Opinion

Forecasting the future of life in Antarctica

Gabrielle Koerich, [1,*,@] Ceridwen I. Fraser, [2] Charles K. Lee, [3] Fraser J. Morgan, [4,5] and Jonathan D. Tonkin [1,5,6,7,*,@]

Antarctic ecosystems are under increasing anthropogenic pressure, but efforts to predict the responses of Antarctic biodiversity to environmental change are hindered by considerable data challenges. Here, we illustrate how novel data capture technologies provide exciting opportunities to sample Antarctic biodiversity at wider spatiotemporal scales. Data integration frameworks, such as point process and hierarchical models, can mitigate weaknesses in individual data sets, improving confidence in their predictions. Increasing process knowledge in models is imperative to achieving improved forecasts of Antarctic biodiversity, which can be attained for data-limited species using hybrid modelling frameworks. Leveraging these state-of-the-art tools will help to overcome many of the data scarcity challenges presented by the remoteness of Antarctica, enabling more robust forecasts both near- and long-term.

Harnessing new tools to track and forecast rapid changes in Antarctica

Antarctica remains one of the most pristine locations on Earth [1], but is increasingly threatened by a suite of global change stressors [2]. Forecasting ecological responses to ongoing anthropogenic change is fundamental to guiding effective protection of biodiversity in Antarctica. However, challenges hampering ecological forecasts elsewhere are amplified in Antarctica: existing data are sparse due to the logistical challenges of sampling, and often lack relevant biological, spatial, and temporal information [3,4]. To improve benchmarking and forecasting of biodiversity, using techniques that can harness existing data in all their forms is imperative. However, as the climate moves away from what species have historically experienced and novel biotic interactions become possible, models that explicitly incorporate ecological processes are becoming increasingly necessary to improve long-term forecasts [5]. Opportunely, recent advancements in quantitative techniques are providing a rapidly expanding suite of tools for fusing different types of information and modelling approaches.

Here, we make a case for how these techniques can be implemented to increase the capability and accuracy of biodiversity forecasts in Antarctica, strengthening evidence-based management of the continent. We propose that forecasts should leverage available data, however sparse, but increasingly focus on incorporating ecological processes into flexible, hybrid modelling tools to improve predictions.

Antarctic ecosystems: isolated, rapidly changing, and data poor

Antarctic ecosystems are shaped by the significant challenges imposed on life in an extremely dry continent covered in ice, with long periods of limited or no sunlight [3]. These challenges force the majority of terrestrial biodiversity into the few existing ice-free areas, and result in relatively simple ecosystems that are sensitive to climate-driven environmental changes [3,6]. One such change of particular importance in these water-limited systems will be altered ice and snow melt conditions, including the temporal availability of meltwater [3]. Ice-free areas are predicted to expand [6], possibly connecting historically isolated, highly endemic populations [3,7]. Moreover, the

Highlights

Antarctic ecosystems are important contributors to global climatic and biogeochemical cycles, and represent simplified global sentinels that can provide insights into how biodiversity will respond to global change.

New techniques to capture biodiversity data offer promising solutions to overcome the considerable logistical constraints faced by biologists in Antarctica.

Advancements in data and model integration allow the combination of limited biodiversity data sets to improve data coverage and, therefore, forecasts of Antarctic ecosystems.

Leveraging available data, however sparse, while increasingly focusing on incorporating ecological processes into flexible, hybrid modelling frameworks can increase the capability and accuracy of biodiversity forecasts in Antarctica.

[1]School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[2]Department of Marine Science, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[3]International Centre for Terrestrial Antarctic Research, School of Science, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
[4]Manaaki Whenua – Landcare Research, Auckland 1072, New Zealand
[5]Te Pūnaha Matatini, Centre of Research Excellence in Complex Systems, Auckland, New Zealand
[6]Bioprotection Aotearoa, Centre of Research Excellence, Canterbury, New Zealand
[7]Website: http://tonkinlab.org

*Correspondence:
gabrielle.koerich@pg.canterbury.ac.nz
(G. Koerich) and
jonathan.tonkin@canterbury.ac.nz
(J.D. Tonkin).
@Twitter: @gabkoerich (G. Koerich) and
@jdtonkin (J.D. Tonkin).

projected loss of sea ice will directly impact marine ecosystems via habitat and connectivity changes, and alter the productivity of lower latitude ecosystems [8]. Habitat changes of this nature are concerning because ocean circulation models and genomics have revealed that organisms frequently cross the Antarctic Circumpolar Current, with establishment likely only precluded by unfavourable conditions [9,10]. Escalating human activity on the continent is also expected to expedite the establishment of invasive species [11–13], thereby altering ecosystem processes by filling new functional roles [3].

Understanding Antarctic biodiversity responses to ongoing change is imperative because this biodiversity has a critical role in global climatic and biogeochemical cycles, and food webs [8,14]. Moreover, the simplicity of biological communities and the sharp thermal gradients found in terrestrial Antarctica can provide insights into how species may fare under climate change [15]. However, the current state of understanding of responses to both ongoing and future changes leaves considerable room for improvement. To forecast how these climate induced changes will impact Antarctic biodiversity and wider ecosystem processes, we must first gather diverse ecological data and estimates of their uncertainty.

Data challenges in Antarctica

The extreme environmental conditions of Antarctica exacerbate the widespread data scarcity challenge faced by conservationists and ecological modellers alike. Most of the ice-free areas are difficult to access, and sampling is typically restricted to a short 3-month period during the summer, making data spatially, temporally, and taxonomically restricted. A very small portion of the continent has ever been visited or surveyed by researchers, and there remains significant room for discovery. The Southern Ocean is equally challenging, with large tracts remaining unsurveyed [4]. Furthermore, much of the biological information available has resulted from opportunistic visits and, therefore, is relatively patchy and typically presence only, enabling little or no understanding of the mechanisms that drive species distributions [16]. Although collecting more process-informed data is possible using novel sampling technologies (as detailed in the following section), these existing **unstructured data** (see Glossary) hold vital information that can be extracted using state-of-the-art flexible statistical modelling frameworks.

