 Collect. Vol. Sci. Pap. ICCAT, 75(5): 940-947 (2018)

STANDARDIZING US BLUE MARLIN LONGLINE CPUE USING HABITAT COVARIATES

C.P. Goodyear,[1] F. Forrestal,[2] M. Schirripa,[3] and M. Lauretta[2]

SUMMARY

The relative value of species habitat at the location of a longline set can be estimated with a species distribution model (SDM). These data can then be used as covariates in GLM CPUE standardizations to replace intra-annual spatiotemporal strata (area, season). We compared the two approaches using blue marlin (Makaira nigricans) data from US longline logbooks. GLM covariates included 1) gear features, month and area, or 2) gear features and either a habitat coefficient (w) or habitat relative density (H). Habitat relative densities were obtained from either of two SDMs, and habitat coefficients were estimated from H using hook depths of individual gears. All GLM standardized abundance predictions differed from the trend in nominal catch rates. The trends predicted with the habitat-based covariates for the baseline SDM were essentially the same as those from the standard approach using month and area (r^2=0.98, n=30). Those with covariates from the alternative SDM were also very similar (r^2=0.95-0.96). SDM-derived habitat covariates could obviate problems with statistical imbalance and improve standardizations in many situations.

RÉSUMÉ

*La valeur relative de l'habitat des espèces à l'emplacement d'un mouillage de palangres peut être estimée à l'aide d'un modèle de distribution des espèces (SDM). Ces données peuvent ensuite être utilisées comme covariables dans les standardisations avec GLM de la CPUE pour remplacer les strates spatio-temporelles intra-annuelles (zone, saison). Nous avons comparé les deux approches en utilisant les données sur le makaire bleu (*Makaira nigricans*) provenant des carnets de pêche des palangriers américains. Les covariables du GLM comprenaient 1) les caractéristiques des engins, le mois et la zone, ou 2) les caractéristiques des engins et un coefficient d'habitat (w) ou une densité relative de l'habitat (H). Les densités relatives de l'habitat ont été obtenues à partir de l'un des deux SDM, et les coefficients de l'habitat ont été estimés à partir de H en utilisant la profondeur des hameçons des différents engins. Toutes les prévisions d'abondance standardisées avec le GLM différaient de la tendance des taux de capture nominaux. Les tendances prédites avec les covariables basées sur l'habitat pour le SDM de référence étaient essentiellement les mêmes que celles de l'approche standard utilisant le mois et la zone (r^2 = 0,98, n = 30). Celles prédites avec des covariables du SDM alternatif étaient également très similaires (r^2 = 0,95-0,96). Les covariables de l'habitat obtenues du SDM pourraient éviter les problèmes de déséquilibre statistique et améliorer les standardisations dans de nombreuses situations.*

RESUMEN

*El valor relativo del hábitat de las especies en la localización de un lance de palangre puede estimarse con un modelo de distribución de especies (SDM). Estos datos pueden usarse como covariables en las estandarizaciones de CPUE con GLM para sustituir los estratos espaciotemporales dentro del año (área, temporada). Comparamos los dos enfoques utilizando los datos de aguja azul (*Makaira nigricans*) de los cuadernos de pesca del palangre estadounidense. Las covariables del GLM incluían: 1) características del arte, mes y área o 2) características del arte y bien un coeficiente de hábitat (w) o bien la densidad relativa del hábitat (H). Las densidades relativas del hábitat (H) se obtuvieron mediante uno u otro de los dos SDM, y los coeficientes del hábitat se estimaron a partir de H utilizando profundidades de anzuelo de*

[1] 1214 North Lakeshore Drive, Niceville, Florida 32578, USA. phil_goodyear@msn.com

[2] CIMAS/RSMAS University of Miami, 4600 Rickenbacker Cwy. Miami, FL 33149. fforrestal@rsmas.miami.edu

[3] NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA. Michael.Schirripa@noaa.gov

artes individuales. Todas las predicciones de abundancia estandarizada con GLM diferían de la tendencia en las tasas de captura nominal. Las tendencias predichas con las covariables basadas en el hábitat para el SDM de referencia eran esencialmente las mismas que las del enfoque estándar utilizando mes y área (r^2=0,98, n=30). Aquellas con covariables del SDM alternativo eran también muy similares (r^2=0,95-0,96). Las covariables del hábitat derivadas de SDM podrían obviar problemas con el desequilibrio estadístico y mejorar las estandarizaciones en muchas situaciones.

