



OPEN

Modeling present and future distribution of plankton populations in a coastal upwelling zone: the copepod *Calanus chilensis* as a study case

Reinaldo Rivera^{1✉}, Rubén Escribano^{1,2}, Carolina E. González¹ & Manuela Pérez-Aragón^{1,2}

Predicting species distribution in the ocean has become a crucial task to assess marine ecosystem responses to ongoing climate change. In the Humboldt Current System (HCS), the endemic copepod *Calanus chilensis* is one of the key species bioindicator of productivity and water masses. Here we modeled the geographic distribution of *Calanus chilensis* for two bathymetric ranges, 0–200 and 200–400 m. For the 0–200 m layer, we used the Bayesian Additive Regression Trees (BART) method, whereas, for the 200–400 m layer, we used the Ensembles of Small Models (ESMs) method and then projected the models into two future scenarios to assess changes in geographic distribution patterns. The models were evaluated using the multi-metric approach. We identified that chlorophyll-a (0.34), Mixed Layer Depth (0.302) and salinity (0.36) explained the distribution of *C. chilensis*. The geographic prediction of the BART model revealed a continuous distribution from Ecuador to the southernmost area of South America for the 0–200 m depth range, whereas the ESM model indicated a discontinuous distribution with greater suitability for the coast of Chile for the 200–400 m depth range. A reduction of the distribution range of *C. chilensis* is projected in the future. Our study suggests that the distribution of *C. chilensis* is conditioned by productivity and mesoscale processes, with both processes closely related to upwelling intensity. These models serve as a tool for proposing indicators of changes in the ocean. We further propose that the species *C. chilensis* is a high productivity and low salinity indicator at the HCS. We recommend further examining multiple spatial and temporal scales for stronger inference.

The knowledge and understanding of oceanographic drivers influencing patterns of distribution of zooplankton over large-scale domains, such as ocean basins, are poor, limiting our capacity to predict changes in diversity and spatial distribution under a changing ocean^{1,2}. The species distribution of species forming the zooplankton is known to be affected by changes in oceanographic conditions, such as oxygenation, temperature, salinity, and stratification^{3,4}. This reveals the strong dependence of these organisms on oceanographic variables due to their limited migration capacity, which reflects the strong effects of hydrographic patterns on their distribution^{3,5}. Specifically, marine copepods respond to changes in oceanic environmental conditions, and thus might act as indicators of natural perturbations affecting the whole system⁶.

The Humboldt Current System (HCS) is recognized as one of the most productive areas of the world, sustaining a high production of pelagic fishes⁶, closely associated with upwelling events driven by wind at different intensities and frequencies along the South American coast⁷. The HCS is characterized by a predominant northward flow of surface waters of subantarctic origin and by a strong upwelling of cold, nutrient-rich subsurface waters of equatorial origin⁸. These dynamic and heterogeneous conditions of HCS support a high diversity of species^{8,9}.

Within this large ecosystem, one of the most characteristic species is the planktonic copepod *Calanus chilensis* (Brodsky, 1959) (Calanoida: Calanidae), an endemic species⁶ and a typical herbivorous copepod⁶. It is distributed from ~1 to 23°S and is one of the most abundant species among copepods^{10–13}. However, in the last decades,

¹Millennium Institute of Oceanography (IMO), University of Concepcion, 4030000 Concepcion, Chile. ²Department of Oceanography, Faculty of Natural and Oceanographic Sciences, University of Concepcion, 4030000 Concepcion, Chile. ✉email: reijavier@gmail.com

observations show *C. chilensis* is being replaced by *C. australis* toward oceanic waters¹⁴. Additionally, *C. chilensis* is associated with upwelling centers, reproducing continuously throughout the year^{12,15–17}, and showing high abundance near the coast at the upper 250 m of the water column, which coincides with the Poleward Undercurrent¹³.

Given its high abundance and extensive geographic distribution, *C. chilensis* appears to be an important contributor to secondary production and thus a crucial link between primary production and fish production^{6,18}. For its prominent ecological role, *C. chilensis* is recognized as key species worldwide in terms of secondary production, and this has motivated considerable research on its life cycle and secondary productivity^{6,12,13}. Despite its wide latitudinal distribution, information on its geographic distribution is fragmentary¹⁹, and even scarce in relation to the environmental drivers that explain its wide geographic distribution. This is certainly basic knowledge for evaluating the importance of this species to the production of the whole system⁶. Additionally, global climate change is driving changes in the phenology, distribution and abundance of species^{20–22}, with a significant impact on marine ecosystems²³. Nevertheless, for many taxa, there is a limited understanding of how geographic distribution patterns will be affected or changed²¹. Therefore, predicting potential habitat changes would require knowing how global climate change may affect populations and communities in the medium and long term.

Given this background, *C. chilensis* appears to provide an opportunity to test hypotheses on the response of pelagic populations to oceanographic conditions⁶. One way to assess its patterns, causes and consequences is the use of the wide availability of reliable biological and environmental information, which enables the generation of predictive models of the geographic distribution, as well as allows to evaluate hypotheses and generate predictive models of the geographic distribution of the species^{24,25}. The current availability of global environmental databases (e.g. Copernicus, bio-Oracle), geo-referenced records of species (e.g. OBIS, GBIF), and the development of the geographic information system (GIS) have allowed significant advances in the study and testing of hypotheses related to the spatial distribution of species using new approaches and ecological models^{25,26}. Among these tools, species distribution modeling (SDM) that uses spatially explicit information has been widely implemented to study the distribution patterns and environmental factors that explain and predict the distribution of species and populations^{27,28}, in the present as well as in other temporal scenarios²⁹.

For the marine environment, there are numerous examples of SDM implemented to answer questions regarding current species distribution patterns^{30–32}, bycatch^{33,34}, marine conservation planning^{35,36}, range shifts^{37,38}, biological invasions^{39–41} and climate change^{42,43} (see Ref.⁴⁴ for recent reviews). Nevertheless, fish, mollusks and mammals are the main taxa studied⁴⁴. The latter denotes that pelagic organisms such as zooplankton are less studied in relation to vertebrates and invertebrates at a macroscale^{1,45–49}.

The lack of macro-scale studies on zooplankton makes evident the scarce knowledge of the patterns and factors modulating the distribution of organisms that due to their trophic position and/or their status as indicators of oceanic conditions are considered ecologically relevant^{13,50}. Particularly, a macroscale change in temperature has modified the geographical range of some copepod species^{3,51,52}. For this reason, knowing the mechanisms that regulate biogeographical patterns at the macroscale allow to predict variation in species distribution under the effect of natural or anthropogenic impacts^{53,54}. Therefore, under this background, the HCS may constitute a suitable model to improve our understanding of underlying mechanisms modulating distribution patterns of species with high indicator values such as *C. chilensis*.

In this research, we assessed the distribution of the copepod species *C. chilensis* in the HCS and the environmental drivers shaping it, under present climatic conditions and under projected global climate change. We thus aimed at understanding the influence of oceanographic conditions over large-scale patterns and evaluating their value for an indicator species in the HCS. For the estimation of environmental effects, we used *C. chilensis* occurrence data and environmental variables to predict the potential geographic distribution of this species and identify its drivers, by applying Bayesian Additive Regression Trees (BART) and ensembles of small models (ESM) for two distinct bathymetric ranges (see Fig. 1 and Fig. 3). The models were projected under a climate change scenario to study the geographic distribution of this species in the future.