A roadmap for better predictions in Antarctica

Consistently collect and report data

To monitor biodiversity change in a way that best informs biodiversity forecasts in Antarctica, researchers must collect, archive, and share new data in a standardised manner, including spatially and temporally explicit metadata [17]. Consistent data structure and format, which results from such approaches, reduces information loss when incorporating data from different monitoring programs and research groups into models [18]. Biodiversity observatory networks, such as Group on Earth Observations Biodiversity Observation Network (GEO BON) and Genomic Observatories Meta-Database (GEOME), can have a key role in such initiatives by promoting the coordination and delivery of standardised biodiversity and ecosystem services data [18,19]. Such a network to supplement biodiversity data already exists for the Southern Ocean [the Southern Ocean Observing System (SOOSⁱ)], and one for terrestrial communities is being proposed by the Antarctic Near-shore and Terrestrial Observation System (ANTOS) [20]. Although cataloguing the diversity of species across different habitats in Antarctica remains an important task, repeated sampling of key species or communities at the same locations will bring benefits for the quantification of trends and drivers.

Leverage novel technologies to improve data coverage

Remote sensing and environmental DNA (eDNA) offer promising opportunities to overcome the logistical limitations of traditional sampling in Antarctica and expand the scope of what is possible.

For example, satellite imagery can be used to monitor some biological communities in near-real time, providing a continuous measure in space of habitat structure and extent, with on-site measurements being used to validate and complement satellite data [21]. For biological variables that require *in situ* measurements (e.g., community composition), remote sensing can extrapolate from point data to larger scales [22]. In Antarctica, very high-resolution (VHR) satellite images are enabling the non-invasive monitoring of penguins and marine mammals at large spatial scales [23,24] (Box 1). Although remote sensing is more limited in its ability to detect smaller life forms, efforts are underway in Antarctica to understand microbial mat communities and terrestrial vegetation by using multispectral satellite and drone imagery [25–27].

Autonomous devices that can remotely collect and analyse eDNA offer further exciting opportunities for Antarctic biodiversity assessment in hard-to-reach areas and periods, including monitoring for non-native species [28]. Although eDNA methods have only recently been applied to Antarctic systems (e.g., [29,30]), major collaborative research programmes, such as New Zealand's Antarctic Science Platform[ii], are combining genomic sequencing and whole-ecosystem analysis to understand biogeographic processes. Moreover, the Antarctic Treaty encourages the exchange and sharing of scientific observations and results, and the Scientific Committee on Antarctic Research (SCAR) has supported collaborative efforts that collect and share eDNA samples (e.g., the Antarctic Genetic Archive[iii]) and data (e.g., the SCAR Antarctic Biodiversity Portal[iv] or GenBank). To circumvent the challenges of discerning contemporary from ancient diversity signals in the dry, cold Antarctic, companion approaches, such as eRNA [31], could be used. However, the current paucity of genetic reference data for many Antarctic taxa (particularly invertebrates, plants, and fungi) precludes the reliable identification of many species from eDNA [29,32], although it can still be a useful tool for broad-scale diversity comparisons or for identifying the presence of well-studied vertebrates.

Although ongoing data collection efforts are essential to advance Antarctic forecasts, historical data continues to hold value for elucidating biodiversity responses to environmental change. Machine learning tools, including natural language processing, are being developed to search for, classify, and extract relevant information from historical publications more efficiently [33]. Despite the need for quality control of data, these tools can reduce the amount of time spent by researchers synthesising data. However, gathered data will be highly patchy and incomplete, and very limited in process information, due to the difficulties of sampling in Antarctica. Fortunately, with the growing field of data-model integration [34], we are increasingly equipped to ensure such valuable information can be put to better use.

Harness all forms of data to generate fit-for-purpose models

Generating fit-for-purpose forecasts of Antarctic biodiversity first requires weighing the modelling objectives (e.g., near- versus long-term) against the limitations of available data. The ecological modelling tools available, such as those focussed on predicting species distributions, vary from drawing correlations between biodiversity patterns and the environment to simulating the underlying processes that drive the observed patterns [35–37]. Thus, a state-to-process continuum of models and data exists [36], which is typically inversely correlated with data availability (Figure 1).

At one end of the state-to-process continuum are **correlative models**, which describe ecological patterns making direct connections between the state of the ecological system (e.g., species presence, presence/absence, or abundance) and what are considered to be the drivers [36,37]. Given their more lenient data requirements, these models are more widely applied but do not capture the underlying dynamics of species responses to the environment, or other critical mechanisms (Box 2) and, therefore, provide mixed results in their ability to predict range dynamics

Glossary

Correlative model: model describing biological states, including species distributions, via statistical relationships between the biological state and the environment.

Ensemble forecasting: combination/averaging of different model predictions; widely used in species distribution models.

Forecast skill: how precise the predictions of a model are.

Hierarchical model: in a general sense, a model with conditional dependencies between different levels. These levels are often submodels, either data, process, or parameter models. Hierarchical models can be both frequentist and Bayesian. A hierarchical model can also be a model for nested data, in which parameters vary among groups and, thus, the model aims to simulate this variation probabilistically; often named a 'mixed-effects' model.

Hybrid model: model that combines two different submodels, often based on different data types, to predict species distributions. These models can either be developed sequentially or link different submodels, in which at least one model is process based.

Integrated population model: models that unify several data sets to estimate population dynamics and/or demographic rates.

Integrated species distribution model: models that combine multiple data sets via data integration to estimate species climatic envelopes (i.e., multiple observation models).

Joint likelihood: integrates different calculated likelihoods into a single estimator (e.g., the latent state) to infer a shared set of parameters. Likelihoods, functions describing the probability of the data according to the set of parameters from a chosen model, can be calculated via point process models.

Latent state: property that cannot be directly observed and, thus, is inferred through a mathematical model, such as species distributions.

Point process model: distribution of points in a determined space, in which the probability of a point arising in a determined location is independent, and determined by the intensity of the underlying process (i.e., the density of points in a certain location). A point process model is a type of state-space model (and, therefore, a type of

into novel conditions [38,39]. For instance, species distribution models (SDMs) link species ranges with climatic envelopes and have been widely used in ecology [37]. In Antarctica, SDMs have demonstrated that invasive terrestrial species will increasingly find suitable habitats in the Antarctic Peninsula under climate change [12]. SDMs have also identified important feeding areas of seabirds and marine mammals to inform marine reserve delineations in sub-Antarctic islands [40], and estimated climate sensitivities of different barnacle life stages [41]. However, research using SDMs is typically focussed on more widely studied taxa, with, for instance, terrestrial Antarctic species and marine invertebrates rarely modelled using SDMs (but see [42–44] for examples). Although SDMs can incorporate data collected in either **structured** or unstructured forms, most are built with unstructured presence-only data and, therefore, inferences are limited due to possible sampling biases and observational uncertainty [45]. Methods have been developed to strengthen the inference of correlative models, including combining different model predictions in **ensemble forecasts** [46]. However, integrating data offers a way to bolster predictions by retaining the individual strengths of different data sets [47].