KEYWORDS

Blue Marlin, Longline, Catchability, Gear coefficient, Habitat coefficient, Stock assessment, CPUE Standardization, Statistics, GLM, Population modeling

1. Introduction

Catchability for a longline can be modeled as a joint function of a habitat coefficient (*w*) and an essential gear coefficient (*k,* Goodyear et al. 2017, 2018). The habitat coefficient can be estimated for each set using knowledge of the fishing characteristics of the gear and relative value of habitat *H* around each hook at the time and location of the set:

$$w = \overline{\overline{H}_h}$$

where: $\overline{H}_h$ = the average habitat value around the hook and $\overline{\overline{H}_h}$ is the average around all hooks on the set. The values of *H* are expressed as density at a unit of population and are derived from an appropriate species distribution model (see Goodyear et al. 2018 for more detail). The values of *H* and *w* are continuous variables proportional to density of the fish at hooks (*w*) or in the vicinity of (*H*) longline sets. Appropriately estimated, these variables should be potent correlates of species catch rates. Here we compare the species abundance trends from GLM standardizations using traditional covariates for intra-annual spatiotemporal variability with the species abundance trends from GLMs that use these habitat covariates. The comparisons were based on the US longline logbooks using blue marlin species distribution models.

2. Method

2.1 Species distribution model

The SDM used in this study is a detailed model of the four-dimensional distribution of blue marlin (Goodyear 2016). Oceanographic data and species habitat preferences are used to distribute the population in time and space. The current implementation partitions the Atlantic from 50 S to 55 N latitude at a spatial resolution of 1º latitude and 1º longitude with 46 depth layers. The oceanographic data were monthly values from the Earth System Model from 1956 to 2012 and matched the spatial resolution of the SDM. The oceanographic data were provided by colleagues at the US National Atlantic Oceanographic and Meteorological Laboratory (AOML). At the time of this study 2012 was the last year that oceanographic data were available. The 2012 values were substituted as needed to provide oceanographic data through 2015. This convention accounts for the large month-to-month variability in oceanographic conditions but omits any effects from annual trends that may have been important in the last few years.

Two implementations of the SDM were employed in the analyses here. The baseline was the model for Atlantic blue marlin described in Goodyear (2016) but with year by year monthly oceanography. An alternative was used to explore the sensitivity of the CPUE standardizations to error in the SDM predictions. This model substituted a temperature preference profile in which the species prefers higher temperatures **(Figure 1**). This approach (the thermophilic SDM) adjusted the observed PSAT-tag data by the average volume of habitat within the observed temperature bins (see the discussion in Goodyear 2016). At the highest temperatures (>30C) the predicted relative densities are much elevated by the thermophilic model assumption. However, the volumes of ocean strata within the temperature extremes are relatively very small. As a consequence, the population fractions for the two models within cooler strata (below 30º C) are not as different as might be inferred from **Figure 1.** Nonetheless, the predicted densities in ocean strata at the highest temperatures are much higher than for the baseline assumption.

2.2 Fishery data

The US Longline Logbook data evaluated here covered the period 1986-2015. These data were used in a related study of the performance of standardization methods that used simulated longline catch data (Forrestal et al. 2017). That study described 128 discrete gear types in the fishery. These gear profiles were adopted unchanged for the current analyses. The features of the gear that are included were known to be important because of the results of prior studies (Forrestal et al. 2017). A scan of the longline catch-effort data revealed errors in values recorded for the number of hooks on some sets. As a consequence, records with fewer than 220 hooks per set (mean-2SD) were removed. The resulting final data set contained about 290 thousand records (97.7% of the total).

2.3 Data compilation

The habitat covariates were compiled for each set using the protocol described in Goodyear et al. (2018). The information for each set from the logbook file identifies the gear, month, year and location (latitude and longitude) and the numbers of blue marlin caught. Data from an SDM provides estimates of the species relative densities (H) based on species behavior and habitat considerations for the year and month and latitude and longitude of the set. The estimated probability distributions for the hooks are read from gear files. This pre-processing step reads a catch-effort record from the simulated logbook and computes w from the SDM and hook data. Each output record contains the original CPUE data and adds w, and the average surface to 100m habitat relative density, H_{100}.

