Received: 14 April 2025 | Revised: 12 October 2025 | Accepted: 30 October 2025

DOI: 10.1002/ecy.70270

STATISTICAL INNOVATION

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Longer time series with missing data improve parameter estimation in a state-space model in coral reef fish communities

Alfonso Ruiz-Moreno[1,2,3] | **Michael J. Emslie**[2] | **Sean R. Connolly**[1,3]

[1]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

[2]Australian Institute of Marine Science, PMB 3 Townsville MC, Townsville, Queensland, Australia

[3]Smithsonian Tropical Research Institute, Panama City, Panama

Correspondence
Alfonso Ruiz-Moreno
Email: alfonso.ruizmoreno@my.jcu.edu.au

Funding information
Smithsonian Tropical Research Institute; Centre of Excellence Program, Grant/Award Number: CE140100020; James Cook University; Australian Institute of Marine Science

Handling editor: Lorenzo Ciannelli

Abstract

Analysis of time series data is fundamental in ecology for understanding community dynamics, and the mechanisms driving such dynamics. However, ecological time series commonly contain missing values, which can arise due to methodological changes in monitoring programs, poor weather conditions, logistical constraints, or human error. State-space models are a useful suite of techniques for analyzing ecological time series with missing data. Such models can estimate the unobserved true abundances as latent variables, even for putative census occasions that lack observations. Nevertheless, the impact of missing data on parameter accuracy and precision in these state-space models remains poorly investigated, particularly in species-rich systems. We evaluated the performance of a multivariate process-based state-space model for time series of varying lengths and sampling frequencies, using simulated data informed by empirical counts of reef fish communities from the Great Barrier Reef in Australia. We found that despite containing missing data, the models fitted to longer time series produced more accurate and precise parameter estimates compared to shorter, complete-case time series. Additionally, higher spatial replication with uneven census intervals enhanced parameter precision more than maintaining census frequency at the expense of reducing the number of sites. Analysis of reef fish community data revealed attenuation in intrinsic population growth parameters and weaker intraspecific density dependence when modeling longer time series whose inter-census intervals change, while estimates of variability in species' population growth parameters due to environmental fluctuations remained consistent regardless of sampling design. Nonetheless, analyses of shorter, complete-case time series still produced reliable parameter estimates. Our findings highlight the robustness of Bayesian state-space models to missing data, demonstrating that flexibility in long-term monitoring programs does not compromise ecological inference from these models.

KEYWORDS
Bayesian modeling, community dynamics, long-term monitoring, missing data, multivariate time series, reef fish, state-space model

INTRODUCTION

Time series data are crucial in ecology because they provide opportunities to investigate temporal changes in biodiversity patterns and their drivers. For example, analysis of time series can reveal species richness and turnover changes over time (Blowes et al., 2019; Dornelas et al., 2014), forecast future temporal trends (Dornelas et al., 2013) and make predictions about extinction risk (Ferguson & Ponciano, 2014). Furthermore, fitting process-based autoregressive models to abundance time series has shown the importance of different mechanisms in driving community dynamics, such as intra- and interspecific interactions and species responses to environmental fluctuations (Mutshinda et al., 2009; Ovaskainen et al., 2017; Ruiz-Moreno et al., 2024), as well as the contribution of such mechanisms to community stability (Thibaut et al., 2012). These models typically assume evenly spaced censuses, but ecological time series often contain missing data due to uneven census intervals arising from changes in methodology, poor weather, logistical and financial constraints or human error. A common way of dealing with missing data in ecology is to perform a complete-case analysis, where the dataset is analyzed without the missing values (Nakagawa, 2015). A disadvantage of this approach is the loss of information and potential bias in parameter estimation (Humbert et al., 2009; Nakagawa, 2015).

