



**HAL**  
open science

## Massive and infrequent informed emigration events in a species threatened by climate change: the emperor penguins.

Jimmy Garnier, Gemma Clucas, Jane Younger, Bilgecan Sen, Bilgecan Sen, Christophe Barbraud, Michelle Larue, Alexander D. Fraser, Sara Labrousse, Stéphanie Jenouvrier

### ► To cite this version:

Jimmy Garnier, Gemma Clucas, Jane Younger, Bilgecan Sen, Bilgecan Sen, et al.. Massive and infrequent informed emigration events in a species threatened by climate change: the emperor penguins.: Rare but massive informed dispersion. 2022. hal-03822288v1

**HAL Id: hal-03822288**

**<https://hal.science/hal-03822288v1>**

Preprint submitted on 20 Oct 2022 (v1), last revised 27 Mar 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genetic and demographic data reveal dispersal processes: Informed, massive and infrequent emigration events in Emperor penguins.

Jimmy Garnier<sup>a,1</sup>, Gemma Clucas<sup>b</sup>, Jane Younger<sup>c</sup>, Bilgecan Sen<sup>d,e</sup>, Christophe Barbraud<sup>f</sup>, Michelle LaRue<sup>g,h</sup>, Alexander D. Fraser<sup>i</sup>, Sara Labrousse<sup>j</sup>, and Stéphanie Jenouvrier<sup>e</sup>

<sup>a</sup>CNRS, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LAMA, 73000 Chambéry, France

<sup>b</sup>Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Rd. Ithaca, NY 14850, USA

<sup>c</sup>Milner Centre for Evolution, University of Bath, Bath UK BA2 7AY

<sup>d</sup>Institute for Advanced Computational Science, Stony Brook University, Stony Brook, New York, USA

<sup>e</sup>Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

<sup>f</sup>Centre d'Etudes Biologiques de Chizé, CNRS-La Rochelle University UMR7372, 79360 Villiers en Bois, France

<sup>g</sup>School of Earth and Environment, University of Canterbury, NZ

<sup>h</sup>School of Earth and Environmental Sciences, University of Minnesota, USA

<sup>i</sup>Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, nipaluna / Hobart 7001, Tasmania, Australia

<sup>j</sup>Laboratoire d'Océanographie et du Climat: Expérimentations et approches numériques (LOCEAN), UMR 7159 Sorbonne-Université, CNRS, MNHN, IRD, IPSL 75005, Paris, France

**Keywords:** dispersal distance | dispersal kernel | dispersal range | emigration rates<sup>1 2</sup>

**Significance statement** In the face of climate change, dispersal is a central process for species to survive but dispersal rates, range, and behaviors remain difficult to measure. By developing new models combining demographic and genetic data, we propose to unravel species dispersal behaviors. For Emperor penguins, living in the most remote place on earth where measurements of dispersal are nonexistent. Here, we found that penguins likely depart from their colony when the habitat becomes unsuitable when food resources are low and disperse over short distances relative to their movement capacity. Although, emigration rates remain low on average, occasional massive movements can occur. This approach can be applied to any species and data-poor system and promises to transform our understanding of dispersal processes.

**Abstract** Dispersal is an ubiquitous phenomenon which affects the population dynamics and evolution of natural populations, hence it is a fundamental process in driving biodiversity responses to rapid global change. However, dispersal rates and dispersal range are difficult to measure in most species, and remain unknown for many. In addition, informed behaviors, whereby individuals leave their natal area and select a breeding habitat non-randomly, may play an important role in species' responses to global change, and are even more difficult to comprehend. Here, we develop a new mathematical function combining demographic and genetic data to determine the dispersal distance, emigration rates and dispersal behaviors (random or informed emigration and establishment). We apply our approach to the Emperor penguin, a species threatened by climate change. We found that Emperor penguins have a short distance of dispersal compared to their capacity to cover large distances during migration. On average, emigration rates are small and Emperor penguins leave their colonies when the habitat becomes unsuitable (informed emigration). However, for some regions, massive emigration events can occur. Specifically, emigration is more likely to occur for habitat with low food availability that can not sustain large populations. Our model opens the doors to estimate both the mean

<sup>1</sup>Author contributions: SJ and JG conceptualized the research. JG developed the mathematical function. JG and SJ performed the dispersal analysis and BS, JG and SJ performed the covariate analysis. JY, GC, CB, MLR, DI, BF, SL curated the genetic, demographic or environmental data. SJ and JG wrote the initial draft and all co-authors provided critical revision. SJ and JG acquired the financial support for the project. AF and SL provided and processed fast ice data.

<sup>2</sup>Author declaration: No competing interests.

dispersal distance, emigration rates, and dispersal behaviours across the tree of life to transform our understanding of dispersal processes and their consequences for populations and ecosystems.

Dispersal among suitable habitats influences the dynamics of populations (e.g. refs [1]), their gene flow and genetic structure [2], and hence the ecological and evolutionary processes driving biodiversity [3]. The rate and range of dispersal of plant propagules and animal individuals are usually characterized by tracking individual movements and population redistribution (e.g using abundance data [4] or “mark-recapture/resighting” techniques in animal studies [5]). But such movement data are extremely challenging to collect. Genetic markers naturally present in populations offer unique opportunities to study dispersal [6]. However, such genetic methods (e.g. long-term frequency-based approach using population structure from FST) estimate effective dispersal over several generations, rather than dispersal processes relevant for the temporal scales at which ecological and demographic processes occur.

Recently, many methods have been developed to assess the dispersal kernel over one generation based on genetic data, especially to estimate seed dispersal kernels [7]. These genetic methods often rely on simple dispersal assumptions. For instance, the classical methods based on Euclidian distances or least-cost distances (e.g. in models of isolation by distance [8, 9]) assume a single and optimal migration path for individuals, while individuals may change their migration route during dispersal [10, 11]. More realistic approaches based on resistance networks [12, 13], which capture the relative cost of dispersing across particular landscape compare to some reference condition, have been developed. However, their implementation is time-consuming and the estimation of dispersal parameters, e.g. by maximum likelihood, generally lacks accuracy [14]. Furthermore, genetic data alone may not provide enough information on demographic processes because dispersal processes may depend on the environment [15]. Here, we combine genetic methods with environment-dependent meta-population models to propose a new likelihood function enabling the quantification of dispersal rate, distance and behaviors.

Specifically, to go beyond previous approaches, we expand a new method linking movement and demographic dynamics with genetic data [16]. This novel method is based on a mechanistic-statistical approach [17, 4] in the framework of state-space models [18]. It has been developed theoretically to characterize diffusion rates from genetic data of insects over one generation [16], but has yet to be applied to other species. In addition, this method has ignored reproductive and dispersal behaviors. The latter is particularly important as some species use personal and social information to decide whether to leave a natal or current breeding site and where to settle (e.g. [19]). Such ‘informed dispersal’ behavior [11] enables individuals to settle in habitats of better quality, potentially improving their fitness, hence increasing population viability and species persistence, especially in the face of global changes [20]. Here we develop a novel likelihood function that includes reproductive behaviors, including informed departure and settlement, that we apply to a wild animal species and a sparse data system. Specifically, we unravel, for the first time, the dispersal processes of Emperor penguins (*Aptenodytes forsteri*), a seabird species living in extreme environment currently threatened by climate change [21].

Emperor penguins are the only species to breed on coastal sea ice around Antarctica during winter. Due to the logistical challenges of monitoring populations in these conditions, very little is known about their dispersal behaviors. In fact, Emperor penguins have only been marked at one site (Pointe Géologie), with no recaptures elsewhere. The recent advent of satellite tracking has allowed for enhanced understanding of emperor penguin movement over large spatial scales. However, this approach is not suitable for ascertaining dispersal between colonies due to the limited life-span of these devices [22].

Like many seabirds, Emperor penguins are considered highly philopatric. However, this traditional view has been challenged with advances in genetic analyses and very high-resolution satellite imagery (VHR) that suggest that movements between colonies occur [23]. Indeed, genetic studies have characterized at least four genetically different metapopulations that are located in different geographical regions of Antarctica and variously connected through gene flow, while colonies within each appear to be panmictic via dispersal of individuals among colonies [24]. In addition, recent work using VHR satellite imagery has recently documented colony movements, disappearances and relocations [25]. For example, a dramatic decline of the world’s second largest Emperor penguin colony occurred at Halley Bay, while the nearby Dawson-Lambton

colony, 55 km to the south, has seen a more than tenfold increase in penguin numbers during the same period [26]. Halley Bay has suffered three years of almost total breeding failure caused by a shift in the local environment and sea ice conditions, and those unfavourable conditions may have forced penguins to relocate to Dawson-Lambton [26]. The colony had been present at Halley Bay since at least 1956, persisting for 60 years before the major environmental perturbation led to a massive population decline and emigration event. This suggests that Emperor penguin movements may be triggered by major environmental perturbations and that individuals leave their current breeding site using information about their habitat quality, such as the presence of a stable and suitable ice habitat to breed. These dispersal behaviors correspond to an informed emigration.

To understand the role of dispersal in Emperor penguin responses to climate change, demographers have developed a theoretical meta-population model including detailed dispersal mechanisms [27]. Specifically, the model decomposed dispersal into three distinct behavioural stages: the decision to leave the resident patch (emigration), movement between patches (transfer), and settlement into a new patch (establishment that makes a colonization event successful). In addition, the model integrates dispersal distance, informed dispersal behaviors, and density-dependent emigration and establishment rates within a structured habitat. Indeed, individuals may gather and exchange information during these different dispersal stages (informed dispersal decisions [11]). For example, individuals may preferentially leave poor-quality habitat (e.g. climate deteriorated or exceeding carrying capacity) and settle in higher-quality habitat by relying on environmental cues or by assessing habitat quality through the breeding success or presence of conspecifics [10]. In parallel, geneticists have assessed Emperor penguin genetic population structure using 4,596 genome-wide single nucleotide polymorphisms (SNPs), characterized in 110 individuals (10–16 per colony) from eight colonies around Antarctica [24]. Here, by coupling this meta-population model with those genetic data, we can address the following questions: 1. What are the most likely dispersal behaviors in emigration and establishment of Emperor penguins (informed versus random)? 2. What is the mean dispersal distance for this species? 3. What are the emigration rates? 4. What environmental and demographic factors drive emigration rates?

To understand the environmental and demographic drivers of emigration, we use a random forest algorithm to quantify the role of several variables on the emigration probability [28]. Specifically, we use novel time series of Emperor penguin populations for all known colonies estimated from VHR satellite imagery (LaRue et al. submitted), novel landfast sea ice (hereafter simply “fast ice”) data (sea ice that is “fastened” to the coastline or grounded icebergs and is an important platform to breed for Emperor penguins) [29], and unique sea ice and food web dynamics variables obtained from a forced ocean-sea-ice (FOSI) configuration of the the Community Earth System Model (CESM2) ocean ecosystem model [30].

## Results and Discussion

Using a new mathematical function combining demographic and genetic data we found that Emperor penguins leave poor habitat colonies but settle randomly into another colony; show small emigration rates – albeit massive emigration events can occur for some regions – and short dispersal distances. Here we present and discuss in details our new likelihood function to characterize behavior, ranges and rates of dispersal of any species, and the results of our model selection for Emperor penguins. Specifically, our model selection shows that the most likely behaviors supported by the genetic data is the semi-informed dispersal behavior (Table 1), whereby Emperor penguins are likely to depart from colonies with poor habitat (i.e. with negative intrinsic population growth; informed emigration), but settle randomly into another colony (random establishment).

### A new likelihood function to characterize behavior, ranges and rates of dispersal

Based on a mechanistic-statistical approach, we develop a numerically-tractable likelihood function associated with the demographic characteristics of the Emperor penguin, including dispersal behavior, mean dispersal distance, and emigration rates per colony, among others (see SI1 for details), given the genotypes  $\mathcal{G}_{i,\tau}$  of

Table 1: Model selection based on minimization of four selection criteria: the Bayesian Information Criteria (BIC), two Deviance Information Criteria (DIC) and a predictive Information Criteria (IC) for the three different dispersal behaviour: Random dispersal (random emigration and establishment), Semi-Informed dispersal (informed emigration but random establishment) and Informed dispersal (informed emigration and establishment).

Dispersal behavior	<i>BIC</i>	<i>DIC</i> <sub>1</sub>	<i>DIC</i> <sub>2</sub>	<i>IC</i>
Random	678	679	684	674
Semi-informed	-43.6	-828	-41	-1615
Informed	674	605.5	676	535

the sampled individuals. Our approach extends previous work in several directions by including overlapping generations and complex discursive processes allowing this approach to be applied to any species. Specifically, our model allows for the first time to estimate both the mean dispersal distance and emigration rates, and most importantly dispersal behaviors (e.g. informed versus random). Furthermore, our method is extremely flexible as it allows the inclusion of data sampled at various times and locations. More precisely, we combine an environmental dependent meta-population model projecting the demographic dynamics of the Emperor penguins over Antarctica [27] with a genetic based statistical model conditional on these demographic dynamics. Thanks to classical genetic assignment approaches [31], and dispersal analysis from spatially sampled data (see methods and SI1.E for details), we compute the likelihood function associated with the unknown parameters  $\Theta$  of our demographic and statistical model as:

$$\mathcal{L}(\Theta | \mathcal{G}_{i,\tau}) = \prod_{\tau=1}^J \prod_{i=1}^{G_\tau} \sum_{r=1}^R \mu_\tau^r(t_i) \gamma_{i,\tau}^r \quad (1)$$

Since assuming that the allele frequencies in each genetic cluster are known [24], the likelihood only depends on the proportion  $\mu_\tau^r(t_i)$  of the genetic cluster  $r$  in colony  $\tau$  at time  $t_i$  reflecting dispersal and population dynamics, and the probability  $\gamma_{i,\tau}^r$  that a given genotype sampled in location  $\tau$  belongs to a genetic cluster  $r$ , reflecting the genetic data.