2.4 Analyses

The GLM's were run in R using the glmmADMB library (R Core Team, 2015). The standardized annual abundance predictions combined separate GLM's for the successful sets and the catch rates of those that were successful. There was no attempt to select the variables for each fit based on any performance-based criteria or make judgments about the quality of the fits to the simulated data. The habitat coefficients (w) and habitat relative densities (H) were included as numerical variables. Factors included year, month, area, the use of light sticks, hook type, bait type, and hooks between floats (hbf), The area assignments used the ICCAT billfish spatial strata. The models included covariates as follows:

1. year, month, area, lightstick, hooktype, baittype, hbf
2. year, lightstick, hooktype, baittype, hbf, w
3 year, lightstick, hooktype, baittype, hbf, H_{100}

Models 2, and 3 were repeated for both SDM assumptions. The differences between results with the traditional factors for intra-annual variations in habitat and with habitat-based covariates were inspected visually in scattergrams and by correlation the annual abundances estimated with the two approaches.

3. Results

The nominal CPUE and relative abundances for each standardization are presented in **Table 1** and **Figure 2.** Though the protocol used here did not concern identifying models with the best statistical properties, covariates included as factors in each analysis were significant ($p<0.05$). That result is unsurprising because the suite of variables selected for inclusion here were already known to influence catch rates (Forrestal et al. 2017). The basic outcome of each standardization was to diminish the magnitude of the decline in CPUE evident in the nominal (unstandardized) series (**Figure 2A**). The values predicted by the SDM alternatives **(Figure 2C-2F**) were similar to the values estimated with month-area as factors (**Figure 2B**). The estimates for each of the methods using the two SDM-derived variables (w or H) and distribution model assumptions (baseline or thermophilic) were similar to one another (**Figure 2C-2F**). The abundance indices estimated with month-area factors were strongly correlated with those estimated using w or H calculated with the baseline SDM (**Figure 3**). The same was true for indices derived with w and H using the thermophilic SDM (**Figure 4**).

4. Discussion

The results indicate that for the US longline dataset the two approaches yield the same basic answer. Inspection of **Figure 2** suggests the numerical differences in the annual abundance estimates for the different approaches examined here are meaningless (**Table 1**). This view is supported by very strong correlations between the annual abundance estimates using traditional covariates and those derived with variables computed using the SDM relative densities (**Figures 3 and 4)**. Both methods resulted in data equally suitable for inclusion as indices of abundance in assessment modelling activities. This result suggests that the standardization protocol using the SDM-derived covariates was as successful at standardizing the CPUE as partitioning the catch by area and month to account for intra-year variability in CPUE. The SDM-derived habitat covariates are continuous variables computed with ancillary data. They are not affected by the vagaries of binning required to estimate the effects of factors that might explain intra-annual variability in relative abundance. This approach could obviate problems with statistical imbalance and improve standardizations in many situations.

Acknowledgements

CPG's contribution to this study was supported by The Billfish Foundation. FCF's contribution was supported by the NOAA Southeast Fisheries Science Center via the University of Miami Cooperative Institute of Marine Science. The authors wish to thank Sang-ki Lee and Yanyun Liu of NOAA Atlantic Oceanographic and Meteorological Laboratory for gathering and compiling oceanographic data.

References

Forestall, F.C., Goodyear, C.P., Schirripa, M., Babcock, E., Lauretta, M., and Sharma, R. 2017. Testing robustness of CPUE standardization using simulated data: findings of initial blind trials. Collect. Vol. Sci. Pap. ICCAT, 74(2): 391-403.

Goodyear, C. P., Luo, J., Prince, E. D., Hoolihan, J. P., Snodgrass, D., Orbesen, E. S., Serafy, J. E. 2008. Vertical habitat use of Atlantic blue marlin (*Makaira nigricans*): interaction with pelagic longline gear. Mar. Ecol.: Prog. Ser. 365, 233–245.

Goodyear, C.P. 2016. Modeling the time-varying density distribution of highly migratory species: Atlantic blue marlin as an example. Fisheries Research, 183: 469-481.

Goodyear, C.P., Schrippa, M. and Forrestal, F. 2017. Longline data simulation: A paradigm for improving cpue standardization. Collect. Vol. Sci. Pap. ICCAT, 74(2): 379-390.

Goodyear, C.P., Schrippa, M. and Forrestal F., and Lauretta M. 2018. Habitat covariates for standardizing longline CPUE: and example with blue marlin. SCRS.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org.