State-space models are a popular framework in ecology, which in principle can deal with missing data. State-space models involve two components, a process sub-model and an observation sub-model (Calder et al., 2003). In population and community time series, the process sub-model describes the true unobserved abundances (e.g., true fish abundances in a community), while the observation sub-model describes the effects of sampling error (e.g., observed fish count abundances) (Auger-Méthé et al., 2021; Clark & Bjørnstad, 2004). These hidden true abundances are assumed to be autocorrelated, so the abundance of a species at a given time step will affect the abundance at one or more subsequent time steps (Auger-Méthé et al., 2021). In contrast, the observed abundances are independent of one another (conditional on the true abundances), and only reflect the observation error of the hidden true abundances (Auger-Méthé et al., 2021). The advantage of state-space models is that the hidden true unobserved abundances can still be estimated as latent variables even for the times when there is no observed abundance data available constraining them (Clark & Bjørnstad, 2004). Furthermore, state-space models can be fitted to ecological time series data for a wide range of model complexity, from normal linear models to more complex likelihoods, Frequentist to Bayesian, and univariate to multivariate (Clark & Bjørnstad, 2004; Dennis et al., 2006; Hampton et al., 2013; Holmes et al., 2012; Newman et al., 2023).

Although the state-space framework can accommodate missing values (Clark & Bjørnstad, 2004; Holmes et al., 2012; Newman et al., 2023), studies have yet to examine how missing data affect the bias and precision of parameter estimates, particularly in multivariate models for species-rich systems. For example, Clark and Bjørnstad (2004) used a univariate Bayesian state-space approach to explore the influence of observation and process error on parameter uncertainty but did not explicitly compare estimates from complete versus incomplete time series. Indeed, prior studies fitting Bayesian or frequentist state-space models to time series with missing data have noted that missing observations increase parameter uncertainty (Auger-Méthé et al., 2021; Clark & Bjørnstad, 2004) due to unobserved true abundances not being constrained by the observed data. However, to date, these studies have not explicitly addressed the potential magnitude or direction of bias in parameter estimates. This is especially relevant for adaptive monitoring programs, which evolve in response to environmental or management changes rather than following a fixed sampling design (Henrys et al., 2024; White & Bahlai, 2021). Such adaptive monitoring frameworks allow for dynamic monitoring across locations and habitats, improving detection of shifts in species distributions and habitat condition (Henrys et al., 2024; Lindenmayer & Likens, 2010). However, it remains unclear how changes in monitoring designs influence parameter estimation in state-space models, particularly in species-rich systems where missing data may result from variable census frequency or incomplete spatial coverage. For example, a coral reef fish monitoring program may expand to mangrove habitats as rising temperatures affect coral reef-associated species that use mangroves as nursery habitats. Budgetary or logistical constraints may limit the number of sites sampled per census, requiring trade-offs such as halving coral reef sampling to include mangrove sites or alternating census between habitats, resulting in more spatial replication at the expense of temporal gaps. Understanding how missing data affect state-space model inferences is crucial for optimizing monitoring frequency and site selection while balancing temporal and spatial coverage.

Here, we evaluate how missing data can impact parameter estimation in a Bayesian state-space model describing temporal community dynamics of reef fishes. Firstly, we simulate communities and analyze 11, 18 and 25-year time series without missing data and a 25-year time series with missing data using parameters previously estimated from fish communities on the Great Barrier Reef (GBR) in Australia (Ruiz-Moreno et al., 2024). We fit the

model and evaluate how missing data affect the accuracy and precision of parameter estimates compared to fitting the model to complete-case subsets (11, 18-year and complete 25-year). Additionally, we analyze speciose reef fish assemblage time series data on the GBR. We fit the state-space model to both a complete-case 11-year portion and the full 25-year time series, with missing data, to evaluate the changes in the parameter estimates and measure the differences in uncertainties.

