Partitioning prediction uncertainty in climate-dependent population models

Gilles Gauthier¹, Guillaume Péron²,³, Jean-Dominique Lebreton⁴, Patrick Grenier⁵ and Louise van Oudenhove¹,⁶

¹Département de Biologie and Centre d’Études Nordiques, Université Laval, 1045 avenue de la Médecine, Quebec, Quebec, Canada G1V 0A6
²Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630, USA
³UMR CNRS 5558 — LBBE ‘Biométrie et Biologie Évolutive’ UCB Lyon 1, Bât. Grégor Mendel, 43 bd du 11 novembre 1918, 69622 Villeurbanne Cedex, France
⁴UMR 5175, Centre d’écologie fonctionnelle et évolutionaire, CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
⁵Groupe Scénarios et services climatiques, Ouranos, 550 rue Sherbrooke Ouest, Montréal, Quebec, Canada H3A 1B9
⁶UMR 1355, INRA, Institut Sophia Agrobiotech, 400 Route des Chappes, 06903 Sophia Antipolis, France

The science of complex systems is increasingly asked to forecast the consequences of climate change. As a result, scientists are now engaged in making predictions about an uncertain future, which entails the efficient communication of this uncertainty. Here we show the benefits of hierarchically decomposing the uncertainty in predicted changes in animal population size into its components due to structural uncertainty in climate scenarios (greenhouse gas emissions and global circulation models), structural uncertainty in the demographic model, climatic stochasticity, environmental stochasticity unexplained by climate–demographic trait relationships, and sampling variance in demographic parameter estimates. We quantify components of uncertainty surrounding the future abundance of a migratory bird, the greater snow goose (Chen caerulescens atlantica), using a process-based demographic model covering their full annual cycle. Our model predicts a slow population increase but with a large prediction uncertainty. As expected from theoretical variance decomposition rules, the contribution of sampling variance to prediction uncertainty rapidly overcomes that of process variance and dominates. Among the sources of process variance, uncertainty in the climate scenarios contributed less than 3% of the total prediction variance over a 40-year period, much less than environmental stochasticity. Our study exemplifies opportunities to improve the forecasting of complex systems using long-term studies and the challenges inherent to predicting the future of stochastic systems.

1. Introduction

Prediction is difficult, especially about the future (attributed to Niels Bohr)

Climate warming is already impacting individuals, populations and food webs across most ecosystems of the Earth [1,2] and this is likely to accelerate in coming decades. To anticipate impacts, researchers often rely on models to generate plausible future trajectories. Such models can be classified into two broad categories: phenomenological and mechanistic models [3,4]. Phenomenological models are derived from statistical models (e.g. regressions) and are not based on a mechanistic understanding of the ecological processes driving the observed changes. An example of that is the forecasting of changes in species distributions based on anticipated changes in the location of the climate isotherms that delimit their current range [3,6]. Mechanistic models, by contrast, are based on biological processes describing pathways through which climate affects ecosystem function [3,7]. Physiological tolerance models are an example of mechanistic models for forecasting species distributions [8]. Models based on
demographic processes are also used to forecast change in population size and distribution. An early example of that is a study with the emperor penguin (Aptenodytes forsteri) where the climate was treated as a two-state variable with demographic parameters specific to each of these two states [9]. More recent and sophisticated attempts to couple population dynamic models with climate warming scenarios are reviewed by Jenouvrier [10]. However, such studies are still scarce because they require long-term demographic data coupled with a detailed knowledge of the functional relationships between climatic factors and vital rates, and a good understanding of the underlying mechanisms [10–12].

In both approaches, any predicted trajectory derived from ecological models involves uncertainties, an important issue that has not always received full attention [13]. Failure to adequately measure or consider uncertainties associated with model-based predictions of future change in animal abundance can lead to misguided decisions by policymakers or managers [3,14,15]. Pragmatically, partitioning the sources of uncertainties can also be used to direct future research aimed at improving the precision of model predictions. However, the magnitude and relative contributions of various sources of uncertainty to the overall uncertainty of the prediction are rarely quantified in ecological models [13,16,17]. A rare example is provided by a study that assessed the relative contribution of uncertainties in future greenhouse gas emissions and in global circulation models (GCMs) to forecasted changes in the abundance of a bryophyte species, while also considering but not specifically analysing the uncertainty created by the demographic model [18].