Results

0–200 m depth range. The geographic prediction of the BART model in the 0–200 m depth range using the Copernicus database presented a high performance [Area Under the Curve (AUC) = 0.933; True Skill Statistic (TSS) = 0.783]. The threshold that maximizes TSS was 0.446, and it was used as the threshold to generate the binary prediction. The variables that contributed most to the predicted distribution of *C. chilensis* were chlorophyll-a (0.34), Mixed Layer Depth [MLD (0.302)], and salinity (0.36). The geographical prediction of the model indicates high probability of presence mostly near the coasts from the northern region of the HCS (off Ecuador) to Tierra del Fuego (off southern Chile/Argentina), with an increase in the predicted probability in oceanic areas between 30 and 41° S (Fig. 1A). The binary prediction revealed a continuous longitudinal strip of predicted presence from Ecuador (1° S) to the southernmost part of Chile and Argentina (56° S) (Fig. 1B). The credible interval width (measure of spatial uncertainty) indicates that the highest uncertainty of prediction is located at the geographic extremes of the HCS, being more pronounced in front of the Magallanes region (Fig. 1C). The partial dependence curves revealed that chlorophyll-a has a positive effect on the presence of *C. chilensis* (Fig. 2A). On the other hand, the probability of presence decreases at higher values of MLD (Fig. 2B) and salinity, with lower probability of presence of *C. chilensis* from values over 34 psu (Fig. 2C).

The BART model approach for the 0–200 m depth using the bio-ORACLE database revealed an AUC of 0.972 and TSS of 0.875. The geographical prediction indicated that the highest probability of presence is located continuously from ~ 17° to 41° S, mainly in coastal areas of the HCS (Fig. S1).

200–400 m depth range. The ESM for the 200–400 m depth range using the Copernicus database suggests that *C. chilensis* has a distribution restricted to the coast of Chile (Fig. 3A), which extends to the south beyond

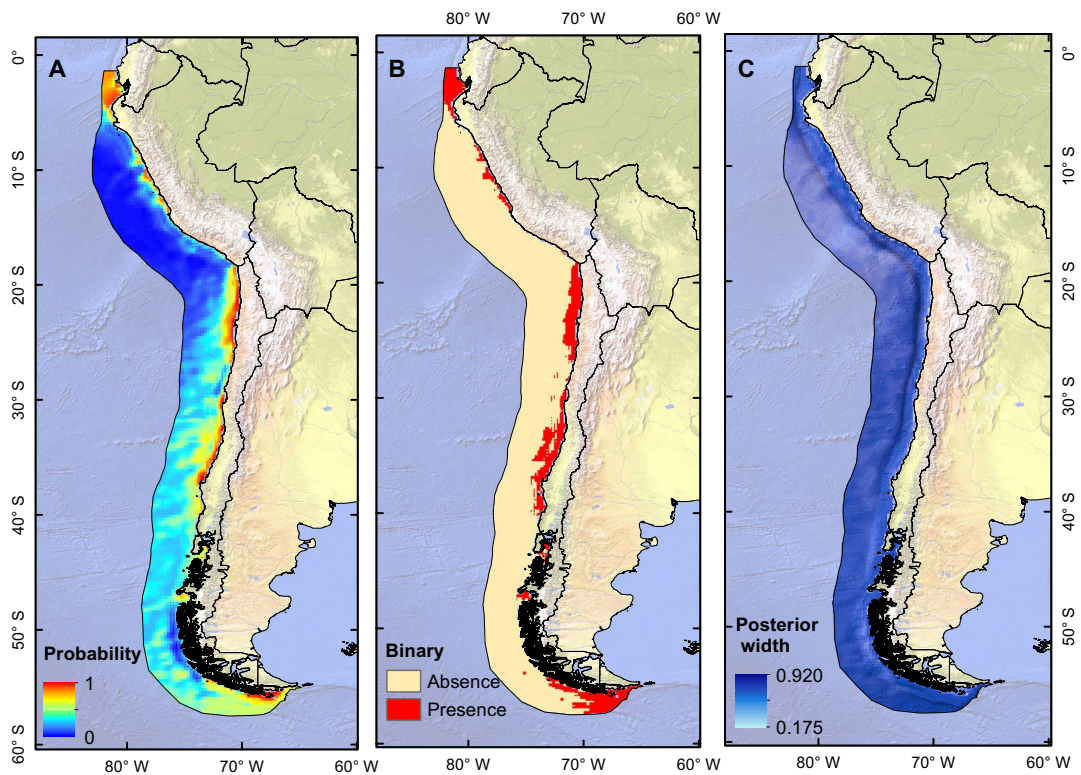


Figure 1. Potential current distribution maps generated by BART approach for the 0–200 m depth range. (A) Median of the posterior probability of the presence of *C. chilensis*, (B) binary model, and (C) posterior width (95% credible interval). This figure was generated using ArcGIS 10.4.1 (ESRI, Redland, CA; www.esri.com). Map projection is WGS84 (EPSG 4326).

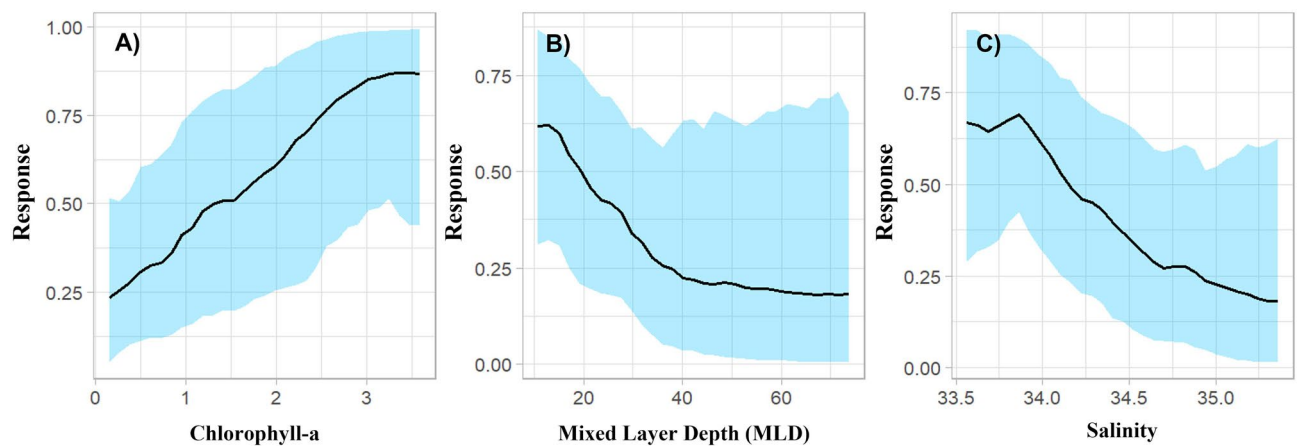


Figure 2. Partial dependence plot which shows the effect of each variable on the probability of presence of *C. chilensis* (A) chlorophyll-a, (B) mixed layer depth, and (C) salinity. Light blue = 95% credible interval. 0–200 m depth range. Figure done using R (<https://www.r-project.org>).

the range of its known distribution (Fig. 3A,B). The binary prediction revealed a continuous distribution from northern Chile (~18°S) to the Los Lagos region (~41°S), whereas further south, the probability of presence is discontinuous until 49°S (Fig. 3B). The performance of the ESMs showed high values of AUC and Continuous Boyce Index (CBI) (0.897 and 0.66 respectively), indicating the good performance of the models (Table 1). The variables that contributed the most to the ESM were Eddy Kinetic Energy (EKE) (0.263%), chlorophyll-a (0.253%), and Net Primary Productivity (NPP) (0.25%) (Table 2). The ranking of each technique until the final assemblage and the contribution of each variable are presented in Tables 1 and 2 respectively.