METHODS

Data

We used two versions of the same reef fish abundance data, which were collected by the Australian Institute of Marine Science's Long Term Monitoring Program (LTMP). Firstly, we used the LTMP data from Ruiz-Moreno et al. (2024), which contains count data for 40 species (for species list see Appendix S1: Table S1 in Ruiz-Moreno et al., 2024) of reef fishes at 41 reefs across an 11-year period (between 1995 and 2005). After 2005, the sampling regime changed to biennial sampling (from 2006 to 2019) (Figure 1A). Therefore, we used this full time series, including years with missing data (i.e., no sampling in 2006, 2008, 2010, 2012, 2014, 2016 and 2018), as the second version of the data. This version contained count data for the same 40 species of reef fishes at 46 reefs across a 25-year period. The change in the number of reefs is due to some reefs being removed from the 11-year complete-case data subset, due to bad weather conditions that prevented sampling at some reefs from 1995 to 2005. Therefore, the 11-year dataset contains 41 reefs with no missing data, while the 25-year dataset contains all 46 reefs (Figure 1B, Appendix S1: Figure S1). Data were collected using underwater visual surveys. At each reef, there were three sites in a standard reef slope habitat, usually on the north-east flank of the reef. Each site contained five permanently marked 50-m transects, approximately parallel to the reef crest between 6 m and 9 m