In the present work, we describe the population dynamics of Emperor penguin using a metapopulation model developed by [27], which comprises the population size in each known colony. Although the local population dynamics account for stage and age structure, the dispersal processes do not, and emigration rates estimated here are averaged at the colony level. Furthermore, we assume that the metapopulation is composed of fixed genetic clusters which are at Hardy-Weinberg equilibrium, allowing us to focus only on the evolution of the number of individuals belonging to each cluster. Another possibility for other systems, more computationally demanding, is to compute the dynamics of each allele frequency at any time and any location. This dynamic can be modeled by adapting the methods of [32], but requires simulating a  $N + 1$  dynamical system where  $N$  is the total number of alleles in the population, which would be very large in our case study. Another alternative approach might be to use multilocus likelihoods for hybrids and backcrosses developed by [33]. This approach needs to track at any time step the dynamics of the hybrids and to compute the associated likelihood.

## Dispersal ranges

Figure 1(a) shows the posterior distribution of the mean dispersal distance for the best supported model, and suggests a short distance of dispersal relative to the potential dispersal distance that could be inferred based on tracked juveniles and adults [22, 34]. Indeed, we found that the most likely dispersal distance of the Emperor penguins is around 428 km. On the other hand, tracking movement studies have shown that individuals can cover incredible distances during their migration routes. In the Ross Sea, non-breeders travel up to 9,000 km [35] and after the moult adults covered more than 2,000 km on their return journey to their colonies [34]. In East Antarctica, one juvenile covered more than 7,000 km during the first eight

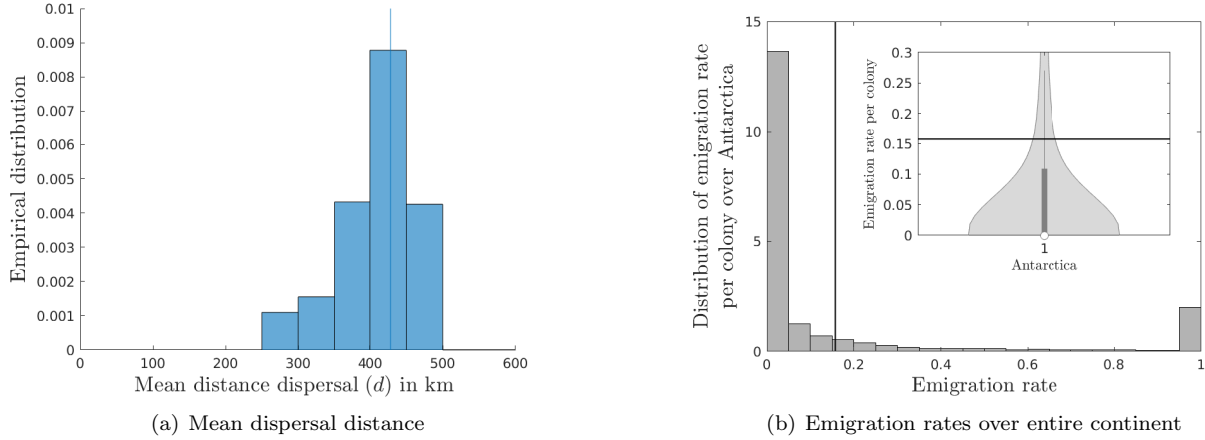


Figure 1: Posterior distributions of the mean distance dispersal  $d$  per individuals and the emigration rate per colony per year for the entire Antarctic continent. The plain lines are the mean of the distributions: (a) 428 km (blue) and (b) 0.157 (black).

months after leaving its natal colony in Terre Adélie [22]. However, the potential individual dispersal can be reduced by specific behaviors. For example, seabirds exhibit specific behavioral traits, such as a high degree of philopatry [36] and the importance of social cues in the recruitment of new breeders [37], which may reduce individual dispersal distance relative to individuals potential movement [38, 39]. For instance, the colonies in the Ross Sea are genetically distinct from the rest of the colonies [40], suggesting that, despite their large migration potential during the non-breeding season, Emperor penguin dispersal distance could be somewhat limited. In addition, after the demise of Halley Bay, many of the birds from Halley Bay may have relocated to nearby Dawson-Lambton colony, while the formation of new colonies elsewhere or movement to other colony locations further away are considered less likely [26].

## Dispersal rates

Figure 1(b), summarizes the emigration rates over the entire continent and supplementary Figure SI2 details the emigration rates at each colony. We found that although some massive emigration events can occur, such events remain rare. Indeed, the median is zero as more than 50% of the emigration rates calculated at each colony are zero. The averaged rate of emigration is 15.7% per year. Previous studies have debated the magnitude of emigration rates with some studies arguing for large emigration rates [23], while others report low emigration rates [41, 42].

Large emigration rates are expected as massive movements between colonies have been documented in the past two decades from satellite imagery: 1. Some colonies are known to ‘blink’ (disappear in some years, reappear in others) [25]; 2. Others are known to relocate onto icebergs or ice shelves during late formation of sea ice in the autumn [43]; and 3. Some colonies have shown dramatic declines whilst the nearby colonies have markedly increased in size [26]. Those blinking, relocation and movement events remain somewhat infrequent. Spatially, 17% of colonies are known to blink. Some colonies blinked several times in the last decade, so accounting for this temporal variation, there is only a 4% probability that a colony is absent in a given year. On the other hand, low emigrations rates are also expected because there is a high return rate of marked chicks at Pointe Géologie ([42] and C Barbraud personal communication).

However, it is important to define the temporal and spatial scales at which those rates occur, and here we propose that the averaged annual proportion of individuals moving from one breeding site to another is relatively low for Emperor penguins, while en masse emigration may occur occasionally at some locations (see Fig. 2(a) and Fig.SI2). This pattern has been observed in many seabird and bird species, and is

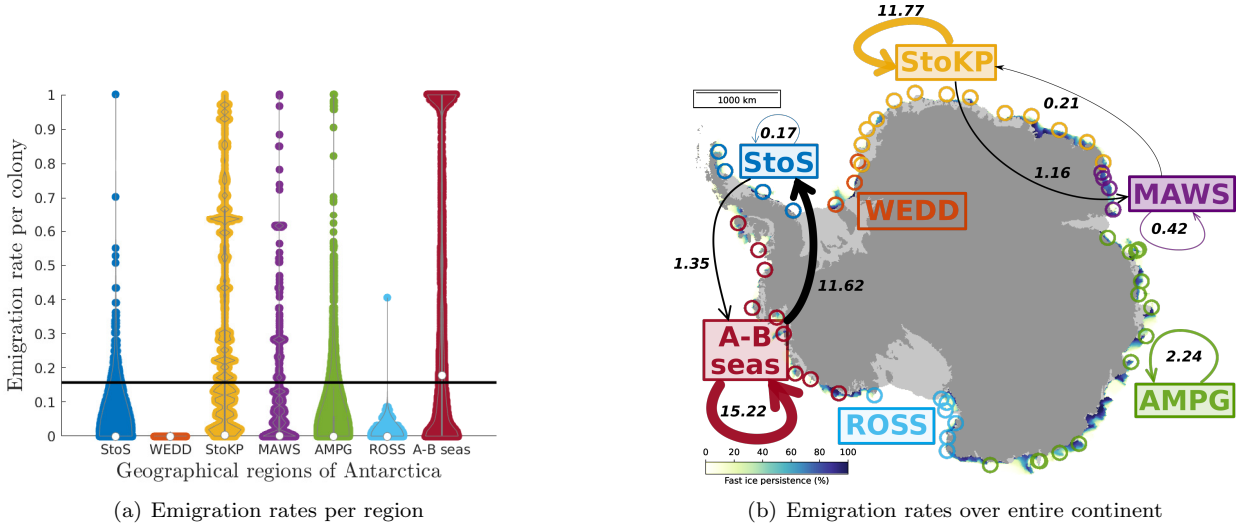


Figure 2: Emigration rates per year per colony (panel (a)) and between and among the seven regions of Antarctica (panel (b)), from 2009 to 2014: from Smith to Snowhill Island in the Wedell sea (StoS), Weddell sea (Gould Bay to Halley Bay colonies) (WEDD), from Stancomb to Kloa point (StoK), Mawson Bay (Fold Island to Cape Darnley colonies) (MAWS), from Amanda Bay to Pointe Geologie colonies (AMPG), the Ross sea (Cape Washington and Cape Crozier) (ROSS) and Admunsen and Bellingshausen seas (Ledda bay to Rothschild Island) (A-B seas). In panel (a), white dots corresponds to the median of the posterior distributions of the emigration rates per colony for each region and the black line is the mean emigration rate for the entire Antarctic continent (0.157). In panel (b), the map background shows the distribution of fast ice persistence, expressed as a percentage of time covering each unit (square of 6.25km) from March 2000 to March 2018 (the map is extracted from [29]).

coherent with the philopatric behaviour of those species. For example, greater flamingos are long lived and philopatric like Emperor penguins. On average they show low emigration rates but when conditions for breeding are unfavorable at their colony (typically low water levels) they move en masse to another breeding site (see [44, 45]).

In addition, our framework focuses on emigration rates at the population level. However, in many vertebrate species, especially in seabirds, juvenile dispersal is higher than adult dispersal [46]. For Emperor penguins, it is likely that massive emigration events consist mainly of adults in some regions while the low background levels of emigration is likely dominated by juvenile dispersal. Further work should include this age structure into the dispersal demographic model in order to disentangle dispersal rates among adults from those among juveniles. However, that would require understanding the detailed mechanisms of density dependence on those two age classes, which are unknown for Emperor penguins [47].

Figure 2(b) shows that Emperor penguins mainly move towards nearby colonies of the same regions with an average rate per year that varies among regions: 15% in colonies of the Amundsen and Bellington seas (A-B seas) to 0.17% in colonies from Smith to Snowhill Island in the Wedell sea regions (StoS) (see Fig.S13 for more details). However, massive emigration is also likely to occur between different regions especially between the A-B seas regions and StoS regions (11% to 1.35%) and colonies from Stancomb to Kloa Point (StoKP) and Mawson bay (MAWS) (1.16% to 0.21%).

## Dispersal behaviors

Our models selection shows that Emperor penguins are likely to depart from colonies with poor habitat (i.e. with negative intrinsic population growth; informed emigration), but settle randomly into another colony

(random establishment), a behavior referred as semi-informed dispersal behavior (Table 1 and SI1.G for details of the computations). In our meta-population model a poor habitat reflects poor environmental conditions (as measured by sea ice concentrations at large spatial scales) that can not support persistent populations, and/or a colony that has surpassed its carrying capacity.

A random establishment uncorrelated to habitat quality suggest that the individuals have a relatively poor knowledge of alternative breeding sites. Indeed, during the non-breeding season, individuals molt and then need to acquire resources to regain the necessary body condition to breed (by the end of molt penguins have lost nearly half their body mass). There is little time left to prospect for other breeding sites, and it might be extremely difficult to assess the habitat quality in such a variable environment. Therefore, individuals may settle randomly in a new colony before making the decision whether to stay or leave.

An informed emigration behavior in Emperor penguin is supported by the pattern of colony movement from analysis of VHR satellite imagery. Indeed, massive colony movements have been suggested to occur in poor habitat quality, specifically, unstable ice habitat caused by poorly-formed fast-ice or late formation and early loss of the fast-ice habitat on which a colony is located. For example, late formation of sea ice in the autumn may result in the relocation of a colony onto icebergs or ice shelves, which often results in a longer commute to the nearest foraging grounds with important consequences for breeding productivity [43]. The calving of glaciers, ice tongues and ice shelves can also compromise the stability of the ice habitat, forcing colonies to relocate [23]. Other atmospheric or oceanographic perturbations affecting the ice habitat have also rendered sites uninhabitable (e.g. Halley Bay [26]). Species relying on personal and social information for dispersal decisions may cope better with future climate change because individuals will escape poor local conditions and settle in higher quality patches, thus increasing their fitness [48, 19, 20].

## Potential drivers of dispersal rates

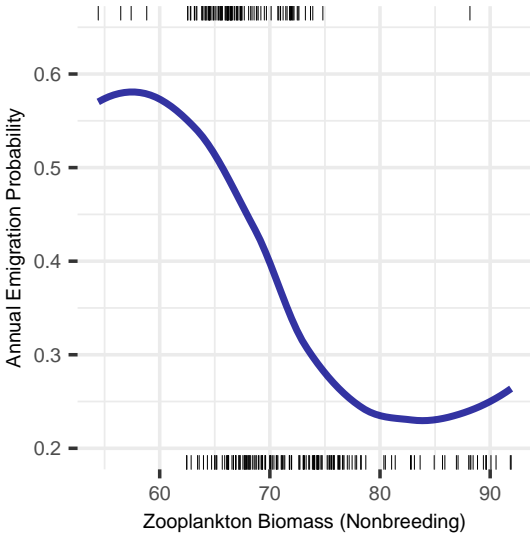
To quantify if fast ice and other environmental and demographic drivers affect emigration rates, we used independent (not included in our meta-population model) environmental factors to model annual presence of emigration, which include fast ice data [49], other ocean-sea-ice variables and food web dynamics variables produced from a forced ocean-sea-ice (FOSI) configuration of CESM2 (see Methods for details). Contrary to our expectations, we did not detect an important effect of fast-ice variables on the probability of emigration using a random forest analysis. Instead, we found that zooplankton biomass (Fig. 3(a)) and net primary productivity (Fig. 3(b)) during the non-breeding period (January to March) are the two most important variables affecting annual emigration probability (probability of observing non-zero emigration rate in a given year and colony). These variables represent the lower food web dynamics of the Antarctic ecosystem and are proxies for the food sources of the Emperor penguins. Both of these factors are negatively related with annual emigration probability, suggesting that Emperor penguins are less likely to emigrate from their colony if food sources are abundant prior to breeding.