The power of data integration to improve forecasts depends on the approach to integrate data (Figure 1). The simplest method is to pool different data sets by assuming disparities are insignificant enough, reducing data sets to their lowest common denominator (e.g., abundance data to presence-only) [47]. By contrast, linking different data sets while retaining their strengths and uncertainties is possible using **state-space models**. The actual distribution of a species is an unobservable phenomenon (i.e., a **latent state**) from which we receive snapshots of individual occurrences, restrained in their location and time of collection [47]. In a state-space model, the observations (occurrences) are separated from the underlying true dynamics of the system (the latent state; here, the distribution). One such approach is to used **integrated species distribution models** (IDMs), which use two key tools based on a state-space modelling framework to integrate different forms of data: **point process models** and **joint likelihood**. Point process models can merge different data sets of the same state, such as abundance data, to generate, for example, a common point process of abundance [45]. Then, point process models from different data types (e.g., presence-only and presence–absence) can be linked via joint likelihood because these share a common latent state (the distribution) [48]. By integrating different data types and sources, IDMs reduce biases and increase the predictive performance of correlative models [45]. Joint-likelihood approaches can also be used to infer important ecological parameters [47]. For instance, **integrated population models** were used to estimate survival and fecundity (vital rates) of emperor penguins (*Aptenodytes forsteri*) through the combination of state-space models of population counts and estimated likelihoods of demographic data (capture–recapture and number of breeders) [49]. These joint-likelihoods were then integrated into a **hierarchical model** to quantify climate-change threats to emperor penguins [49].

Hierarchical models offer the flexibility needed to overcome many of the diverse challenges associated with Antarctic biodiversity data. For example, they can borrow strength across species by leveraging data from abundant species to estimate vital rates of rarer species [37]. Thus, hierarchical models (often Bayesian) can provide a framework to incorporate different types of model and data, fill data gaps, and estimate important parameters on ecological processes driving species distributions [37,50]. Furthermore, the iterative nature of Bayesian methods (i.e., models can be updated as new data arrives) makes them particularly well suited to forecasts in Antarctica [34], where currently sparse data are set to rapidly expand due to emerging data capture technologies (Box 3).

At the other end of the state-process continuum are **process-based models** (PBMs). These models explicitly simulate the processes driving the range and population dynamics of species and, thus, are able to establish causal relationships between the underlying biology of a species

hierarchical model) because the intensity defines the latent state.

Process-based/mechanistic model: model describing a biological state, such as species distributions, or rate through causal links between processes and observed patterns. For simplicity, we use these terms interchangeably, but some have argued that these two terms represent different levels of complexity [35].

State-space model: model comprising a true (latent or hidden) state and an observation model describing the data generated by the latent state. A state-space model is hierarchical in the sense that the latent state (e.g., distribution) conditions the observation (occurrence of individuals) of the data. The term 'hidden process model' is also used to refer to a state-space model.

Structured data: data derived from a standardised sampling effort and design with a defined protocol, making data comparable across time and locations.

Transferability: ability of a model trained in a certain set of conditions to make accurate and precise predictions under a different set of conditions. Models can be transferred in space, time, and taxonomy.

Unstructured data: data collected without a defined sampling protocol. Data are usually the result of opportunistic sampling, without a standardised effort and design.

Box 1. Monitoring penguins and seals from space

Although data are limited for most Antarctic species, some groups have been the target of various surveys and provide examples of how novel technologies can be used to cost-effectively monitor population trends on greater spatial and temporal scales [24]. The application of VHR imagery (0.31–0.60 m spatial resolution) was first used in Antarctica for censuses of emperor penguins (*Aptenodytes forsteri*) in 2007; since then, this technology has further supported the estimation of breeding pairs of Adélie (*Pygoscelis adeliae*) and chinstrap (*Pygoscelis antarctica*) penguins [24] (Figure I). Combined with long-term mark–recapture data, spatial distribution of colonies assessed via VHR imagery was used to build demographic and metapopulation models of how future changes in sea ice will affect emperor penguin colonies [59,60]. Outputs from these models demonstrated that human decisions on greenhouse gas emissions can directly affect extinction rates of emperor penguin colonies in Antarctica: ~80% quasi extinction was predicted under business-as-usual emissions, in contrast to 44% if the Paris Agreement 2°C target is met [59].

Satellite imagery does not come without challenges, particularly in the amount of information that needs to be processed. To overcome this challenge, the 'Satellites over Seals' project recruited citizen scientists to help detect and enumerate colonies of Weddell seals (*Leptonychotes weddellii*) in large numbers of VHR images across Antarctica [23]. Besides increasing the spatial range of data collection of this species, this approach was also able to expand the historical data series to regions where traditional sampling was intermittent or rare [61]. In doing so, the project achieved the first-ever global distribution of Weddell seals and determined that the population size is smaller than previously estimated [23]. Models based on these data, together with crabeater seal (*Lobodon carcinophaga*) data also collected from VHR images, revealed a higher sensitivity of crabeaters to climate change due to their reliance on sea ice for breeding and krill for their diet [62,63].

