



**HAL**  
open science

## Coupling genetic and demographic data to reveal dispersal processes in emperor penguins

Jimmy Garnier, Gemma Clucas, Jane Younger, 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, Christophe Barbraud, et al.. Coupling genetic and demographic data to reveal dispersal processes in emperor penguins. 2025. hal-03822288v3

**HAL Id: hal-03822288**

**<https://hal.science/hal-03822288v3>**

Preprint submitted on 29 Jan 2025

**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.

# Coupling genetic and demographic data to reveal dispersal processes in emperor penguins.

Jimmy Garnier<sup>1,\*</sup>, Gemma Clucas<sup>2</sup>, Jane Younger<sup>3</sup>, Bilgecan Sen<sup>4,5</sup>,  
Christophe Barbraud<sup>6</sup>, Michelle LaRue<sup>7,8</sup>, Alexander D. Fraser<sup>9</sup>, Sara  
Labrousse<sup>10</sup>, Stéphanie Jenouvrier<sup>5</sup>

## Affiliations:

1. CNRS, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LAMA, 73000 Chambéry, France,
2. Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Rd. Ithaca, NY 14850, USA
3. Milner Centre for Evolution, University of Bath, Bath UK BA2 7AY
4. University of Maryland Center for Environmental Science, Frostburg, USA
5. Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
6. Centre d'Etudes Biologiques de Chizé, CNRS-La Rochelle University UMR7372, 79360 Villiers en Bois, France
7. School of Earth and Environment, University of Canterbury, NZ
8. School of Earth and Environmental Sciences, University of Minnesota, USA
9. Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, nipaluna / Hobart 7001, Tasmania, Australia
10. Laboratoire d'Océanographie et du Climat: Expérimentations et approches numériques (LOCEAN), UMR 7159 Sorbonne-Université, CNRS, MNHN, IRD, IPSL 75005, Paris, France

**Corresponding author:** Jimmy Garnier

E-mail: jimmy.garnier@univ-smb.fr

ORCID: 0000-0002-0145-1028

Telephone: +33 (0)4 79 75 87 28

Université Savoie Mont-Blanc,

Laboratoire de Mathématiques (LAMA) - EDPs<sup>2</sup> (UMR 5127), Bât. Le Chablais, Campus Scientifique, 73376 Le Bourget du Lac, France

**Competing interests statement:** Authors declare no conflict of interest.

**Data and code accessibility statement:** The SNP data set is available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.4s7t3>. The codes for the likelihood function and forecasts of emperor penguin population are available from the GitHub repository [https://github.com/garnieji/EP\\_demographic\\_genetic](https://github.com/garnieji/EP_demographic_genetic) and the code for the importance of climatic and demographic covariates is available from the GitHub repository [https://github.com/bilgecansen/Emperor\\_dispersal](https://github.com/bilgecansen/Emperor_dispersal).

**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.

**Acknowledgments:** 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. J.G. acknowledges support from French ANR project ReaCh (ANR-23-CE40-0023-01).

**Type of article:** Research article

## Abstract

Dispersal is an ubiquitous phenomenon that affects the dynamics of the population and the evolution of natural populations; however, it is challenging to measure in most species. Furthermore, the influence of informed dispersal behaviors, referring to the non-random selection of breeding habitats by individuals, on species' responses to rapid global change is substantial but difficult to comprehend.

Here, we present a modeling framework to assess the dispersal characteristics and behaviors of a metapopulation, when observations provide information on its neutral genetic structure for a restricted sampling of locations. Our mechanistic-statistical model couples a deterministic model capturing the spatio-temporal dynamics of four genetic clusters across all breeding colonies by integrating demographic processes with genetic projections, with a probabilistic observation model describing the probability to sample an individual from a given genetic cluster.

We apply this new framework to the emperor penguin, a species living in Antarctica and currently experiencing habitat loss. The model estimates the species' dispersal distance, rates of emigration, and behaviors associated with dispersal (informed or random). By incorporating these estimations with satellite censuses of breeding colonies, we can identify environmental and demographic factors that influence the dispersal of emperor penguins. Finally, we provide new global population forecasts for emperor penguins that can inform conservation actions in Antarctica.

**Keywords:** dispersal distance | dispersal kernel | dispersal range | emigration rates

# Introduction

Dispersal between suitable habitats influences the dynamics of populations (e.g. refs [1, 2, 3, 4, 5]), their gene flow and genetic structure [6, 7, 8, 9], and hence the ecological and evolutionary processes driving biodiversity [10, 11, 12, 13]. The rate and range of dispersal of plant propagules and animal individuals are commonly characterized by tracking individual movements and population redistribution (e.g using abundance data [14] or "mark recapture / sighting" techniques in animal studies [15, 16, 17, 18]). However, such movement data are extremely challenging to collect, especially for endangered species or animals living in remote places on Earth. Genetic markers naturally present in populations offer unique opportunities to study dispersal [19, 20, 21]. However, such genetic methods (e.g. long-term frequency-based approach using population structure described by the FST fixation index) 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 distance kernel over one generation based on genetic data, especially to estimate seed dispersal kernels [22, 23, 24]. Although these methods are accurate [25], they often rely on simple dispersal assumptions. For example, classical methods based on Euclidian distances or least-cost distances (e.g., in models of isolation by distance [26, 27, 28]) assume a single and optimal movement path for individuals, while individuals may change their route during dispersal [29, 30]. Newer methods have been developed that are based on resistance networks [31, 32]. These methods consider the relative cost of dispersal in a specific landscape compared to a reference condition. However, their implementation is time-consuming and the estimation of dispersal parameters, for example, by maximum likelihood, generally lacks accuracy [33]. Furthermore, genetic data alone may not provide enough information on demographic processes because dispersal processes may depend on the environment [34], the population sizes in different environments may vary, dispersal may occur at short or long distance and dispersal might also depend on individual choice [25].

Here, we integrate genetic methods with environment-dependent metapopulation models to develop a new likelihood function that quantifies dispersal rates, distances, and behaviors. This novel approach advances previous methods by linking movement and demographic patterns with genetic data [35]. Specifically, it is based on a mechanistic-statistical approach [36, 37, 16, 17, 14, 38] in the framework of state-space models [39, 40]. It has been developed theoretically to characterize the diffusion rates of insects genetic data over one generation [35], but it 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. [41]). Such 'informed dispersal' behavior [30] enables individuals to settle in habitats of better quality, potentially improving their fitness, therefore increasing population viability and species persistence, especially in the face of global changes [42].

In this study, we present a likelihood function for a metapopulation mechanistic-statistical model that integrates reproductive and dispersal behaviors, including informed departure

and settlement decisions. We apply this model to Emperor penguins (*Aptenodytes forsteri*), an Antarctic seabird that is increasingly threatened by climate change [43]. Due to the logistical challenges of monitoring populations in extreme environmental conditions, very little is known about their dispersal behaviors. In fact, emperor penguins have only been marked at one site (Pointe Géologie [44]), with no recaptures elsewhere. The recent advent of satellite telemetry tags has allowed for an enhanced understanding of the movement of emperor penguins on large spatial scales within a season. However, this approach is not suitable for determining dispersal between colonies due to the limited life span of these devices [45].

Like many seabirds, emperor penguins are considered highly philopatric [46]. However, this traditional view has been challenged by advances in genetic analyses and very high-resolution satellite imagery (VHR), suggesting that movements between colonies occur [47]. In fact, genetic studies have identified at least four distinct genetic clusters among emperor penguins [48]. Each cluster is located in a different geographic region of Antarctica, some spanning thousands of kilometers of coastline and comprising multiple breeding colonies. While there is some degree of gene flow connecting these clusters, they remain genetically distinct from one another. However, within each cluster, dispersal of individuals between breeding colonies is sufficient to maintain panmixia [48]. In addition, VHR satellite imagery has recently documented colony movements, disappearances, and relocations [49]. For example, a dramatic decline in the world's second largest emperor penguin colony occurred at Halley Bay, while the nearby Dawson-Lambton colony, 55 km to the south, saw a more than tenfold increase in penguin numbers during the same period [50]. Halley Bay has suffered three years of almost complete breeding failure caused by a change in the local environment and sea ice conditions, and those unfavorable conditions may have forced penguins to relocate to Dawson-Lambton [50]. The colony had been present at Halley Bay since at least 1956, persisting for 60 years before the major environmental disturbance led to a massive population decline and emigration event. This suggests that emperor penguin movements may be triggered by major environmental disturbances 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 informed emigration.

Using a mechanistic-statistical approach, we have developed a likelihood function that links the demographic characteristics of the emperor penguin to genetic data. It enables us to: (1) determine the most likely dispersal behaviors in emigration and establishment of emperor penguins (informed versus random); (2) estimate the mean dispersal distance for this species and the emigration rates between colonies; (3) highlight environmental and demographic factors, that drive emigration rates by combining our approach with independent demographic and environmental data; and (4) project the global population of emperor penguin change into the future using the most recent large ensemble of climatic projection in Antarctica (CESM2-LENS [51]).