We also used independent demographic factors from VHR satellite imagery alongside environmental variables to model colony-level average emigration probability. Specifically, we expected that smaller and declining colonies, which blink more often would exhibit larger emigration rates. We found little support of an impact of the frequency of blinking and population growth rate on the average emigration probability, but larger colonies have lower overall emigration probability (Fig. 3(c)). Despite, large uncertainties, the size of the colonies without emigration are almost twice as large than colonies with emigration on average (3,741 penguins versus 1,880 penguins for emigration) (see Fig.SI5). However, environmental factors (zooplankton biomass and net primary productivity) remain the most important variables (Fig. 3(d)). Although more work is needed to elucidate the proximate factors of suitable habitat and emigration rates of Emperor penguins, those results converge to suggest that massive emigration may occur for habitat conditions with low food availability that cannot sustain large populations.

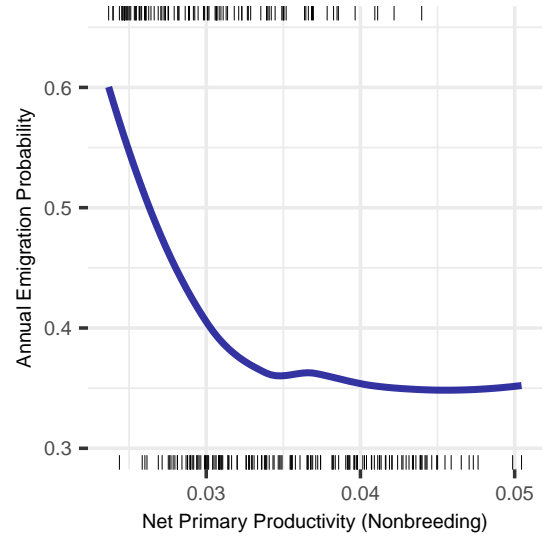
## Consequences of dispersal behaviors

Previously, [27] have shown that high emigration rates and long distance dispersal accelerate the projected global population decline of Emperor penguins and decrease the global population size by 65% by 2100

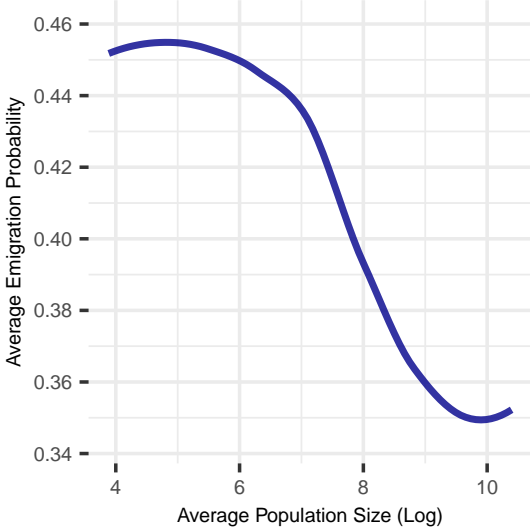




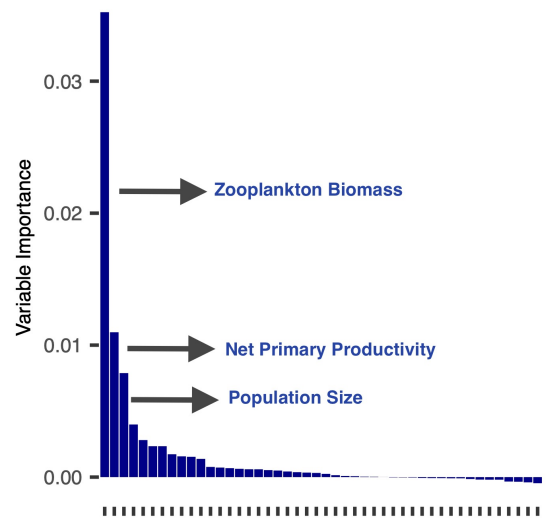
(a) Partial dependence plot for zooplankton Biomass



(b) Partial dependence plot for net primary productivity



(c) Partial dependence plot for population size



(d) Conditional variable importance

Figure 3: Impact of environmental and demographic factors on emigration rates of Emperor penguins. Zooplankton biomass (a) and net primary productivity (b) during the non-breeding period (January to March) have a negative effect on annual emigration probability (probability of observing non-zero median emigration rate in a given year and colony). Lines at the top and bottom of the plots show the presence (median emigration rate $\neq$ 0) and absence (median emigration rate=0) of emigration, respectively, in a given year and colony. Colony size (c) have negative effects on average emigration probability (proportion of years with non-zero median emigration rates between 2009 and 2013 in a colony). (d) Conditional variable importance scores of random forests modelling average emigration probability. Only the top 3 variables are shown.

compared to a scenario without dispersal. However, here we show that high emigration rates and long distance dispersal are unlikely for Emperor penguins. We found a relatively short distance of dispersal and low average emigration rates that will slightly increase the global population relative to a scenario without dispersal (see Fig.SI6(a)-(b)). However, the impact of dispersal behavior, distance and emigration rate on the future global population size is relatively small compared with the impact of climate change mitigation [50] (see Fig.SI6(c)). At the end of the century, there will be no suitable habitat if greenhouse gas emissions continue their current course, resulting in large global population declines, regardless of dispersal processes [50]. The most important action to ensure refugia for the Emperor penguin and halt dramatic global population declines is to limit temperature increases to well below  $2^{\circ}\text{C}$  [21].

## Conclusion

In conclusion, using a novel likelihood for a model linking genetic data to meta-population dynamics, we revealed and quantified the heretofore unknown dispersal behaviors, rates and range of Emperor penguins. Interestingly, we show that Emperor penguins are more likely to depart from poor habitat (informed emigration) with low food resources that can not sustain large colonies. Although this informed behavior suggests that Emperor penguins can assess the quality of their breeding site, our study stresses that individuals have a relative poor knowledge of alternative breeding site because they more likely establish randomly in a new colony. Unraveling those dispersal processes will also reduce uncertainties in future population projections of Emperor penguins necessary for ongoing conservation and management actions. This modeling approach can be applied to any species and data-poor system to reveal dispersal processes, and promises to transform our understanding of dispersal ranges, rates and behaviors beyond the charismatic Emperor penguin to better predict how species will cope with future global changes.

**Acknowledgment** S.J. and J.G. acknowledges support from Mission Blue, S.J. from NSF OPP 2037561 and NASA 80NSSC20K1289. We also acknowledge the Institut Paul Emile Victor (Project IPEV 109 OrnithoEco) and Terres Australes et Antarctiques Françaises for supporting the long-term program on Emperor penguins at Pointe Géologie. This project received grant funding from the Australian Government as part of the Antarctic Science Collaboration Initiative program.

## References

- [1] A Hastings. Can spatial variation alone lead to selection for dispersal? *Theor. Popul. Biol.*, 24:244–251, 1983.
- [2] M Slatkin. Gene flow and the geographic structure of natural populations. *Science*, 236(4803):787–792, 1987.
- [3] M. W. Cadotte. Dispersal and species diversity: A meta-analysis. *Am. Nat.*, 167(6):913–924, 2006.
- [4] Lionel Roques, Samuel Soubeyrand, and Jérôme Rousselet. A statistical-reaction–diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology*, 274(1):43–51, 2011.
- [5] P Turchin. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer Associates, 1998.
- [6] R. Nathan, G. Perry, J. T. Cronin, A. E. Strand, and M. L. Cain. Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273, 2003.
- [7] Etienne K Klein, A Bontemps, and Sylvie Oddou-Muratorio. Seed dispersal kernels estimated from genotypes of established seedlings: does density-dependent mortality matter? *Methods in Ecology and Evolution*, 4, 2013.

- [8] Sewall Wright. Isolation by distance. *Genetics*, 28(2):114–138, 1943.
- [9] T Broquet, N Ray, E Petit, J M Fryxell, and F Burel. Genetic isolation by distance and landscape connectivity in the american marten (*Martes americana*). *Landscape Ecol*, 21:877–889, 2006.
- [10] J A Stamps. Habitat selection by dispersers: proximate and ultimate approaches. In J Clobert, E Danchin, A Dhondt, and J Nichols, editors, *Dispersal*. New York: Oxford University Press, 2001.
- [11] J Clobert, J Le Galliard, J Cote, S Meylan, and M Massot. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol. Lett.*, 12(3):197–209, 2009.
- [12] Brad H. McRae. Isolation by resistance. *Evolution*, 60(8):1551–1561, 2006.
- [13] Tabitha A. Graves, Richard B. Chandler, J. Andrew Royle, Paul Beier, and Katherine C Kendall. Estimating landscape resistance to dispersal. *Landscape Ecology*, 29:1201–1211, 2014.
- [14] Tabitha A. Graves, Paul Beier, and J. Andrew Royle. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, 22(15):3888–3903, 2013.
- [15] W H Lowe and F W Allendorf. What can genetics tell us about population connectivity? *Molecular Ecology*, 19(15):3038–3051, 2010.
- [16] L. Roques, E. Walker, P. Franck, S. Soubeyrand, and E. K. Klein. Using genetic data to estimate diffusion rates in heterogeneous landscapes. *Journal of Mathematical Biology*, 73(2):397–422, 2016.
- [17] T.R.E. Southwood and P.A. Henderson. *Ecological Methods*. Wiley, 2009.
- [18] Toby A. Patterson, Len Thomas, Chris Wilcox, Otso Ovaskainen, and Jason Matthiopoulos. State-space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2):87–94, 2008.
- [19] B Doligez, E Danchin, and J Clobert. Public information and breeding habitat selection in a wild bird population. *Science*, 297(5584):1168–1170, 2002.
- [20] A Ponchon, R Garnier, D Grémillet, and T Boulinier. Predicting population response to environmental change: the importance of considering informed dispersal strategies in spatially structured population models. *Divers. Distrib.*, 21:88–100, 2015.
- [21] Stephanie Jenouvrier, Judy Che-Castaldo, Shaye Wolf, Marika Holland, Sara Labrousse, Michelle LaRue, Barbara Wienecke, Peter Fretwell, Christophe Barbraud, Noah Greenwald, et al. The call of the emperor penguin: Legal responses to species threatened by climate change. *Global change biology*, 27(20):5008–5029, 2021.
- [22] J B Thiebot, A Lescroël, C Barbraud, and C A Bost. Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarct. Sci.*, 25(4):536–544, 2013.
- [23] Michelle A LaRue, Gerald Kooyman, Heather J Lynch, and Peter Fretwell. Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography*, 38(2):114–120, 2015.
- [24] J. L. Younger, G. V. Clucas, D. Kao, A. D. Rogers, K. Gharbi, T. Hart, and K. J. Miller. The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Molecular Ecology*, 26(15):3883–3897, 2017.
- [25] Peter T Fretwell and Philip N Trathan. Discovery of new colonies by sentinel2 reveals good and bad news for emperor penguins. *Remote Sensing in Ecology and Conservation*, 7(2):139–153, 2021.

- [26] Peter T Fretwell and Philip N Trathan. Emperors on thin ice: Three years of breeding failure at halley bay. *Antarctic Science*, 31(3):133–138, 2019.
- [27] S Jenouvrier, J Garnier, F Patout, and L Desvillettes. Influence of dispersal processes on the global dynamics of emperor penguin, a species threatened by climate change. *Biol. Conserv.*, 212:63 – 73, 2017.
- [28] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 2007.
- [29] A. D. Fraser, R. A. Massom, M. S. Handcock, P. Reid, K. I. Ohshima, M. N. Raphael, J. Cartwright, A. R. Klekociuk, Z. Wang, and R. Porter-Smith. Eighteen-year record of circum-antarctic landfast-sea-ice distribution allows detailed baseline characterisation and reveals trends and variability. *The Cryosphere*, 15(11):5061–5077, 2021.
- [30] Matthew C. Long, J. Keith Moore, Keith Lindsay, Michael Levy, Scott C. Doney, Jessica Y. Luo, Kristen M. Krumhardt, Robert T. Letscher, Maxwell Grover, and Zephyr T. Sylvester. Simulations With the Marine Biogeochemistry Library (MARBL). *Journal of Advances in Modeling Earth Systems*, 13(12), 2021.
- [31] J. K Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
- [32] L Roques. *Modèles de réaction-diffusion pour l'écologie spatiale*. Editions Quae, 2013.
- [33] E. C. Anderson and E. A. Thompson. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, 160(3):1217–1229, 2002.
- [34] G L Kooyman, D Siniff, I Stirling, and J Bengtson. Moulting habitat, pre-and post-moulting diet and post-moulting travel of Ross Sea emperor penguins. *Mar. Ecol. Prog. Ser.*, 267:281–290, 2004.
- [35] K T Goetz, B I McDonald, and G L Kooyman. Habitat preference and dive behavior of non-breeding emperor penguins in the eastern ross sea, antarctica. *Mar Ecol Prog Ser*, 593:155–171, 2018.
- [36] N. J. Aebischer and J. C. Coulson. Survival of the kittiwake in relation to sex, year, breeding experience and position in the colony. *Journal of Animal Ecology*, 59(3):1063–1071, 1990.
- [37] J. Michael Reed, Thierry Boulinier, Etienne Danchin, and Lewis W. Oring. *Informed Dispersal*, pages 189–259. Springer US, 1999.
- [38] S Dean Kildaw, David B Irons, David R Nysewander, and C Loren Buck. Formation and growth of new seabird colonies: the significance of habitat quality. *Marine Ornithology*, 33:49–58, 2005.
- [39] J Matthiopoulos, J Harwood, and L Thomas. Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology*, 74(4):716–727, 2005.
- [40] J L Younger, G V Clucas, G Kooyman, B Wienecke, A D Rogers, and *et al.* Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Global Change Biol.*, 21(6):2215–2226, 2015.
- [41] J. Prevost. *Expéditions polaires françaises*, chapter Ecologie du manchot empereur. Hermann Press, Paris, France, 1961.
- [42] J L Mougin and M Van Beveren. Structure et dynamique de la population de manchots empereurs Aptenodytes Forsteri de la colonie de l'archipel de Pointe Géologie, terre adélie. *C.R. Acad. Sci.*, 289(2):157–160, 1979.