These examples demonstrate the possibilities that novel sampling methodologies present for collecting important information on Antarctic ice-dependent vertebrate populations, and highlight their potential utility for monitoring biodiversity in other remote ecosystems [24]. However, the collection of these demographic data from satellite imagery was only possible due to knowledge of the life histories of these species, and considerable work to correct counts and ground-truth data [23]. Thus, the application of VHR for other taxa needs careful consideration of phenologies and life history traits.
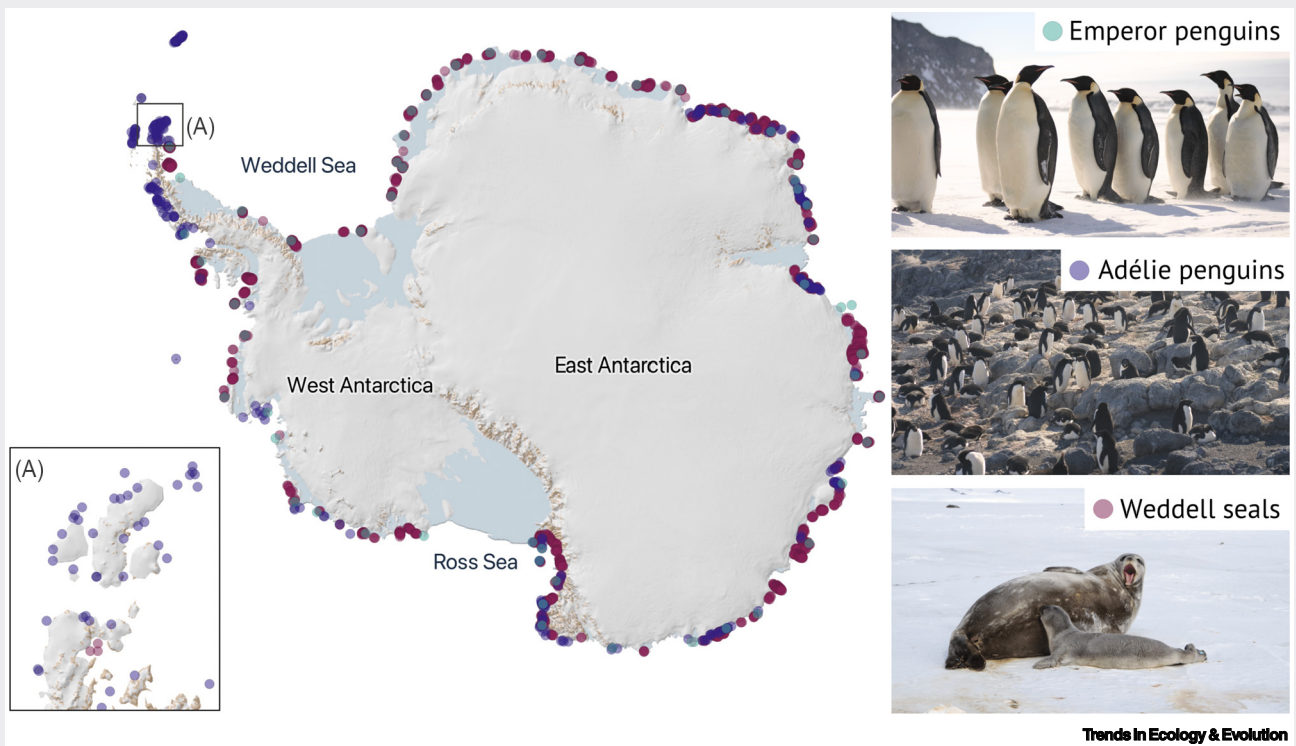

Trends in Ecology & Evolution

Figure I. Remote sensing is being used to census Antarctic populations. Map of Antarctica showing the locations of data collected through very high-resolution (VHR) satellite imagery for emperor (*Aptenodytes forsteri;* blue points) and Adélie penguins (*Pygoscelis adeliae;* purple points), and Weddell seals (*Leptonychotes weddellii;* pink points). (A) The location of individual data points on the tip of the Antarctic Peninsula. Data from [64] (Adélie penguins), [65] (emperor penguins), and [23] (Weddell seals). Photos reproduced with permission from M. LaRue.

and their environment [5]. Therefore, PBMs are better at overcoming the challenge of nonstationarity, which hampers the performance of correlative models, and their **transferability** to novel conditions, such as those expected in Antarctica under climate change, is likely higher [37] (Box 2). Characterising climate change impacts on species distributions will require
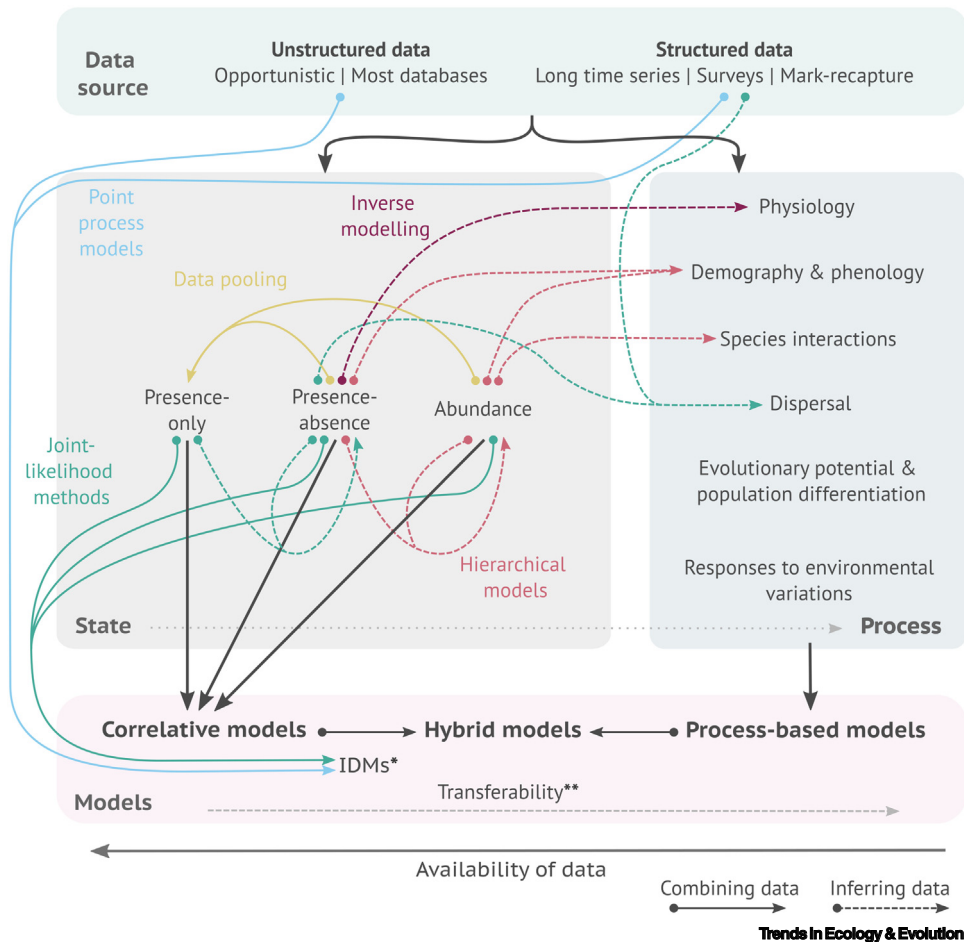
Figure 1. Modelling across the state-to-process continuum. A wide range of approaches are available to leverage diverse data types in forecast development. Dark-grey arrows depict the modelling workflow: the two main data sources (unstructured and structured) feed state and process data (categories according to Urban *et al.* [55]), which are then used in correlative and process-based models. Hybrid models are the combination of process-based and correlative models (thin-grey arrows). Coloured arrows demonstrate how data can be combined (solid arrows) and inferred (dashed arrows) to estimate another type of data; most move up in the state-to-process continuum (dashed light-grey arrow), except for data pooling (yellow arrows). Integrated species distribution models (IDMs) are developed using the combination of different forms of the same data through point process models (e.g., opportunistic occurrence data and presence surveys; blue arrow), or by combining different state variables via joint likelihood (green arrows) [45,47]. Joint likelihood can also be used to complement presence–absence data with presence-only in a multispecies model (example in [47]). Hierarchical models (pink arrows) can infer a range of process parameters (examples taken from [37,45,47]). Inverse modelling (dark-pink arrow) has been used to estimate physiological parameters from presence–absence data [37,54]. Combining and inferring data are not limited to the example connections depicted. *Position does not reflect the state-to-process continuum; IDMs are purely correlative. **Transferability of models (dashed light arrow) is expected to increase as models become more process informed. In the case of IDMs, data integration methods can increase transferability.