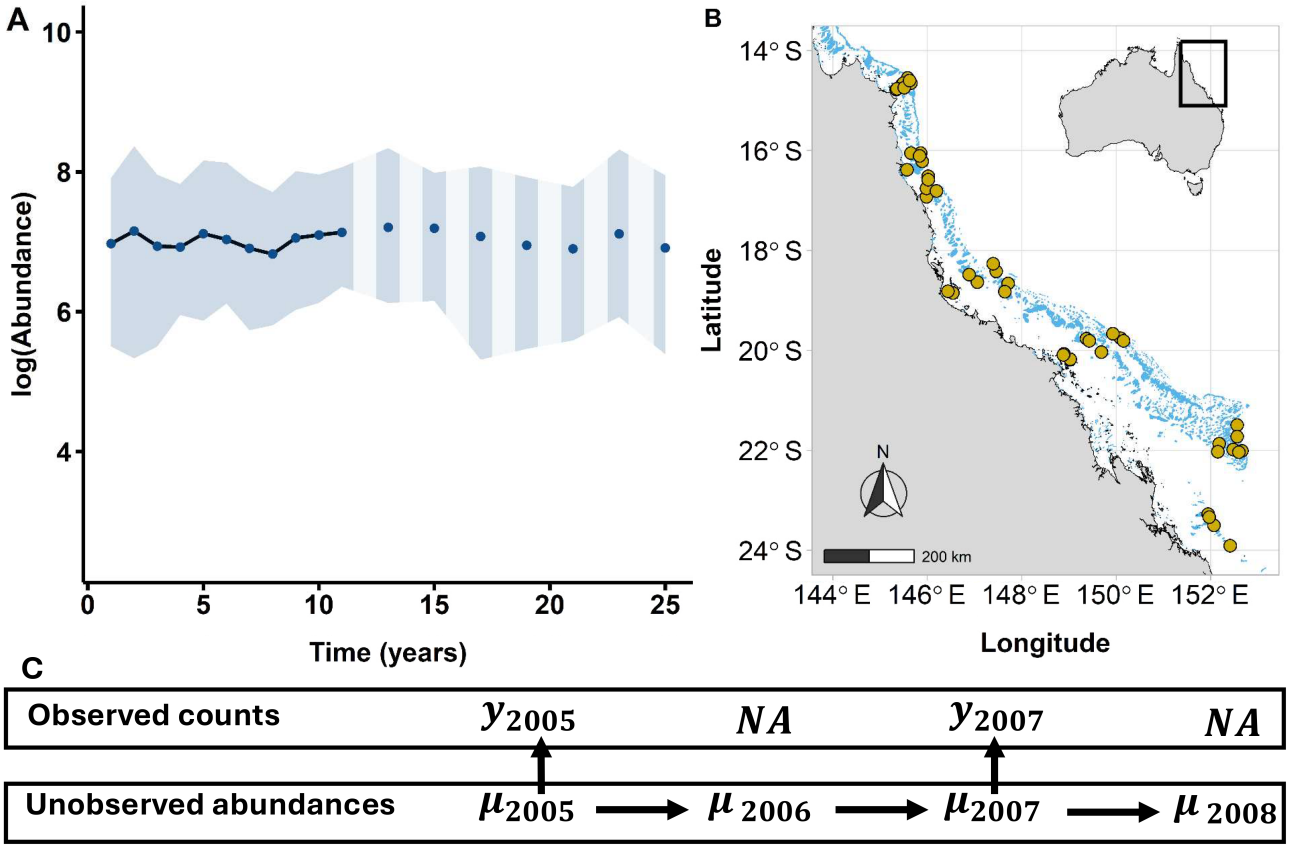

FIGURE 1 Panel (A) shows the time series of mean log(total species counts + 1) across 46 reefs in the time series as dark blue points. Blue polygons represent the 95% percentile range across reefs. White spaces after year 11 indicate years with no sampling across all reefs due to changes in the Long Term Monitoring Program sampling regime. Panel (B) shows a map of the Great Barrier Reef with the 46 sampled reefs as yellow points. Mainland Australia and islands are represented in gray and coral reefs and cays in light blue. Panel (C) shows observed counts with missing data and the underlying unobserved abundances estimated with the process sub-model (adapted from Clark & Bjørnstad, 2004).

($n = 15$ transects reef^{-1} year^{-1}). Observers recorded abundances of the reef fishes estimated to be at least 1 year old to reduce the inherent variability in reef fish numbers due to high mortality in early life history stages. All families were counted on 50 × 5 m transects, except pomacentrids, which were counted on 50 × 1 m transects, due to their small body sizes and site-attached life habits (see Emslie & Cheal, 2018 for detailed methodology).

The model

We used a hierarchical state-space model from Ruiz-Moreno et al. (2024), in which the unobserved true abundances follow the multivariate Gompertz model (Ives et al., 2003):

$$\log(\boldsymbol{\mu}_t) = \boldsymbol{a} + \boldsymbol{B}\log(\boldsymbol{\mu}_{t-1}) + \boldsymbol{e}_t, \quad (1)$$

where $\log(\boldsymbol{\mu}_t)$ is a vector containing species' unobserved true log-abundances at time t. $\boldsymbol{a}$ is a vector containing the estimated species' intrinsic (i.e., density-independent) population growth parameters. $\mathbf{B}$ is a species-by-species interaction matrix whose off-diagonal elements, b_{ij}, indicate the effects of interspecific interactions (e.g., competition, facilitation), and whose diagonal elements, b_{ii}, represent the effects of intraspecific density dependence. $\mathbf{e_t}$ is a vector of process errors for each species, which has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. To reduce the dimensionality of the covariance matrix $\boldsymbol{\Sigma}$, we used a latent variable approach (see Ruiz-Moreno et al., 2024 and Appendix S2). The multivariate Gompertz model assumes stationarity of the stochastic-dynamics process, which is satisfied by the data as most species show no or small differences in median abundances at the beginning and end of time series (Appendix S1: Figure S2).

One of the key findings in Ruiz-Moreno et al. (2024) was that the effects of interspecific interactions were negligible in driving reef fish community dynamics. Therefore, we used leave-one-out cross-validation (LOO-CV) (Vehtari et al., 2017) to test the performance of two alternative models: using the regularized horseshoe prior to estimate the interspecific interaction terms, as in Ruiz-Moreno et al. (2024), and setting all interspecific interaction terms to zero. We found that there was slightly stronger support for a model without interspecific terms versus a model including them (Appendix S1: Table S1). Therefore, we assume $\mathbf{B}$ is a diagonal matrix whose off-diagonal elements are set to zero and the diagonal elements, b_{ii}, represent the effect of the abundance of species i on its own population growth ($b_{ii} = 1$ for density-independent growth; $0 < b_{ii} < 1$ implies compensatory density dependence, $b_{ii} < 0$ implies over-compensatory density dependence).

