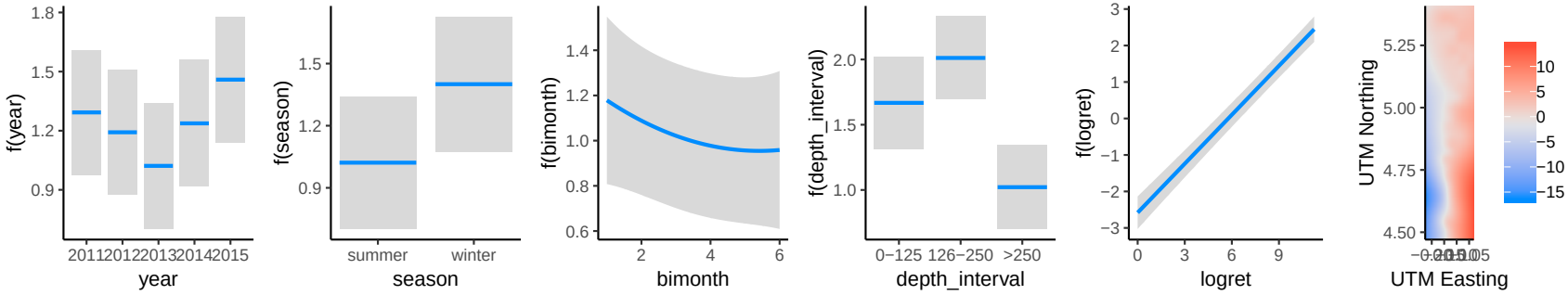
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 3041.369
          % Var explained: 18.98
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.103 Deviance explained = 27.3% -REML = 74265 Scale est. = 44.927 n = 35440



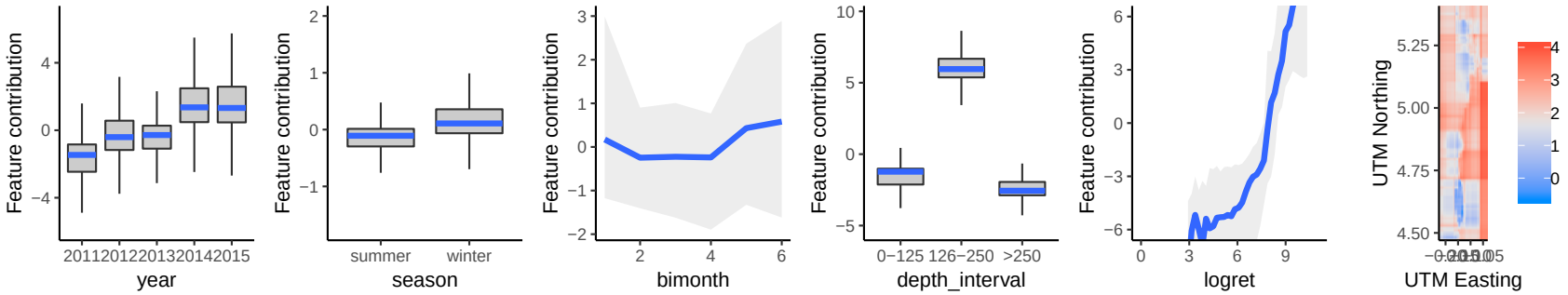
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	47.7	48.9	43.8	0	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-1.62	0.239	-6.78	1.2e-11	***
seasonwinter	0.379	0.0713	5.31	1.08e-07	***
bimonth	-0.124	0.0959	-1.3	0.195	
I(bimonth^2)	0.0115	0.0129	0.891	0.373	
year2012	-0.101	0.0565	-1.79	0.0742	.
year2013	-0.271	0.0563	-4.82	1.47e-06	***
year2014	-0.0556	0.0595	-0.935	0.35	
year2015	0.166	0.0576	2.89	0.00386	**
depth_interval0-125	0.646	0.0996	6.49	8.95e-11	***
depth_interval126-250	0.99	0.0524	18.9	3.35e-79	***
logret	0.447	0.0203	22.1	3.85e-107	***