incorporating particular features of Antarctic ecosystems, such as the limited connectivity between terrestrial communities [3]. Changes in environmental conditions alone are unlikely the best predictor of future species ranges if species are incapable of reaching new suitable habitats, or withstanding novel biotic interactions, for example. However, these models are hungry for data that are less frequently collected and, thus, PBMs of Antarctic species are limited in number. While there are exceptions [51], marine species dominate this space. For instance, dynamic energy budget models were parameterised to describe the life cycle of *Laternula elliptica*, a large

Box 2. Toward more mechanistic forecasts of Antarctic biodiversity

Models forecasting biodiversity change often ignore underlying mechanisms, which can be broadly classified into six major categories: (i) species interactions; (ii) demography, life history, and phenology; (iii) responses to environmental variation; (iv) evolutionary potential and population differentiation; (v) physiology; and (vi) dispersal, colonisation, and range dynamics (Figure I) [55]. Incorporating these mechanisms into models can help to improve forecasts of responses to climate change, particularly as conditions go beyond those that have been observed previously [55]. This challenge of nonstationarity particularly hampers correlative approaches, which rely on associations between historical states and not necessarily the underlying processes that drive those states [66]. By directly simulating biological processes, process-based models are also more straightforward to interpret and interrogate [35], which stands in contrast to the 'black-box' nature of machine learning models, helping with dissemination to managers and policymakers.

Some mechanisms are more fully captured by existing studies in Antarctica than are others. For instance, detailed studies have been carried out on the demographic rates of large, charismatic, and visible-from-space species (Figure IB), helping to identify threats from climate change to critical life-history stages, such as emperor penguins [67]. However, less is known about the diverse invertebrate fauna of Antarctica, both terrestrial and marine, and, although demographic parameters can often be estimated from less informative data (see Figure 1 in the main text), other parameters may be less straightforward. Thus, the Antarctic data challenge applies equally, if not more, to mechanistic data.

To mechanistically forecast Antarctic biodiversity will require a renewed effort to target basic natural history information and fill many of these data gaps. One particularly important threat to Antarctic biodiversity is the increasing rates of non-native species arrival and establishment [68]. Thus, incorporating novel biotic interactions (Figure IA) into models will be increasingly important to predict the potential effects of invaders on resident biodiversity. Beyond surviving in place in response to climate change, which can be captured in models by mechanisms reflecting physiological tolerances (Figure IE), local environmental preferences (Figure IC), and evolutionary potential and local adaptation [7] (Figure ID), endemic Antarctic species will need to move in space to overcome climate change threats. Therefore, integrating dispersal rates of species into mechanistic forecasts is increasingly important (Figure IF). Despite the challenges associated with collecting these data in Antarctica, the potential gains in our ability to anticipate and prevent biodiversity and ecosystem functionality decline cannot be discounted.

(A) Species interactions
(B) Demography
(C) Responses to environment
(D) Evolution
(E) Physiology
(F) Dispersal
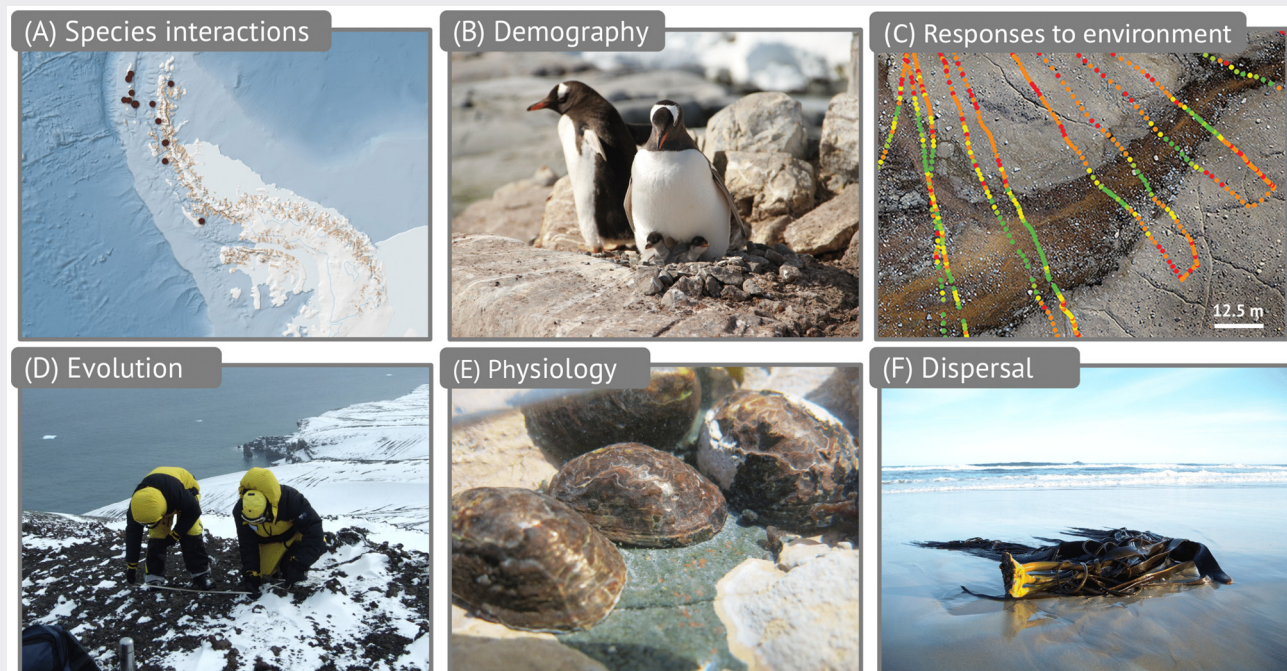
Trends in Ecology & Evolution