Table 1. Indices of annual abundance of blue marlin based on alternative treatments of the US Longline Logbook data. Each series has been normalized by dividing by its mean. The column labeled nominal represents the mean catch per hook. Each of the other columns contains values predicted from a general linear model (GLM) fitted to habitat-derived covariates and/ or ancillary fishery-related data. The traditional GLM applied covariates normally used in stock assessments. A species distribution model (SDM) was used to compute values for habitat-based covariates that replaced month and area in the traditional GLM approach. These include *H*, a habitat density in the surface 100m of the water column, and *w*, the habitat coefficient which additionally incorporates characteristics of the fishing gear. Habitat data for the habitat-based covariates were predicted using either the Baseline or an alternative, "Thermophilic," SDM.

			Species distribution model			
			Baseline		Thermophilic	
year	Nominal	Traditional GLM Covariates (month, area, Gear features)	Gear & *w*	Gear & *H*	Gear & *w*	Gear & *H*
1986	1.640	1.184	1.339	1.286	1.213	1.211
1987	2.251	1.901	1.812	1.783	1.708	1.707
1988	1.489	1.332	1.268	1.253	1.187	1.184
1989	1.410	1.019	0.938	0.940	0.876	0.875
1990	1.695	1.210	1.108	1.106	1.036	1.036
1991	1.411	0.977	0.951	0.958	0.924	0.927
1992	1.607	1.351	1.261	1.275	1.202	1.199
1993	1.864	1.568	1.580	1.595	1.549	1.551
1994	1.902	1.617	1.670	1.689	1.696	1.695
1995	1.537	1.471	1.466	1.471	1.506	1.504
1996	1.607	1.579	1.660	1.648	1.714	1.716
1997	1.240	1.322	1.314	1.306	1.369	1.373
1998	0.800	0.889	0.906	0.899	0.911	0.910
1999	0.786	0.864	0.870	0.868	0.893	0.895
2000	0.751	0.839	0.861	0.857	0.867	0.873
2001	0.457	0.560	0.577	0.573	0.580	0.581
2002	0.814	0.996	1.006	0.992	1.011	1.013
2003	0.432	0.549	0.554	0.547	0.549	0.548
2004	0.505	0.642	0.670	0.675	0.681	0.681
2005	0.469	0.684	0.660	0.677	0.687	0.687
2006	0.361	0.531	0.521	0.524	0.543	0.543
2007	0.478	0.658	0.661	0.668	0.681	0.680
2008	0.523	0.712	0.717	0.735	0.752	0.754
2009	0.719	1.009	0.985	0.987	1.023	1.021
2010	0.423	0.633	0.655	0.668	0.678	0.678
2011	0.419	0.588	0.630	0.634	0.652	0.651
2012	0.556	0.741	0.735	0.737	0.766	0.765
2013	0.479	0.694	0.685	0.688	0.717	0.716
2014	0.492	0.686	0.709	0.717	0.745	0.744
2015	0.884	1.195	1.232	1.246	1.283	1.282