## Material and methods

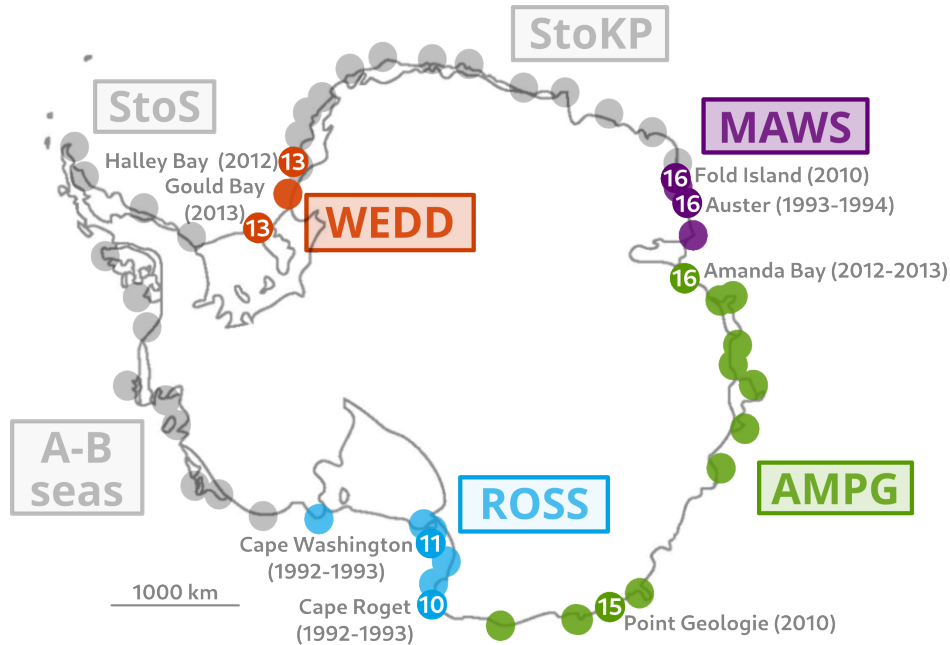
The model presented in this study is based on the combination of a mechanistic metapopulation model that describes the population dynamics, and a stochastic model that accounts for the collection of genetic measurements based on the population dynamics (section "Mechanistic-statistical model") [52]. Previous theoretical studies have demonstrated the relevance of this approach for estimating dispersal parameters using genetic data [35]. We begin by describing our case study, the emperor penguin (section "Case study: emperor penguin"). Then we present the available genetic data for emperor penguins (section "Genetic data"), followed by an explanation of our mechanistic-statistical approach and the statistical inference of the model parameters (section "Statistical inference"). Finally, we compare the different dispersal behavior (section "Comparison of dispersal behaviors"), we evaluate the impact of demographic and environmental factors on emigration rates (section "Impact of demographic and environmental factors on emigration rates") and project the future dynamics of the global emperor penguin population until 2100 (section "Forecasts of emperor penguin global population").

### Case study: emperor penguin

Emperor penguins are seabirds that live in Antarctica. They breed annually during the Antarctic winter in one of the 66 breeding colonies around Antarctica (see circles in Fig. 1). In March, adults settle in a colony to mate, lay a single egg, and raise their chick until December. Adults and juveniles leave colonies in December/January and disperse into the Southern Ocean. After a northward migration following departure from their natal colony, juveniles return close to the Antarctic sea ice in April/May (see [53] and observations from Argos tracking [54, 55]). The first breeding starts at three years. During this period, from fledgling to first breeding, individuals can prospect [53, 46] and eventually assess the habitat quality of a potential colony to settle and breed.

A recent genetic study identified four genetic clusters among emperor penguin that are significantly genetically differentiated, with some degree of gene flow connecting these clusters [48]. Although genetic clustering (STRUCTURE) [56] and FST analyzes support the presence of four genetic clusters, genetic differentiation between emperor penguin colonies is subtle and hierarchical, and does not follow a typical isolation-by-distance (IBD) pattern of differentiation.

The four identified genetic clusters coincide with distinct geographical regions composed of several colonies: (WEDD) Weddell sea (Gould Bay to Halley Bay colonies), (MAWS) Mawson Bay (Fold Island to Cape Darnley colonies), (AMPG) Amanda Bay to Pointe Geologie colonies, and (ROSS) Ross sea (Cape Washington and Cape Crozier colonies) (see Fig. 1). However, these four geographical regions do not cover all the 66 colonies around Antarctica. Three geographical regions remain genetically uncharacterized: (StoS) from Smith to Snowhill Island in the Wedell sea colonies, (StoK) from Stancomb to Kloa point colonies and (A-B seas) Ledda bay to Rotschild colonies (Admunsen and Bellingshausen seas) (see Fig. 1). Although it is possible that there are more than four genetic clusters across the emperor penguin's range, a complete sampling of all colonies is logistically infeasible due to the remote distribution of the species in one of the harshest climates on



**Figure 1:** The four genetic clusters detected around Antarctica that characterize four geographic regions: (WEDD) Weddell sea (Gould Bay to Halley Bay colonies) red dots, (MAWS) Mawson Bay (Fold Island to Cape Darnley colonies) purple dots, (AMPG) Amanda Bay to Pointe Geologie colonies, green dots and (ROSS) Ross sea (Cape Washington and Cape Crozier colonies) blue dots. Grey and coloured dots indicate all the 66 known emperor penguins colonies around Antarctica. The three grey regions corresponds to are without genetic characterization: (StoS) from Smith to Snowhill Island in the Wedell sea colonies, (StoK) from Stancomb to Kloa point colonies and (A-B seas) Ledda bay to Rotschild colonies (Admunsen and Bellingshausen seas). The white numbers indicated the number of individuals sampled from this colonies and the year of sample.

Earth. This limitation is common in wildlife studies, where representative sampling is often the only viable option. Our estimations of dispersal parameters are therefore conditional on the assumption that the emperor penguin population is structured into these four identified genetic clusters.

### Genetic data

We used genetic data collected from 1992 to 2013 in eight colonies around Antarctica by [48] (see Fig. 1). After filtering steps [57], 4.596 neutral genome-wide single nucleotide polymorphisms (SNPs) were retained. Specifically, we here use only SNPs SNPs that are present in all sampled colonies (parameter -p 8 in the Stacks pipeline) and that are present in at least 80% of individuals per colony (parameter -r 0.8 in the Stacks pipeline). We keep SNPs that are neutral and polymorphic at the metapopulation level. However, some allele frequencies in genetic clusters might be equal to 0 because some alleles are private in the sense that they appear only in one genetic cluster. A total of 110 individuals (10-16 per colony) were successfully genotyped at these loci.



## Mechanistic-statistical model

The mechanistic model characterizes the spatio-temporal changes of emperor penguin populations across the 66 colonies in Antarctica. It combines a demographic model, describing the metapopulation dynamics and incorporating various parameters such as mean dispersal distance, emigration rates, and dispersal behaviors (section "Demographic model"), with a genetic population model projecting the number of individuals originating from one of the four genetic clusters for each year and in each colony (section "Genetic population dynamics"). The stochastic model, on the other hand, includes a probabilistic sampling approach to estimate the likelihood of sampling an individual from a given genetic cluster (section "Probabilistic sampling model") and a statistical genetic model that predicts the probability of observing a particular genotype based on the individual's genetic cluster of origin.

**Demographic model.** We use the metapopulation model developed by [58] to project the population vector  $\mathbf{n}$ , that comprises the population size  $n_i$  in each colony  $i$ , from year  $t$  to year  $t + 1$  (we only look at the female here):

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

It incorporates two phases of possibly different duration: a motionless density-dependent reproduction phase ( $\mathbf{F}$ ) followed by a dispersal phase including natal or breeding dispersal ( $\mathbf{D}$ ). The reproduction matrix  $\mathbf{F}$  follows a Ricker model, where the intrinsic growth rate  $r_i(t)$  of each colony varies in time due to sea ice concentration (SIC) variations, described by climatic projection in Antarctica (CESM2-LENS[51]), while carrying capacities of colonies are constant over time. The dispersal phase  $\mathbf{D}$  comprises three stages: (1) emigrating from the resident colony at a rate  $m_i(t)$ , (2) searching for a new colony among other colonies with an average dispersal distance  $d$  (transfer), and (3) settling in a new colony. During the emigration and settling stages, two possible behaviors (informed versus random) can occur:

- an informed emigration: individuals only emigrate from poor quality breeding sites when the habitat quality is not viable (i.e. negative intrinsic population growth  $r_i(t < 0)$ ), they emigrate at a rate  $m_i(t) = \max(1 - r_i(t)/((1 - p_{m_i})r_c^*), 1)$ , where  $r_c^* < 0$  is the worst growth rate and  $p_{m_i}$  is a sensitivity parameter to bad quality habitat;
- a random emigration: individuals leave the colony regardless of the habitat quality at a fixed rate  $m_i = p_{m_i}$ ;
- an informed establishment: individuals select the most suitable habitats (i.e. maximize intrinsic population growth) within their dispersal range of size  $d$ ;
- random establishment: individuals pick a colony in their dispersal range of size  $d$  randomly and regardless of the habitat quality of the colonies.

In our analysis, we only consider three dispersal behaviors: the random dispersal behavior (R) with random emigration and establishment; the semi-informed dispersal behavior (SI) with informed emigration but random establishment; and the informed dispersal behavior (I) with both informed emigration and establishment.