Lastly, we assumed that the observed abundances follow a Poisson distribution:

$$y_{i,t} \sim \text{Poisson}(\mu_{i,t}), \quad (2)$$

where $y_{i,t}$ is the observed count of species i at time t, $\mu_{i,t}$ is the (unobserved) abundance of species i at time t from Equation (1). Due to the use of different transect sizes to count Pomacentrids (50 m × 1 m transect) and non-Pomacentrids (50 m × 5 m transect), we modeled the Pomacentrid counts as

$$y_{(\text{pom})\,i,t} \sim \text{Poisson}\left(\mu_{(\text{pom})\,i,t}/5\right), \quad (3)$$

where $y_{(\text{pom})i,t}$ is the observed number of fish of the pomacentrid species i at time t, $\mu_{(\text{pom})i,t}$ is the mean abundance of the pomacentrid species i at time t per 250 m^2 (the area of the larger transect) (see Ruiz-Moreno et al., 2024 for further details).

The advantage of using a state-space model for ecological time series with missing data is that the unobserved abundances ($\boldsymbol{\mu_t}$ in Equation 1) can still be estimated for years with no observed counts (Figure 1C), without having to perform data imputation (Clark & Bjørnstad, 2004). For example, for the year 2006 there is no observed count, but the underlying unobserved abundance can be estimated by using the estimated previous unobserved abundance in 2005 and the estimated next unobserved abundance in 2007, so the uncertainty in the unobserved abundance estimates depends on the adjacent years and does not require observed counts in every year (Figure 1C). Thus, years with missing observations contribute no term to the likelihood. Missing counts are not imputed, although the underlying unobserved (true) abundances are estimated through the process sub-model.

Simulation study comparing length of time series

Because the species' underlying unobserved abundances are not constrained by the observed count data at multiple years, we wanted to assess the effects of missing data on the parameter estimates and inferences produced in our analyses. In particular, we wanted to assess whether fitting a model to a longer time series with missing data will increase or decrease the uncertainty of the parameter estimates and accuracy versus fitting the model to shorter time series without missing data. Consequently, we simulated data with known parameter values based on empirical data of reef fish abundances from the GBR

(Ruiz-Moreno et al., 2024), and then we fitted our model to four different subsets of the simulated data:

1. *Short fit*: 11 years without missing data.
2. *Missing-data fit*: 25 years of data with biennial sampling observations after year 11 (i.e., no counts for years 12, 14, 16, 18, 20, 22, and 24).
3. *Intermediate fit*: 18 years without missing data. This and the previous scenario both have 18 years of observations, allowing us to assess whether shorter consecutive time series yield more reliable parameter estimates than longer time series with missing data.
4. *Full fit*: 25 years without missing data.

We estimated the accuracy and precision of parameter estimates across scenarios (See Appendix S2 for further details).

Simulation study comparing number of sampled reefs

Additionally, we wanted to assess the effects of changing the number of reefs being monitored on the model parameter estimates. Therefore, we simulated the dynamics of 20 communities using the same parameters as in the previous simulation study and analyzed the last 25 years. However, this time we simulated the data from 40 reefs and then fitted the model to two different subsets of the data (Appendix S1: Figure S3):

1. *40 reefs fit*: We fitted the model to all of the final 25 years of data, but with missing data in the same structure as the Missing-data fit above (i.e., no counts for years 12, 14, 16, 18, 20, 22, and 24).
2. *20 reefs fit*: We fitted the model to all the final 25 years, but we sampled the 40 reefs just for the first 11 years, and only 20 of those reefs were in years 12 to 25.

Both scenarios contain the same total amount of surveys across reefs and years. Accuracy and precision were calculated as in the previous simulation study (Appendix S2).