The ESM model approach for the 200–400 m depth range using the bio-ORACLE database indicated a high performance of the model (AUC 0.955, TSS 0.914, and CBI 0.841) (Table S1). The geographic prediction

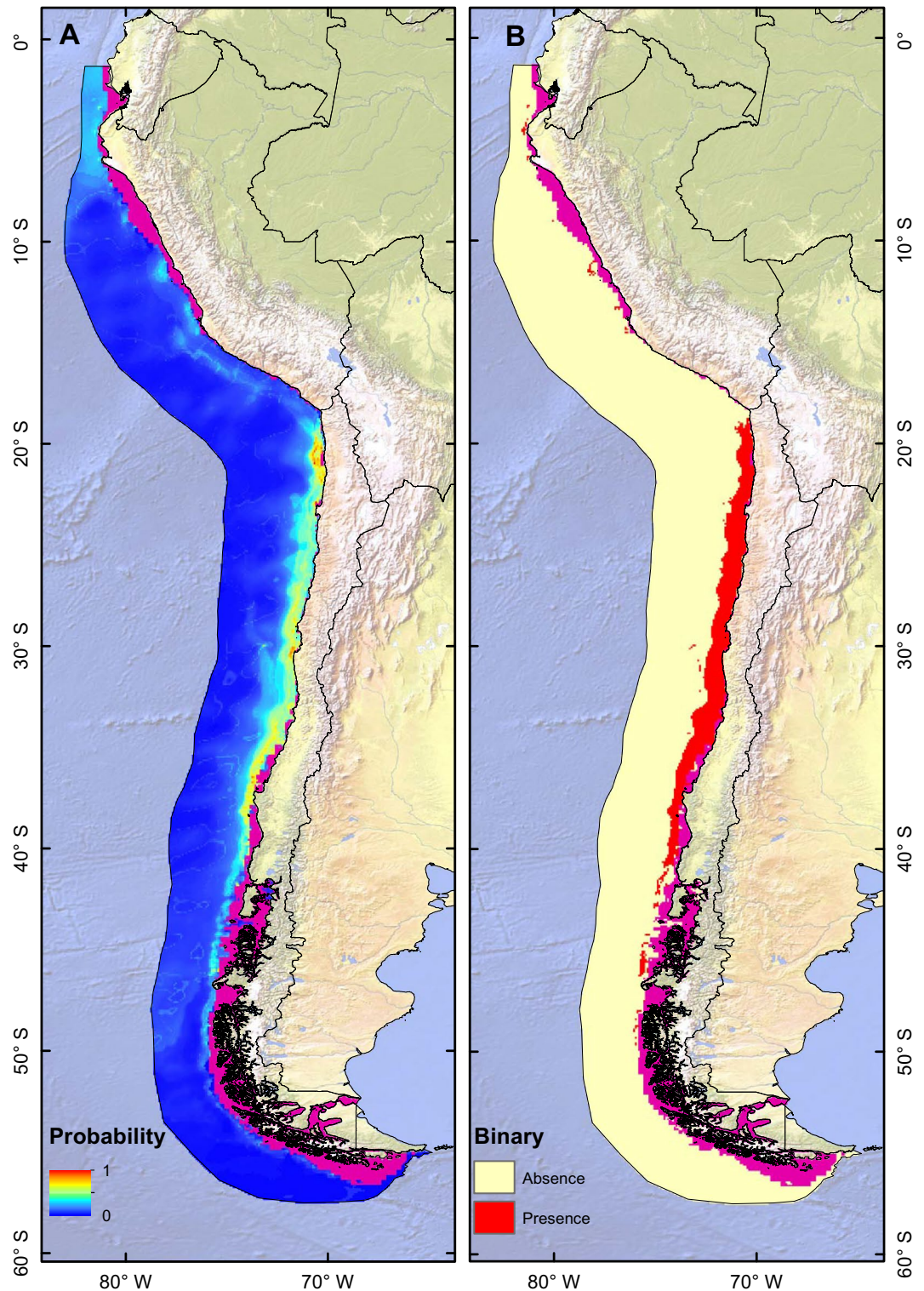


Figure 3. Potential current distribution maps generated through the Ensemble Small Models (ESM) approach for *Calanus chilensis* in the 200–400 m depth range of the Humboldt Current System. (A) Ensemble model suitability; (B) Ensemble binary model. Magenta color indicates depths greater than 200 m. This figure was generated using ArcGIS 10.4.1 (ESRI, Redland, CA; www.esri.com). Map projection is WGS84 (EPSG 4326).

indicated a lower predicted probability in relation to the upper bathymetric range (Fig. S2). The binary prediction of the ESM revealed a more heterogeneous and fragmented distribution in deeper layers (Fig. S2). The variable contribution for each modeling technique used in ESM and final ensemble projections indicated that

Model	Kappa	AUC	TSS	MPA	CBI
GAM	0.171	0.815	0.654	0.179	0.770
GBM	0.113	0.880	0.684	0.065	0.800
GLM	0.102	0.890	0.797	0.530	0.364
RF	0.287	0.802	0.382	0.025	0.780
Ensemble forecasting	0.121	0.897	0.805	0.236	0.664

Table 1. Models' performance of ensemble small models modelling (ESM) method for each technique runs and final ensemble forecasting (200–400 m depth range). *GAM* generalized additive model, *GBM* generalized boosted models, *GLM* generalized linear models, *RF* random forest, *AUC* area under curve, *TSS* true skill statistic, *MPA* minimum predicted area, *CBI* continuous boyce index.

	GAM	GBM	GLM	RF	Ensemble
Chlorophyll <i>a</i>	0.282	0.253	0.242	0.238	0.253
Eddy kinetic energy	0.212	0.247	0.314	0.271	0.263
Net primary productivity	0.279	0.247	0.218	0.267	0.250
Salinity	0.227	0.253	0.226	0.225	0.233

Table 2. Variable contribution (weighted means) for each modeling technique used in ESM and final ensemble projections (200–400 m depth range). *GAM* generalized additive model, *GBM* generalized boosted models, *GLM* generalized linear models, *RF* random forest.

Eddy Kinetic Energy (EKE) and Net Primary Productivity (NPP) (Table S2) are the main drivers of *C. chilensis*, generally coinciding with the prediction generated based on the variables of the Copernicus database (Table 2).

Projections to future scenarios. As shown in Table 3, the models indicate that under a low-impact scenario (Representative Concentration Pathway 2.6 (RCP2.6) there would be a smaller number of lost areas in relation to a high-impact scenario (RCP8.5). Similarly, under benign scenarios there is a greater number of gained areas; that is, pixels which are currently not occupied by the given species but predicted to be in the future; however, in the long term (years 2090–2100), the pixel gain is less. Regarding the percentage of gain (percentage of new sites considering the species' current distribution size) and loss (percentage of currently occupied sites to be lost), there is a greater gain of areas only in low-impact scenarios and in the short term, whereas in the long term the gain decreases and the loss of areas increases. The general result of the projection (species range change) revealed a generalized loss of areas in the geographic distribution range both in low and high-impact scenarios, mainly for the long term (Fig. S3, Table 3).

MESS analysis identified areas where no analogs or novel climates were present. Dissimilarity values were relatively low within suitable areas for *C. chilensis*, not showing severe extrapolation from the models (Fig. S4).

	Near future (2040–2050)		Far future (2090–2100)	
	RCP2.6	RCP8.5	RCP2.6	RCP8.5
Loss	431	690	921	1503
Stable 0	24,294	24,450	25,008	25,379
Stable 1	9522	9263	9032	8450
Gain	1417	1261	703	332
Percentage loss	4.3	6.9	9.2	15.1
Percentage gain	14.2	12.6	7.1	3.3
Species range change	9.9	5.7	– 2.19	– 11.7
Current range size	9953	9953	9953	9953

Table 3. Predicted loss, expansion, and areas of no change (in pixels, resolution = ~9.2 km) for the distribution of *C. chilensis* into the near future (2040–2050) and far future (2090–2100) and under scenarios of low (RCP2.6) and high impact (RCP8.5). Loss = corresponds to the percentage of currently occupied sites to be lost. Stable 0 = number of pixels which are not currently occupied by the given species and not predicted to be. Stable 1 = number of pixels currently occupied by the given species, and predicted to remain occupied into the future.