**Genetic population dynamics.** The population of emperor penguin comprises 4 genetic clusters characterized by their allele frequencies  $\mathcal{F}_{r\lambda} = (p_{r\lambda a})_{a=1,\dots,A_\lambda}$  at the sampled SNPs  $\lambda$  (with possibly some allele frequencies equal to 0 due to the presence of private alleles in some genetic clusters). The frequencies are assumed constant over the sampling interval because they represent less than two generations for emperor penguins (which is 16 years) [59]. Our demographic model describes the survival, reproduction and dispersal of individuals regardless of their native genetic cluster. In order to track the native genetic cluster of individuals, we derive from the demographic model [9, 35], a genetic model that projects the population vector  $\mathbf{n}^r$  of individuals that originate from one of the four genetic cluster  $r$ . The vector  $\mathbf{n}^r$  comprises the number of individuals  $n_\tau^r(t)$  in each colony  $\tau$  originating from cluster  $r$ , and satisfies the following dynamics

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

Initially, we have the following repartition

$$n_i^r(0) = \mu_i^r n_i(0), \quad \text{for all } i \in \{1, \dots, 66\}, \quad (2)$$

where  $\mu_i^r$  is the initial proportion of individuals native from genetic cluster  $r$  within the colony  $i$  ( $\sum_{r=1}^R \mu_i^r = 1$  for all  $i$ ). We assume that for colonies belonging to a region where we have genetic information, the proportion is 1 if the geographical region of the colony matches the genetic cluster and 0 otherwise. For instance, if the colony  $i$  belongs to the region that matches genetic cluster 1, then

$$\mu_i^1 = 1 \quad \text{and} \quad \mu_i^2 = \mu_i^3 = \mu_i^4 = 0.$$

However, for the colonies belonging to the three geographical regions without genetic information, the parameters  $\boldsymbol{\mu} = (\mu_i^1, \mu_i^2, \mu_i^3, \mu_i^4)$  are unknown parameters that we have to estimate; In order to simplify the estimate, we assume that  $\boldsymbol{\mu}$  is the same among colonies of a geographical region. The genetic and demographic dynamics are linked by:

$$n_\tau(t) = \sum_{r=1}^4 n_\tau^r(t).$$

More precisely, since the set of SNPs in our genetic data is selectively neutral, individuals within a colony share the same dispersal and reproduction characteristics independent of their genetic background. In particular, an individual within a colony  $\tau$  and originating from the genetic cluster  $r$  produces offspring at the same rate as any individual within the colony  $\tau$ . Although newborns result from the mating of a male and a female, we assume that they inherit the native genetic cluster of the female.

**Probabilistic sampling model associated with the demographic model.** During the year  $t$  that ranges from 1992 to 2013, individuals from colonies  $J_t$  were genotyped. The sampling of individuals in a given year  $t$ , in the colony  $\tau$  is random among the individuals observed at the colony. Given that our dataset has an average sample size per colony  $G_\tau$  that

is relatively small (10-16 individuals per colony) in comparison to the typical population size of emperor penguin colonies (which ranged from 100 to 25,000 individuals in 2009 [60]), the count of genotyped individuals in  $\tau$  originating from genetic cluster  $r$  follows a multinomial distribution characterized by the parameters  $G_\tau$ , the sample size, and  $(\mu_\tau^1(t), \dots, \mu_\tau^R(t))$ , the proportions of individuals in colony  $\tau$  at time  $t$ , that originate from one of the four genetic clusters. The probability that a genotyped individual  $i$  observed at time  $t$  in colony  $\tau$ , originates from genetic cluster  $r$ , is

$$\mathbb{P}(\text{indiv. } i \text{ originates from cluster } r) = \mu_\tau^r(t) = \frac{n_\tau^r(t)}{n_\tau(t)}.$$

The proportion  $\mu_\tau^r$  corresponds to the ratio between the number of individuals  $n_\tau^r(t)$ , that originate from the genetic cluster  $r$  and the number  $n_\tau(t)$  of individuals alive in the colony  $\tau$  projected by the metapopulation model.

**Statistical genetic assignment approach.** Emperor penguins are diploid organisms, thus their genotypes write  $\mathcal{G} = \{(a_\lambda^1, a_\lambda^2)\}_{\lambda=1, \dots, \Lambda}$ . Since we use single-nucleotide polymorphisms (SNPs), each locus has two alleles, corresponding to the two possible nucleotide variations in the DNA sequence. Using the linkage equilibrium among loci and the Hardy–Weinberg equilibrium assumption within a genetic cluster, the conditional probability for the genotype  $\mathcal{G}_{i,\tau}$  is:

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

where  $k_i$  is the number of heterozygous loci in  $\mathcal{G}_{i,\tau}$  and  $p_{r\lambda a^1}$  and  $p_{r\lambda a^2}$  are the allele frequencies within the genetic cluster  $r$  of the alleles  $a^1$  and  $a^2$  of individual  $i$  at locus  $\lambda$ .

## Statistical inference

**Computation of the likelihood function.** The unknown parameters  $\Theta$  of our mechanistic-statistical model are the mean dispersal distance  $d$ , the emigration sensitivity parameters  $\mathbf{p}_m = (p_{m_1}, \dots, p_{m_\tau})$  and the initial proportions of each genetic clusters,  $\boldsymbol{\mu} = (\mu_i^1, \mu_i^2, \mu_i^3, \mu_i^4)_{i \in \{1,2,3\}}$ . For the three different dispersal behaviors (random, semi-informed or informed dispersal) and with unknown parameters  $\Theta$  of our model, the likelihood function is:

$$\mathcal{L}(\Theta | \mathcal{G}_{i,\tau}) = \prod_{t=1992}^{2013} \prod_{\tau=1}^{J_t} \prod_{i=1}^{G_\tau} \mathbb{P}(\mathcal{G}_{i,\tau})$$

where  $\mathbb{P}(\mathcal{G}_{i,\tau})$  is the probability to sample the genotype  $\mathcal{G}_{i,\tau}$  in colony  $\tau$  at time  $t_i$ , that can be decomposed as follows

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

For each dispersal behavior (random (R), semi-informed (SI) or informed (I)), we estimate the posterior distribution of the parameters  $\Theta$  using the likelihood function and an importance sampling algorithm with prior distribution of parameters  $d$  and  $\mathbf{p}_m$  given by uniform distribution with the following constraints:

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

and prior of the parameter  $\boldsymbol{\mu}$  given by a Dirichlet distribution of order  $R = 4$  with parameters all equal to 1:

$$\mu_h^r \in (0, 1)^4 \quad \text{and} \quad \sum_{r=1}^4 \mu_h^r = 1 \quad \text{for all } h \in \{1, \dots, 66\}.$$

We performed the statistical inference with Matlab version R2021a. The code is available online [https://github.com/garnieji/EP\\_demographic\\_genetic](https://github.com/garnieji/EP_demographic_genetic).

**Confidence intervals and goodness-of-fit.** The model's goodness-of-fit was evaluated by determining the 95% confidence regions for the observed genotypes in each colony and year of observation. To do this, we calculated the probability of each possible observed genotype based on the frequencies of each genetic cluster predicted by the mechanistic model using the estimated parameters that maximize the likelihood function. We then checked if the observed genotypes fell within the 95% confidence regions that represent the most likely outcomes.

### Comparison of dispersal behaviors

In order to determine the most probable dispersal behavior, we conducted a model selection process using four criteria and our likelihood function. These criteria include the Bayesian Information Criteria (BIC), two Deviance Information Criteria (DIC), and a predictive Information Criteria (IC). The BIC is defined by

$$BIC = -2 \log[\mathcal{L}(\Theta^*)] + k \log(I) \quad (3)$$

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

$$DIC = \widehat{\mathcal{D}} + p_{eff} \quad (4)$$

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 [61]:

$$p_{eff} = \widehat{\mathcal{D}} - \mathcal{D}(\widehat{\Theta}) \quad (5)$$

where  $\widehat{\Theta}$  is the posterior mean of  $\Theta$ . The second version has been introduced by [62]:

$$p_{eff} = \frac{1}{2} \mathbb{V}(\mathcal{D}(\Theta)) \quad (6)$$

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

The IC developed by [63] satisfies:

$$IC = 3\widehat{\mathcal{D}} - 2\mathcal{D}(\widehat{\Theta}) \quad (7)$$

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.

### Impact of demographic and environmental factors on emigration rates

We model emigration rate as two separate responses: 1) Annual emigration probability, in which colonies are split into two categories each year with "No emigration" colonies when their median emigration rate is 0 and "Emigration" colonies when their median emigration rate is positive; and 2) Average emigration probability, that is the proportion of years with "Emigration" categorization between 2009 and 2013 for each colony.

When modeling annual emigration probability we only considered environmental variables around each colony, comprising the biomass of zooplankton ( $\text{mmol } C/m^2$ ) and the distance between the colony and the closest edge of fast ice ( $m$ ). They are obtained using the novel landfast sea ice data (hereafter simply "fast ice") data at different scale [64] and the unique sea ice and food web dynamics variables obtained from a forced ocean-sea ice (FOSI) configuration of the ocean ecosystem model of the Community Earth System Model (CESM2) [65]. When applicable, we calculate the average value of each variable for different breeding periods: non-breeding (January to March), laying (April and May), incubation (June and July), and chick-rearing (August to December).

For average emigration probability, each environmental variable was averaged between 2009 and 2013 and we used three additional demographic factors independent of our metapopulation analysis: the size of colony, the growth rate per colony, and the frequency of blinking that corresponds to the relative number of years a colony disappears over a period of 10 years, from 2009 to 2018. They were calculated from the colony presence and population counts of emperor penguins from VHR satellite imagery [66, 67].

For both analysis we use a random forest algorithm to quantify the role of environmental and demographic variables on the probability of emigration [68]. We used the R package "party" for fitting the conditional random forests [68] and "permimp" [69] for computing the variable importance scores. Codes are available online [https://github.com/bilgecansen/Emperor\\_dispersal](https://github.com/bilgecansen/Emperor_dispersal).