Figure I. Ecological mechanisms are increasingly important in Antarctic biodiversity forecasts. (A) The Antarctic Peninsula is a hotbed for the arrival of invasive species (points represent known distribution of non-native species; data from [68]), creating novel biotic interactions for resident biota. (B) Demographic models are increasingly being applied to understand responses of different penguin life stages to climate change [59,60,67] (gentoo penguins pictured). (C) Remote sensing using hyperspectral imaging identifies locations with high concentrations of microbial mats in stream channels in the McMurdo Dry Valleys (colours represent band absorption identifying high concentrations of microbial mats) [26]. (D) Sampling to understand the genetic variation of soil communities on Deception Island. (E) Dynamic Energy Budget models are one way to incorporate physiology into forecasts (here of the limpet *Nacella concinna*) [69]. (F) Dispersal is an important mechanism under climate change with the kelp *Durvillaea antarctica* (pictured) increasingly reaching Antarctica [9]. Photos reproduced with permission from M. LaRue (B); J. Levy (C); C. Fraser (D); M. Sissini (E); and C. Fraser (F).

Box 3. Toward robust iterative ecological forecasts

Although data requirements often dictate model choice, forecast skill, uncertainty, and transferability of models should be front of mind when determining how appropriate a model is [34,70]. Temporal and spatial forecast horizons can, in turn, help to determine the limits of predictions (i.e., how far out robust predictions can be made) [71]. Knowing such limits is imperative for supporting decision-making in the near- and long-term [70]. Arguably most important, however, is that predictions should account for, clearly quantify, and communicate uncertainty, a critical step in helping managers make informed decisions [34]. Uncertainty can arise from a range of sources, including data collection, the initial conditions chosen for the model, model structure, parameter estimation, and stochasticity [34]. In Antarctica, one particularly important source of uncertainty is driver uncertainty in the form of climate projections, because these are hampered by a shortage of instrumental records from the continent and the Southern Ocean [72].

Capturing, partitioning, and understanding the main sources of uncertainty in models and associated data are jobs particularly well suited to Bayesian hierarchical models [34]. Such models are increasingly being applied to near-term ecological forecasts that form an iterative forecast-analysis cycle, in which predictions are updated iteratively as new evidence is made available [34]. This iterative framework stands in stark contrast to most forecasts, which are performed once and then forgotten. Ecological forecasts can be more widely embedded in an iterative cycle that incorporates adaptive management to sequentially reduce uncertainties [34]. The immediate feedback available from iterative near-term forecasting offers a distinct advantage in our ability to not only understand the system, but also explore the immediate effects of decision-making on the near-term state of the system [34]. With remotely sensed data becoming increasingly available in Antarctica, automated iterative near-term ecological forecasts are well within the scope of research in the near future. In addition to data and modelling capabilities, we are seeing rapid developments in the cyberinfrastructure required to support such systems [34].

suspension-feeding bivalve, identifying adaptations to low temperatures and long starvation that make this species sensitive to changes in seasonal patterns of food abundance [52]. Important fisheries, such as the Antarctic toothfish (*Dissostichus mawsoni*), use historic catch data to estimate parameters, such as mortality and growth, to assess maximum sustainable yield using age-structured process-based population models (e.g., [53]). Most of these PBMs are fitted via forward parameterisation, in which direct measurements of a parameter, such as plant growth in an experiment, are used to describe the relationship between the environment and a mechanism [54]. Inferring parameters inversely (inverse modelling; using PBMs to estimate parameters based on observations of species) is also possible as a way to increase process information [37,54].

Hybrid models offer another key solution to the process data scarcity challenge by combining the strengths of both correlative and mechanistic models, thereby helping to maintain the elevated model transferability of PBMs into novel environmental states [38,55,56]. Building a hybrid model can be as simple as using one model output as a limiting factor in another model [36]. This works in both directions: a correlative model can be used to estimate habitat suitability and then a PBM uses this information to build predictions; alternatively, spatial variables with mechanistic information can be used as input in correlative models [56]. Models can also be generated simultaneously and integrated in a wider 'metamodel', based on a Bayesian hierarchical framework, in which predictions are constrained by the different submodels [57]. In Antarctica, Fabri-Ruiz *et al.* compared current and future predictions of a correlative and a dynamic energy budget model of a sea urchin (*Sterechinus neumayeri*) and demonstrated how one model can in fact complement the other, and how the mechanistic model forecasts were more biologically realistic [58]. Therefore, combining both modelling techniques via a hybrid framework can incorporate in forecasts both unstructured, opportunistic range data and parameters reflecting the specific mechanistic adaptations of Antarctic species that will drive their responses to climate change.

Concluding remarks

Improved forecasts of Antarctic biodiversity are necessary to inform decision-making, but are impaired by the characteristics of the data. We have illustrated how leveraging emerging data capture and integration tools, and state-of-the-art modelling frameworks, can improve such predictions. Traditional correlative modelling can greatly benefit from combining data sets to develop

Outstanding questions

How process-informed do models need to be to provide accurate forecasts in Antarctica? Few studies in fact test how model complexity affects **forecast skill** gain, and research exploring model parsimony and uncertainty is necessary to determine the most cost-effective approaches, both in data collection and modelling effort. This challenge is particularly important given the extreme nature of Antarctica and the rate of potential change in the environment that climate change will bring.

How can we most efficiently process the increasing volume of data being collected? Collecting new Antarctic data and rescuing historical data will likely lead to vast volumes of information, and correctly managing and interpreting data require the use of tools that most ecologists are not familiar with. The rapidly developing field of ecological synthesis provides an exciting framework to explore how these data can help uncover drivers and patterns of Antarctic biodiversity.

What are the limits of remote sensing tools for Antarctica? Most terrestrial species in Antarctica are too small for satellite data, except for photosynthetic organisms, which can be detected using multispectral imaging, but remote sensing can provide context around environmental constraints and be used to monitor the extent of communities. How much more spatial resolution can be achieved to allow better measurements of such communities and environmental conditions at finer scales remains to be seen.