Discussion

Given that the distribution of *C. chilensis* is explained by a reduced number of oceanographic variables, linked to productivity (chlorophyll-a) and physical variables such as MLD, our model (Fig. 1) is consistent with that reported in the literature. Indeed, large calanoid species that are common in shelf waters were more abundant within nearshore eddies rich in chlorophyll-a and other cyclonic eddies far offshore⁵⁵. These environmental features are present in water mass typical of intensive upwelling¹³. According to Morales et al.⁵⁵ the eddy field alone did not explain the geographic distribution of *C. chilensis* at mesoscale⁵⁵, where the availability of chlorophyll-a and waters with low salinity is key to explaining the distribution of this species.

Historically, *C. chilensis* is reported in the HCS all along the Chilean and Peruvian coast from approximately 21 to 45°S, with no data available for more southern or northern regions¹³. In earlier studies conducted in Chilean waters, *C. chilensis* was found inhabiting oxygenated waters over the oxygen minimum zone (OMZ)⁵⁶, although sometimes smaller portions of the population were also found below the oxycline⁵⁷, through passive sinking⁵⁸, diel vertical migration (DVM)^{59,60}, or combining DVM (daily vertical migration) and predation. The latter and the fact that this species is restricted to oxygenated waters above the OMZ⁵⁶, represents an important food resource for the larger animals populating this zone, makes *C. chilensis* a species of ecological importance in this system. It is also characterized by being the only species of the family Calanidae where not only a resting stage can tolerate hypoxia¹³, and, due to its large size and abundance, *C. chilensis* play an important role for the carbon flux into the OMZ¹³, due to very active diel vertical migration⁵⁷.

Geographic predictions by the Bayesian BART method for the range 0–200 m indicate that *C. chilensis* is distributed latitudinally from 1 to 55°S in areas near the coast (Fig. 1). Similarly, several studies off Chile found *C. chilensis* close to the shore⁶¹ with the highest abundances observed (ca. 44,000 ind. m²) in a narrow band within Cold Coastal Water, which coincides with the Poleward Undercurrent¹³. In particular, a predicted high-probability zone lies further offshore in central Chile, where a high abundance of *C. chilensis* has been reported, which can be interpreted as the result of advection by eddies^{55,61}. The probability of presence of *C. chilensis* in the HCS appeared to decrease with winds flowing from east to west (between 5° and 25°S) and increase with winds flowing from the west and its meridional components (between 30° and 60°S) (Fig. S5).

The model prediction for the bathymetric range (200–400 m; Fig. 3) revealed a pattern similar to that reported for the 0–200 range; however, there would be a greater geographic discontinuity. As *C. chilensis* is known for inhabiting the upper 250 m¹³, the occurrences reported at greater depths may be thus circumstantial or correspond to sink populations from higher strata that act within the context of a source-sink dynamic^{62,63}. In this stratum, the variables that explain the spatial distribution were related to EKE and NPP. EKE reveals the presence of mesoscale eddies⁶⁴; therefore, the distribution of *C. chilensis* is conditioned by kinetic energy, whereas chlorophyll-a as a proxy of productivity is also another important modulator of the presence of *C. chilensis* in deeper areas. The latter confirms that chlorophyll-a is the main factor affecting the distribution of most species in the HCS⁵⁵, as, for example, large copepod species which are dominant in the upwelling area are associated with the presence of chlorophyll-a and cyclonic eddies⁶⁵.

For the prediction of 0–200 m, the high probability in areas close to 35°S, may be due to a greater historical sampling effort and the wide distribution of copepod species over the shelf/slope in that region, which is explained by physical and biological mechanisms that could be acting to extend the productive area of the coastal upwelling zone⁵⁵.

Although both the BART and the ESM models generally predict an almost continuous distribution from Ecuador to Magellan, the biological characteristics of *C. chilensis* indicate that this species, although it reproduces throughout the year, does so less frequently in central Chile^{66,67} due to the existing variability within the HCS, such as seasonality in food availability and advection, which are some of the main drivers of their seasonal occurrence and abundance¹³. This reveals that seasonal and mesoscale processes are important when interpreting model outputs on the geographic distribution of marine organisms, since such models do not incorporate these variables, as they are not available in repositories such as bio-Oracle or MARSPEC (e.g. upwelling regimes⁶⁸).

The use of a Bayesian approach to model the distribution of *C. chilensis* represented a methodological advantage, because it allows quantifying the uncertainty of the prediction⁶⁹. In general, the use of Bayesian spatial models can help in the analysis of data with geographically uneven levels of sampling effort⁷⁰, a common situation in biodiversity data⁷¹, and especially in ecosystems such as the HCS. Therefore, given the presence of bias in the data, the most suitable analysis strategy is the use of approximations that allow this bias to be reduced, reducing its influence on the parameter estimates⁷⁰. Another difficulty when modeling the distribution of endemic species such as *C. chilensis* is obtaining enough occurrences^{72–74}, which is a recurrent situation for species that are rare, endemic or with biased sampling⁷⁵. To take into account the low number of occurrences for the 200–400 m strata, we used ESMs, an approach described to date as the most suitable for getting robust predictions even when modeling rare species or with a reduced number of occurrences^{76,77}.

In a future scenario (medium and long term), *C. chilensis* would experience a drastic reduction in its geographic range, because of its high dependence on oceanographic conditions which may change with the global warming going on, mainly the availability of nutrients, variations in salinity, and depth of the mixed layer (see Fig. 2). In this regard, this species, being endemic to an ecoregion, with particular oceanographic characteristics, is more sensitive and would be mostly exposed to these fluctuations in the future. However, this trend would be greater in the models that simulate the “worst” climate scenario, and in the long term (2090–2100), being less affected by environmental variability in the short term and under “benign” scenarios (e.g. RCP26; see Table 3). These reported trends are consistent with the effects that anthropogenic global climate change would produce, mainly with respect to changes or modifications in the latitudinal distribution ranges^{23,78}. Whatever the case, altered distribution of this species, may have important ecological and biogeochemical consequences for the functioning of the HCS. *C. chilensis* is a key secondary producer in the upwelling zone¹¹, and has an important

role in nutrient recycling and downward flux of C⁵⁷. Constrained populations may thus affect the food web structure and have implications for the C cycling of the upwelling system.

In this study, we demonstrated that the distribution of *C. chilensis* is conditioned by productivity and mesoscale events which drive its spatial distribution in the HCS. In other regions of the world ocean, predicted changes in productivity, driven by global warming, are expected to alter the distribution of secondary producers over large spatial scales^{79,80}, including dominant copepod species, such as *C. finmarchicus* in the North Atlantic⁸¹ and *Centropages typicus* in coastal areas of the Atlantic Ocean⁸². However, over regional scales, the maintenance mechanism used by *C. chilensis* and probably many other endemic zooplankton species⁸³ present in coastal upwelling systems certainly needs further investigation.

Our model results and predictions are certainly subject to potential biases, derived from data limitations (sampling gaps), and other processes occurring over spatial–temporal scales smaller than model resolution. For instance, the mesoscale and sub-mesoscale advective processes may influence the distribution of drifting planktonic copepods⁵⁵. Also, the spatial arrangement of metapopulations may be subject to differential responses to a highly heterogeneous environment, and so producing distinctive patches of the species over mesoscale to large-scale domains².