- [43] P. T. Fretwell, P. N. Trathan, B. Wienecke, and G. L. Kooyman. Emperor penguins breeding on iceshelves. *PLoS ONE*, 9:e85285, 2014.
- [44] Özge Balkiz, Arnaud Béchet, Lauriane Rouan, R Choquet, Christophe Germain, Juan A. Amat, Manuel Rendón-Martos, Nicola Baccetti, Sergio Nissardi, Uygur Özesmi, and Roger Pradel. Experience-dependent natal philopatry of breeding greater flamingos. *Journal of Animal Ecology*, 79(5):1045–1056, 2010.
- [45] Alan Johnson and Frank Cézilly. *The greater flamingo*. A & C Black, 2007.
- [46] Jean Clobert, Rolf Anker Ims, and François Rousset. Causes, mechanisms and consequences of dispersal. In *Ecology, genetics and evolution of metapopulations*, pages 307–335. Elsevier, 2004.
- [47] Stéphanie Jenouvrier, Marika Holland, Julianne Stroeve, Christophe Barbraud, Henri Weimerskirch, Mark Serreze, and Hal Caswell. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology*, 18:2756–2770, Jul 2012.
- [48] T Boulinier and E Danchin. The use of conspecific reproductive success for breeding patch selection in terrestrial migratory species. *Evol. Ecol.*, 11(5):505–517, 1997.
- [49] Alexander D Fraser, Robert A Massom, Kay I Ohshima, Sascha Willmes, Peter J Kappes, Jessica Cartwright, and Richard Porter-Smith. High-resolution mapping of circum-antarctic landfast sea ice distribution, 2000–2018. *Earth System Science Data*, 12(4):2987–2999, 2020.
- [50] Stéphanie Jenouvrier, Marika Holland, David Iles, Sara Labrousse, Laura Landrum, Jimmy Garnier, Hal Caswell, Henri Weimerskirch, Michelle LaRue, Rubao Ji, et al. The paris agreement objectives will likely halt future declines of emperor penguins. *Global change biology*, 26(3):1170–1184, 2020.
- [51] G. Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2), February 2020.
- [52] Dries Debeer and Carolin Strobl. Conditional permutation importance revisited. *BMC Bioinformatics*, 21(1), 2020.
- [53] L Roques, J Garnier, F Hamel, and E K Klein. Allee effect promotes diversity in traveling waves of colonization. *Proc Natl Acad Sci USA*, 109(23):8828–8833, 2012.
- [54] J Forcada and P. N. Trathan. Penguin responses to climate change in the southern ocean. *Global Change Biology*, 15(7):1618–1630, 2009.
- [55] D. Paetkau, W. Calvert, I. Stirling, and C. Strobeck. Microsatellite analysis of population structure in canadian polar bears. *Mol. Ecol.*, 4(3):347–354, 1995.
- [56] G Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [57] D J Spiegelhalter, N G Best, B P Carlin, and A Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- [58] A Gelman, J Carlin, H Stern, D Dunson, A Vehtari, and D Rubin. *Bayesian Data Analysis*. Chapman & Hall / CRC, 2003.
- [59] T Ando. Predictive bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31:13 – 38, 2011.

# Materials and methods

## Computation of the likelihood function and parameter estimations

To understand the dispersal behavior of Emperor penguins, we use genetic data collected from 1993 to 2013 in eight colonies around Antarctica (see [24] and SI1.A for details on the genetic structure of the Emperor penguins). In each sampled colonies, we have captured from 10 to 16 individuals that we have genotyped at the same  $\Lambda$  loci (with possibly some allele frequencies equal to 0). Since Emperor penguins are diploids, we describe their genotypes by  $\mathcal{G} = \{(a_\lambda^1, a_\lambda^2)\}_{\lambda=1, \dots, \Lambda}$ . Since we use single nucleotide polymorphisms (SNPs) there are two alleles which corresponds to the two possible nucleotide variations of the DNA.

The population of Emperor penguins comprises 4 genetic clusters whose allele frequencies  $\mathcal{F}_{r\lambda} = (p_{r\lambda a})_{a=1, \dots, A_\lambda}$  can be assumed constant over the sampling interval because it represents less than two generations for Emperor penguins (which is 16 years).

We use classical genetic assignment approaches [31] and dispersal analysis from spatially sampled data, to compute the genotype likelihood. More precisely, the probability to sample the genotype  $\mathcal{G}_{i,\tau}$  in colony  $\tau$  at time  $t_i$  is given by the law of total probability

$$\mathbb{P}(\mathcal{G}_{i,\tau}) = \sum_{r=1}^4 \mathbb{P}(\mathcal{G}_{i,\tau} | \text{indiv. } i \text{ belongs to cluster } r) \times \mathbb{P}(\text{indiv. } i \text{ belongs to cluster } r) \quad (2)$$

Using the linkage equilibrium among loci, the conditional probability  $\gamma_{i,\tau}^r$  for the genotype  $\mathcal{G}_{i,\tau}$  is:

$$\gamma_{i,\tau}^r = \mathbb{P}(\mathcal{G}_{i,\tau} | \text{indiv. } i \text{ belongs to } r) = 2^{k_i} \prod_{\lambda=1}^{\Lambda} p_{r\lambda a^1} p_{r\lambda a^2} \quad (3)$$

where  $k_i$  is the number of heterozygous loci in  $\mathcal{G}_{i,\tau}$  (see details in SI1.C). Using probability approximations, we can compute the probability that the genotype  $i$  observed in colony  $\tau$  belongs to cluster  $r$ :

$$\mathbb{P}(\text{indiv. } i \text{ belongs to } r) = \mu_\tau^r(t_i) \quad (4)$$

where  $\mu_\tau^r(t_i)$  is the frequency of the genetic cluster  $r$  in the colony  $\tau$  at time  $t_i$ . This frequency is computed from a genetic-demographic model based on meta-population model constructed by [27], which describes the different dispersal behaviors through some parameters  $\Theta$ , which comprise the mean distance dispersal  $d$  and parameters  $\mathbf{p}_m$  quantifying the emigration rate (see SI1.B-D for details).

Finally, we obtain the likelihood function defined by (1). For each dispersal behavior, the parameters have been obtained by minimizing the logarithm of the inverse Likelihood, that is  $-\log(\mathcal{L}(\Theta))$ . The minimization was performed using a simple Bayesian method, where the prior of the parameters are given in SI1.F.

## Comparison of dispersal behaviors

In our modeling framework, we have considered three dispersal behaviors based on a combination of

- an informed emigration: individuals only emigrate from poor quality breeding sites when the habitat quality is not viable (i.e. negative intrinsic population growth);
- a random emigration: individuals leave the colony regardless of the habitat quality;
- an informed establishment: individuals select the most suitable habitats (i.e. maximize intrinsic population growth) within their dispersal range;
- random establishment: movements are undirected with respect to habitat quality.

Specifically, the random dispersal behavior (R) is a random emigration and establishment; the semi-informed dispersal behavior (SI) is an informed emigration but random establishment; and the informed dispersal behavior (I) is an informed emigration and establishment. To assess which behavior is more likely to occur, we have performed a model selection based on four criteria and our likelihood function (see SI1.G for details).

## Comparison with demographic and environmental factors

We used three demographic factors independent from our meta-population analysis: the size of colony, the growth rate per colony and the frequency of blinking that corresponds to the relative number of year a colony disappear over a period of 10 years, from 2009 to 2018. These demographic factors were calculated from unpublished data of colony presence and population counts of Emperor penguins from high-resolution satellite imagery (Larue et al. submitted). Those estimates represent the portion of the colony in attendance on the fast ice every year and thus available for surveying through aerial counts or satellite images during the chick-rearing season (Larue et al. submitted).

We also considered different environmental variables around each colony, which are characterized by the following descriptors: the fast ice area around a colony ( $m^2$ ), the distance between the colony and the nearest edge of fast ice ( $m$ ), the emergence and breaking date of the fast ice, zooplankton biomass ( $\text{mmol } C/m^2$ ), mixed layer depth ( $m$ ), net primary production ( $\text{mmol } C/m^2/day$ ), upper ocean temperature (top 10m,  $^{\circ}C$ ), and surface wind ( $m/s$ ). We also used new fast ice variables from a recent analysis of Emperor penguin habitat (Labrousse, in prep) describing the persistence and the magnitude of fast ice annual cycle. Fast ice variables were computed using continuous, high-spatio-temporal resolution time series of circum-Antarctic fast-ice extent from [49]. Other reanalysis environmental products were computed in a forced ocean sea ice (FOSI) configuration of the Community Earth System Model (CESM2,  $1^{\circ}$  resolution) [51]. When applicable, we calculated the average value of each variable for a given breeding period, specifically, non-breeding (January to March), laying (April and May), incubation (June and July), and chick-rearing (August to December). Additionally, we used the average within a 100 km buffer around each colony when calculating fast ice variables, 800 km buffer for other variables during the non-breeding period, and a 650 km buffer for other periods. First, we split for every year the colonies into two categories: "No emigration" colonies when their median emigration rate is 0 during this year and "Emigration" colonies when their median emigration rate is positive. We modelled the proportion of emigration years across colonies with both demographic variables and environmental variables using conditional random forests [28], whose best model had an  $R^2$  value of 0.55. Second, we refine our analysis using environmental variables only by modeling the annual emigration probability using the same framework conditional random forests. The best model had an area under the receiving operator characteristic curve (AUC) of 0.82, which was calculated with a 10-fold cross validation, signifying good classification performance. Variable importance scores were calculated with conditional permutation importance [52] (see Fig.SI4).

# Supporting Information

## A Mathematical models and data

### A.1 Genetic data

In our paper, we use the genetic data from Younger et al. [24]. Specifically, we use 4,596 genome-wide single nucleotide polymorphisms (SNPs), characterized in 110 individuals (10 to 16 per colony) from eight colonies around Antarctica (Ammanda Bay, Pointe Geologie, Fold Island, Auster, Cape Roget, Cape Washington, Gould Bay and Halley Bay). This analysis have shown that the global population of emperor penguin comprises 4 genetic clusters.

Moreover, Younger et al. have linked these genetic cluster to specific geographic regions of Antarctica. Hereafter, we name the four genetic cluster according to the four geographic regions that they have characterized:

1. Weddell sea (Gould Bay to Halley Bay colonies) (WEDD),
2. Mawson Bay (Fold Island to Cape Darnley colonies) (MAWS),
3. Amanda Bay to Pointe Geologie colonies (AMPG) and
4. Ross sea (Cape Washington and Cape Crozier colonies) (ROSS).

To characterize the structure at a circum-Antarctica scale, we combine Davis Bay and Mertz Glacier colonies with the AMPG cluster, and Cape Colbeck and Rupert Coast with the ROSS cluster.

Out of theses four geographic regions, we further include three geographic regions for which no genetic data is available:

5. from Smith to Snowhill Island in the Wedell sea colonies (StoS),
6. from Stancomb to Kloa point colonies (StoK) and
7. Ledda bay to Rotschild colonies (Admunsen and Bellingshausen seas, A-B seas).

We thus obtain 7 different geographical regions and 4 genetic clusters.

### A.2 Demographic model for the emperor penguin

The population dynamics of emperor penguin is described using the meta-population model developed by [27]. It projects the population vector  $\mathbf{n}$ —comprising the population size  $n_i$  in each colonies  $i$ —from time  $t$  to  $t + 1$ . We write

$$\mathbf{n}(t + 1) = \mathbf{D}[t, \mathbf{n}(t)]\mathbf{F}[t, \mathbf{n}(t)]\mathbf{n}(t) \quad (5)$$

to indicate that the projection interval is divided into two main phases of possibly different duration: the motionless reproduction phase ( $\mathbf{F}$ ) followed by the dispersal phase ( $\mathbf{D}$ ). The projection matrices  $\mathbf{D}$  and  $\mathbf{F}$  depend on both the current population density  $\mathbf{n}(t)$  and time  $t$  because the habitat conditions vary among patches and over time.

#### The reproduction matrix $\mathbf{F}$

The reproduction matrix  $\mathbf{F}$  is constructed using the Ricker model. It includes the intrinsic population growth rate  $\mathbf{r}(t)$  that may vary in time because it depends on sea ice concentrations (SIC),  $\mathbf{r}(SIC_t)$ . For each projection interval  $t$ , we calculate the growth rate of each colony  $r_i(t)$  using the median of the stochastic population size projected by a sea-ice dependent population model without density dependence to account for uncertainties related to both climate and demographic processes [27]. This Ricker model also includes the carrying capacity of the colonies  $\mathbf{K}$  which are assumed to be constant over time.