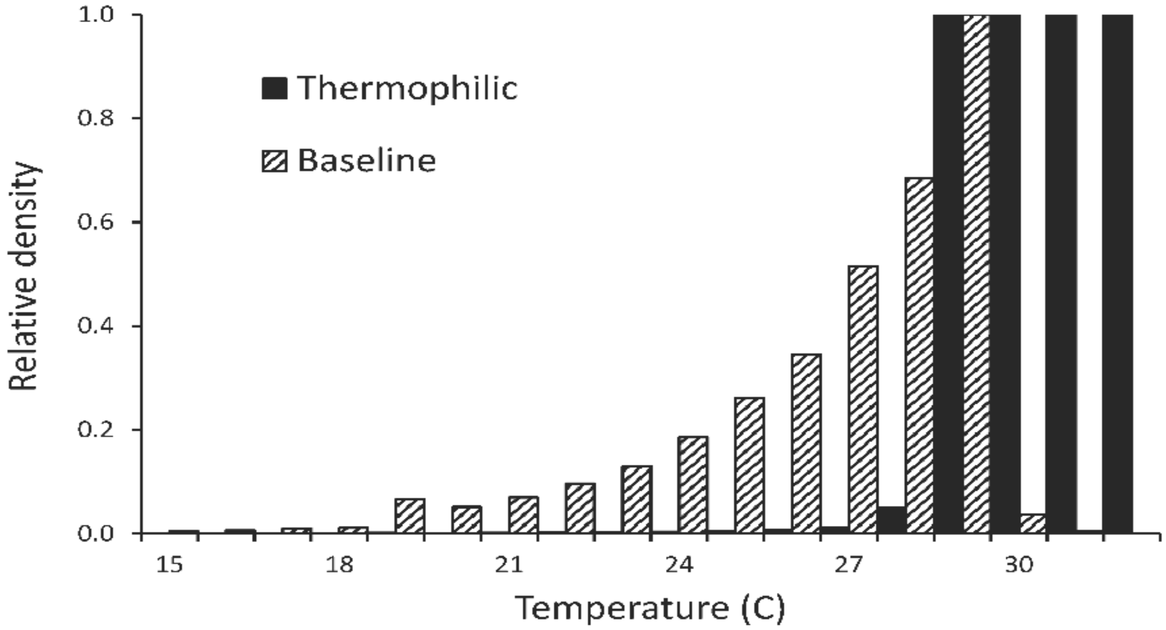

Figure 1. Temperature suitability curves for the alternative species distribution models used in this analysis.

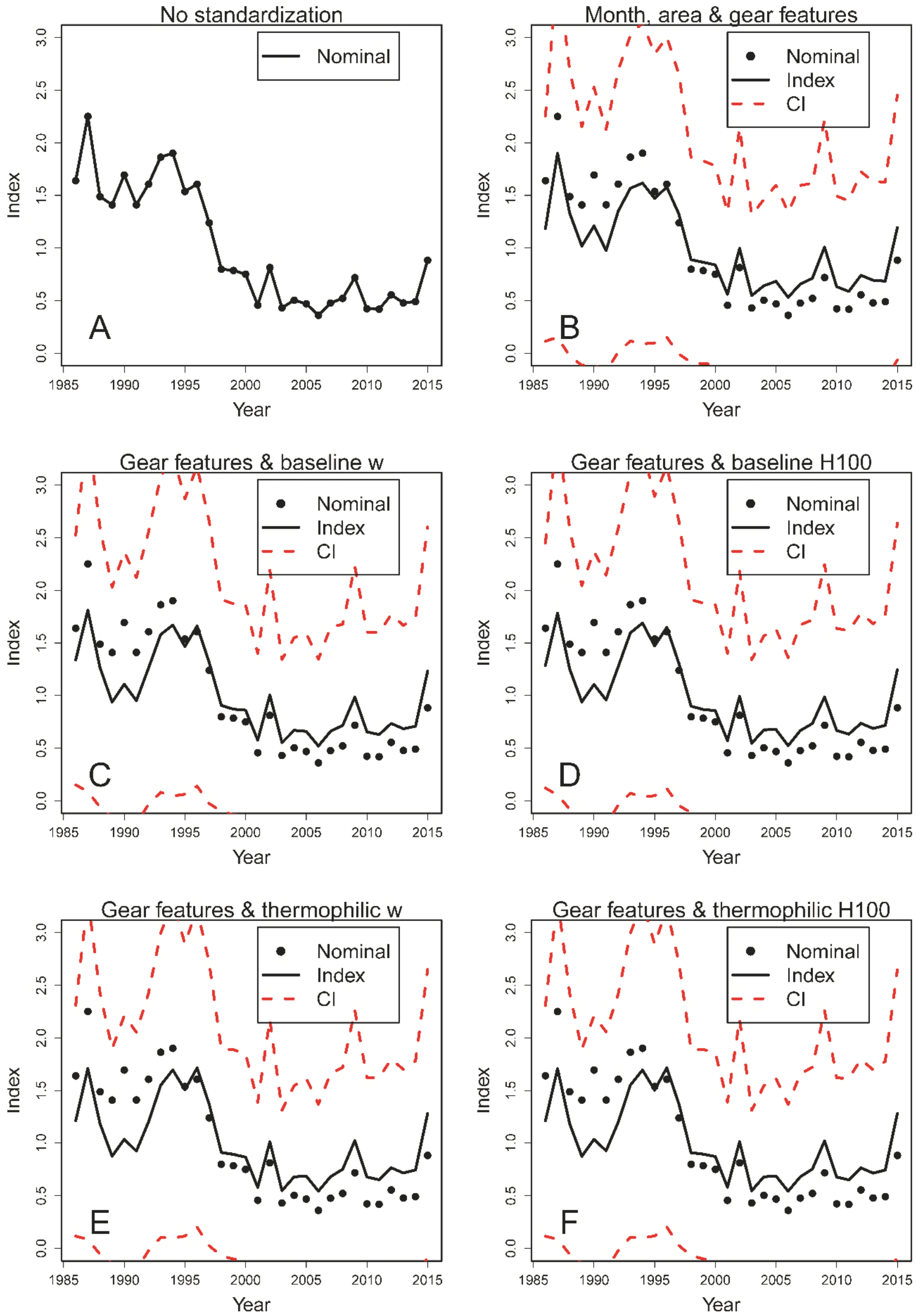

Figure 2. Scattergrams of the data in **Table 1** and with computed 95% confidence intervals for GLM results.