Methods

Study area. The study area comprised the Humboldt-Current large marine ecosystem ranging from Peru to the southern zones of Chile⁸⁴. These regions encompass the currently known geographic distribution of *C. chilensis*⁶. However, we extended the northern limit of the HCS by three degrees to consider new occurrences off the coast of Ecuador. Since the extension of the geographical area can influence the performance of SDM⁸⁵, the spatial scale of the study area must consider the dispersal capacity of the target species²⁶. To address this, we considered the HCS given that *C. chilensis* has been described as endemic to this ecosystem⁶⁶ and it is also restricted to the upwelling zone^{10,11,86}. These areas were considered to be accessible world (M) according to the theoretical framework proposal of Soberón & Peterson⁸⁷.

Occurrences and quality control procedures. We examined 270 records of *C. chilensis* obtained from the Ocean Biodiversity Information System (OBIS) and Global Biodiversity Information Facility (GBIF), as well as records available in specialized literature. All the above online databases were accessed on October 15, 2021 (Suppl. Table S3). Species occurrence records were obtained using the *robis*⁸⁸ and *rgbif* packages⁸⁹ implemented in the R software [Ref.⁹⁰]. After retrieving the data, we eliminated records without information on the geographic coordinates, coordinates equal to zero, or records located inside continents. We selected only records at the level of species and excluded duplicate records. In addition, to avoid spatial biases in the sampling effort, which are common when literature and databases are used⁷⁵. The temporal resolution of the species occurrences data was from the years 1960 to 2021. We carried out a spatial thinning approach to eliminate records with a minimum distance of 5 km from each other using the *spThin* R package⁹¹.

From the 82 thinned occurrence records of *C. chilensis*, models were made for two bathymetric ranges: 0–200 m and 200–400 m. For 0–200 m depth range, 51 occurrences were obtained; and, after spatial thinning, a total of 24 records were recovered. For 200–400 m depth range, 31 occurrences were obtained; and, after spatial thinning, a total of 8 records were recovered (Fig. S6). The occurrences of *C. chilensis* are available in Supporting Information Table S3 and the Figshare repository <https://doi.org/10.6084/m9.figshare.19747618.v1>.

Environmental database. The remotely sensed satellite data used for analyses were obtained from Copernicus Marine Environment Monitoring Service (<https://marine.copernicus.eu/>) to a resolution of 0.083 and 0.25 degrees. The temporal resolution of the data was 1993–2019. Each environmental layer represents annual average values per cell and corresponds to an integrated average from 0 to 200 m and 200–400 m depth respectively. Variables with a resolution of 0.25 degrees were resampled using Cubic Convolution Interpolation in ArcGIS 10.4.1 [Ref.⁹²].

We downloaded seven variables that have a close relationship with the biology of *C. chilensis*⁵⁰: chlorophyll-a (mg m^{-3}), dissolved oxygen (mmol m^{-3}), MLD (m), Net Primary Productivity (NPP) ($\text{mg m}^{-3} \text{ day}^{-1}$), pH, salinity and temperature ($^{\circ}\text{C}$). The Eddy Kinetic Energy (EKE) corresponds to the Sea Level Anomaly intensity and was calculated by its impact on the upwelling^{7,12}. Surface data in terms of geostrophic sea water velocity for the zonal (east-speed U) and meridional (north-speed V) components were extracted from the Copernicus database. Calculations were made using the formula, $\text{EKE} = 1/2 \times (\text{U}^2 + \text{V}^2)$, expressed in cm^2/s^2 .

Considering that the correlation between variables may affect the performance of SDM^{93,94}, we used the variance inflation factor (VIF) to detect collinearity between predictors. Using the *usdm* package⁹⁵, variables with a VIF > 3 were excluded⁹⁶. Complementarily, we evaluated the correlation and eliminated variables with Spearman's rank correlation coefficient ($\rho > 0.7$). Thus, the variables used for posterior analyses were chlorophyll-a, MLD, EKE and salinity (See Supplementary Figs. S7, S8 and Table S4).

Species distribution modelling. To evaluate the potential distribution of *C. chilensis* based on oceanographic predictors, we used two approaches, BART and ESM for 0–200 m and 200–400 m depth ranges respectively, given the larger number of occurrences available for the former and the fewest occurrences available for the latter.

BART is a tree-based method of machine learning which is based on a Bayesian approach to classification and regression trees (CART). BART is defined by a prior distribution and a likelihood for returning occurrence predictions that enable the quantification of uncertainty around the predictions and the estimation of the marginal effects of the covariates⁶⁹. The BART method is a technique that deals with non-linear and non-monotonic

relationships between response and predictor variables, and allows estimating the probability of presence of a species or its populations. BART models were run with default parameters using 200 trees and 1000 back-fitting Markov chain Monte Carlo (MCMC) with 90 iterations, discarding 20% as burn-in through *embarcadero* R package⁶⁹. Model performance was evaluated using the AUC of the receiver operating characteristic curve (ROC) and True Skill Statistics (TSS). To estimate the potential distribution of *C. chilensis*, the predictions (i.e. probability of presence) were converted into binary predictions using a threshold at which TSS is maximized (max TSS)^{97,98}. The Bayesian spatial method allows the incorporation of spatial correlation of the variables and the uncertainty of the parameters in the modeling process, resulting in a better quantification of the uncertainty (credible intervals)^{99,100}. Bayesian spatial models may also aid data analyses with geographically uneven levels of survey effort, which reduces its influence on estimates of the effects of environmental variables⁷⁰.

Given that accurate species distribution models require a minimum occurrence record⁷³, we used the ESMs' approach to identify suitable areas for the presence of *C. chilensis* and estimate its geographic distribution at the 200–400 m depth range, since it is suitable for species with a low number of occurrence records^{76,77}. The ESMs' method is a technique that allows obtaining statistically robust habitat suitability models from combinations of bivariate models. It has been proved that this method can characterize reliable suitability models with less than 25 occurrence records⁷⁶. The final ESM projection assembly was obtained by calibrating four modeling techniques, generalized additive model (GAM), generalized linear model (GLM), generalized boosted regression modeling (GBM) and random forest (RF). We selected default model tunings within *ecospat* package¹⁰¹. For each ESMs' run, the presences were partitioned into 70% for training and 30% for testing. We use 1000 random pseudo-absences in the calibration and projection areas; the prediction ensemble was obtained by averaging 10 runs weighted according to the AUC values. To calibrate and project the ESMs, four uncorrelated environmental predictors were used to model the distribution of *C. chilensis*: chlorophyll-a, MLD, EKE and salinity (Supplementary Table S4, Fig. S8). Finally, the projected ESMs were transformed into binary prediction using the minimum predicted area (MPA)¹⁰².

Model performance of ESM was evaluated using the multi-metric approach to determine the variability among estimates, the AUC of the ROC, TSS, Cohen's kappa, and the Continuous Boyce Index (CBI). The AUC ranges from 0 to 1, where scores lower than 0.5 indicate discrimination worse than random, and a 1 score indicates perfect discrimination¹⁰³. TSS ranges from -1 to $+1$, where $+1$ indicates perfect agreement between predictions and observations, and values of 0 or less indicate agreement no better than random classification. Cohen's kappa statistic ranges from -1 to $+1$, where $+1$ indicates perfect agreement and values of zero or less indicate a performance no better than random⁹⁷. Finally, the CBI varies from -1 for an inverse model to 0 for a random model to 1 for a perfect model^{104,105}. The analyses were performed in the *ecospat* package¹⁰¹.