Model fitting

We implemented our model through a Bayesian framework using Stan (Stan Development Team, 2024) in R (R Core Team, 2024), through the package *rstan*. For the simulation study comparing the length of the time series, the simulation study comparing the number of sampled reefs, and the model fits to the reef fish data, we ran four chains, each with 10,000 iterations, 5000 iterations as warm up and 5000 as sampling. This left 20,000 samples in the posterior distribution of each parameter. For the intrinsic population growth term ($\boldsymbol{a}$) and the intraspecific density dependence (diagonal elements of matrix $\mathbf{B}$) we implemented a hierarchical approach, and we used weakly informative priors (see Appendix S3). Model convergence was monitored by checking that the potential scale reduction factor (R-hat) was close to 1 for all parameters and checking that the effective sample sizes were large (Appendix S1: Figure S4). Model fit was further assessed by posterior predictive checks (Appendix S1: Figure S4).

RESULTS

Results of the simulation study comparing length of time series

Comparing the accuracy in the intrinsic growth parameters revealed there was a consistent overestimation of the true parameter values across all model fits (Figure 2A). The accuracy decreased sequentially from the *Full fit* (25-year), *Missing-data fit* (25-year with missing data), *Intermediate fit* (18-year), and *Short fit* (11-year). Similarly, model accuracy for intraspecific density dependence (b_{ii}) decreased with shorter time series (Figure 2B), but in this case, all models underestimated the true value from the simulations, where lower values indicate stronger density dependence. As a result, models fitted to the shorter time series tended to overestimate (i.e., b_{ii} values further below 1) the strength of intraspecific density dependence. This pattern was maintained but weaker for the sensitivity to environmental fluctuation estimates, with all the models overestimating the true values (Figure 2C) despite their posterior medians being closer to the true values than either intrinsic growth or intraspecific density dependence parameters. Environmental correlation parameters were similarly accurate across all model fits and centered around the true values used to simulate the data (Figure 2D). The carrying capacity accuracy estimates showed the same pattern of decreasing accuracy with time series length and there was a slight overestimate in the true parameter values for the 18-year and 11-year models, while the 25-year and 25-year with missing data models produced medians very close to the true values (Figure 2E). There was a similar pattern for the precision estimates for all the parameters, where the 25-year model fit had the higher precision across all parameter estimates, followed by the 25-year with missing data, 18-year and 11-year models (Figure 2F–J).