## The dispersal phase $\mathbf{D}$

The dispersal phase ( $\mathbf{D}$ ) is decomposed into three stages: (1) emigrating from the resident patch, (2) searching for new patch among other patches with an average dispersal distance  $d$  (transfer), and (3) settling in a new patch. During this dispersal event, individuals may select the habitat with highest quality (informed search) or settle in a random habitat (random search). The dispersal projection matrix  $\mathbf{D}$  is thus decomposed as follows

$$\mathbf{D} := \mathbf{S}[t; d] \mathbf{M}[t, \mathbf{n}(t); \mathbf{p}_m] \quad (6)$$

to indicate that matrices for searching behavior,  $\mathbf{S}$ , and emigration,  $\mathbf{M}$ , depend on the population size ( $\mathbf{n}$ ) as well as the environmental conditions which depend on time  $t$  and the coastal distance between the colonies. We contrast three dispersal behaviors: (1) an informed dispersal behavior where individuals leave poor habitat colony (informed emigration) and perform informed search, (2) a random dispersal behavior where individuals randomly leave and randomly search for a new colony and (3) a semi-informed dispersal behavior where individuals leave only poor habitat colony (informed emigration) but settle randomly among reachable colonies (randomly search).

**Migration matrix  $\mathbf{M}$ .** For the **random emigration behavior**, the emigration rate  $m_i$  for each colony  $i$  depends only on the proportion  $p_{m_i}$  of individuals that leave the colony. Thus the migration matrix is a diagonal matrix with diagonal coefficient  $m_i = p_{m_i}$ .

For the **informed emigration behavior**, the emigration rate depends on the quality of the habitat, measured through the realized population growth  $r_i^*(t)$  and a sensitivity parameter  $\mathbf{p}_m = (p_{m_1}, \dots, p_{m_7})$  measuring the intensity of the emigration. The realized population growth rate  $r_i^*(t)$  is a function of both the intrinsic growth rate  $r_i(t)$ , and the carrying capacity of the colonies  $K_i$ :

$$r_i^*(t) = \begin{cases} (1 + r_i(t)) \exp\left(1 - \frac{N_i(t)}{K_i}\right) - 1 & \text{if } r_i(t) > 0 \\ r_i(t) & \text{if } r_i(t) \leq 0 \end{cases}$$

We assume that the emigration rate  $m$  increases linearly from  $m = 0$  at  $r^* \geq 0$  to  $m = 1$  at critical value  $r_c^* < 0$ . Thus, a critical threshold  $r_c^*$  close to 0 corresponds to high migration, while a larger threshold reflects low migration. Here, we will estimate the emigration rate parameter  $\mathbf{p}_m = (p_{m_1}, \dots, p_{m_7})$ , which quantifies the critical value in each region  $i$ :

$$r_{i,c}^* = (1 - p_{m_i})r_m^*$$

where  $r_m^*$  is the lowest intrinsic growth rate. Hence, the emigration matrix  $\mathbf{M}$  only depends on the ratio  $r_i^*(t)/(1 - p_{m_i})r_m^*$

$$m_i \left[ \frac{r_i^*(t)}{(1 - p_{m_i})r_m^*} \right] = \begin{cases} 1 & \text{if } r_i^*(t) < (1 - p_{m_i})r_m^* \\ 1 - \frac{r_i^*(t)}{(1 - p_{m_i})r_m^*} & \text{if } (1 - p_{m_i})r_m^* \leq r_i^*(t) \leq 0 \\ 0 & \text{if } r_i^*(t) > 0 \end{cases} \quad (7)$$

**Searching matrix  $\mathbf{S}$ .** Once individuals have left their colonies, we assume that they search for a new colony. The searching matrices  $\mathbf{S}[\mathbf{x}]$  is

$$\mathbf{S}_{ij}[\mathbf{x}, d] := \mathcal{S}(j|i, \mathbf{x}(t), d), \quad \text{for } j \neq i \quad \text{and} \quad \mathbf{S}_{ii}[\mathbf{x}, d] := - \sum_{j \neq i} \mathbf{S}_{ij}[\mathbf{x}, d],$$

indicating that the probability of settlement in a colony  $j$  depends on leaving colony  $i$ , the characteristics of the habitat in the colony  $j$  ( $\mathbf{x}(t)$ ), and the dispersal ability of the individuals  $d$ .

**For the random search** individuals can move randomly across landscape according to a dispersal kernel  $k(x)$  which describes the probability of traveling a distance  $x$ . This probability distribution can take various

forms according to the dispersal ability of the species. In our simulation, we use a uniform kernel because emperor penguins have the ability to cover incredible distances, thus all colonies are potentially connected:

$$k_{unif}(x) := \frac{1}{d} \mathbf{1}_{[0,d]}(x), \quad \text{for all } x \in [0, +\infty),$$

where  $d$  represents the mean distance dispersal of the species and  $\mathbf{1}_{[0,d]}(x)$  is the characteristic function of the interval  $[0, d]$ . Thus under the random search, the probability  $S_R(j|i, \mathbf{x}(t), d)$  of moving to colony  $j$  given that individual left its resident colony  $i$  at time  $t$  is defined by

$$S_R(j|i, \mathbf{x}(t), d) := \frac{k(\text{dist}(i, j))}{\sum_{j \neq i} k(\text{dist}(i, j))}. \quad (8)$$

where  $\text{dist}(i, j)$  corresponds to the landscape topography, specifically the coastal distance between colonies in our case study. With the random search, individuals may settle in a new colony of lower quality than their resident colony.

**For the informed search** individuals select their habitat using the fitness of conspecifics as a source of public information on colony quality. Thus the colony quality is described through the realized growth rate  $\mathbf{r}^*(t)$ . The probability  $S_I(j|i, \mathbf{x}(t), d)$  of moving to a new colony  $j$  given that individual left its resident colony  $i$  at time  $t$  is

$$S_I(j|i, \mathbf{x}(t), d) := \begin{cases} 1 & \text{if } r_j^*(t) = \max(r_k^*(t) | \text{dist}(i, k) \leq d), \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where  $d$  represents the mean distance dispersal of the species. Unlike individuals using the random search, individuals will move to the most favorable colony they can reach, i.e.  $\max r_k^*(t)$  given that  $\text{dist}(i, k) \leq d$ .

### A.3 A genetic-demographic statistical framework to estimate dispersal parameters

#### Number of individuals available for genetic sampling for each colony

The sampling of individuals at a given time  $t$  at a colony  $\tau$  is random over the individuals observed at the colony. Thus the expected number of individuals that can be potentially captured  $C_\tau(t)$  is proportional to the number of individuals alive in the colony  $\tau$  at time  $t$ :

$$C_\tau(t) = \beta_\tau n_\tau(t),$$

where  $\beta_\tau$  is the capture rate at colony  $\tau$ .

#### Number of individuals of each cluster within a colony

Within a colony, the population is structured into different neutral genetic cluster  $r$ , with the number of individuals within colony  $i$  from cluster  $r$  denoted by  $n_i^r(t)$ . The total number of individuals that belongs to the cluster  $r$  across colonies is  $\mathbf{n}^r(t)$ . Since we are looking at neutral set of loci, all individuals are supposed to share the same dispersal and reproduction characteristics independently of their clusters. Thus the number of individuals  $\mathbf{n}^r(t)$  satisfies [53]:

$$\mathbf{n}^r(t+1) = \mathbf{D}[t, \mathbf{n}(t)] \mathbf{F}[t, \mathbf{n}(t)] \mathbf{n}^r(t) \quad (10)$$

and initially,

$$n_i^r(0) = \mu_i^r n_i(0), \quad \text{for all } i \in \{1, \dots, 54\}, \quad (11)$$

where  $\mu_i^r$  is the initial proportion of clusters within a colony ( $\sum_{r=1}^R \mu_i^r = 1$  for all  $i$ ).

For the emperor penguin case study, the metapopulation  $\mathbf{n}$  is described by four observed cluster  $R = 4$ . The genetic and demographic dynamics are linked by:

$$\mathbf{n}(t) = \sum_{r=1}^R \mathbf{n}^r(t). \quad (12)$$

Thus, the expected number of individuals  $C_\tau^r(t)$  belonging to a cluster  $r$  that can be sampled at colony  $\tau$  at time  $t$  is given by

$$C_\tau^r(t) = \beta_\tau n_\tau^r(t). \quad (13)$$

## A.4 Parameters and data

We aim at estimating two dispersion parameters: the mean distance dispersal  $d$  and the parameter  $\mathbf{p}_m$  which determines the emigration rate. Auxiliary unknown parameters are the initial proportion of each cluster within a colony  $\mu_i^r$ . The parameter  $\boldsymbol{\mu}^r$  is unknown because we do not have an exhaustive sampling of the region.

We expect to recover the unknown parameters from the genetic data collected from 1993 to 2013 in some colony around Antarctica. The genetic clusters are at Hardy-Weinberger equilibrium inside the metapopulation (see [24] for more details). Thus, we assume that each colony in the metapopulation is composed of all the 4 observed clusters (but note that the proportion or frequency of a cluster could be null).

The location of the sampled colonies is known in our analysis. Moreover, since the sampling interval represents less than two generations for the emperor penguins [54], we can assume that the allele frequencies of each cluster is constant over this period. Thus for each cluster and each locus  $\lambda$  out of  $\Lambda$  loci, the allele frequencies of  $A_\lambda$  alleles are known and described by:

$$\mathcal{F}_{r\lambda} = (p_{r\lambda a})_{a=1, \dots, A_\lambda}. \quad (14)$$

The loci included in the data set have been genotyped by at least 80% of the individuals per population. We can thus assume that the individuals were genotyped at the same  $\Lambda$  loci (with possibly some allele frequencies equal to 0). The emperor penguins are diploids, thus we describe their genotype by:

$$\mathcal{G} = \{(a_\lambda^1, a_\lambda^2)\}_{\lambda=1, \dots, \Lambda}. \quad (15)$$

Since we are dealing with single nucleotide polymorphisms (SNPs) there are two alleles which correspond to the two possible nucleotide variations of the DNA. Hereafter, we denote by  $\mathcal{G}$  the genotype of an individual that includes all typed loci.

## A.5 Computation of Likelihood

The computation of the genotype likelihoods involves classical genetic assignment approaches [55, 31] and dispersal analysis from spatially sampled data. In each sampled colonies  $\tau$  we have captured and genotyped  $G_\tau$  individuals. The number of genotyped individuals  $G_\tau$  ranges from 10 to 16 individuals depending on the sampled colony  $\tau$ . Thus we have  $G_\tau$  genotypes  $\mathcal{G}_{i,\tau}$  with  $i = 1, \dots, G_\tau$ .

The conditional probability that an individual  $i$  carries alleles  $(a^1, a^2) \in \{1, \dots, A_\lambda\}^2$  at locus  $\lambda$  given that this individual comes from cluster  $r$ , can be deduced from the allele frequencies of each cluster (see section A.3). From the Hardy-Weinberg equilibrium assumption within a cluster, the two-alleles are independent and allele frequencies are:

$$\mathbb{P}((a^1, a^2) | \text{indiv. } i \text{ belongs to cluster } r) = 2^{k_\lambda} p_{r\lambda a^1} p_{r\lambda a^2} \quad (16)$$

where  $k_\lambda = 0$  if the individual is homozygous at locus  $\lambda$ , that is  $a^1 = a^2$ , and  $k_\lambda = 1$  otherwise. Using the linkage equilibrium among loci, the conditional probability for the genotype  $\mathcal{G}_{i,\tau}$  is:

$$\mathbb{P}(\mathcal{G}_{i,\tau} | \text{indiv. } i \text{ belongs to cluster } r) = 2^{k_i} \prod_{\lambda=1}^{\Lambda} p_{r\lambda a^1} p_{r\lambda a^2} \quad (17)$$

where  $k_i$  is the number of heterozygous loci in  $\mathcal{G}_{i,\tau}$ . Hence, the probability for the genotype  $\mathcal{G}_{i,\tau}$  sampled in colony  $\tau$  at time  $t_i$  is given by the law of total probability

$$\mathbb{P}(\mathcal{G}_{i,\tau}) = \sum_{r=1}^R \mathbb{P}(\mathcal{G}_{i,\tau} | \text{indiv. } i \text{ belongs to cluster } r) \mathbb{P}(\text{indiv. } i \text{ belongs to cluster } r) \quad (18)$$

In Section A.3, we show that the expected number of individuals observed in colony  $\tau$  is given by  $C_\tau$ , and the expected number of individuals belonging to cluster  $r$  in the colony  $\tau$  is given by  $C_\tau^r$ . Since the genotyping process corresponds to a sampling without replacement, the number of genotyped individuals in  $\tau$  belonging to cluster  $r$  follows a multivariate hypergeometric distribution with parameters  $C_\tau$ ,  $C_\tau^r$  and  $G_\tau$ . In our data collection, the number of sampled individuals per colony ranges from 10 to 16. From previous studies [50], the estimated number of individuals in the colonies, where data have been collected, ranges from 213 to 22510 in 2009. Thus  $C_\tau$  is quite large and it is larger than the average number of sampled individuals per colony. Thus the multivariate hypergeometric distribution can be approximated by a multinomial distribution with parameters  $G_\tau$  and  $(C_\tau^1/C_\tau, \dots, C_\tau^R/C_\tau)$ . Using this approximation, we compute the probability that a genotyped individual  $i$  observed in colony  $\tau$  belongs to cluster  $r$ :

$$\mathbb{P}(\text{indiv. } i \text{ belongs to } r) = \frac{C_\tau^r}{C_\tau} = \frac{n_\tau^r(t_i)}{n_\tau(t_i)} = \mu_\tau^r(t_i). \quad (19)$$

where  $\mu_\tau^r(t_i)$  is the frequency of the genetic cluster  $r$  in the colony  $\tau$  at time  $t_i$ . We deduce the probability of the genotype  $\mathcal{G}_{i,\tau}$

$$\mathbb{P}(\mathcal{G}_{i,\tau}) = \sum_{r=1}^R \mu_\tau^r(t_i) \left[ 2^{k_i} \prod_{\lambda=1}^{\Lambda} p_{r\lambda a^1} p_{r\lambda a^2} \right] \quad (20)$$

Finally, we compute the likelihood function associated with the unknown parameters, given the genotyped  $\mathcal{G}_{i,\tau}$  and the sampling times  $(t_i)$  as:

$$\begin{aligned} \mathcal{L}(\Theta) &= \prod_{\tau=1}^J \prod_{i=1}^{G_\tau} \mathbb{P}(\mathcal{G}_{i,\tau} | \Theta) \\ &= \prod_{\tau=1}^J \prod_{i=1}^{G_\tau} \sum_{r=1}^R \mu_\tau^r(t_i) \left[ 2^{k_i} \prod_{\lambda=1}^{\Lambda} p_{r\lambda a^1} p_{r\lambda a^2} \right] \end{aligned} \quad (21)$$

## A.6 Estimation of parameters

We aim to infer the following parameter vector  $\Theta = (d, \mathbf{p}_m, \boldsymbol{\mu})$ .