In addition to the models fitted using the oceanographic layers retrieved from the Copernicus database, we fitted a second set of SDMs using layers downloaded (except for MLD and EKE) from the bio-ORACLE database version 2.2⁷⁷ at a resolution of 0.08 degrees (~ 9.2 km at the equator). Then, to evaluate the variability in model predictions depending on the used environmental dataset, outputs resulting from the BART and ESM models based on Copernicus and bio-ORACLE data were compared through Schoener's D overlap index, which ranges from 0 (no overlap) to 1 (complete overlap). Analyses were carried out with ENMTTools package¹⁰⁶ in R [Ref.⁹⁰]. Results indicated a high similarity between predictions for the 0–200 m depth range (BART approach) ($D = 0.86$; Spearman rank correlation = 0.81), and a lower similarity between predictions for the 200–400 m depth range (ESM approach) ($D = 0.69$; Spearman rank correlation = 0.42). For this, we present the models built with the variables from the Copernicus database, as they can be downloaded separately for distinct depth ranges, which is not possible with the bio-ORACLE database. The BART and ESM models for the depth range of 0–200 and 200–400 m built with bio-ORACLE variables are shown in Supplementary Fig. S1, S2, Table S4.

Projection of models to the future. To identify future potential shifts in the distribution range of *C. chilensis* we modeled the distribution in the context of global climate change, projecting the models in two periods: the near future (2040–2050) and the far future (2090–2100). In both scenarios, two representative concentration pathways (RCP) scenarios were used, RCP2.6 and RCP8.5. RCP2.6 represents a peak-and-decline scenario ending with very low greenhouse gas concentration levels by the end of the twenty-first century, whereas RCP8.5 is a scenario of increasing emissions over time leading to high greenhouse concentration levels (see Ref.¹⁰⁷). To assess the extrapolation risk, we performed a Multivariate Environmental Similarity Surface (MESS) analysis to determine novel climatic conditions under future climate scenarios¹⁰⁸. Negative values indicate localities that are environmentally dissimilar from the reference region. Positive values indicate climate similarity¹⁰⁸. The MESS analyses were performed in *ntbox* package¹⁰⁹. The projections were made using the BART approach only for the 0–200 m, given the greater availability of occurrence records. Since variables such as MLD and EKE are not available for the future, we model using salinity, chlorophyll-a and keeping MLD constant, since they correspond to the variables selected as the most important to explain the geographic distribution in the present (Fig. 2). The oceanographic variables were obtained from bio-ORACLE v 2.1¹⁰⁷ at a resolution of 5 arcminutes (~ 9.2 km) through the *SDMpredictors* package¹¹⁰. We used a TSS threshold to convert maps into binary presence/absence to calculate changes in the size of suitable habitats using *biomod2* package¹¹¹.

We followed best-practice standards in SDM regarding guidelines for response and predictor variables, model building and evaluation¹¹². We provide as Supplementary Table S5 a description of the modeling steps following the ODMAP (Overview, Data, Model, Assessment and Prediction) protocol¹¹². The geographic data were processed in ArcGIS 10.4.1 [Ref.⁹²] and all statistical analyses were performed in the R 4.0.1 [Ref.⁹⁰].

Data availability

The datasets generated analysed during the current study are available in the Figshare repository <https://doi.org/10.6084/m9.figshare.19747618.v1>.

Received: 14 November 2022; Accepted: 6 February 2023

Published online: 23 February 2023

References

1. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).
2. González, C. E., Medellín-Mora, J. & Escribano, R. Environmental gradients and spatial patterns of calanoid copepods in the southeast Pacific. *Front. Ecol. Evol.* **8**, 1–16 (2020).
3. Rombouts, I. *et al.* Global latitudinal variations in marine copepod diversity and environmental factors. *Proc. R. Soc. B Biol. Sci.* **276**, 3053–3062 (2009).
4. Brandão, M. C. *et al.* Macroscale patterns of oceanic zooplankton composition and size structure. *Sci. Rep.* **11**, 1–19 (2021).
5. McClain, C. R. & Barry, J. P. Habitat heterogeneity, disturbance, and productivity work in concert to regulate biodiversity in deep submarine canyons. *Ecology* **91**, 964–976 (2010).
6. Escribano, R. & Rodríguez, L. Life cycle of *Calanus chilensis* Brodsky in Bay of San Jorge, Antofagasta Chile. *Hydrobiologia* **292–293**, 289–294 (1994).
7. Strub, P. T., Mesías, M. J., Montecino, V., Rutllant, J. & Salinas, S. Coastal ocean circulation off western South America coastal segment. *Sea* **11**, 273–313 (1998).
8. Montecino, V. & Lange, C. The Humboldt current system: Ecosystem components and processes, fisheries, and sediment studies. *Prog. Oceanogr.* **83**, 65–79 (2009).
9. Miloslavich, P. *et al.* Marine biodiversity in the Atlantic and Pacific coasts of South America: Knowledge and gaps. *PLoS ONE* **6**, e14631 (2011).
10. Marín, V., Espinoza, S. & Fleminger, A. Morphometric study of *Calanus chilensis* males along the Chilean coast. *Hydrobiologia* **292**, 75–80 (1994).
11. Escribano, R. & McLaren, I. Production of *Calanus chilensis* in the upwelling area of Antofagasta Northern Chile. *Mar. Ecol. Prog. Ser.* **177**, 147–156 (1999).
12. Escribano, R. & Hidalgo, P. Spatial distribution of copepods in the north of the Humboldt Current region off Chile during coastal upwelling. *J. Mar. Biol. Assoc. U. K.* **80**, 283–290 (2000).
13. Hirche, H. J., Barz, K., Ayon, P. & Schulz, J. High resolution vertical distribution of the copepod *Calanus chilensis* in relation to the shallow oxygen minimum zone off northern Peru using LOKI, a new plankton imaging system. *Deep Res. I Oceanogr. Res. Pap.* **88**, 63–73 (2014).
14. Sabatini, M., rez, F. & Martos, P. Distribution pattern and population structure of *Calanus australis* Brodsky, 1959 over the southern Patagonian Shelf off Argentina in summer. *ICES J. Mar. Sci.* **57**, 1856–1866 (2000).
15. Escribano, R. Population dynamics of *Calanus chilensis* in the Chilean Eastern Boundary Humboldt Current. *Fish. Oceanogr.* **7**, 245–251 (1998).
16. Hidalgo, P. *et al.* Patterns of copepod diversity in the Chilean coastal upwelling system. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **57**, 2089–2097 (2010).
17. Hidalgo, P., Escribano, R., Fuentes, M., Jorquera, E. & Vergara, O. How coastal upwelling influences spatial patterns of size-structured diversity of copepods off central-southern Chile (summer 2009). *Prog. Oceanogr.* **92–95**, 134–145 (2012).
18. Giraldo, A., Escribano, R. & Marín, V. Spatial distribution of *Calanus chilensis* off Mejillones Peninsula (northern Chile): Ecological consequences upon coastal upwelling. *Mar. Ecol. Prog. Ser.* **230**, 225–234 (2002).
19. Gonzalez, A. & Marín, V. Distribution and life cycle of *Calanus chilensis* and *Centropages brachiatus* (Copepoda) in Chilean coastal waters: A GIS approach. *Mar. Ecol. Prog. Ser.* **165**, 109–117 (1998).
20. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
21. Parmesan, C. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Glob. Change Biol.* **13**, 1860–1872 (2007).
22. Visser, M. E. & Both, C. Shifts in phenology due to global climate change: The need for a yardstick. *Proc. R. Soc. B Biol. Sci.* **272**, 2561–2569 (2005).
23. Chaudhary, C., Richardson, A. J., Schoeman, D. S. & Costello, M. J. Global warming is causing a more pronounced dip in marine species richness around the equator. *Proc. Natl. Acad. Sci.* **118**, e2015094118 (2021).
24. Ferrier, S., Drielsma, M., Manion, G. & Watson, G. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodivers. Conserv.* **11**, 2309–2338 (2002).
25. Jetz, W., Wilcove, D. S. & Dobson, A. P. Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biol.* **5**, e157 (2007).
26. Peterson, A. T. *et al.* *Ecological Niches and Geographic Distributions (MPB-49)* (Princeton University Press, 2011). <https://doi.org/10.2307/j.ctt7stnh>.
27. Franklin, J. *Spatial Inference and Prediction. Mapping Species Distributions* Vol. 141 (Cambridge University Press, 2010).
28. Guisan, A., Thuiller, W. & Zimmermann, N. E. *Habitat Suitability and Distribution Models: With Applications in R. Ecology Biodiversity and Conservation* (Cambridge University Press, 2017). <https://doi.org/10.1017/9781139028271>.
29. Freer, J. J., Partridge, J. C., Tarling, G. A., Collins, M. A. & Genner, M. J. Predicting ecological responses in a changing ocean: The effects of future climate uncertainty. *Mar. Biol.* **165**, 7 (2017).
30. Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E. & Lundquist, C. J. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Front. Mar. Sci.* **4**, 421 (2017).
31. Pennino, M. G. *et al.* Accounting for preferential sampling in species distribution models. *Ecol. Evol.* **9**, 653–663 (2019).
32. Coll, M., Pennino, M. G., Steenbeek, J., Sole, J. & Bellido, J. M. Predicting marine species distributions: Complementarity of food-web and Bayesian hierarchical modelling approaches. *Ecol. Model.* **405**, 86–101 (2019).
33. Stock, B. C. *et al.* Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. *Can. J. Fish. Aquat. Sci.* **77**, 146–163 (2019).
34. Lezama-Ochoa, N. *et al.* Spatio-temporal distribution of the spinetail devil ray *Mobula mobular* in the Eastern tropical Atlantic ocean. *Endanger. Species Res.* **43**, 447–460 (2020).
35. Marshall, C. E., Glegg, G. A. & Howell, K. L. Species distribution modelling to support marine conservation planning: The next steps. *Mar. Policy* **45**, 330–332 (2014).
36. Hunt, T. N., Allen, S. J., Bejder, L. & Parra, G. J. Identifying priority habitat for conservation and management of Australian humpback dolphins within a marine protected area. *Sci. Rep.* **10**, 1–14 (2020).
37. Champion, C., Brodie, S. & Coleman, M. A. Climate-driven range shifts are rapid yet variable among recreationally important coastal-pelagic fishes. *Front. Mar. Sci.* **8**, 1–13 (2021).