### Forecasts of emperor penguin global population

Coupling our new estimated dispersal parameters with the meta-population model developed by [43], we project the total population size of emperor penguin over the century for different climate scenarios. Previous studies have provided a more detailed description of this forecasting approach, which yields a robust forecast by incorporating various sources of uncertainties( [70, 59, 71, 43] for emperor penguins and [72, 73] for a general approach).

Climate scenarios, which are labeled based on the projected global warming increase ( $^{\circ}$  C) above preindustrial levels, are discussed in greater depth in [74]. These scenarios include

**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 behavior: 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

an increase of 4.3°C [RCP8.5], of 2.6°C [new scenario], of 2.4°C [RCP4.5], of 2°C [Paris 2°C] and of 1.5°C [Paris 1.5°C]. The new scenario developed by [74] is intended to demonstrate probable effects on sea ice and therefore emperor penguins by 2100 if governments act now to control greenhouse gas emissions by 2050.

We compare the result of this updated model (semi-informed dispersal), with the projections of the model without dispersal (see Figure 5(a)). Finally, we compare projections for different dispersal behaviors and between climate scenarios (see Figure 5(b)-(c)).

## Results

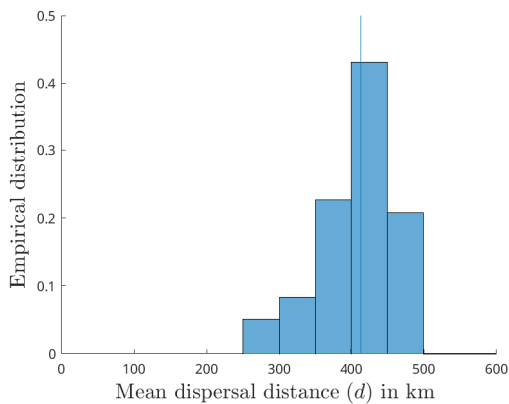
### Goodness-of-fit and convergence

From the three dispersal behaviors (random (R), semi-informed (SI) and informed (I)), we obtained three different set of parameters that maximize the likelihood function. On average, 99% of the observed genotypes fell within the 95% confidence regions, indicating that models with different dispersal behaviors accurately represented the data (109 of the observed genotypes over a total of 110). The three scenarios exhibited a peak in likelihood at the mean dispersal distance parameters, suggesting that these parameters optimize the likelihood for all three scenarios. With respect to the remaining parameters, that are the emigration sensitivity parameters  $p_{m_i}$  and the initial genetic cluster proportions  $\mu_i^r$ , the likelihood function exhibits a smoother distribution around the optimized parameters, indicating that the likelihood is less responsive to changes in these parameters.

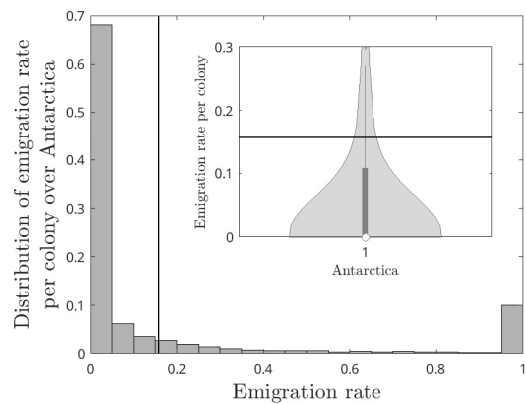
### Dispersal processes

**Dispersal behaviors.** Based on our model selection, it is evident that the genetic data strongly suggest the prevalence of semi-informed dispersal behavior among emperor penguins (see Table 1). This behavior indicates that these penguins are more likely to leave colonies with unfavorable habitat conditions. (i.e., with negative intrinsic population growth; informed emigration), but settle randomly into another colony (random establishment).

**Dispersal ranges.** Figure 2(a) shows the posterior distribution of the mean dispersal distance for the best supported model. It indicates a relatively short dispersal distance of approximately 414 km. This mean dispersal distance is modest compared to the potential movement range derived from the tracking of juveniles and adults at sea (tracking



(a) Mean dispersal distance



(b) Emigration rates over entire continent

**Figure 2:** Posterior distributions of the mean dispersal distance  $d$  per individuals and the emigration rate per colony per year for the entire Antarctic continent. Plain lines are the mean of the distributions: (a) 414 km (blue) and (b) 0.157 (black).

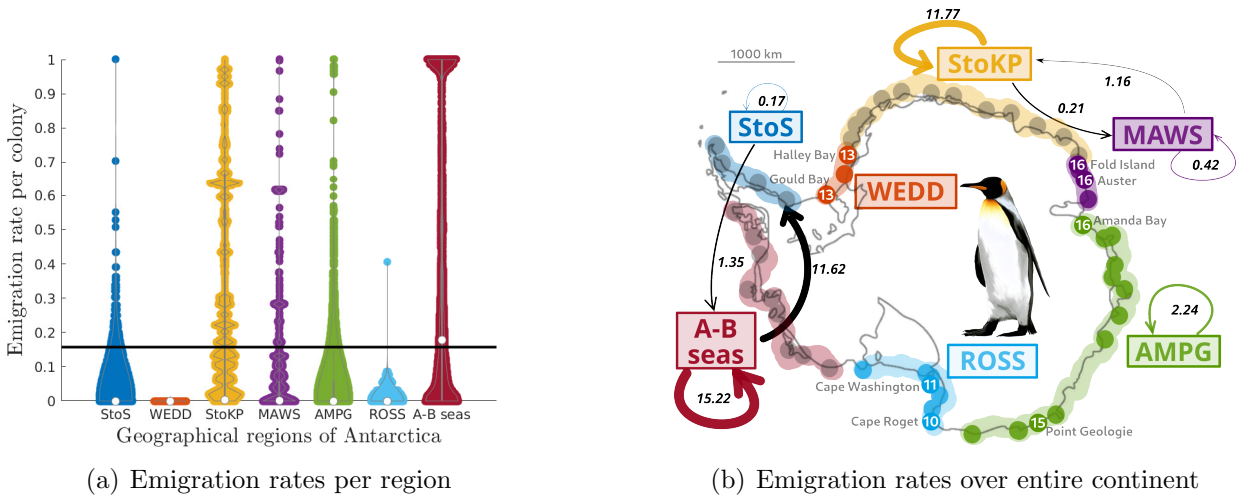
studies report traveling distance from 2,000 km to 7,000, with extreme distance of 9,000 km) [75, 45, 76].

**Dispersal rates.** Figure 2(b), summarizes the emigration rates throughout the continent and the supplementary Figure S3 details the emigration rates at each colony. The occurrence of dispersal events is not common, as indicated by a median of zero, since more than 50% of the calculated emigration rates in each colony are zero. However, some regions may experience massive emigration events with averaged rate of emigration that is not negligible: 15.7% per year.

Figure 3(a) shows the rate of emigration per colony, organized into clusters based on regions. Furthermore, Figure 3(b) provides an overview of the average emigration rate both within and between these regions. Emperor penguins move mainly to nearby colonies in the same regions with an average rate per year that varies between regions: 15% in colonies of the Amundsen and Bellingshausen seas (A-B seas) to 0.17% in colonies from Smith to Snowhill Island in the Wedell sea regions (StoS) (see Figure S4 for more details). However, massive emigration is also likely to occur between different regions, especially between the A-B seas regions and the StoS regions (11% to 1.35%) and colonies from Stancomb to Kloa Point (StoKP) and Mawson bay (MAWS) (1.16% to 0.21%).

### Potential drivers of dispersion

**Average rate of emigration.** The zooplankton biomass in the non-breeding season (January to March) stands out as the primary factor influencing the average probability of emigration (the proportion of years with non-zero median emigration rates between 2009 and 2013 in a colony) among all the evaluated environmental and demographic factors (see Figure 4(d)). Subsequently, the size of the colony was identified as the second most important factor influencing the average probability of emigration (Figure 4(d)). Both factors



**Figure 3:** 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 correspond 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 dots correspond to the 66 colonies of emperor penguins around Antarctica and the color shading indicate their geographical region. Grey dots indicate colonies without genetic information, while coloured dots corresponds to the four genetic clusters detected by [48]. The white numbers indicate the number of individuals sampled in colonies included in our study..



have a positive impact on the rate of emigration (Figure 4(a) and 4(c)). Despite considerable uncertainties, colonies without emigration were found to have an average size almost twice as large as colonies with emigration (3,741 penguins compared to 1,880 penguins for emigration, as shown in Figure S6).

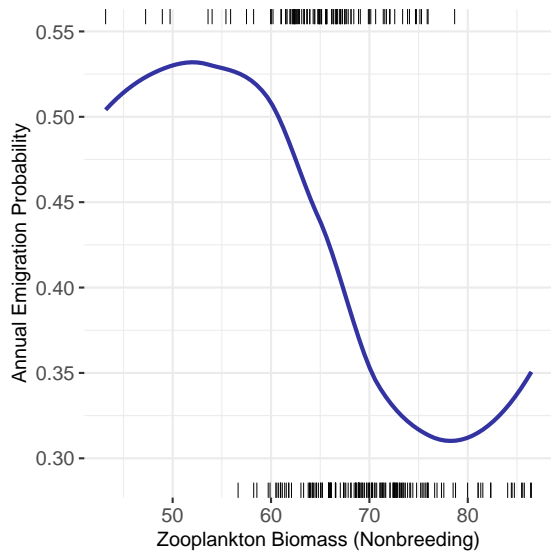
**Annual emigration probability.** Next, we refine our analysis by focusing solely on the annual probability of emigration (the probability of observing a non-zero median emigration rate in a given year and colony). Our findings indicate that proximity to the nearest fast ice edge positively affects the annual probability of emigration (refer to Figure 4(b)), ranking second in importance after zooplankton biomass (refer to Figure S5).