In this paper, we develop a time-varying matrix population model for a migratory bird with temperature dependencies in demographic parameters coupled with climate warming scenarios. Our primary aim was to hierarchize the various sources of uncertainty contributing to the error in model predictions and assess their relative importance. This allowed us to project the expected change in population size of our study species over four decades under contrasted climate scenarios and their associated error. Here the terms ‘prediction’ and ‘forecast’ refer to population trajectories issued from our model and projected into the future; each separate prediction should be interpreted as a plausible trajectory among a large ensemble of such trajectories. The background for this work is the management of the growing greater snow goose population (Chen caeculescens atlantica), a population currently exposed to a strong warming trend [19,20], especially on its breeding ground due to the amplification of climate warming in the Arctic [21,22].

2. Material and methods

(a) Study system

The greater snow goose is a migratory bird that breeds in the High Arctic, primarily in the eastern Canadian Arctic Archipelago, winters along the US Atlantic coast and stages twice a year in southern Canada along the St Lawrence River in the province of Quebec [19]. The population is hunted during the autumn and winter in southern Canada and the USA. A photographic census conducted every spring in Quebec has provided a relatively accurate estimate of population size each year [23]. The population size has increased from approximately 25,000 birds in 1965 to 1,000,000 birds in 1999, partly due to food-subsidy obtained from feeding in farmland in winter and spring [19]. Because of its high abundance and potential for inducing ecological imbalance on its Arctic breeding grounds, exceptional measures were implemented to stop the growth of this population, starting in 1999 (liberalized hunting regulations and introduction of a spring hunt, first in southern Canada and more recently in the USA [19,24]).

The demography of this population has been studied since 1990 at its largest known breeding colony in the Canadian Arctic, Bylot Island, Nunavut [25]. Reproductive success has been determined annually from a sample of several hundred nests and several thousand birds have been captured and marked every year in mass-banding drives when they are molting. All birds were leg-banded and a sample of adult females was also marked with a neck collar [26] but in this study we did not use the information from neck-collared females. Several vital rates (survival, recruitment and breeding propensity) were estimated using the resulting capture-recapture data combining live recaptures in the summer and dead recoveries from hunters on the staging and wintering grounds in previous analyses [25,27].

(b) Population model

We built a pre-breeding stage-based population model [28] with five pre-breeding stages (PB1–PB5 for ages 1–5; all individuals that skipped breeding were forced back into breeding) and one non-breeding stage (NB). We modelled only the female component, assuming an equal sex-ratio at birth. The transition from year t (columns) to year t + 1 (lines) was described by the following transition matrix:

\[
\begin{pmatrix}
PB1 & PB2 & PB3 \\
PB1 & 0 & 0 & 0 \\
PB2 & S_1 \cdot (1 - b_2) & 0 & 0 \\
PB3 & 0 & S_2 \cdot (1 - b_3) & 0 \\
PB4 & 0 & 0 & S_3 \cdot (1 - b_{4.5}) \\
PB5 & 0 & 0 & 0 \\
B & S_1 \cdot b_2 & S_2 \cdot b_3 & S_2 \cdot b_{4.5} \\
NB & 0 & 0 & 0 \\
\end{pmatrix}
\]

where \(F\) denotes fecundity and was decomposed into the product of clutch size, proportion of nests where at least one egg survives to hatch, egg survival in nests where at least one egg survives to hatch, hatching probability of surviving eggs and gosling survival from hatching to near fledging; \(b_a\) denotes age-specific recruitment (probability of breeding for the first time at age \(a\) conditional on survival); \(\hat{k}\) denotes the probability to skip a breeding season for a female that had already bred (individuals that skipped breeding were forced back into breeding the following year; see electronic supplementary material,
Table 1. Decomposition of the sources of variation in projecting future population size when coupling a demographic model to climate scenarios. \(N_0\), population size vector; \(M\), transition matrix model; \(X_t\), vector of climatic covariates linking specific vital rates (elements \(m_{ij}\) of the matrix \(M\)) to temperature (typically of the form \(m_{ij} = \beta_0 + \beta_1 \cdot X_t\)).