Species 10: Sandpaper Skate

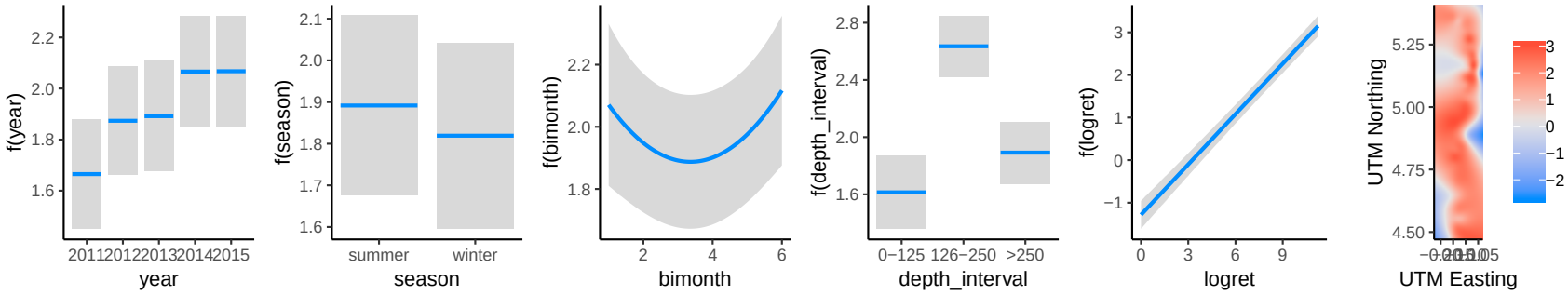
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 385.3107
          % Var explained: 22.37
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.125 Deviance explained = 20% -REML = 92531 Scale est. = 20.414 n = 35440



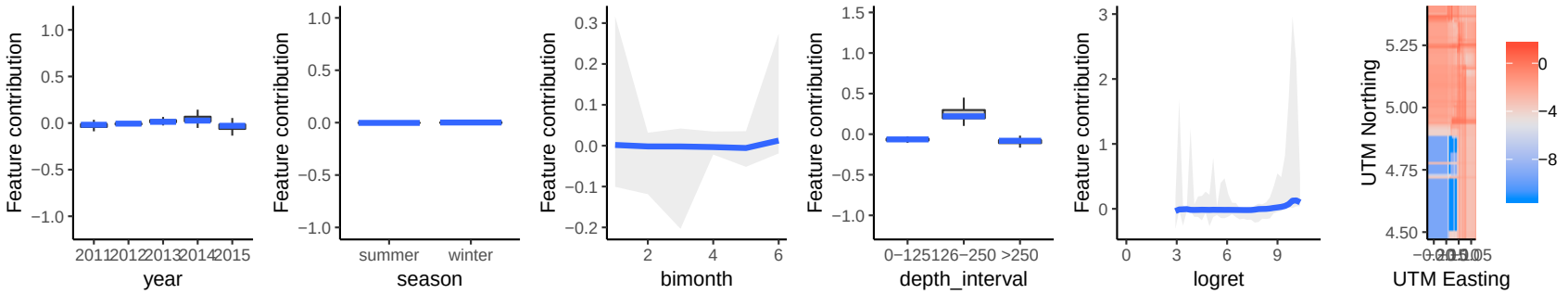
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48.3	49	28.5	1.56e-256	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-1.17	0.19	-6.13	8.83e-10	***
seasonwinter	-0.0725	0.0552	-1.31	0.189	
bimonth	-0.221	0.0756	-2.92	0.00345	**
I(bimonth^2)	0.0329	0.0102	3.24	0.00121	**
year2012	0.209	0.0462	4.52	6.3e-06	***
year2013	0.227	0.0449	5.05	4.55e-07	***
year2014	0.401	0.0471	8.52	1.64e-17	***
year2015	0.403	0.0476	8.47	2.65e-17	***
depth_interval0-125	-0.278	0.0803	-3.46	0.00054	***
depth_interval126-250	0.744	0.0387	19.2	8.38e-82	***
logret	0.394	0.0162	24.3	2.52e-129	***