### Forecasts of emperor penguin global population

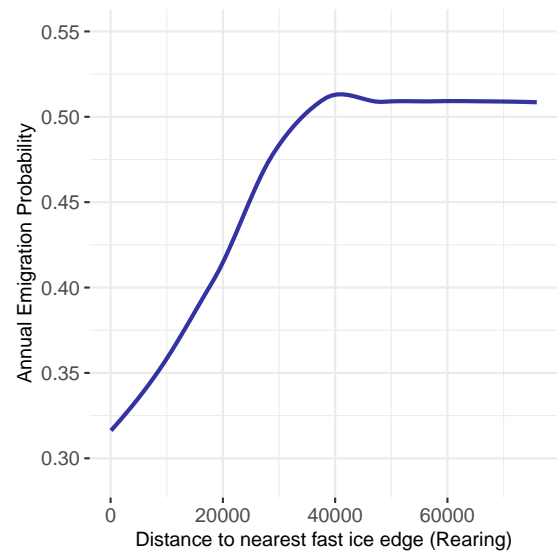
The emperor penguins are most likely to disperse through semi-informed dispersal with a small mean distance of 414 km and small emigration rates. This dispersal process is estimated to result in a greater global population, up to 7%, compared to a scenario without dispersion when climate scenarios lead to significant population declines (from scenario 4.3°C and scenario 2°C, as shown in Figure 5). However, under a climate scenario of 1.5°C [Paris 1.5°C], which causes lesser declines, our newly estimated dispersal processes do not improve the global population size compared to a scenario without dispersion, but may actually reduce it. Ultimately, the impact of dispersal on future global population size is relatively small compared to the impact of climate change mitigation [74, 43] (Figure 5).

## Discussion

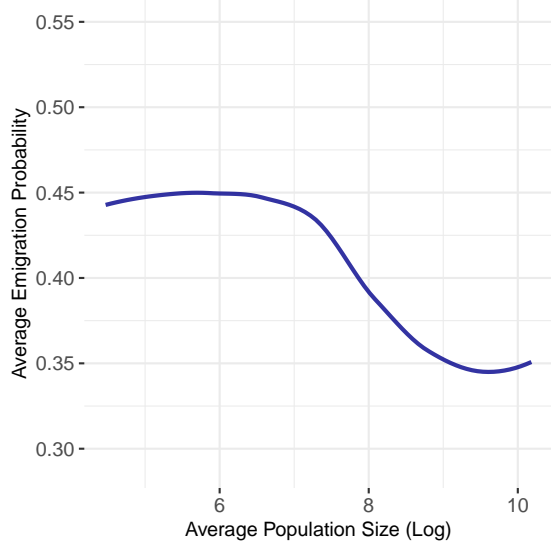
By developing an innovative mechanistic-statistical approach that integrates demographic and genetic information, we (1) elucidated the most common dispersal behavior used by emperor penguins to move from colony to colony around Antarctica, (2) quantified the mean distance at which individuals move to settle in a new colony (3) estimated the average emigration rates per colony. Our research showed that emperor penguins often abandon colonies with less favorable conditions and move randomly to other colonies. Although they generally exhibit low emigration rates, it is important to note that large-scale emigration events can occur in certain regions. Furthermore, their dispersal distances are relatively short. We discuss these novel estimations of dispersal processes in the context of potential individual movements documented via satellite telemetry tags and colony movements captured via VHR satellite imagery. We note that neither satellite telemetry tags nor VHR satellite imagery allow direct characterization of dispersal rates, distances, and behaviors but are the only information available to date to speculate about dispersal processes. Unraveling those dispersal processes will also reduce uncertainties in future population projections of emperor penguins necessary for ongoing conservation and management actions. By incorporating this new understanding of dispersion mechanisms into projection models [77], the prediction of the global population trend for emperor penguins, under different climate scenarios, reveals that the influence of dispersion is minimal compared to the influence of climate change mitigation.



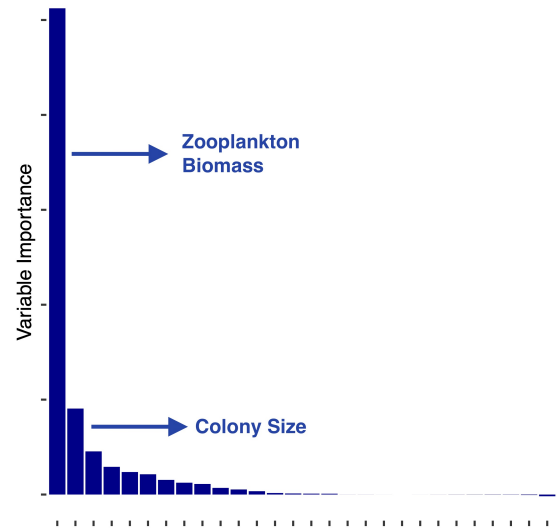
(a) Partial dependence plot for zooplankton Biomass



(b) Partial dependence plot for distance to nearest fast ice edge

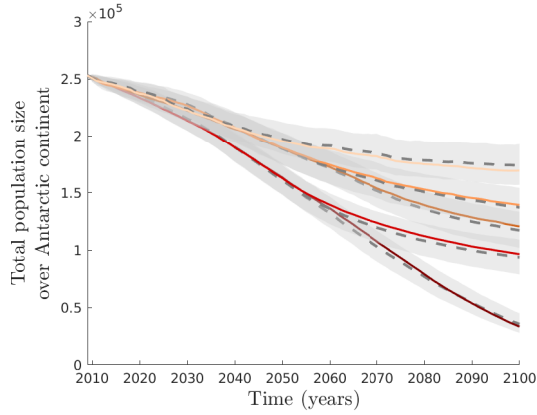


(c) Partial dependence plot for population size

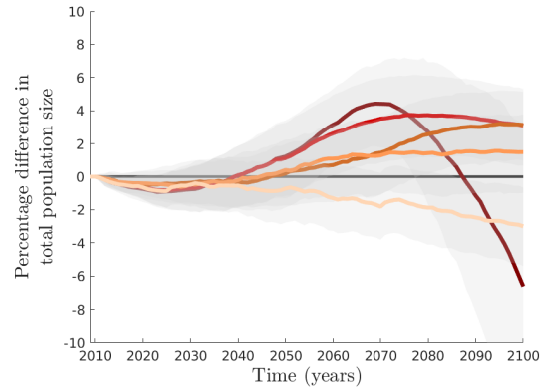


(d) Conditional variable importance

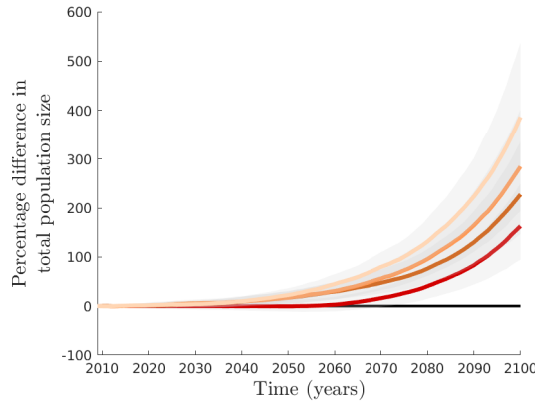
**Figure 4:** Impact of environmental and demographic factors on emigration rates of emperor penguins. (a) Zooplankton biomass ( $\text{mmol } C/m^2$ ) during the non-breeding period (January to March) have a negative effect on annual emigration probability (probability of observing a nonzero median emigration rate in a given year and colony); (b) Distance to nearest fast ice edge (meters) has a positive effect. Lines at the top and bottom of the graphs show the presence (median emigration rate  $> 0$ ) and absence (median emigration rate  $= 0$ ) of emigration, respectively, in a given year and colony; (c) the colony size has a negative effect on the average probability of emigration (proportion of years with non-zero median emigration rates between 2009 and 2013 in a colony); (d) Conditional variable importance scores of random forests modeling average emigration probability. Only the top 2 variables are shown.



(a) Total population size of EP



(b) Difference in dispersal behavior



(c) Difference in climate scenarios



**Figure 5:** Projection of the total population size of the emperor penguin from 2009 to 2100 using the metapopulation demographic model (see SI-A.2 ) 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 gray regions correspond to the confidence intervals 1% 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.

## New likelihood function

The accuracy of inferring how landscape affects dispersal depends on how different types of habitats or geographical elements influence movement [25]. Classical genetic methods based on simple dispersal assumptions (e.g., Euclidean distances or least-cost distances approaches [26, 27, 28]) can provide reliable estimates when the landscape features present

strong contrasts in permeability, even with limited sample sizes. However, as the landscape becomes more complex or when different features have similar effects on movement, the power and accuracy of these methods diminish rapidly. In particular, genetic distance-based methods are more effective at detecting variables that impede dispersal (e.g., barriers) than those that facilitate it, as impeding variables create a stronger signal of reduced gene flow. Consequently, in complex landscapes with diverse environmental structures, combining genetic information with detailed demographic descriptions is crucial to accurately evaluate the impact of landscape variables on dispersal [25]. This approach helps mitigate the limitations of genetic methods alone, especially when there are low contrasts between landscape variables, which can obscure the true effects of the landscape on dispersal pathways.