38. Przeslawski, R., Falkner, I., Ashcroft, M. B. & Hutchings, P. Using rigorous selection criteria to investigate marine range shifts. *Estuar. Coast. Shelf Sci.* **113**, 205–212 (2012).
39. Januario, S. M., Estay, S. A., Labra, F. A. & Lima, M. Combining environmental suitability and population abundances to evaluate the invasive potential of the tunicate *Ciona intestinalis* along the temperate South American coast. *PeerJ* **3**, e1357. <https://doi.org/10.7717/peerj.1357> (2015).
40. Pinochet, J., Rivera, R., Neill, P. E., Brante, A. & Hernández, C. E. Spread of the non-native anemone *Anemonia alicemartinae* Häussermann & Försterra, 2001 along the Humboldt-current large marine ecosystem: An ecological niche model approach. *PeerJ* <https://doi.org/10.7717/peerj.7156> (2019).
41. Lh, G., Rj, R. & Brante, A. One step ahead of sea anemone invasions with ecological niche modeling: Potential distributions and niche dynamics of three successful invasive species. *Mar. Ecol. Prog. Ser.* **690**, 83–95 (2022).
42. Allynid, A. J. *et al.* Comparing and synthesizing quantitative distribution models and qualitative vulnerability assessments to project marine species distributions under climate change. *PLoS ONE* **15**, 1–28 (2020).
43. Pennino, M. G. *et al.* Current and future influence of environmental factors on small pelagic fish distributions in the northwestern mediterranean sea. *Front. Mar. Sci.* **7**, 1–20 (2020).
44. Melo-Merino, S. M., Reyes-Bonilla, H. & Lira-Noriega, A. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecol. Model.* **415**, 108837 (2020).
45. Rosa, R., Dierssen, H. M., Gonzalez, L. & Seibel, B. A. Ecological biogeography of cephalopod molluscs in the Atlantic Ocean: Historical and contemporary causes of coastal diversity patterns. *Glob. Ecol. Biogeogr.* **17**, 600–610 (2008).
46. Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J. & Follows, M. J. Patterns of diversity in marine phytoplankton. *Science* **327**, 1509–1511 (2010).
47. Rodríguez-Ramos, T., Marañón, E. & Cermeño, P. Marine nano- and microphytoplankton diversity: Redrawing global patterns from sampling-standardized data. *Glob. Ecol. Biogeogr.* **24**, 527–538 (2015).
48. Righetti, D., Vogt, M., Gruber, N., Psomas, A. & Zimmermann, N. E. Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Sci. Adv.* **5**, eaau6253 (2022).
49. Busseni, G. *et al.* Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean. *Glob. Ecol. Biogeogr.* **29**, 1915–1928 (2020).
50. Ruz, P. M., Hidalgo, P., Yáñez, S., Escribano, R. & Keister, J. E. Egg production and hatching success of *Calanus chilensis* and *Acartia tonsa* in the northern Chile upwelling zone (23°S) Humboldt Current System. *J. Mar. Syst.* **148**, 200–212 (2015).
51. Ashlock, L., García-Reyes, M., Gentemann, C., Batten, S. & Sydeman, W. Temperature and patterns of occurrence and abundance of key copepod taxa in the Northeast Pacific. *Front. Mar. Sci.* **8**, 1–10 (2021).
52. Campbell, M. D. *et al.* Testing Bergmann's rule in marine copepods. *Ecography* **44**, 1283–1295 (2021).
53. Soberón, J. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* **10**, 1115–1123 (2007).
54. Soberón, J. & Nakamura, M. Niches and distributional areas: Concepts, methods, and assumptions. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19644–19650 (2009).
55. Morales, C. E. *et al.* Mesoscale structure of copepod assemblages in the coastal transition zone and oceanic waters off central-southern Chile. *Prog. Oceanogr.* **84**, 158–173 (2010).
56. Gonzalez, R. R. & Quiñones, R. A. Ldh activity in *Euphausia mucronata* and *Calanus chilensis*: Implications for vertical migration behaviour. *J. Plankton Res.* **24**, 1349–1356 (2002).
57. Escribano, R., Hidalgo, P. & Krautz, C. Zooplankton associated with the oxygen minimum zone system in the northern upwelling region of Chile during March 2000. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **56**, 1083–1094 (2009).
58. Fernández-Urruzola, I. *et al.* Plankton respiration in the Atacama Trench region: Implications for particulate organic carbon flux into the hadal realm. *Limnol. Oceanogr.* **66**, 3134–3148 (2021).
59. Steinberg, D. K. & Landry, M. R. Zooplankton and the ocean carbon cycle. *Ann. Rev. Mar. Sci.* **9**, 413–444 (2017).
60. Tutas, P. & Escribano, R. Zooplankton diel vertical migration and downward-C flux into the oxygen minimum zone in the highly productive upwelling region off northern Chile. *Biogeosciences* **17**, 455–473 (2020).
61. Gonzalez, A. & Marín, V. H. Distribution and life cycle of *Calanus chilensis* and *Centropages brachiatus* (Copepoda) in Chilean coastal waters: A GIS approach. *Mar. Ecol. Prog. Ser.* **165**, 109–117 (1998).
62. Pulliam, H. R. Sources, sinks, and population regulation. *Am. Nat.* **132**, 652–661 (1988).
63. Dias, P. C. Sources and sinks in population biology. *Trends Ecol. Evol.* **11**, 326–330 (1996).
64. Ding, M., Lin, P., Liu, H., Hu, A. & Liu, C. Lagrangian eddy kinetic energy of ocean mesoscale eddies and its application to the Northwestern Pacific. *Sci. Rep.* **10**, 12791 (2020).
65. Morales, C. E. *et al.* The distribution of chlorophyll-a and dominant planktonic components in the coastal transition zone off Concepción, central Chile, during different oceanographic conditions. *Prog. Oceanogr.* **75**, 452–469 (2007).
66. Escribano, R. & Rodriguez, L. Life cycle of *Calanus chilensis* Brodsky in Bay of San Jorge, Antofagasta Chile. *Hydrobiologia* **292**, 289–294 (1994).
67. Hidalgo, P. & Escribano, R. Coupling of life cycles of the copepods *Calanus chilensis* and *Centropages brachiatus* to upwelling induced variability in the central-southern region of Chile. *Prog. Oceanogr.* **75**, 501–517 (2007).
68. Sobarzo, M., Bravo, L., Donoso, D., Garcés-Vargas, J. & Schneider, W. Coastal upwelling and seasonal cycles that influence the water column over the continental shelf off central Chile. *Prog. Oceanogr.* **75**, 363–382 (2007).
69. Carlson, C. J. embarcadero: Species distribution modelling with Bayesian additive regression trees in r. *Methods Ecol. Evol.* **11**, 850–858 (2020).
70. Gelfand, A. *et al.* Explaining species distribution patterns through hierarchical modeling. *Bayesian Anal.* <https://doi.org/10.1214/06-BA102> (2006).
71. Hortal, J. *et al.* Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 523–549 (2015).
72. Wisz, M. S. *et al.* Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **14**, 763–773 (2008).
73. van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J. & Raes, N. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* **39**, 542–552 (2016).
74. Gaul, W. *et al.* Data quantity is more important than its spatial bias for predictive species distribution modelling. *PeerJ* **8**, e10411 (2020).
75. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* **19**, 10–15 (2014).
76. Breiner, F. T., Guisan, A., Bergamini, A. & Nobis, M. P. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol. Evol.* **6**, 1210–1218 (2015).
77. Breiner, F. T., Nobis, M. P., Bergamini, A. & Guisan, A. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods Ecol. Evol.* **9**, 802–808 (2018).
78. Yasuhara, M. *et al.* Past and future decline of tropical pelagic biodiversity. *Proc. Natl. Acad. Sci.* **117**, 12891–12896 (2020).
79. Richardson, A., Schoeman, D., Richardson, A. J. & Schoeman, D. S. Climate impact on plankton ecosystems in the Northeast Atlantic. *Science* **305**, 1609–1612 (2004).
80. Chiba, S., Sugisaki, H., Nonaka, M. & Saino, T. Geographical shift of zooplankton communities and decadal dynamics of the Kuroshio-Oyashio currents in the western North Pacific. *Glob. Change Biol.* **15**, 1846–1858 (2009).