Species 11: Rosethorn Rockfish

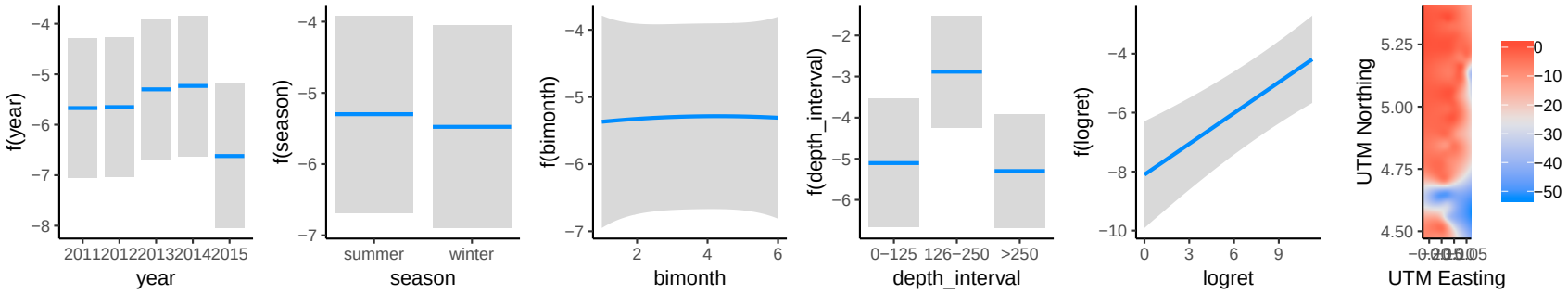
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 7.098818
          % Var explained: 17.73
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.00397 Deviance explained = 52.8% -REML = 8403.7 Scale est. = 63.336 n = 35440



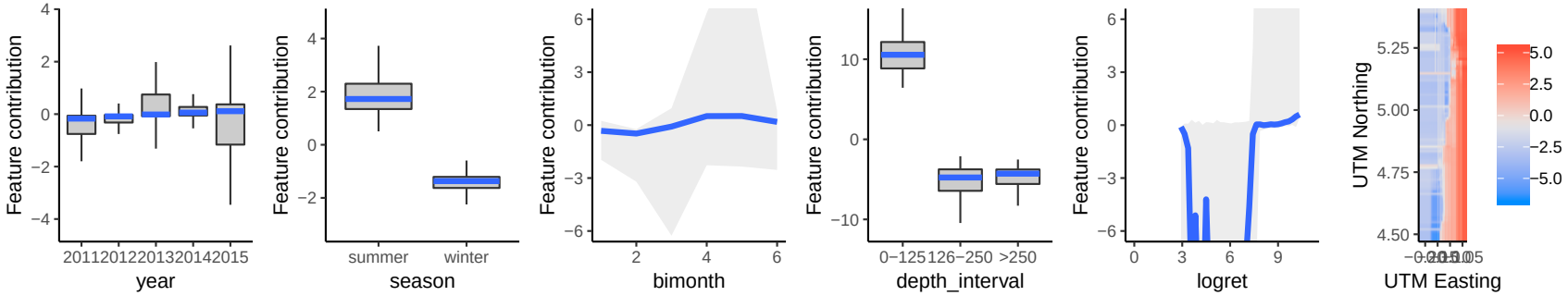
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	45.9	47.2	9.94	4.45e-69	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-11.1	1.46	-7.61	2.86e-14	***
seasonwinter	-0.178	0.291	-0.611	0.541	
bimonth	0.0664	0.394	0.168	0.866	
I(bimonth^2)	-0.00776	0.0533	-0.146	0.884	
year2012	0.0185	0.253	0.0731	0.942	
year2013	0.371	0.234	1.58	0.114	
year2014	0.439	0.257	1.71	0.0873	.
year2015	-0.951	0.301	-3.16	0.00157	**
depth_interval0-125	0.195	0.443	0.44	0.66	
depth_interval126-250	2.42	0.211	11.4	2.85e-30	***
logret	0.348	0.0767	4.54	5.74e-06	***