First, we assume that the vector of parameters  $\mathbf{p}_m$  has only 7 different values corresponding to the seven geographical regions defined in section A.1. In addition, the initial proportions  $\boldsymbol{\mu}$  of clusters are only unknown in the colonies belonging to the three regions without genetic characterisation (StoS, StoK and A-B sea).

For each behavior (informed, random or semi-informed see section A.2), the estimates of dispersal parameters  $(\hat{d}, \hat{\mathbf{p}}_m)$  and the initial cluster proportions  $(\hat{\boldsymbol{\mu}})$  have been obtained by minimizing the logarithm of the inverse Likelihood, that is  $-\log(\mathcal{L}(d, \mathbf{p}_m, \boldsymbol{\mu}))$ .

The minimization algorithm is performed using a Bayesian method, where the prior of the parameters  $d$  and  $\mathbf{p}_m$  are assumed to be uniformly distributed with the following constraints:

$$(\hat{d}, \hat{\mathbf{p}}_m) \in (250, 6500) \times (0, 1)^7$$

and the prior of the parameter  $\boldsymbol{\mu}$  follows a Dirichlet distribution of order  $R = 4$  with parameters all equal to 1, thus we have:

$$\hat{\mu}_h^r \in (0, 1)^4 \quad \text{and} \quad \sum_{r=1}^4 \hat{\mu}_h^r = 1 \quad \text{for all } h \in \{1, \dots, 54\}.$$

## A.7 Choice of the dispersal behavior with a model selection procedure

We explore 3 dispersal behaviors (informed, random or semi-informed see section A.2). In order to choose the most likely behavior, we perform a model selection using the different selection model criteria: the Bayesian Information Criteria (BIC) [56], two Deviance Information Criteria (DIC) [57, 58] and a predictive Information Criteria (IC) [59]. The BIC is defined by

$$BIC = -2 \log[\mathcal{L}(\Theta^*)] + k \log(I) \quad (22)$$

where  $I$  is the sample size,  $k$  the number of parameters and  $\Theta^*$  is the maximum likelihood estimate of the parameter vector  $\Theta$ , that is  $\Theta^* = \operatorname{argmax}(\mathcal{L}(\Theta))$ . In our study,  $k$  and  $I$  are the same for all the models.

The DIC satisfies

$$DIC = \widehat{\mathcal{D}} + p_{eff} \quad (23)$$

where  $\widehat{\mathcal{D}}$  is the posterior mean of the deviance  $\mathcal{D}(\Theta) = -2 \log[\mathcal{L}(\Theta^*)]$  and  $p_{eff}$  is the effective number of parameters of the model. We use two different versions of the DIC, which rely on different definitions of  $p_{eff}$ . The first version has been developed by [57]:

$$p_{eff} = \widehat{\mathcal{D}} - \mathcal{D}(\widehat{\Theta}) \quad (24)$$

where  $\widehat{\Theta}$  is the posterior mean of  $\Theta$ . The second version has been introduced by [58]:

$$p_{eff} = \frac{1}{2} \mathbb{V}(\mathcal{D}(\Theta)) \quad (25)$$

where  $\mathbb{V}(\mathcal{D}(\Theta))$  is the posterior variance of  $\mathcal{D}(\Theta)$ .

We also use the IC developed by [59]:

$$IC = 3\widehat{\mathcal{D}} - 2\mathcal{D}(\widehat{\Theta}) \quad (26)$$

In practice, the posterior mean and variance, which appear in our four criteria, are computed with their empirical values using the weighted posterior sample  $\{\Theta_m, w_m\}$  provided by our minimization algorithm.

## B Estimated parameters

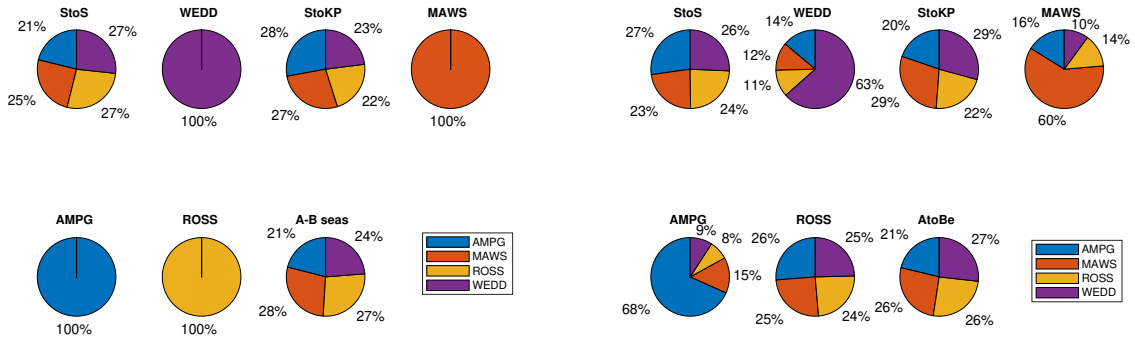
### B.1 Estimation of the initial proportion of cluster within a colony.

The model permits to estimate the initial proportion of cluster within colonies. Since the clusters are associated to geographical regions (AMPG, WEDD, ROSS and MAWS), we assume that for the colonies belonging to those regions, the initial proportion of the corresponding cluster is equal to 1 (Fixed observed cluster on Fig. 4).

For the semi-informed dispersal (Fig. 4), we found that the colonies

- from Snowhill to Smith (StoS) are mostly composed of individuals from WEDD cluster (geographically closest cluster) and AMPG cluster.
- from Stancomb to Kloa Point (StoKP) are mostly composed of MAWS and WEDD clusters;
- from Ledda bay to Rotschild (AtoBe) seems panmictic because all the proportion are almost equal.

This pattern is confirmed even when we do not constrain the initial proportion of the observed cluster (Estimated observed cluster on Fig. 4b). In that case the highest proportion of a given cluster matches with the observed cluster. This correct assignment suggests that our model reproduces well the expected pattern.



(a) Fixed observed cluster – Semi-Informed behavior (b) Estimated observed cluster – Semi-Informed behavior

Figure 4: Estimation of the proportion of the four genetic clusters (AMPG, MAWS, ROSS and WEDD) across the seven geographical regions in Antarctica.

## B.2 Emigration rate per colony

In the main text we gather colonies per geographical regions. However, we quantify the emigration rates for each colony (see Fig. 5). We observe that the emigration rates vary among colonies belonging to a similar geographical regions. In addition, we provide the averaged migration flux between colonies of Antarctica from 2009 to 2014 (Fig. 6) for the following regions: Smith to Snowhill Island in the Wedell sea (StoS), Weddell sea (Gould Bay to Halley Bay colonies) (WEDD), from Stancomb to Kloa point (StoK), Mawson Bay (Fold Island to Cape Darnley colonies) (MAWS), from Amanda Bay to Pointe Geologie colonies (AMPG), the Ross sea (Cape Washington and Cape Crozier) (ROSS) and Admunsen and Bellinghausen seas (Ledda bay to Rothschild Island) (A-B seas)

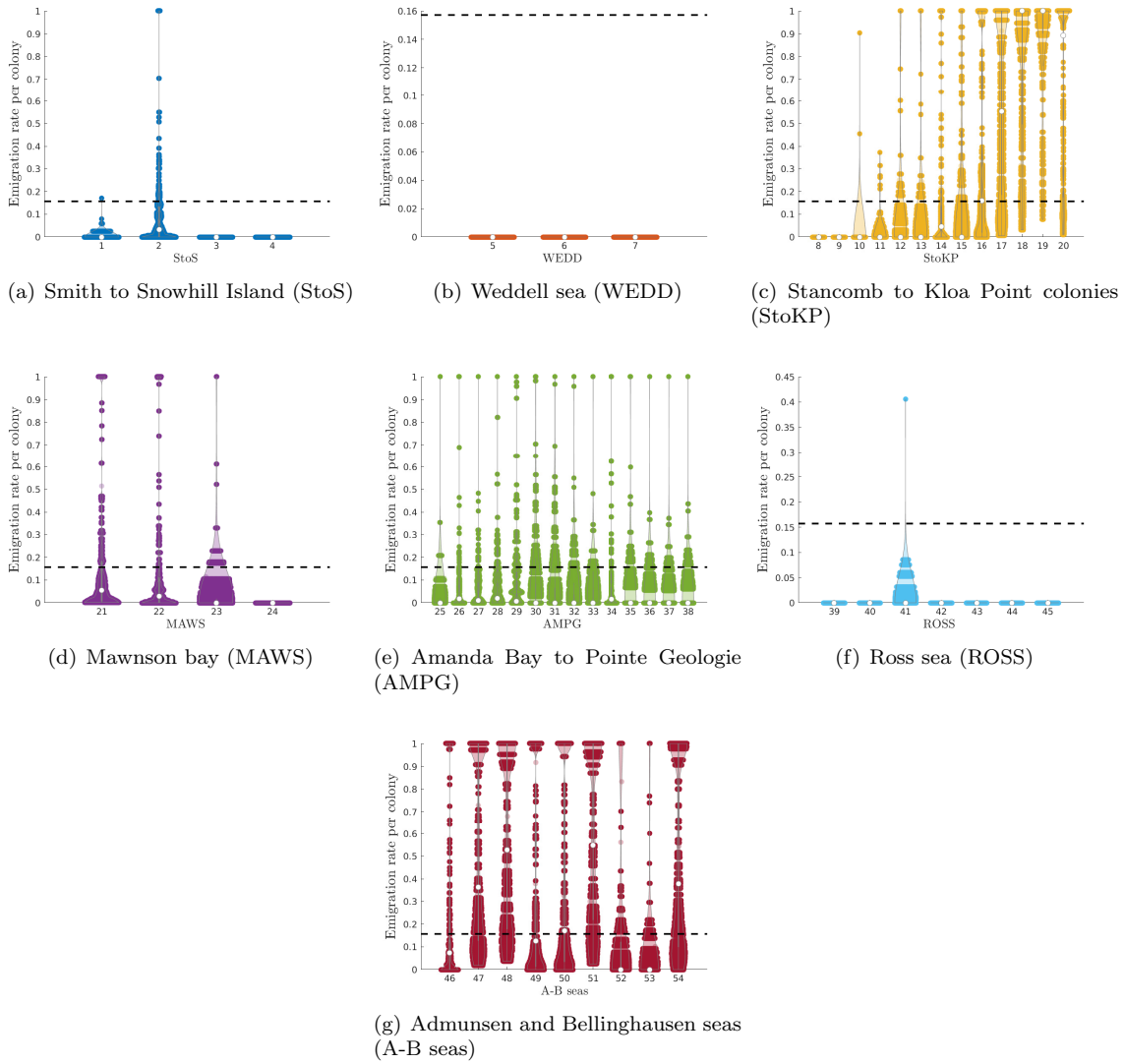


Figure 5: Posterior distributions of the emigration rates per colony over the entire Antarctic continent (a) and for the seven regions of Antarctica (b). White dots represent the median of the distributions and the black line is the mean emigration rate over all colonies in Antarctica (0.157)

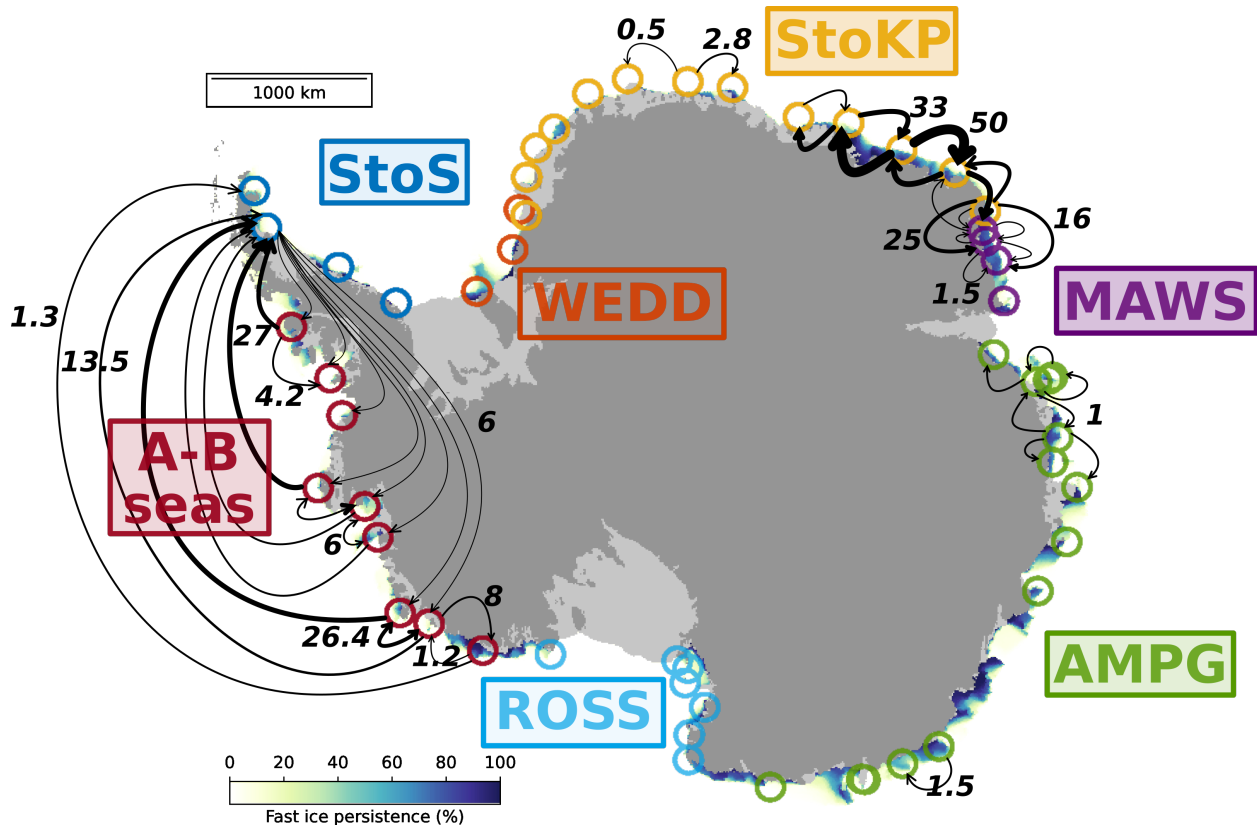


Figure 6: Migration flux estimation averaged between colonies of Antarctica from 2009 to 2014. Map background shows the distribution of fast ice persistence, expressed as a percentage of time covering each unit (square of 6.25km) from March 2000 to March 2018 (the map is extracted from [29]).