81. Reygondeau, G. & Beaugrand, G. Future climate-driven shifts in distribution of *Calanus finmarchicus*. *Glob. Change Biol.* **17**, 756–766 (2011).
82. Beaugrand, G., Lindley, J. A., Helaouet, P. & Bonnet, D. Macroecological study of *Centropages typicus* in the North Atlantic Ocean. *Prog. Oceanogr.* **72**, 259–273 (2007).
83. Hirche, H. J., Barz, K., Ayon, P. & Schulz, J. High resolution vertical distribution of the copepod *Calanus chilensis* in relation to the shallow oxygen minimum zone off northern Peru using LOKI, a new plankton imaging system. *Deep Sea Res. I Oceanogr. Res. Pap.* **88**, 63–73 (2014).
84. Spalding, M. D. *et al.* Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
85. Barve, N. *et al.* The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* **222**, 1810–1819 (2011).
86. Riquelme-Bugueño, R. *et al.* The influence of upwelling variation on the spatially-structured euphausiid community off central-southern Chile in 2007–2008. *Prog. Oceanogr.* **92–95**, 146–165 (2012).
87. Soberón, J. & Peterson, A. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers. Inform.* <https://doi.org/10.17161/bi.v2i0.4> (2005).
88. Provoost, P. & Bosch, S. robis: Ocean Biodiversity Information System (OBIS) Client (2020).
89. Chamberlain, S. & Oldoni, D. rgbif: Interface to the Global Biodiversity Information Facility API (2021).
90. R Core Team. R: A Language and Environment for Statistical Computing (2021).
91. Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B. & Anderson, R. P. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38**, 541–545 (2015).
92. ESRI. *ArcGIS Desktop: Release 10.4.1* (Environmental Systems Research Institute, 2016).
93. De Marco, P. & Nóbrega, C. C. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS ONE* **13**, e202403 (2018).
94. Feng, X. *et al.* A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.* **3**, 1382–1395 (2019).
95. Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. Where is positional uncertainty a problem for species distribution modelling?. *Ecography* **37**, 191–203 (2014).
96. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).
97. Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).
98. Pinto-Ledezma, J. N. & Cavender-Bares, J. Predicting species distributions and community composition using satellite remote sensing predictors. *Sci. Rep.* **11**, 16448 (2021).
99. Ellison, A. M. Bayesian inference in ecology. *Ecol. Lett.* **7**, 509–520 (2004).
100. Pennino, M. G., Muñoz, F., Conesa, D., López-Quílez, A. & Bellido, J. M. Bayesian spatio-temporal discard model in a demersal trawl fishery. *J. Sea Res.* **90**, 44–53 (2014).
101. Di Cola, V. *et al.* ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40**, 774–787 (2017).
102. Engler, R., Guisan, A. & Rechsteiner, L. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* **41**, 263–274 (2004).
103. Pearce, J. & Ferrier, S. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **133**, 225–245 (2000).
104. Boyce, M. S., Vernier, P. R., Nielsen, S. E. & Schmiegelow, F. K. A. Evaluating resource selection functions. *Ecol. Model.* **157**, 281–300 (2002).
105. Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* **199**, 142–152 (2006).
106. Warren, D. & Dinnage, R. ENMTools: Analysis of Niche Evolution using Niche and Distribution Models (2020).
107. Assis, J. *et al.* Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* **27**, 277–284 (2018).
108. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).
109. Osorio-Olvera, L. *et al.* ntbox: An R package with graphical user interface for modelling and evaluating multidimensional ecological niches. *Methods Ecol. Evol.* **11**, 1199–1206 (2020).
110. Bosch, S., Tyberghein, L. & De Clerck, O. 'sdmpredictors': Species distribution modelling predictor datasets. R package version 0.2.6. *R Packag. version 0.2.6* (2018).
111. Thuiller, W., Georges, D., Engler, R. & Breiner, F. biomod2: Ensemble platform for species distribution modeling (2020).
112. Zurell, D. *et al.* A standard protocol for reporting species distribution models. *Ecography* **43**, 1261–1277 (2020).

Acknowledgements

This work was funded by Grant the Millennium Institute of Oceanography (Grant ICN12_019). RR thanks Magdalena Sanhueza and Paloma Nuñez for their constructive and critical comments on the manuscript.

Author contributions

R.R. and R.E. conceived and designed the study, R.R., R.E., M.P.-A. performed the analyses and wrote the paper. R.R., R.E., M.P.-A. and C.G. participated in technical discussions and co-wrote the paper. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29541-9>.

Correspondence and requests for materials should be addressed to R.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023