Species 12: Slender Sole

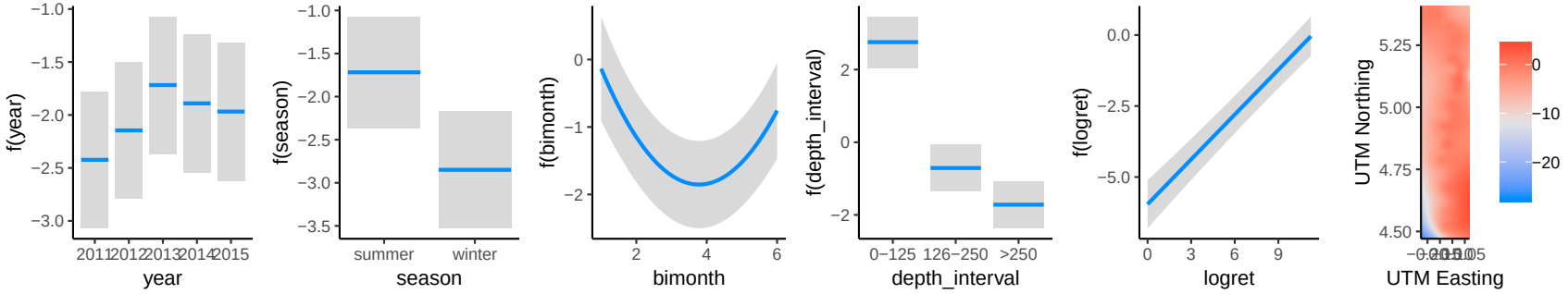
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 2193.15
          % Var explained: 28.43
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.129 Deviance explained = 47.3% -REML = 51504 Scale est. = 47.418 n = 35440



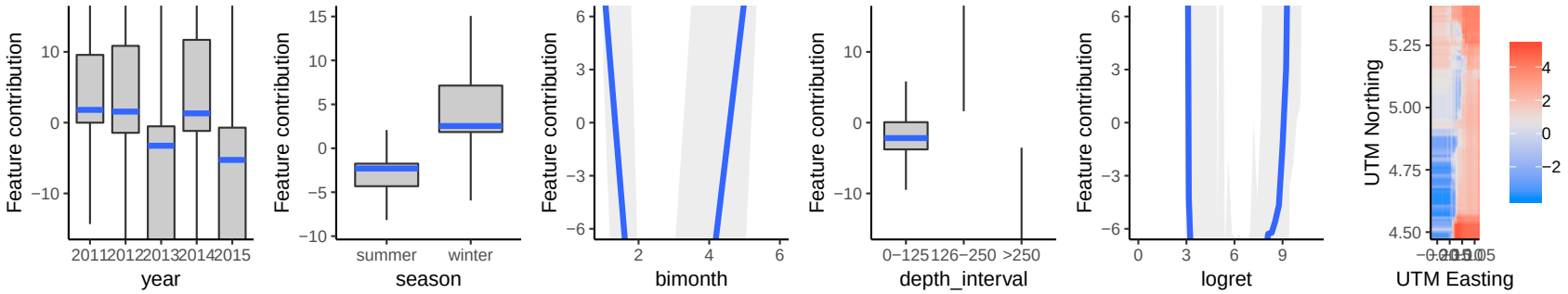
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48	48.9	13.1	9.18e-104	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-3.53	0.495	-7.12	1.08e-12	***
seasonwinter	-1.13	0.159	-7.11	1.14e-12	***
bimonth	-1.68	0.218	-7.69	1.47e-14	***
I(bimonth^2)	0.222	0.0292	7.62	2.52e-14	***
year2012	0.278	0.114	2.44	0.0147	*
year2013	0.705	0.11	6.44	1.23e-10	***
year2014	0.533	0.116	4.57	4.78e-06	***
year2015	0.456	0.117	3.88	0.000103	***
depth_interval0-125	4.47	0.199	22.5	4.96e-111	***
depth_interval126-250	1.01	0.122	8.28	1.31e-16	***
logret	0.526	0.0359	14.7	1.51e-48	***