Recently, new mechanistic-statistical approaches, including reproduction, have been developed to estimate the dispersion of pest species ([14] for processionary moths and [78] for a bacteria) from genetic data. These approaches have the advantage of remaining relevant even if the degree of differentiation is low or the quantity of observed data is scarce [35]. However, these approaches do not account for temporal variations or dispersal behaviors, which is particularly important since some species use personal and social information to decide whether to leave a natal or current breeding site and where to settle (e.g. [41]). Such ‘informed dispersal’ behavior [30] enables individuals to settle in better quality habitats, potentially improving their fitness, thus increasing population viability and species persistence, especially in the face of global changes [42].

Here, we develop a novel likelihood function that incorporates informed departure and settlement behaviors, based on temporal and spatial variations in reproductive strategies and population dynamics. To achieve this level of complexity, the model is conditional on the assumption of a metapopulation being composed of four fixed genetic clusters at Hardy-Weinberg equilibrium, allowing us to focus solely on the evolution of the number of individuals within each cluster. However, we acknowledge that given our representative sampling approach, there could be more than four genetic clusters in the emperor penguin population, and our framework is based on this set of conditions.

Alternatively, for other systems, it may be more appropriate to compute the dynamics of allele frequencies at any time and location, though this approach is more computationally intensive. This dynamic can be modeled by adapting the methods of [9], but requires simulating a system that scales with the total number of alleles in the population – an unfeasible task for our study. Another option is to use multilocus likelihoods for hybrids and backcrosses as developed by [79], which involves tracking hybrid dynamics at each time step and computing the corresponding likelihoods.

Therefore, while these alternative approaches offer more flexibility with respect to population structure and do not necessarily assume fixed genetic clusters, they come with their own set of assumptions and greater computational demands. It is important to note that all models, including ours, are simplifications of reality and are conditional on a set of assumptions tailored to the specific goals and limitations of the study. Our chosen approach balances complexity and feasibility, which is particularly suited for the current research context of data-sparse environments.

## Four genetic clusters

Due to the species' breeding distribution across harsh and inaccessible areas, complete genetic sampling of all colonies is logistically infeasible. This limitation is not unique to emperor penguins, as restricted sampling is often the only viable option for studying the population genetic structure of many wildlife species. Although additional genetic clusters may exist in unsampled regions, previous studies suggest that geographic distance alone is not a reliable predictor of genetic structure in emperor penguins[80] and more broadly Southern Ocean penguins[81]. For instance, in emperor penguin, Amanda Bay and Pointe Géologie, located 3,200 km apart, belong to the same genetic cluster[48].

The observed genetic differentiation among the four detected genetic clusters in emperor penguins is likely explained by historical factors rather than contemporary geographic barriers. Previous studies indicate that only three populations of emperor penguins may have survived during the Last Glacial Maximum (LGM), with the Ross Sea acting as a critical refuge[80]. The LGM has profoundly influenced the genetic structure of many penguin species[81]. Indeed, consistent genome-wide signatures of post-LGM expansion have been detected in penguin species that currently breed south of the LGM sea ice zone, suggesting that many Southern Ocean species retreated to ice-free refugia during the LGM and rapidly recolonized high-latitude shores as the ice receded[81]. These historical refugia have likely shaped the present-day genetic structure of several penguin species, including emperor penguins[80].

We acknowledge that this representative sampling could be seen as a limitation for classical methods, such as BayesAss and coalescent models, which typically require extensive spatial sampling and high genetic differentiation between populations to estimate dispersal rates accurately. Unlike traditional methods that indirectly infer dispersal from genetic structure, our approach explicitly models dispersal dynamics using a combination of a demographic model and a genetic population model. Our demographic model describes the dispersal of all individuals across colonies, while the genetic population model tracks the lineage of individuals over time and assumes that a newborn inherits its genetic cluster from its mother. By integrating these demographic processes and focusing on the explicit movement of individuals, our model overcomes the limitations of traditional genetic methods that rely heavily on genetic differentiation and extensive sampling.

By tracking the movement and lineage of individuals across colonies, our model offers a valuable tool to understand how these populations may respond to ongoing climate change with dispersal behaviors, despite the constraints of limited genetic sampling. This approach sheds light on the complex dispersal dynamics of emperor penguins across Antarctica, contributing to a more comprehensive understanding of their connectivity and resilience in the face of environmental change.

## Dispersal ranges

The posterior distribution of the mean dispersal distance for the best supported model, depicted in Figure 2(a), suggests a short dispersal distance relative to the potential distance that tracked juveniles and adults cover after departing the colony [45, 76]. In fact, we found that the most likely dispersal distance of the emperor penguins is around 414 km

Satellite telemetry studies have shown that penguins can cover incredible distances during their searching routes. In the Ross Sea, non-breeders can travel up to 9,000 km [75] to their wintering grounds, and after the moult, adults covered more than 2,000 km on their return journey to their colonies [76]. In East Antarctica, one juvenile covered more than 7,000 km during the first eight months after leaving its natal colony in Terre Adélie [45].

However, the distance covered during the searching phase does not necessarily reflect the dispersal distance. In fact, individual potential dispersal can be reduced by specific behaviors. For example, seabirds exhibit specific behavioral traits, such as a high degree of philopatry [82] and the importance of social cues in the recruitment of new breeders [83], which can reduce the dispersal distance of individuals relative to possible movement [84, 85]. For example, the colonies in the Ross Sea are genetically distinct from the rest of the colonies [86], suggesting that, despite their large dispersal potential during the non-breeding season, the dispersal distance of the emperor penguin could be somewhat limited. Additionally, after the demise of Halley Bay, many of the birds of Halley Bay may have relocated to the nearby Dawson-Lambton colony, while the formation of new colonies elsewhere or movement to other locations of the colony further away are considered less likely [50].

## Dispersal rates

Previous studies have debated the magnitude of emigration rates in emperor penguins, with some studies arguing for large emigration rates [47], while others for low emigration rates [53, 46]. Although large-scale emigration events are possible, our research indicates that these occurrences are rare.

Large emigration rates producing 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) [49]; 2. Others are known to relocate to icebergs or ice shelves during late sea ice formation in the autumn [87]; 3. Some colonies have shown dramatic declines, while nearby colonies have increased in size markedly [50]. Those blinking, relocation, and massive movement events remain somewhat infrequent.

Spatially, 17% of the colonies are known to blink. In the past decade, certain colonies experienced intermittent periods of absence. Taking into account this fluctuation over time, the likelihood of a colony being absent in any given year is only 4%. Furthermore, it is anticipated that emigration rates will be low due to the significant number of marked chicks that have been observed to return to Pointe Géologie [46].

Nonetheless, it is crucial to establish the specific time and space frames in which these rates take place. In this study, we suggest that the overall yearly percentage of emperor penguins dispersing from one breeding site to another is relatively minimal. However, there might be instances of mass emigration occurring sporadically in certain locations. (see Figure 3(a) and Figure S3). This pattern has been noticed in numerous species of seabirds and birds, and it aligns with the philopatric behavior exhibited by these species. As an example, greater flamingos exhibit similar characteristics to emperor penguins in terms of their long lifespan and tendency to breed in one location (philopatric). Generally, they have a low rate of emigration, but when the conditions for breeding are poor at their colony, such as when water levels are low, they relocate together to another breeding location.

(see [88, 89, 90]).

In addition, our framework focuses on emigration rates at the population level. However, in many vertebrate species, especially in seabirds, juvenile dispersal is greater than adult dispersal [91]. For emperor penguins, massive emigration events are likely to consist mainly of adults in some regions, whereas the low background levels of emigration are likely dominated by juvenile dispersal. Further work should include this age structure in the dispersal demographic model to disentangle the dispersal rates of adults from those of juveniles. However, this would require understanding the detailed mechanisms of density dependence on those two age classes, which are unknown for emperor penguins [92].

Moreover, in cases where emigration and prospecting evolve simultaneously, the emigration strategy that emerges is one in which successful breeders consistently exhibit philopatry, while unsuccessful breeders are more inclined to emigrate, particularly when the breeding success of conspecifics is low [93]. This suggests that large-scale emigration events in emperor penguins may primarily involve unsuccessful breeders. Future research could incorporate a breeding stage structure into the dispersal demographic model to better understand the dispersal rates of both successful and unsuccessful breeders.

Finally, emigration rates vary substantially among regions. For example, the average annual emigration rate per colony between colonies in the A-B region is 15.7%. This high rate is likely driven by the lower habitat quality in the A-B region due to rapid declines in sea ice concentration [94]. The spring season of 2022 saw record low sea ice extent in Antarctica, with the greatest negative anomaly occurring in the central and eastern Bellingshausen Sea, west of the Antarctic Peninsula [95]. Some areas experienced a 100% loss in sea ice concentration during November, leading to widespread breeding failure of emperor penguin colonies [95]. These findings suggest that such extreme environmental changes are influencing movement patterns and demographic connectivity within this region.

This elevated emigration rate may suggest a genetically homogeneous population in the A-B region, potentially representing a new genetic cluster. However, due to the region's inaccessibility and the logistical challenges of sampling, genetic data are scarce, making it difficult to determine whether this population is part of one of the four known genetic clusters or constitutes a distinct cluster. This highlights the need for increased research efforts and enhanced sampling strategies in this understudied region to better understand the genetic structure and demographic connectivity of emperor penguins in the face of rapid environmental change.

### **Potential drivers of dispersal rates**

Zooplankton biomass serves as an indicator of the food sources available to emperor penguins and reflects the dynamics of the lower food web in the Antarctic ecosystem. Our study revealed that it is the main factor that influences dispersion rates (Figure S5). Specifically, we observed a negative relationship between zooplankton biomass during the nonbreeding period and the probability of annual emigration, as shown in Figure 3a. Consequently, when resources are abundant before breeding, emperor penguins are less inclined to leave their colony.