<table>
<thead>
<tr>
<th>step</th>
<th>source of variation</th>
<th>equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>time-averaged deterministic population model</td>
<td>(N_{t+1} = M \times N_t)</td>
</tr>
<tr>
<td>2</td>
<td>deterministic model with time-varying climate-dependent covariates</td>
<td>(N_{t+1} = M(X_t) \times N_t)</td>
</tr>
<tr>
<td>3</td>
<td>structural uncertainty in the demographic model(^a)</td>
<td>(N_{t+1} = M(X_t) \times N_t)</td>
</tr>
<tr>
<td>4</td>
<td>uncertainty in the climate scenarios(^b)</td>
<td>(N_{t+1} = M(X_t) \times N_t)</td>
</tr>
<tr>
<td>5</td>
<td>stochasticity in the climate projection model ((\xi_{t,j}))</td>
<td>(N_{t+1} = M(X_t + \xi_{t,j}) \times N_t)</td>
</tr>
<tr>
<td>6</td>
<td>residual environmental stochasticity ((\epsilon_t))</td>
<td>(N_{t+1} = (M(X_t + \xi_{t,j}) + \epsilon_t) \times N_t)</td>
</tr>
<tr>
<td>7</td>
<td>sampling variance of climate–vital rates relationships ((\xi))</td>
<td>(N_{t+1} = (M(X_t + \xi) + \epsilon_t) \times N_t)</td>
</tr>
<tr>
<td>8</td>
<td>demographic stochasticity ((\eta_t))(^a)</td>
<td>(N_{t+1} = (M(X_t + \xi) + \epsilon_t + \xi_t) \times N_t)</td>
</tr>
</tbody>
</table>

\(^a\)These steps were ignored in our case study.

\(^b\)In our application, this step includes both uncertainty in the structure of the global circulation model and in the representative concentration pathway (future greenhouse gas emissions). \(k = 1, \ldots, 63\) in our application to greater snow geese.

Methods; and \(S_t\) denotes age-specific yearly survival probability (notation \(a+\) represents all individuals of age \(a\) or more). Except for age-specific recruitment, we did not consider improvement in reproductive success with age because it is unknown in this population (i.e. fecundity parameters are averaged over all age classes). Survival through the first year of life (\(S_0\)) was adjusted to a 10-month period in the matrix model (\(S_t^{0.12}\)) because survival was estimated from one late-summer to the next but the matrix model used a pre-breeding census formulation [23]. In this hunted species, survival was decomposed into survival to hunting (\(1 - h_t\), where \(h_t\) = harvest rate) and survival to other sources of mortality (\(s_t\) with \(S_t = s_t(1 - h_t)\) following Lebreton [29]. Survival to hunting was estimated as a linear function of measured harvest rate, which accommodates both partial compensation of hunting mortality [30] and an imperfect measure of harvest rate [25].

Parameter estimates came from a previous capture–recapture analysis [25] and included the following temperature dependencies:

- a positive effect of spring temperature in the Arctic breeding ground on clutch size;
- a positive effect of early summer temperature in the Arctic breeding ground on hatching probability;
- a negative effect of a combined index of spring temperature in the staging area of southern Quebec and in the Arctic breeding ground on the probability to skip breeding;
- a negative effect of a combined index of temperature during the non-breeding season in the staging area of southern Quebec and wintering area the eastern USA on recruitment probability; and
- a positive effect of late-summer temperature in the Arctic breeding ground on first-year natural survival.

Model parameters also included dependencies on harvest rate, as described above, and post-1999 special management measures aimed at increasing harvest rate [25] (electronic supplementary material, Methods). Harvest rate (\(h_t\)) was derived from annual hunter surveys and population size estimates [25]. The model was fitted using the 1990–2011 time series of estimated harvest rates (yearly range for adults: 4–13%, yearlings: 20–50%). In the projections, we used the average of the 1999–2011 values when a spring hunt was present (adults: 10%, yearlings: 34%) and of the 1990–1997 values in the absence of a spring hunt (adults: 7%, yearlings: 33%). Introduction of a spring hunt to control population size had negative effects on fecundity due to hunting disturbance and we accounted for that in our model (electronic supplementary material, Methods). In our projections, the effects of the spring hunt on hunting-related mortality and fecundity were removed each time the population fell below 400,000 females, the population objective, to mimic the management decision process for this population. We did not include further density-dependence in the model because population growth was still exponential when special management actions were introduced in 1999 [31]. Finally, survival probabilities estimated from the capture–recapture data were adjusted upwards because a previous analysis of this dataset found an underestimation of survival in the capture–recapture dataset [23] (electronic supplementary material, Methods).

(c) Sources of uncertainty and simulation strategy

We decomposed the uncertainty surrounding the future abundance of a population in a hierarchical manner (table 1). In our case study, those uncertainty components were estimated in a stepwise, nested fashion, following the order presented in table 1. We started with a purely deterministic matrix population model [28] in Steps 1 and 2.

Step 3 represents the structural uncertainty on the demographic model (i.e. uncertainty regarding the nature of the demographic processes at play in a study population). This step could be ignored in our case study because previous work had brought unambiguous support for a single general population model describing the age structure and seasonality of survival and fecundity variation [25]. Climate–demographic relationships were modelled within the generalized linear mixed model framework [32], leaving little structural uncertainty regarding covariate effects on model parameters. We nonetheless explored the consequences of using an inadequate, biased demographic model on the uncertainty of model predictions in a formal analysis (electronic supplementary material, Theoretical exploration of prediction variance components); we come back to this issue in the discussion.