Species 13: Spiny Dogfish Shark

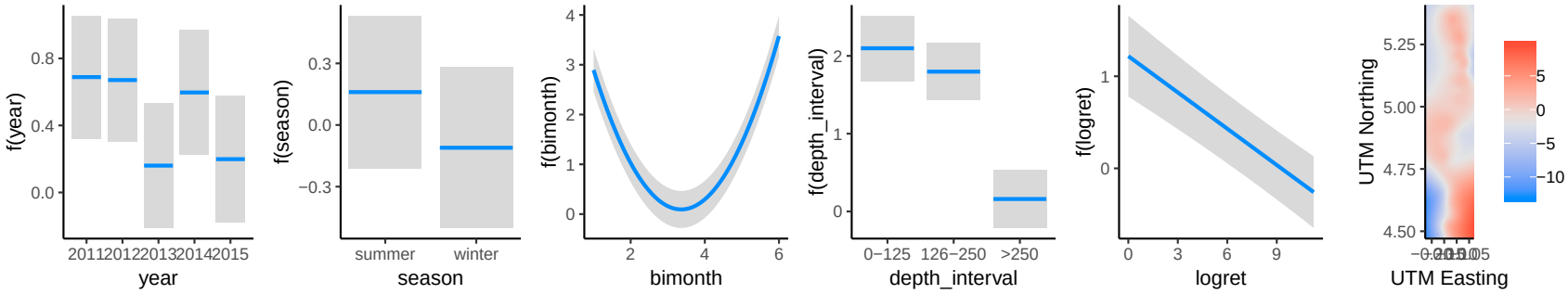
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 444124.3
          % Var explained: 25.06
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.0248 Deviance explained = 40.6% -REML = 1.0483e+05 Scale est. = 64.451 n = 35440



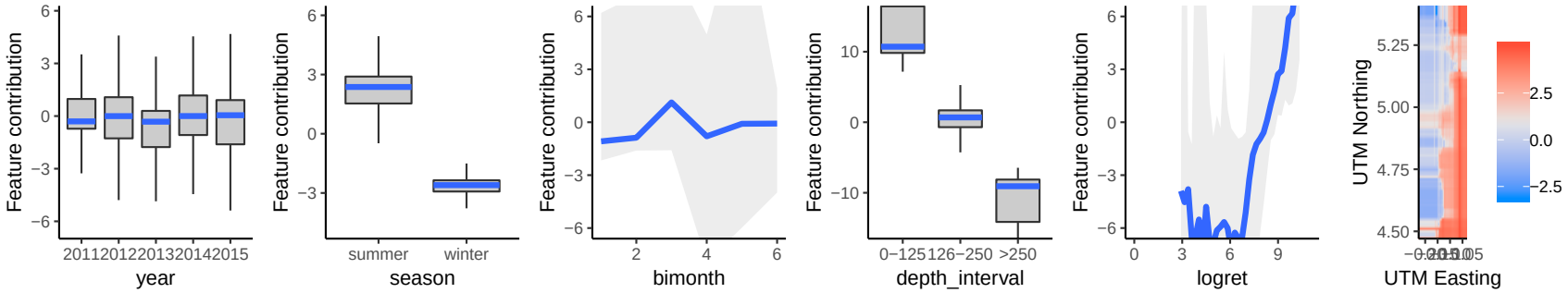
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48.2	49	40.8	0	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	7.67	0.255	30.1	2.23e-196	***
seasonwinter	-0.271	0.0944	-2.87	0.00412	**
bimonth	-3.37	0.124	-27.2	1.46e-161	***
I(bimonth^2)	0.501	0.0169	29.7	5.82e-192	***
year2012	-0.0176	0.0692	-0.255	0.799	
year2013	-0.528	0.069	-7.65	1.98e-14	***
year2014	-0.0918	0.0724	-1.27	0.204	
year2015	-0.49	0.0753	-6.5	8.04e-11	***
depth_interval0-125	1.94	0.124	15.6	8.99e-55	***
depth_interval126-250	1.64	0.0654	25	6.35e-137	***
logret	-0.131	0.0158	-8.31	9.88e-17	***