Variables related to fast ice also play a significant role in determining the likelihood of

emigration. Specifically, the annual probability of emigration is positively influenced by the distance to the nearest fast-ice edge ( Figure 4(b) and Figure S5). Consequently, emperor penguins are more inclined to leave their colony when it becomes more challenging to access open water. This positive relationship between distance to the nearest fast-ice edge and emigration probability has also been documented in relation to breeding success [96].

On the other hand, we observed that demographic factors did not have a significant impact on the average probability of emigration, except for population size (Figure 4(d)). The larger colonies were found to have a lower average probability of emigration (Figure 4(c)) and an annual probability of emigration (Figure S6). Despite expectations, smaller, declining, and frequently blinking colonies do not necessarily have higher emigration rates.

Although more work is needed to elucidate the proximate factors of suitable habitat and emigration rates of emperor penguins, our results suggest that massive emigration events occur in habitats with low food availability that cannot sustain large populations and in colonies that are distant from open water.

## **Dispersal behaviors and their consequences for the dynamics of the global population**

Based on our analysis, it is evident that emperor penguins are prone to migrating from colonies that have unfavorable habitats. They then randomly settle in another colony, a behavior known as semi-informed dispersal behavior [93] (see Table 1).

Previously, [58] 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 compared to a scenario without dispersal. However, here we show that high emigration rates and long-distance dispersal are unlikely for emperor penguins. We observed a limited dispersal distance and low average emigration rates, which will result in a slight increase in the global population compared to a scenario where dispersal does not occur. (see Figures 5(a)-(b)).

Nevertheless, the influence of dispersal behavior, distance, and emigration rate on the future global population size is relatively insignificant compared to the influence of climate change mitigation [74] (see Figure 5(c)). At the end of the century, there will be no suitable habitat if greenhouse gas emissions continue their current course, resulting in a large decline in the global population, regardless of dispersal processes[74]. To mitigate the rapid decline in its worldwide population, it is imperative to limit temperature increases to levels that are considerably below 2°C [43].

## **Conclusion**

By developing an innovative likelihood for a model that connects genetic information with metapopulation dynamics, we have discovered and measured previously unidentified dispersion patterns in emperor penguins from scarce genetic data. The application of this modeling approach has the potential to be used in various species and data-limited systems to uncover dispersal processes. It has the ability to enhance our understanding of the ranges, speeds, and behaviors of dispersal.



## References

- [1] Hastings A. Can spatial variation alone lead to selection for dispersal? *Theor Popul Biol.* 1983;24:244-51.
- [2] Clobert J, Ims RA, Rousset F. 13 - Causes, Mechanisms and Consequences of Dispersal. In: Hanski I, Gaggiotti OE, editors. *Ecology, Genetics and Evolution of Metapopulations*. Burlington: Academic Press; 2004. p. 307-35.
- [3] Bowler DE, Benton TG. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biol Rev.* 2005;80:205-25.
- [4] Levin SA, Muller-Landau HC, Nathan R, Chave J. The Ecology and Evolution of Seed Dispersal: A Theoretical Perspective. *Annual Review of Ecology, Evolution, and Systematics.* 2003;34(1):575-604.
- [5] Cayuela H, Rougemont Q, Prunier JG, Moore JS, Clobert J, Besnard A, et al. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Mol Ecol.* 2018;27(20):3976-4010.
- [6] Slatkin M. Gene Flow and the Geographic Structure of Natural Populations. *Science.* 1987;236(4803):787-92.
- [7] Bohonak AJ. Dispersal, Gene Flow, and Population Structure. *Q Rev Biol.* 1999;74(1):21-45.
- [8] Hewitt GM. The genetic legacy of the Quarternary ice ages. *Nature.* 2000;405:907-13.
- [9] Roques L, Garnier J, Hamel F, Klein EK. Allee effect promotes diversity in traveling waves of colonization. *Proc Natl Acad Sci USA.* 2012;109(23):8828-33.
- [10] McPeck MA, Holt RD. The Evolution of Dispersal in Spatially and Temporally Varying Environments. *Am Nat.* 1992;140(6):1010-27.
- [11] Olivieri I, Gouyon PH. 13 - Evolution of Migration Rate and Other Traits: The Metapopulation Effect. In: Hanski I, Gilpin ME, editors. *Metapopulation Biology*. Academic Press; 1997. p. 293-323.
- [12] Cadotte MW. Dispersal and Species Diversity: A Meta-Analysis. *Am Nat.* 2006;167(6):913-24.
- [13] Ronce O. How Does It Feel to Be like a Rolling Stone? Ten Questions about Dispersal Evolution. *Annual Review of Ecology, Evolution, and Systematics.* 2007;38:231-53.
- [14] Roques L, Soubeyrand S, Rousset J. A statistical-reaction-diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology.* 2011;274(1):43-51.

- [15] Turchin P. Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants. Sinauer Associates; 1998.
- [16] Ovaskainen O, Rekola H, Meyke E, Arjas E. Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. *Ecology*. 2008;89 2:542-54.
- [17] Southwood TRE, Henderson PA. *Ecological Methods*. Wiley; 2009.
- [18] Lagrange P, Pradel R, Bélisle M, Gimenez O. Estimating dispersal among numerous sites using capture—recapture data. *Ecology*. 2014;95(8):2316-23.
- [19] Nathan R, Perry G, Cronin JT, Strand AE, Cain ML. Methods for estimating long-distance dispersal. *Oikos*. 2003;103(2):261-73.
- [20] Hamrick JL, Trapnell DW. Using population genetic analyses to understand seed dispersal patterns. *Acta Oecologica*. 2011;37(6):641-9.
- [21] Robledo-Arnuncio JJ. Joint estimation of contemporary seed and pollen dispersal rates among plant populations. *Mol Ecology Resources*. 2012;12(2):299-311.
- [22] Robledo-Arnuncio JJ, García C. Estimation of the seed dispersal kernel from exact identification of source plants. *Mol Ecol*. 2007;16(23):5098-109.
- [23] Klein EK, Bontemps A, Oddou-Muratorio S. Seed dispersal kernels estimated from genotypes of established seedlings: does density-dependent mortality matter? *Methods in Ecol and Evol*. 2013;4.
- [24] Gelmi-Candusso TA, Bialozyt R, Slana D, Zárata Gómez R, Heymann EW, Heer K. Estimating seed dispersal distance: A comparison of methods using animal movement and plant genetic data on two primate-dispersed Neotropical plant species. *Ecol Evol*. 2019;9(16):8965-77.
- [25] Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N. Inferring landscape effects on dispersal from genetic distances: how far can we go? *Mol Ecol*. 2011;20(4):692-705.
- [26] Wright S. ISOLATION BY DISTANCE. *Genetics*. 1943;28(2):114-38.
- [27] Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 1997;145(4):1219-28.
- [28] Broquet T, Ray N, Petit E, Fryxell JM, Burel F. Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landsc Ecol*. 2006;21:877-89.
- [29] Stamps JA. Habitat selection by dispersers: proximate and ultimate approaches. In: Clobert J, Danchin E, Dhondt A, Nichols J, editors. *Dispersal*. New York: Oxford University Press; 2001. .

- [30] Clobert J, Le Galliard J, Cote J, Meylan S, Massot M. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol Lett.* 2009;12(3):197-209.
- [31] McRae BH. ISOLATION BY RESISTANCE. *Evolution.* 2006;60(8):1551-61.
- [32] Graves TA, Chandler RB, Royle JA, Beier P, Kendall KC. Estimating landscape resistance to dispersal. *Landsc Ecol.* 2014;29:1201-11.
- [33] Graves TA, Beier P, Royle JA. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Mol Ecol.* 2013;22(15):3888-903.
- [34] Lowe WH, Allendorf FW. What can genetics tell us about population connectivity? *Mol Ecol.* 2010;19(15):3038-51.
- [35] Roques L, Walker E, Franck P, Soubeyrand S, Klein EK. Using genetic data to estimate diffusion rates in heterogeneous landscapes. *Journal of Mathematical Biology.* 2016;73(2):397-422.
- [36] Berliner LM. Physical-statistical modeling in geophysics. *J Geophys Res.* 2003;108:8776.
- [37] Wikle CK. Hierarchical models in environmental science. *International Statistical Review.* 2003;71:181-99.
- [38] Soubeyrand S, Roques L. Parameter estimation for reaction-diffusion models of biological invasions. *Popul Ecol.* 2014;56(2):427-34.
- [39] Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J. State-space models of individual animal movement. *Trends in Ecol & Evol.* 2008;23(2):87-94.
- [40] Durbin J, Koopman S. *Time Series Analysis by State Space Methods.* Oxford University Press; 2012.
- [41] Doligez B, Danchin E, Clobert J. Public Information and Breeding Habitat Selection in a Wild Bird Population. *Science.* 2002;297(5584):1168-70.
- [42] Ponchon A, Garnier R, Grémillet D, Boulinier T. Predicting population response to environmental change: the importance of considering informed dispersal strategies in spatially structured population models. *Divers Distrib.* 2015;21:88-100.
- [43] Jenouvrier S, Che-Castaldo J, Wolf S, Holland M, Labrousse S, LaRue M, et al. The call of the emperor penguin: Legal responses to species threatened by climate change. *Global change biology.* 2021;27(20):5008-29.
- [44] Barbraud C, Weimerskirch H. Emperor penguins and climate change. *Nature.* 2001;6834(411):183-6.