## C Importance of climatic and demographic covariates

Here we present the figure of the conditional variable importance scores of random forests modelling annual emigration probability when refining our analysis only with environmental covariates (Fig. 7).

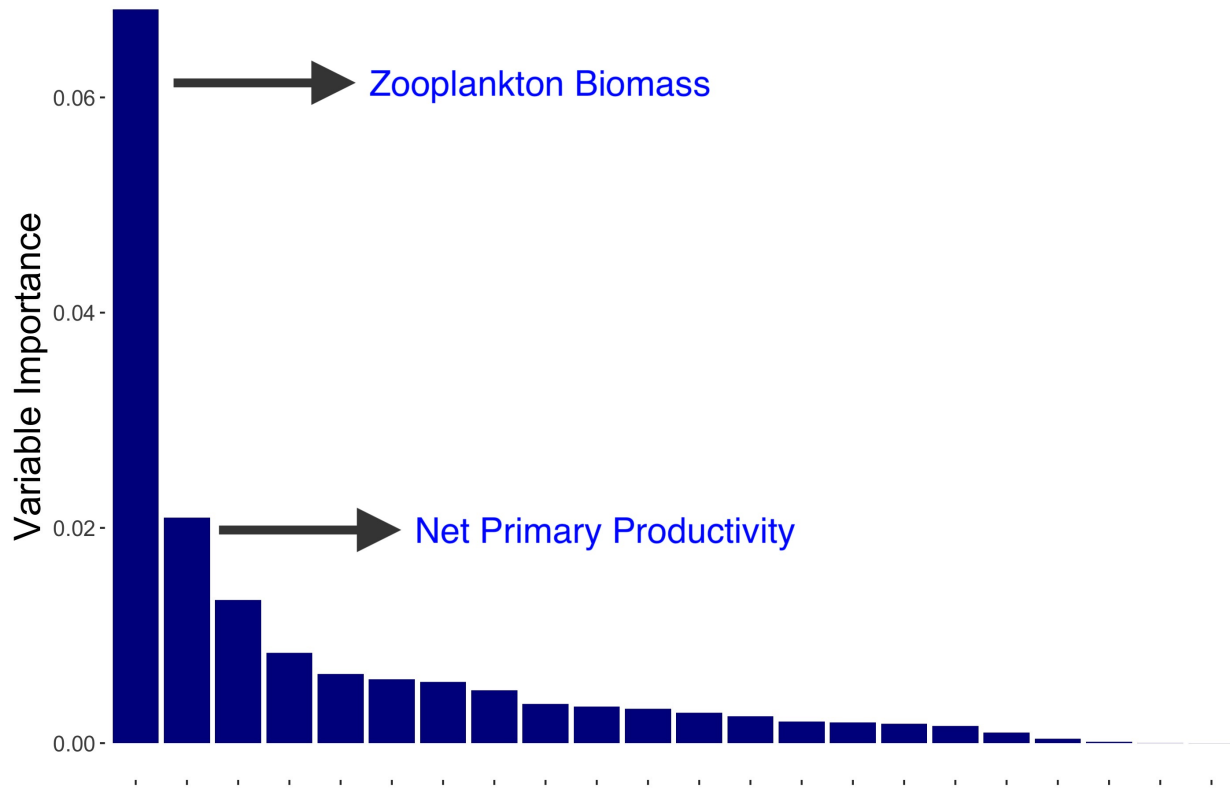


Figure 7: Conditional variable importance scores of random forests modelling annual emigration probability. Only the top 2 variables are shown.

In addition, we show the distribution of the size of the colonies for colonies with or without emigration (Fig. 8).

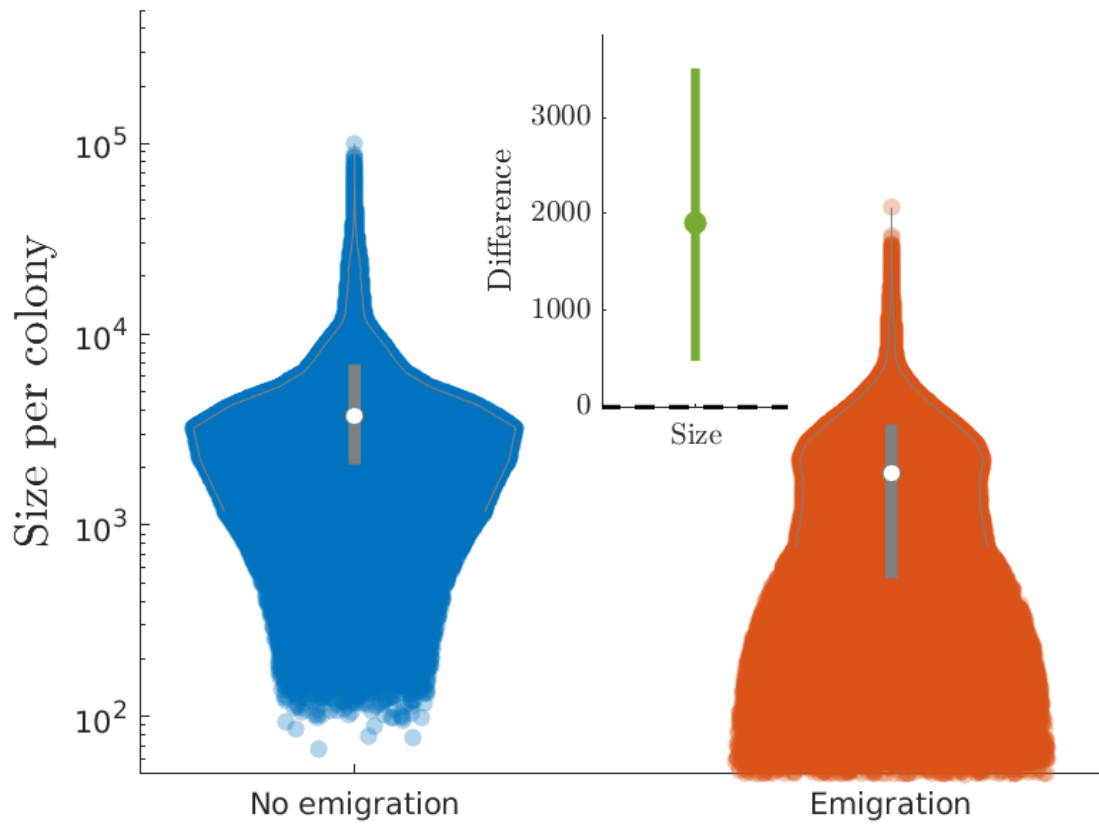


Figure 8: Distribution of the size per colony for colonies with Emigration (orange violinplots) and without emigration (blue violinplots). The subparts represent the boxplot of the difference between demographic covariates in colonies without and with emigration (green corresponds to positive difference). White dots correspond to the median of the distributions.

## D Projection of the total Emperor penguin population size

Plugging the estimated demographic parameters provided by our new analysis into our metapopulation model (section A.2), we project the total population size of Emperor penguin over the century in different climate scenarios: scenario 4.3°C [RCP8.5], scenario 2.6°C [new scenario], scenario 2.4°C [RCP4.5], scenario 2°C [Paris 2°C] and scenario 1.5°C [Paris 1.5°C]. We compare the outcome of this updated model (semi-informed dispersal), with the projections of the model without dispersal (see Fig. 10). We show that the more likely dispersal behavior predicted for the Emperor penguins (semi-informed dispersal with small mean distance dispersal 428km and small emigration rates), results in a greater global population up to 5% compared with a scenario without dispersion when the climate scenarios are unfavorable (from scenario 4.3°C and scenario 2°C). However, under a favorable climate scenario 1.5°C (Paris 1.5°C), this dispersal behavior does not improve significantly the global population size but may attenuate it. We also compared our projection with the inferred dispersal behavior with the projections obtain with a broader range of dispersal behaviors (random, semi-informed or informed dispersal) and various mean dispersal distances and emigration rates. We show that our predicted scenario is not among the more optimistic for the Emperor penguin population size because we project on average a 5% reduction of the population size compared with the other scenario of dispersal. As a conclusion, the impact of dispersal behavior, distance and emigration rate on the future global population size is relatively small compared with the impact of climate change mitigation [50, 21] (see Fig.SI6 and SI7).

## References

- [1] A Hastings. Can spatial variation alone lead to selection for dispersal? *Theor. Popul. Biol.*, 24:244–251, 1983.
- [2] M Slatkin. Gene flow and the geographic structure of natural populations. *Science*, 236(4803):787–792, 1987.
- [3] M. W. Cadotte. Dispersal and species diversity: A meta-analysis. *Am. Nat.*, 167(6):913–924, 2006.
- [4] Lionel Roques, Samuel Soubeyrand, and Jérôme Rousselet. A statistical-reaction–diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology*, 274(1):43–51, 2011.
- [5] P Turchin. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer Associates, 1998.
- [6] R. Nathan, G. Perry, J. T. Cronin, A. E. Strand, and M. L. Cain. Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273, 2003.
- [7] Etienne K Klein, A Bontemps, and Sylvie Oddou-Muratorio. Seed dispersal kernels estimated from genotypes of established seedlings: does density-dependent mortality matter? *Methods in Ecology and Evolution*, 4, 2013.
- [8] Sewall Wright. Isolation by distance. *Genetics*, 28(2):114–138, 1943.
- [9] T Broquet, N Ray, E Petit, J M Fryxell, and F Burel. Genetic isolation by distance and landscape connectivity in the american marten (*Martes americana*). *Landscape Ecol*, 21:877–889, 2006.
- [10] J A Stamps. Habitat selection by dispersers: proximate and ultimate approaches. In J Clobert, E Danchin, A Dhondt, and J Nichols, editors, *Dispersal*. New York: Oxford University Press, 2001.
- [11] J Clobert, J Le Galliard, J Cote, S Meylan, and M Massot. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol. Lett.*, 12(3):197–209, 2009.