Species 14: Spotted Ratfish

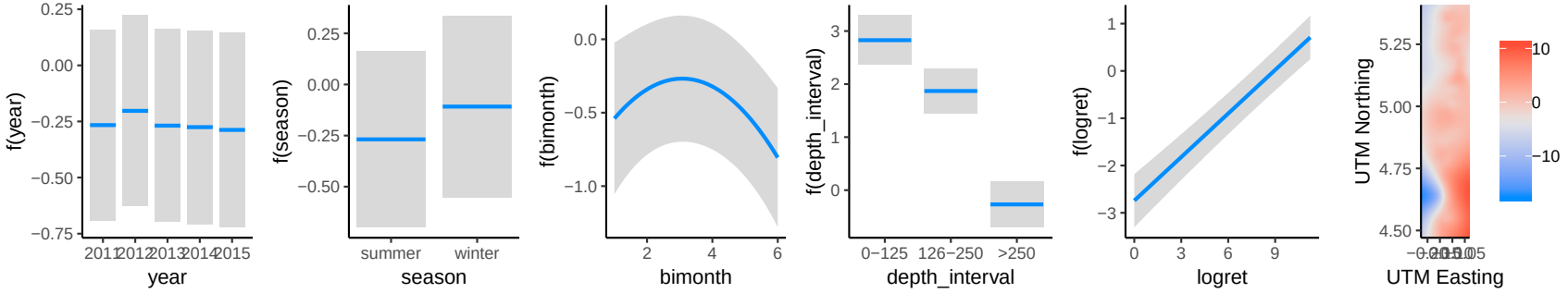
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 6941.341
          % Var explained: 7.86
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.0237 Deviance explained = 34.4% -REML = 88372 Scale est. = 46.446 n = 35440



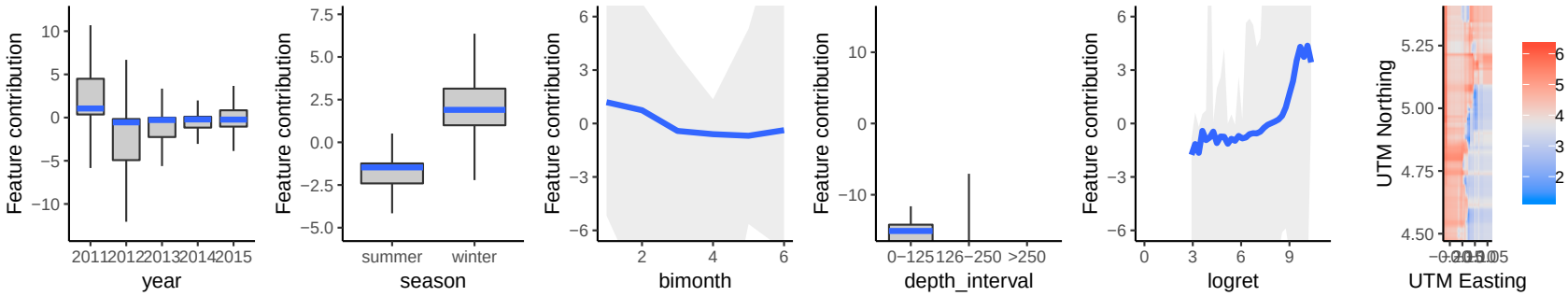
term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48.4	49	23.6	3.53e-205	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	-2.69	0.332	-8.09	5.94e-16	***
seasonwinter	0.16	0.103	1.56	0.118	
bimonth	0.388	0.146	2.66	0.00777	**
I(bimonth^2)	-0.063	0.0194	-3.25	0.00115	**
year2012	0.0635	0.0759	0.836	0.403	
year2013	-0.00229	0.0746	-0.0306	0.976	
year2014	-0.00898	0.0789	-0.114	0.909	
year2015	-0.0214	0.0793	-0.27	0.787	
depth_interval0-125	3.1	0.131	23.6	5.54e-122	***
depth_interval126-250	2.14	0.0778	27.5	2.39e-164	***
logret	0.307	0.024	12.8	1.76e-37	***