In Step 4, we addressed the structural uncertainty pertaining to both the choice of a GCM and to future anthropogenic greenhouse gas emissions (representative concentration pathways or RCP), collectively termed ‘uncertainty in the climate scenario’ for short hereafter. We used a set of 63 climate scenarios for this purpose (information on these scenarios is provided in electronic supplementary material, Climate scenarios).

In Step 5, we represented the inter-annual variability in weather realizations under given climate scenarios, termed ‘climate stochasticity’ for short hereafter. We simulated 40 weather realizations for each climate scenario, using autoregressive
Figure 1. Observed (black line) and predicted (dashed line) population trajectory from 1990 to 2011 (grey area: 95% confidence interval from [23]) and mean projected trajectories from 2012 to 2050 (green, cyan and blue lines) under three scenarios of future anthropogenic greenhouse gas emissions (representative concentration pathways or RCP). Coloured area: 90% projection intervals hierarchically decomposed according to the various sources of uncertainties in a cumulative fashion (see table 1 for definitions), and truncated at 10^6 females to zoom on the area of interest. Each new colour represents the additive variance caused by the focal source of variance (i.e. nested structure). The horizontal grey line is the threshold below which special management actions implemented to control the population (spring hunt) are stopped. The probability density curves in the right panel represent the distributions of projected population sizes in 2050 when considering only the uncertainty in the climate scenario (yellow) or all sources of uncertainty (red). The difference between the modal values of the two distributions is explained by the approximate lognormal distribution of population sizes.

3. Results and discussion

For the observed period (1990–2011; 22 years), the model effectively tracked the observed changes in abundance and remained within the 95% confidence interval of the latter in all years (figure 1). In particular, it discriminated the two phases of the population’s dynamics: exponential growth until 1999 (mean annual population growth rate, observed: 12%, predicted: 10%) and relative stability after the introduction of the special management actions in 1999 (mean annual population growth rate, observed: −1%, predicted: 1%). Observed, but not forecasted population size fluctuated annually after 1999 (figure 1), suggesting that some of the observed variation was due to sampling uncertainty in the population count data. Indeed, the sampling variance of the count is proportional to population size [23] making recent counts less reliable than those from the beginning of the study period.

For the forecasting period 2012–2050, our model based on climate warming scenarios predicted, on average, a slow population increase at a rate of 0.7% per year for the next 40 years, with a tendency for a slowing down of the rate of increase in the later period (figure 1). Despite the strong warming trend that geese should experience throughout their annual range in the coming decades (electronic supplementary material, Climate scenarios), the predicted increase in population size is relatively modest. This is likely because our previous analysis [25] showed that temperature has a positive effect on demographic parameters at some periods of the annual cycle but a negative effect at others. This important result emphasizes the need to consider climatic effects during the full annual cycle of a species when modelling future population changes [11,33]. It also suggests that previous analyses based on phenomological models and suggesting rapid population expansion of some goose populations in response to a warming Arctic climate [34] may have been overly optimistic. Our analysis shows that the uncertainty surrounding future population abundance of greater snow geese accumulated very rapidly (figure 1). The distributions of projected population sizes in 2050 had a broad peak, which further emphasizes the large uncertainty associated with the predictions. As expected from the exponential structure of the model [35] (electronic supplementary material, Theoretical exploration of prediction uncertainty, as in any statistical regression. We resampled in the sampling variance–covariance matrix extracted from the analysis of van Oudenhove et al. [25] with some constraints (electronic supplementary material, Methods). We generated 40 parameter values for each expected value obtained at Step 6, bringing the total number of simulations to 4.032.000.

We did not consider demographic stochasticity in this study (Step 8) because it is negligible compared to the other sources of variation when population size is very large [28]. The mean population trajectory was calculated across all the simulations. To help the interpretation of the results, we expressed the total prediction variance as a sum of the variance components described in table 1.
variability (see table 1 for definition) to the empirical prediction variance.

Figure 2. Change over time in the relative contribution of the various sources of uncertainty to the empirical prediction variance.

never contributed more than 3% of the total prediction variance (figure 2), putting a clear emphasis on our limited ability to predict the demography of wild species for more than a few years ahead based on climate projections alone.