How will improvements in climate models and instrumental records aid biodiversity models? Higher resolution at finer scales is necessary to better predict changes in microenvironments, particularly important for many microscopic communities in Antarctica. The importance of extreme weather events in driving biological communities to alternative stable states is increasingly being recognised and, thus, having this information to incorporate into models should be another target.

integrated species distribution models. When data on processes are necessary to overcome issues of model transferability into future environmental conditions, data integration can help infer mechanisms, and hybrid models can merge the strengths of both correlative and PBMs. Although integrating data into flexible and hybrid modelling frameworks is still relatively unexplored in Antarctica, the approach offers exciting possibilities for addressing a range of questions. Indeed, Antarctic science represents many opportunities for improving the way we achieve ecological forecasts in challenging circumstances.

Although our recommendations focus on better use of existing data and tools as well as increased monitoring efforts, they do not negate the need for greater understanding of biological and ecological processes to improve our ability to mechanistically forecast species populations. Future data collection efforts should focus on informing iterative forecasts (Box 3), and the implementation of international observation networks (as proposed by ANTOS and SOOS) make the ultimate goal of forecasting the future of Antarctic life with well-defined uncertainties attainable. At the national level, funding and support for establishing observation networks and, critically, regular resurveys and long-term monitoring should be prioritised. Challenges remain (see Outstanding questions), but scientists, managers, and policymakers cannot afford to let perfect be the enemy of good. We have the opportunity to learn by doing [34], and to optimise the use of data and tools that are in hand to guide science-based management of Antarctic biodiversity.

Declaration of interests

None declared by authors.

Resources

[i]soos.aq

[ii]antarcticscienceplatform.org.nz

[iii]ictar.aq/antarctic-genetic-archive

[iv]biodiversity.aq

References

1. Leihy, R.I. *et al.* (2020) Antarctica's wilderness fails to capture continent's biodiversity. *Nature* 583, 567–571
2. Rintoul, S.R. *et al.* (2018) Choosing the future of Antarctica. *Nature* 558, 233–241
3. Convey, P. *et al.* (2014) The spatial structure of Antarctic biodiversity. *Ecol. Monogr.* 84, 203–244
4. Kennicutt, M.C. *et al.* (2014) Polar research: six priorities for Antarctic science. *Nature* 512, 23–25
5. Briscoe, N.J. *et al.* (2019) Forecasting species range dynamics with process-explicit models: matching methods to applications. *Ecol. Lett.* 22, 1940–1956
6. Lee, J.R. *et al.* (2017) Climate change drives expansion of Antarctic ice-free habitat. *Nature* 547, 49–54
7. McGaughran, A. *et al.* (2019) Genome-wide SNP data reveal improved evidence for Antarctic glacial refugia and dispersal of terrestrial invertebrates. *Mol. Ecol.* 28, 4941–4957
8. Murphy, E.J. *et al.* (2021) Global connectivity of Southern Ocean ecosystems. *Front. Ecol. Evol.* 9, 454
9. Fraser, C.I. *et al.* (2018) Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nat. Clim. Chang.* 8, 704–708
10. Fraser, C.I. *et al.* (2022) Southern Hemisphere coasts are biologically connected by frequent, long-distance rafting events. *Curr. Biol.* Published online June 1, 2022. https://doi.org/10.1016/j.cub.2022.05.035
11. Duffy, G.A. and Lee, J.R. (2019) Ice-free area expansion compounds the non-native species threat to Antarctic terrestrial biodiversity. *Biol. Conserv.* 232, 253–257
12. Duffy, G.A. *et al.* (2017) Barriers to globally invasive species are weakening across the Antarctic. *Divers. Distrib.* 23, 982–996
13. Bergstrom, D.M. (2022) Maintaining Antarctica's isolation from non-native species. *Trends Ecol. Evol.* 37, 5–9
14. Deppeler, S.L. and Davidson, A.T. (2017) Southern Ocean phytoplankton in a changing climate. *Front. Mar. Sci.* 4, 40
15. McGaughran, A. *et al.* (2021) Evolutionary responses to warming. *Trends Ecol. Evol.* 36, 591–600

16. Chown, S.L. *et al.* (2015) The changing form of Antarctic biodiversity. *Nature* 522, 431–438

17. Toczydlowski, R.H. *et al.* (2021) Poor data stewardship will hinder global genetic diversity surveillance. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2107934118

18. Jetz, W. *et al.* (2019) Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* 3, 539–551

19. Riginos, C. *et al.* (2020) Building a global genomics observatory: using GEOME (the Genomic Observatories Metadatabase) to expedite and improve deposition and retrieval of genetic data and metadata for biodiversity research. *Mol. Ecol. Resour.* 20, 1458–1469

20. Cary, S.C. and Cummings, V. (2020) *Antarctic Near-shore and Terrestrial Observation System (ANTOS), 2018-2020 Report*, Scientific Committee on Antarctic Research

21. Turner, W. (2014) Sensing biodiversity. *Science* 346, 301–302

22. Pereira, H.M. *et al.* (2013) Essential biodiversity variables. *Science* 339, 277–278

23. LaRue, M. *et al.* (2021) Insights from the first global population estimate of Weddell seals in Antarctica. *Sci. Adv.* 7, eabh3674

24. LaRue, M. *et al.* (2022) High-resolution satellite imagery meets the challenge of monitoring remote marine protected areas in the Antarctic and beyond. *Conserv. Lett.* e12884

25. Salvatore, M.R. *et al.* (2020) Remote characterization of photosynthetic communities in the Fryxell basin of Taylor Valley, Antarctica. *Antarct. Sci.* 32, 255–270

26. Levy, J. *et al.* (2020) Detection and community-level identification of microbial mats in the McMurdo Dry Valleys using drone-based hyperspectral reflectance imaging. *Antarct. Sci.* 32, 367–381

27. Turner, D. *et al.* (2018) Assessment of Antarctic moss health from multi-sensor UAS imagery with random forest modelling. *Int. J. Appl. Earth Obs. Geoinf.* 68, 168–179

28. Yamahara, K.M. *et al.* (2019) In situ autonomous acquisition and preservation of marine environmental DNA using an autonomous underwater vehicle. *Front. Mar. Sci.* 6, 373

29. Fraser, C.I. *et al.* (2018) Evidence of plant and animal communities at exposed and subglacial (cave) geothermal sites in Antarctica. *Polar Biol.* 41, 417–421