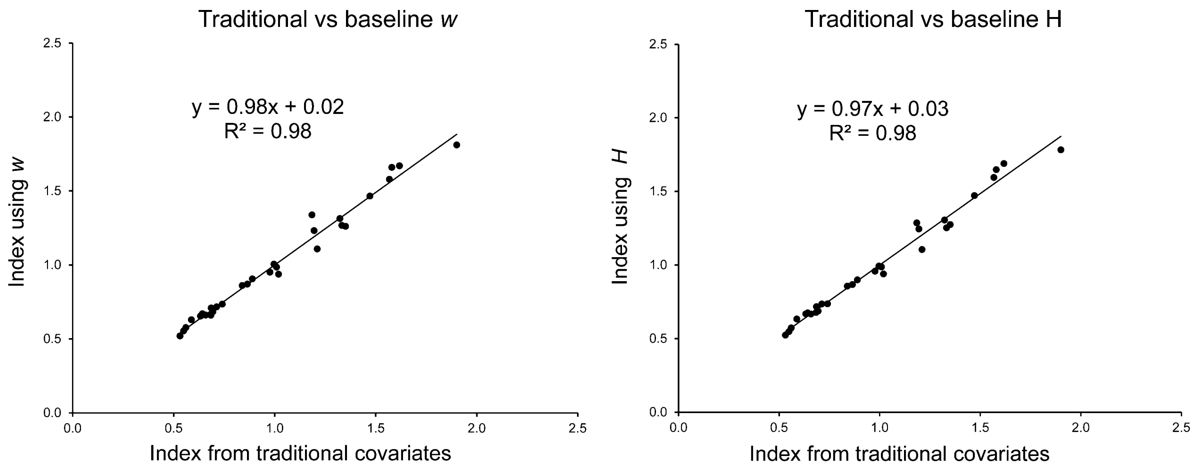

Figure 3. Relationships between the abundances predicted with the traditional GLM and those predicted using the habitat-derived covariates with the baseline SDM.

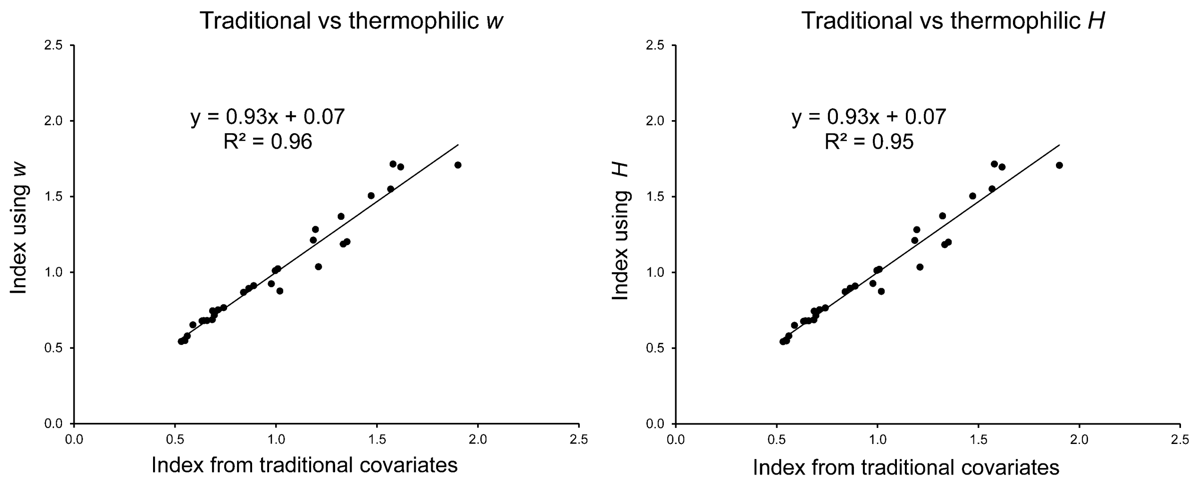

Figure 4. Relationships between the abundances predicted with the traditional GLM and those predicted using the habitat-derived covariates with the thermophilic SDM.