Predictions are obviously conditional on the population model used to generate them and on the range of covariate values available to calibrate it. In our well-studied population [23,25–27], we were able to work with a single model that received strong support from the data. However, this may not always be the case in less intensively studied species, in which several models could represent the available data equally well. This would correspond to Step 3 of our decomposition of prediction variance (table 1). Structural uncertainty in the demographic model, also known as model uncertainty or uncertainty in model selection, can be addressed in several ways. First, it can be accounted for via model averaging, a procedure in which the relative support that each model receives is incorporated into the computation of confidence intervals around demographic parameter estimates [37]. This link between confidence intervals and structural uncertainty shows than one can expect structural (model) uncertainty to produce effects similar to sampling (parameter) uncertainty. A second, more explicit way of representing structural uncertainty is to consider biases in the parameter estimates, as done in our formal analysis (electronic supplementary material, Theoretical exploration of prediction variance components). This analysis shows that the variance associated with structural uncertainty, expressed as biases in the estimates, should indeed increase at the same rate as the sampling variance over time, and should be an additional source of concern when modelling the population dynamics of species with a poorly known demography.

Our insistence on the decomposition of the uncertainty into a series of components that are as comprehensive as possible is critical to direct future efforts to reduce this uncertainty. This is important because scientists dealing with complex systems [38,39] are increasingly asked to forecast the consequences of climate change [3,40,41]. We showed that sampling variance was the largest source of variation but it is also, fortunately, the one upon which researchers have potentially the most control. Reducing this uncertainty could be achieved by increasing sample sizes, increasing recapture rates in capture–recapture studies such as ours, periodically updating climate–demographic relationships as new data becomes available, and effectively combining multiple data sources [23].

Using relationships between demographic parameters and climatic covariates and feeding these relationships into a matrix population model as we did is akin to using transfer functions in time series analysis [42]. Two major differences, however, are that data used to estimate the parameters of the transfer function (i.e. the capture–recapture data) are independent of the time series itself (the population count) and in our study, we scrutinize uncertainty in both the predictors and the transfer function. Another problem common to forecasting efforts is the occurrence of latent relationships, left undetected or confounded with stochastic variation at the model fitting stage, which will only manifest themselves when the climate warms up or population grows beyond some threshold. Predictions are by nature based on some form of status quo assumption, whereas entirely new processes may emerge and completely change the nature of climate–demographic relationships [43]. In other words, the risk is mostly with things you don’t know and that ‘you don’t know you don’t
know’ [42]. With that in mind, adequately computing and reporting known sources of uncertainty in models coupling ecological processes and climate scenarios remains essential to avoid giving a false sentiment of confidence on projections issued from such modelling exercises [10].

4. Conclusion
Predicted changes in animal population size based on climate-dependent demographic models entail a large part of uncertainty, which is often not adequately accounted for or communicated by researchers. Here we empirically demonstrate and exhaustively quantify the contribution of the different sources of uncertainty in demographic projections, one of the main tools for scientists interested in the ecological consequences of climate change. Considerable emphasis is often placed on uncertainty linked to climate scenarios (i.e. greenhouse gas emissions and GCMs) in climate-dependent population models [18,44,45]. However, we show that these sources can be overwhelmed by environmental stochasticity. Sampling variance (i.e. measurement error) also rapidly accumulates in model predictions, even in a well-studied species, and can become the dominant source of uncertainty after a few decades. Efforts should be primarily devoted to reduce the latter source of uncertainty if one wishes to improve confidence in predicted change in animal populations exposed to climate warming based on demographic models. Finally, we also urge researchers to consider the full annual cycle of species when developing predictive population models, especially in long-distance migrants which can be exposed to very different climatic regimes at various times of the year [33].

Data accessibility. Our paper is a modelling paper that builds upon climate-demography relationships that have been previously published in [23] and [25]. The only original data are the population size inventory conducted by the Canadian Wildlife Service (presented in electronic supplementary material, table S3). For constructing the climate scenarios, we used simulations from GCMs (listed in electronic supplementary material, table S2) as well as observation-based products (described in the electronic supplementary material) obtained from public institutions.

Authors’ contributions. G.G. and J.D.L. conceived the project; J.D.L. made the theoretical developments; G.P. conducted the data analysis; L.V.O. contributed analytical tools; P.G. generated the climate scenarios; G.G. and G.P. wrote the paper with inputs from J.D.L., P.G. and L.V.O.

Competing interests. The authors declare no conflict of interest.

Funding. This work was funded by grants from the Natural Science and Engineering Research Council of Canada, the Network of centres of excellence ArcticNet and the Canadian Wildlife Service to G.G.

Acknowledgement. We thank R.F. Rockwell for his helpful comments on this manuscript.

References