30. Howell, L. *et al.* (2021) Environmental DNA as a tool for monitoring Antarctic vertebrates. *N.Z. J. Zool.* 48, 245–262

31. Yates, M.C. *et al.* (2021) Environmental RNA: a revolution in ecological resolution? *Trends Ecol. Evol.* 36, 601–609

32. Czechowski, P. *et al.* (2017) A primer to metabarcoding surveys of Antarctic terrestrial biodiversity. *Antarct. Sci.* 29, 3–15

33. Le Guillarme, N. and Thuiller, W. (2022) TaxoNERD: deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature. *Methods Ecol. Evol.* 13, 625–641

34. Dietze, M.C. *et al.* (2018) Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1424–1432

35. Connolly, S.R. *et al.* (2017) Process, mechanism, and modeling in macroecology. *Trends Ecol. Evol.* 32, 835–844

36. Dormann, C.F. *et al.* (2012) Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeogr.* 39, 2119–2131

37. Evans, M.E.K. *et al.* (2016) Towards process-based range modeling of many species. *Trends Ecol. Evol.* 31, 860–871

38. Fordham, D.A. *et al.* (2018) How complex should models be? Comparing correlative and mechanistic range dynamics models. *Glob. Change Biol.* 24, 1357–1370

39. Rousseau, J.S. and Betts, M.G. (2022) Factors influencing transferability in species distribution models. *Ecography*, e06060

40. Thiers, L. *et al.* (2017) Important marine sectors for the top predator community around Kerguelen Archipelago. *Polar Biol.* 40, 365–378

41. Gallego, R. *et al.* (2017) On the need to consider multiphasic sensitivity of marine organisms to climate change: a case study of the Antarctic acorn barnacle. *J. Biogeogr.* 44, 2165–2175

42. Pertierra, L.R. *et al.* (2017) Global thermal niche models of two European grasses show high invasion risks in Antarctica. *Glob. Change Biol.* 23, 2863–2873

43. Contador, T. *et al.* (2020) Assessing distribution shifts and eco-physiological characteristics of the only Antarctic winged midge under climate change scenarios. *Sci. Rep.* 10, 9087

44. Guillaumot, C. *et al.* (2021) Species distribution modelling of the Southern Ocean benthos: a review on methods, cautions and solutions. *Antarct. Sci.* 33, 349–372

45. Miller, D.A.W. *et al.* (2019) The recent past and promising future for data integration methods to estimate species' distributions. *Methods Ecol. Evol.* 10, 22–37

46. Araujo, M. and New, M. (2007) Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47

47. Isaac, N.J.B. *et al.* (2020) Data integration for large-scale models of species distributions. *Trends Ecol. Evol.* 35, 56–67

48. Fletcher, R.J. *et al.* (2016) Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landsc. Ecol.* 31, 1369–1382

49. Abadi, F. *et al.* (2017) Integrated population modeling reveals the impact of climate on the survival of juvenile emperor penguins. *Glob. Change Biol.* 23, 1353–1359

50. Cressie, N. *et al.* (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.* 19, 553–570

51. Pertierra, L.R. *et al.* (2020) Combining correlative and mechanistic niche models with human activity data to elucidate the invasive potential of a sub-Antarctic insect. *J. Biogeogr.* 47, 658–673

52. Agüera, A. *et al.* (2017) A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism. *PLoS ONE* 12, e0183848

53. Mormede, S. *et al.* (2014) A stock assessment model of Antarctic toothfish (*Dissostichus mawsoni*) in the Ross Sea region incorporating multi-year mark-recapture data. *CCAMLR Sci.* 21, 39–62

54. Higgins, S.I. *et al.* (2020) Predictive ability of a process-based versus a correlative species distribution model. *Ecol. Evol.* 10, 11043–11054

55. Urban, M.C. *et al.* (2016) Improving the forecast for biodiversity under climate change. *Science* 353, aad8466

56. Zurell, D. *et al.* (2016) Benchmarking novel approaches for modelling species range dynamics. *Glob. Change Biol.* 22, 2651–2664

57. Talluto, M.V. *et al.* (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework: Integrated models of species ranges. *Glob. Ecol. Biogeogr.* 25, 238–249

58. Fabri-Ruiz, S. *et al.* (2021) Using correlative and mechanistic niche models to assess the sensitivity of the Antarctic echinoid *Sterechinus neumayeri* to climate change. *Polar Biol.* 44, 1517–1539

59. Jenouvrier, S. *et al.* (2020) The Paris Agreement objectives will likely halt future declines of emperor penguins. *Glob. Change Biol.* 26, 1170–1184

60. Jenouvrier, S. *et al.* (2017) Influence of dispersal processes on the global dynamics of emperor penguin, a species threatened by climate change. *Biol. Conserv.* 212, 63–73

61. LaRue, M.A. *et al.* (2020) Engaging 'the crowd' in remote sensing to learn about habitat affinity of the Weddell seal in Antarctica. *Remote Sens. Ecol. Conserv.* 6, 70–78

62. Wege, M. *et al.* (2020) Citizen science and habitat modelling facilitates conservation planning for crabeater seals in the Weddell Sea. *Divers. Distrib.* 26, 1291–1304

63. Wege, M. *et al.* (2021) Distribution and habitat suitability of Ross seals in a warming ocean. *Front. Mar. Sci.* 8, 468

64. Lynch, H.J. and LaRue, M.A. (2014) First global census of the Adélie Penguin. *Auk* 131, 457–466

65. LaRue, M.A. *et al.* (2015) Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography* 38, 114–120

66. Tonkin, J.D. *et al.* (2019) Prepare river ecosystems for an uncertain future. *Nature* 570, 301–303

67. Jenouvrier, S. *et al.* (2009) Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1844–1847

68. Hughes, K.A. *et al.* (2015) Biological invasions in terrestrial Antarctica: what is the current status and can we respond? *Biodivers. Conserv.* 24, 1031–1055

69. Guillaumot, C. *et al.* (2020) Can DEB models infer metabolic differences between intertidal and subtidal morphotypes of the Antarctic limpet *Nacella concinna* (Strebel, 1908)? *Ecol. Model.* 430, 109088

70. Yates, K.L. *et al.* (2018) Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol.* 33, 790–802

71. Petchey, O.L. *et al.* (2015) The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.* 18, 597–611

72. Auger, M. *et al.* (2021) Southern Ocean in-situ temperature trends over 25 years emerge from interannual variability. *Nat. Commun.* 12, 514