- [45] Thiebot JB, Lescroël A, Barbraud C, Bost CA. Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarct Sci.* 2013;25(4):536-44.
- [46] Mougin JL, Van Beveren M. Structure et dynamique de la population de manchots empereurs *Aptenodytes Forsteri* de la colonie de l'archipel de Pointe Géologie, Terre Adélie. *CR Acad Sci.* 1979;289(2):157-60.
- [47] LaRue MA, Kooyman G, Lynch HJ, Fretwell P. Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography.* 2015;38(2):114-20.
- [48] Younger JL, Clucas GV, Kao D, Rogers AD, Gharbi K, Hart T, et al. The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Mol Ecol.* 2017;26(15):3883-97.
- [49] Fretwell PT, Trathan PN. Discovery of new colonies by Sentinel2 reveals good and bad news for emperor penguins. *Remote Sensing in Ecology and Conservation.* 2021;7(2):139-53.
- [50] Fretwell PT, Trathan PN. Emperors on thin ice: Three years of breeding failure at Halley Bay. *Antarctic Science.* 2019;31(3):133-8.
- [51] Rodgers KB, Lee SS, Rosenbloom N, Timmermann A, Danabasoglu G, Deser C, et al. Ubiquity of human-induced changes in climate variability. *Earth System Dynamics.* 2021;12(4):1393-411.
- [52] Soubeyrand S, Neuvonen S, Penttinen A. Mechanical-statistical modeling in ecology: from outbreak detections to pest dynamics. *Bull Math Bio.* 2009;71:318-38.
- [53] Prevost J. *Ecologie du manchot empereur.* In: *Expeditions polaires francaises.* Hermann Press, Paris, France; 1961. .
- [54] Thiebot JB, Lescroël A, Barbraud C, Bost CA. Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarctic Science.* 2013;25(4):536-44.
- [55] S L, F O, AR S, C B, et al. First odyssey beneath the sea ice of juvenile emperor penguins in East Antarctica. *Mar Ecol Prog Ser.* 2019;609:1-16.
- [56] Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multi-locus Genotype Data. *Genetics.* 2000;155(2):945-59.
- [57] Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol.* 2016;25(13):2967-77.

- [58] Jenouvrier S, Garnier J, Patout F, Desvillettes L. Influence of dispersal processes on the global dynamics of Emperor penguin, a species threatened by climate change. *Biol Conserv.* 2017;212:63-73.
- [59] Jenouvrier S, Holland M, Stroeve J, Serreze M, Barbraud C, Weimerskirch H, et al. Projected continent-wide declines of the emperor penguin under climate change. *Nature Climate Change.* 2014;4(8):715.
- [60] Fretwell PT, Trathan PN. Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography.* 2009;18(5):543-52.
- [61] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2002;64(4):583-639.
- [62] Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. *Bayesian Data Analysis.* Chapman & Hall / CRC; 2003.
- [63] Ando T. Predictive Bayesian Model Selection. *American Journal of Mathematical and Management Sciences.* 2011;31:13-38.
- [64] Fraser AD, Massom RA, Handcock MS, Reid P, Ohshima KI, Raphael MN, et al. Eighteen-year record of circum-Antarctic landfast-sea-ice distribution allows detailed baseline characterisation and reveals trends and variability. *The Cryosphere.* 2021;15(11):5061-77.
- [65] Long MC, Moore JK, Lindsay K, Levy M, Doney SC, Luo JY, et al. Simulations With the Marine Biogeochemistry Library (MARBL). *Journal of Advances in Modeling Earth Systems.* 2021;13(12).
- [66] LaRue M, Brooks C, Wege M, Salas L, Gardiner N. High-resolution satellite imagery meets the challenge of monitoring remote marine protected areas in the Antarctic and beyond. *Conservation Letters.* 2022;15(4):e12884.
- [67] LaRue M, Iles I, Labrousse S, Fretwell P, Ortega D, Devane E, et al. Advances in remote sensing of emperor penguins: first multi-year time series documenting trends in the global population. *Proceedings of the Royal Society B.* 2024.
- [68] Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics.* 2007;8(1).
- [69] Debeer D, Strobl C. Conditional permutation importance revisited. *BMC Bioinformatics.* 2020;21(1).
- [70] Jenouvrier S, Caswell H, Barbraud C, Holland M, Stroeve J, Weimerskirch H. Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proceedings of the National Academy of Sciences.* 2009;106:1844-7.

- [71] Jenouvrier S, Garnier J, Patout F, Desvillettes L. Influence of dispersal processes on the global dynamics of Emperor penguin, a species threatened by climate change. *Biological Conservation*. 2017 Aug;212:63-73. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0006320716306218>.
- [72] DuVivier AK, Molina MJ, Deppenmeier AL, et al. Projections of winter polynyas and their biophysical impacts in the Ross Sea Antarctica. *Clim Dyn*. 2024;62.
- [73] Iles D, Jenouvrier S. 12. In: *Projected population consequences of climate change*. In: *Effects of Climate Change on Birds*. Oxford University Press; 2019. .
- [74] Jenouvrier S, Holland M, Iles D, Labrousse S, Landrum L, Garnier J, et al. The Paris Agreement objectives will likely halt future declines of emperor penguins. *Global change biology*. 2020;26(3):1170-84.
- [75] Goetz KT, McDonald BI, Kooyman GL. Habitat preference and dive behavior of non-breeding emperor penguins in the eastern Ross Sea, Antarctica. *Mar Ecol Prog Ser*. 2018;593:155-71.
- [76] Kooyman GL, Siniff D, Stirling I, Bengtson J. Moulting habitat, pre-and post-moulting diet and post-moulting travel of Ross Sea emperor penguins. *Mar Ecol Prog Ser*. 2004;267:281-90.
- [77] Jenouvrier S, Garnier J, Patout F, Desvillettes L. Influence of dispersal processes on the global dynamics of Emperor penguin, a species threatened by climate change. *Biological Conservation*. 2017;212:63-73.
- [78] Roques L, Desbiez C, Berthier K, Soubeyrand S, Walker E, Klein EK, et al. Emerging strains of watermelon mosaic virus in Southeastern France: model-based estimation of the dates and places of introduction. *Scientific Reports*. 2021;11.
- [79] Anderson EC, Thompson EA. A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics*. 2002;160(3):1217-29.
- [80] Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, 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 Biology*. 2015;21(6):2215-26.
- [81] Cole TL, Dutoit L, Dussex N, Hart T, Alexander A, Younger JL, et al. Receding ice drove parallel expansions in Southern Ocean penguins. *Proceedings of the National Academy of Sciences*. 2019;116(52):26690-6.
- [82] Aebischer NJ, Coulson JC. Survival of the Kittiwake in Relation to Sex, Year, Breeding Experience and Position in the Colony. *Journal of Animal Ecology*. 1990;59(3):1063-71.
- [83] Reed JM, Boulinier T, Danchin E, Oring LW. In: Nolan V, Ketterson ED, Thompson CF, editors. *Informed Dispersal*. Springer US; 1999. p. 189-259.

- [84] Kildaw SD, Irons DB, Nysewander DR, Buck CL. Formation and growth of new seabird colonies: the significance of habitat quality. *Marine Ornithology*. 2005;33:49-58.
- [85] Matthiopoulos J, Harwood J, Thomas L. Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology*. 2005;74(4):716-27.
- [86] Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, *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*. 2015;21(6):2215-26.
- [87] Fretwell PT, Trathan PN, Wienecke B, Kooyman GL. Emperor Penguins Breeding on Iceshelves. *PLoS ONE*. 2014;9:e85285.
- [88] Nager RG, Johnson AR, Boy V, Rendon-Martos M, Calderon J, Cézilly F. Temporal and spatial variation in dispersal in the greater flamingo (*Phoenicopterus ruber roseus*). *Oecologia*. 1996;107(2):204-11.
- [89] Balkiz O, Béchet A, Rouan L, Choquet R, Germain C, Amat JA, *et al*. Experience-dependent natal philopatry of breeding greater flamingos. *Journal of Animal Ecology*. 2010;79(5):1045-56.
- [90] Johnson A, Cézilly F. *The greater flamingo*. A & C Black; 2007.
- [91] Clobert J, Ims RA, Rousset F. Causes, mechanisms and consequences of dispersal. In: *Ecology, genetics and evolution of metapopulations*. Elsevier; 2004. p. 307-35.
- [92] Jenouvrier S, Holland M, Stroeve J, Barbraud C, Weimerskirch H, Serreze M, *et al*. Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology*. 2012 Jul;18:2756-70.
- [93] Ponchon A, Scarpa A, Bocedi G, Palmer SCF, Travis JMJ. Prospecting and informed dispersal: Understanding and predicting their joint eco-evolutionary dynamics. *Ecol Evol*. 2021;11(21):15289-302.
- [94] Abram NJ, Thomas ER, McConnell JR, Mulvaney R, Bracegirdle TJ, Sime LC, *et al*. Ice core evidence for a 20th century decline of sea ice in the Bellingshausen Sea, Antarctica. *Journal of Geophysical Research: Atmospheres*. 2010;115(D23).
- [95] Fretwell PT, Boutet A, Ratcliffe N. Record low 2022 Antarctic sea ice led to catastrophic breeding failure of emperor penguins. *Communications Earth & Environment*. 2023;4(1):273.
- [96] Labrousse S, Fraser AD, Sumner M, Le Manach F, Sauser C, Horstmann I, *et al*. Land-fast ice: A major driver of reproductive success in a polar seabird. *Biology Letters*. 2021;17(6):20210097.