Species 15: Tanneri Tanner Crab

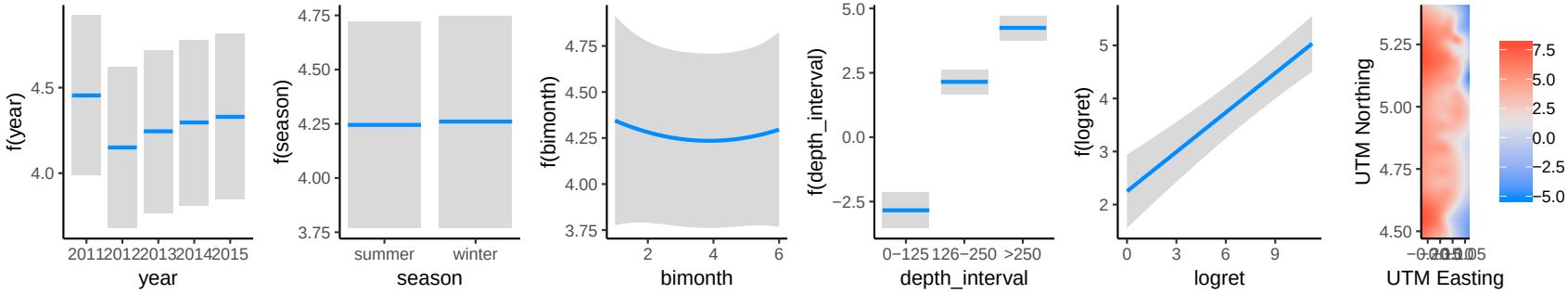
$\mathbf{RF_{strat}}$

```
Mean of squared residuals: 7438.715
          % Var explained: 34.52
```



$\mathbf{GAM_{strat}}$

R-sq.(adj) = 0.176 Deviance explained = 53.9% -REML = 85417 Scale est. = 85.96 n = 35440



term	edf	ref.df	statistic	p.value	signif
s(avg_lat,avg_long)	48	48.9	8.84	2.42e-62	***

term	estimate	std.error	statistic	p.value	signif
(Intercept)	1.73	0.398	4.35	1.38e-05	***
seasonwinter	0.0155	0.121	0.128	0.898	
bimonth	-0.103	0.162	-0.637	0.524	
I(bimonth^2)	0.0134	0.022	0.609	0.543	
year2012	-0.304	0.105	-2.9	0.00378	**
year2013	-0.21	0.101	-2.07	0.0386	*
year2014	-0.158	0.11	-1.43	0.153	
year2015	-0.125	0.113	-1.1	0.271	
depth_interval0-125	-7.08	0.273	-26	1.97e-147	***
depth_interval126-250	-2.1	0.0956	-21.9	7.42e-106	***
logret	0.248	0.0332	7.46	9.18e-14	***