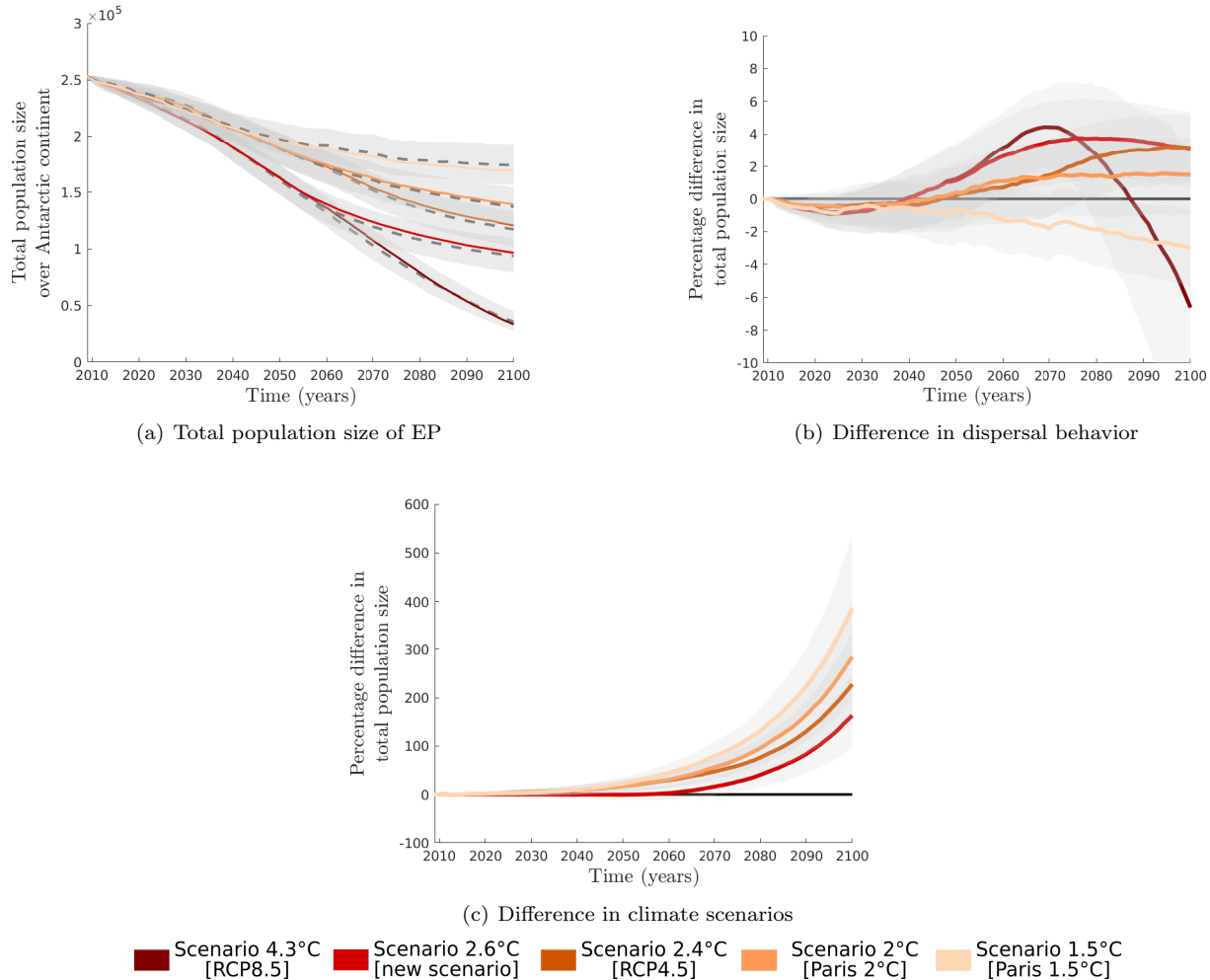


Figure 9: Projection of the total population size of Emperor penguin from 2009 to 2100 using the demographic model (5) with different climate scenarios. In panel (a), the projected total population size without dispersal (dashed curves) and with semi-informed dispersal with the most likely parameters provided by our analysis (plain curves); The grey regions corresponds to the 1% confident intervals around the median. In panel (b), we present the percentage difference of population size between the projection with a semi-informed dispersal and the projection without dispersal for each climate scenario. In panel (c), we present the percentage difference of population size between the projection with the worst climate scenario 4.3°C [RCP8.5] and the other scenarios.

- [12] Brad H. McRae. Isolation by resistance. *Evolution*, 60(8):1551–1561, 2006.
- [13] Tabitha A. Graves, Richard B. Chandler, J. Andrew Royle, Paul Beier, and Katherine C Kendall. Estimating landscape resistance to dispersal. *Landscape Ecology*, 29:1201–1211, 2014.
- [14] Tabitha A. Graves, Paul Beier, and J. Andrew Royle. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, 22(15):3888–3903, 2013.
- [15] W H Lowe and F W Allendorf. What can genetics tell us about population connectivity? *Molecular Ecology*, 19(15):3038–3051, 2010.

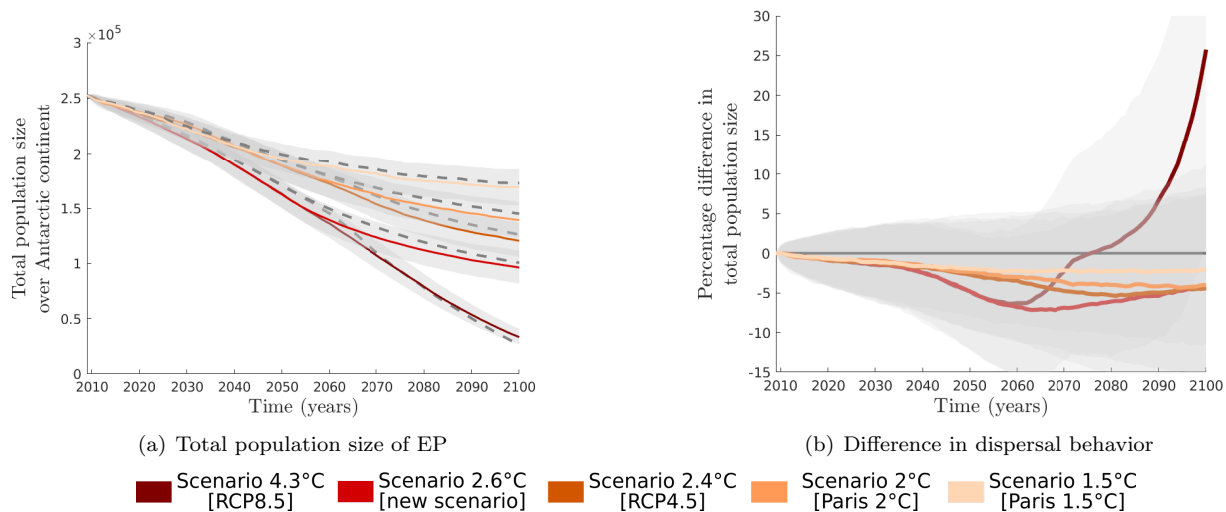


Figure 10: Projection of the total population size of Emperor penguin from 2009 to 2100 using the demographic model (5) with different climate scenarios. In panel (a), the projected total population size with various dispersal behavior (dashed curves) and with semi-informed dispersal with the most likely parameters provided by our analysis (plain curves); The grey regions corresponds to the 1% confident intervals around the median. In panel (b), we present the percentage difference of population size between the projection with a semi-informed dispersal and the projection with various dispersal for each climate scenario.

- [16] L. Roques, E. Walker, P. Franck, S. Soubeyrand, and E. K. Klein. Using genetic data to estimate diffusion rates in heterogeneous landscapes. *Journal of Mathematical Biology*, 73(2):397–422, 2016.
- [17] T.R.E. Southwood and P.A. Henderson. *Ecological Methods*. Wiley, 2009.
- [18] Toby A. Patterson, Len Thomas, Chris Wilcox, Otso Ovaskainen, and Jason Matthiopoulos. State-space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2):87–94, 2008.
- [19] B Doligez, E Danchin, and J Clobert. Public information and breeding habitat selection in a wild bird population. *Science*, 297(5584):1168–1170, 2002.
- [20] A Ponchon, R Garnier, D Grémillet, and T Boulinier. Predicting population response to environmental change: the importance of considering informed dispersal strategies in spatially structured population models. *Divers. Distrib.*, 21:88–100, 2015.
- [21] Stephanie Jenouvrier, Judy Che-Castaldo, Shaye Wolf, Marika Holland, Sara Labrousse, Michelle LaRue, Barbara Wienecke, Peter Fretwell, Christophe Barbraud, Noah Greenwald, et al. The call of the emperor penguin: Legal responses to species threatened by climate change. *Global change biology*, 27(20):5008–5029, 2021.
- [22] J B Thiebot, A Lescroël, C Barbraud, and C A Bost. Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarct. Sci.*, 25(4):536–544, 2013.
- [23] Michelle A LaRue, Gerald Kooyman, Heather J Lynch, and Peter Fretwell. Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography*, 38(2):114–120, 2015.
- [24] J. L. Younger, G. V. Clucas, D. Kao, A. D. Rogers, K. Gharbi, T. Hart, and K. J. Miller. The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Molecular Ecology*, 26(15):3883–3897, 2017.

- [25] Peter T Fretwell and Philip N Trathan. Discovery of new colonies by sentinel2 reveals good and bad news for emperor penguins. *Remote Sensing in Ecology and Conservation*, 7(2):139–153, 2021.
- [26] Peter T Fretwell and Philip N Trathan. Emperors on thin ice: Three years of breeding failure at halley bay. *Antarctic Science*, 31(3):133–138, 2019.
- [27] S Jenouvrier, J Garnier, F Patout, and L Desvillettes. Influence of dispersal processes on the global dynamics of emperor penguin, a species threatened by climate change. *Biol. Conserv.*, 212:63 – 73, 2017.
- [28] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 2007.
- [29] A. D. Fraser, R. A. Massom, M. S. Handcock, P. Reid, K. I. Ohshima, M. N. Raphael, J. Cartwright, A. R. Klekociuk, Z. Wang, and R. Porter-Smith. Eighteen-year record of circum-antarctic landfast-sea-ice distribution allows detailed baseline characterisation and reveals trends and variability. *The Cryosphere*, 15(11):5061–5077, 2021.
- [30] Matthew C. Long, J. Keith Moore, Keith Lindsay, Michael Levy, Scott C. Doney, Jessica Y. Luo, Kristen M. Krumhardt, Robert T. Letscher, Maxwell Grover, and Zephyr T. Sylvester. Simulations With the Marine Biogeochemistry Library (MARBL). *Journal of Advances in Modeling Earth Systems*, 13(12), 2021.
- [31] J. K Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
- [32] L Roques. *Modèles de réaction-diffusion pour l'écologie spatiale*. Editions Quae, 2013.
- [33] E. C. Anderson and E. A. Thompson. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, 160(3):1217–1229, 2002.
- [34] G L Kooyman, D Siniff, I Stirling, and J Bengtson. Moulting habitat, pre-and post-moulting diet and post-moulting travel of Ross Sea emperor penguins. *Mar. Ecol. Prog. Ser.*, 267:281–290, 2004.
- [35] K T Goetz, B I McDonald, and G L Kooyman. Habitat preference and dive behavior of non-breeding emperor penguins in the eastern ross sea, antarctica. *Mar Ecol Prog Ser*, 593:155–171, 2018.
- [36] N. J. Aebischer and J. C. Coulson. Survival of the kittiwake in relation to sex, year, breeding experience and position in the colony. *Journal of Animal Ecology*, 59(3):1063–1071, 1990.
- [37] J. Michael Reed, Thierry Boulinier, Etienne Danchin, and Lewis W. Oring. *Informed Dispersal*, pages 189–259. Springer US, 1999.
- [38] S Dean Kildaw, David B Irons, David R Nysewander, and C Loren Buck. Formation and growth of new seabird colonies: the significance of habitat quality. *Marine Ornithology*, 33:49–58, 2005.
- [39] J Matthiopoulos, J Harwood, and L Thomas. Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology*, 74(4):716–727, 2005.
- [40] J L Younger, G V Clucas, G Kooyman, B Wienecke, A D Rogers, and *et al.* Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Global Change Biol.*, 21(6):2215–2226, 2015.
- [41] J. Prevost. *Expéditions polaires françaises*, chapter Ecologie du manchot empereur. Hermann Press, Paris, France, 1961.

- [42] J L Mougin and M Van Beveren. Structure et dynamique de la population de manchots empereurs *Aptenodytes Forsteri* de la colonie de l’archipel de Pointe Géologie, terre adélie. *C.R. Acad. Sci.*, 289(2):157–160, 1979.
- [43] P. T. Fretwell, P. N. Trathan, B. Wienecke, and G. L. Kooyman. Emperor penguins breeding on iceshelves. *PLoS ONE*, 9:e85285, 2014.
- [44] Özge Balkiz, Arnaud Béchet, Lauriane Rouan, R Choquet, Christophe Germain, Juan A. Amat, Manuel Rendón-Martos, Nicola Baccetti, Sergio Nissardi, Uygur Özesmi, and Roger Pradel. Experience-dependent natal philopatry of breeding greater flamingos. *Journal of Animal Ecology*, 79(5):1045–1056, 2010.
- [45] Alan Johnson and Frank Cézilly. *The greater flamingo*. A & C Black, 2007.
- [46] Jean Clobert, Rolf Anker Ims, and François Rousset. Causes, mechanisms and consequences of dispersal. In *Ecology, genetics and evolution of metapopulations*, pages 307–335. Elsevier, 2004.
- [47] Stéphanie Jenouvrier, Marika Holland, Julianne Stroeve, Christophe Barbraud, Henri Weimerskirch, Mark Serreze, and Hal Caswell. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology*, 18:2756–2770, Jul 2012.
- [48] T Boulinier and E Danchin. The use of conspecific reproductive success for breeding patch selection in terrestrial migratory species. *Evol. Ecol.*, 11(5):505–517, 1997.
- [49] Alexander D Fraser, Robert A Massom, Kay I Ohshima, Sascha Willmes, Peter J Kappes, Jessica Cartwright, and Richard Porter-Smith. High-resolution mapping of circum-antarctic landfast sea ice distribution, 2000–2018. *Earth System Science Data*, 12(4):2987–2999, 2020.
- [50] Stéphanie Jenouvrier, Marika Holland, David Iles, Sara Labrousse, Laura Landrum, Jimmy Garnier, Hal Caswell, Henri Weimerskirch, Michelle LaRue, Rubao Ji, et al. The paris agreement objectives will likely halt future declines of emperor penguins. *Global change biology*, 26(3):1170–1184, 2020.
- [51] G. Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2), February 2020.
- [52] Dries Debeer and Carolin Strobl. Conditional permutation importance revisited. *BMC Bioinformatics*, 21(1), 2020.
- [53] L Roques, J Garnier, F Hamel, and E K Klein. Allee effect promotes diversity in traveling waves of colonization. *Proc Natl Acad Sci USA*, 109(23):8828–8833, 2012.
- [54] J Forcada and P. N. Trathan. Penguin responses to climate change in the southern ocean. *Global Change Biology*, 15(7):1618–1630, 2009.
- [55] D. Paetkau, W. Calvert, I. Stirling, and C. Strobeck. Microsatellite analysis of population structure in canadian polar bears. *Mol. Ecol.*, 4(3):347–354, 1995.
- [56] G Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [57] D J Spiegelhalter, N G Best, B P Carlin, and A Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

- [58] A Gelman, J Carlin, H Stern, D Dunson, A Vehtari, and D Rubin. *Bayesian Data Analysis*. Chapman & Hall / CRC, 2003.
- [59] T Ando. Predictive bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31:13 – 38, 2011.