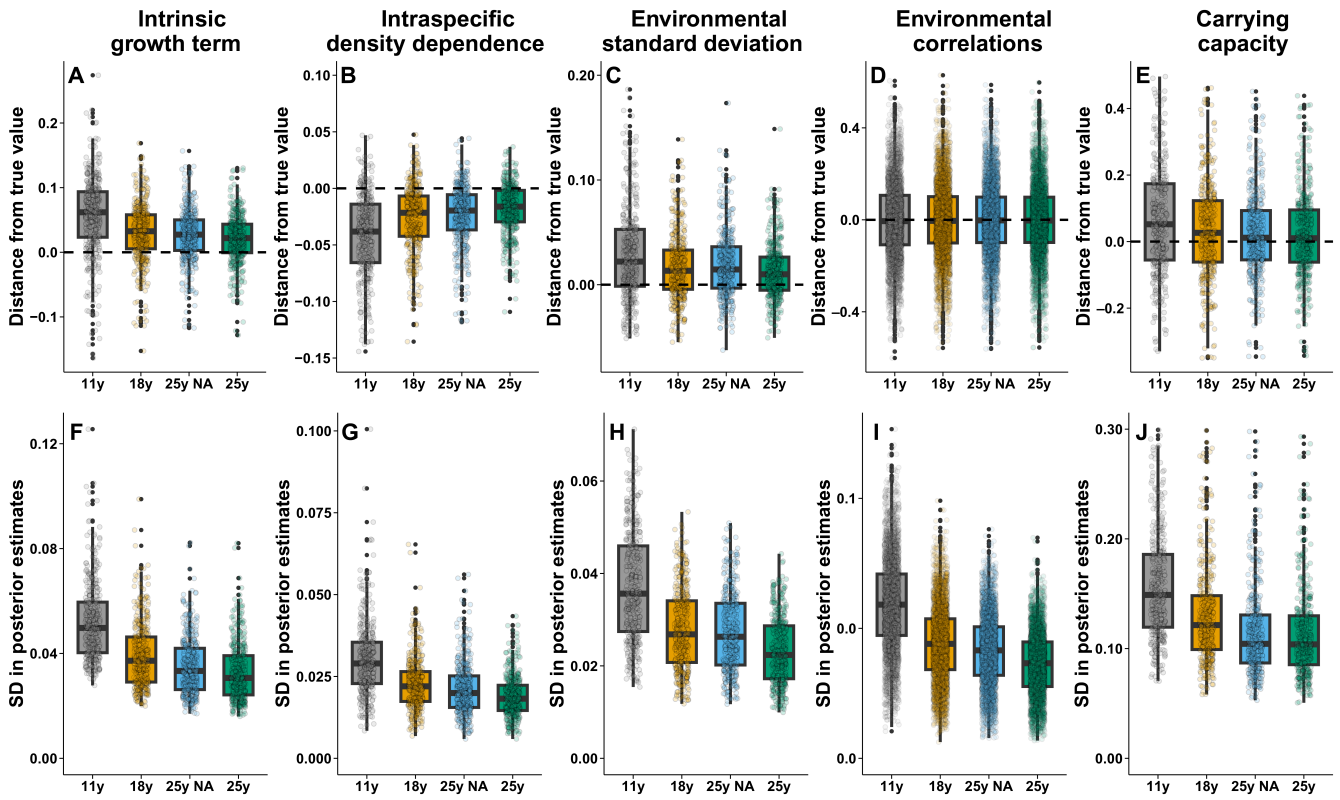

FIGURE 2 Top row panels (A–E) show accuracy estimates by showing the distribution of the models estimated distances from the true simulated values. Within each panel, the distribution of accuracy estimates is shown for the *Short fit* (11y) in gray, the *Intermediate fit* (18y) in orange, the Missing-*data fit* (25y NA) in blue, and the *Full fit* (25y) in green. Each column corresponds to one parameter. The black dashed horizontal line at 0 represents that the model parameter estimate is the same as the true simulated. The bottom row panels (F–J) show precision estimates with values closer to 0 indicating lower uncertainty. The colored dots in each boxplot represent all the accuracy or precision estimates for the corresponding model type. The y-axes were zoomed in for clarity and some outliers fall outside the range (see Appendix S1: Figure S7 for full *y*-axes range).

Results of the simulation study comparing number of sampled reefs

Accuracy was comparable across all model parameters (Environmental standard deviation and correlations) or slightly higher (intrinsic population growth, intraspecific density dependence and carrying capacity) for the *40 reefs fit* (i.e., the simulation with missing years of data) compared to the *20 reefs fit* (Figure 3A–E). As in the preceding simulation study, both the intrinsic growth parameter and the strength of density dependence are overestimated by both models (Figure 3A,B). Similarly, the environmental standard deviation (Figure 3C) and the carrying capacity (Figure 3E) are also overestimated by both the *40 reefs fit* and the *20 reefs fit*, but their median accuracy was closer to the true values than for intrinsic growth and density dependence. Like the previous simulation study, the environmental correlation estimates are centered around the true values (Figure 3D). The *40 reefs fit* estimates are more consistently precise than the *20 reefs fit* estimates across all parameters, although these estimates are very close to each other.

LTMP data results

The biases observed in the simulation study are qualitatively reflected in the LTMP results: intrinsic population growth terms were lower and intraspecific density dependence estimates weaker (closer to 1) under the 25-year model fit than the 11-year model fit, suggesting analyses of shorter time series may overestimate species' ability to recover from low densities, as well as the strength of density regulation (Figure 4A,B). As in the simulations, these shifts resulted in lower carrying capacity estimates for the longer time series (Figure 4E), despite no substantial differences in species' observed mean abundance between 1995 and 2005 (annually sampled), versus 2006 to 2019 (biennial sampling) (Appendix S1: Figure S5). This seems to support the idea that model structure and

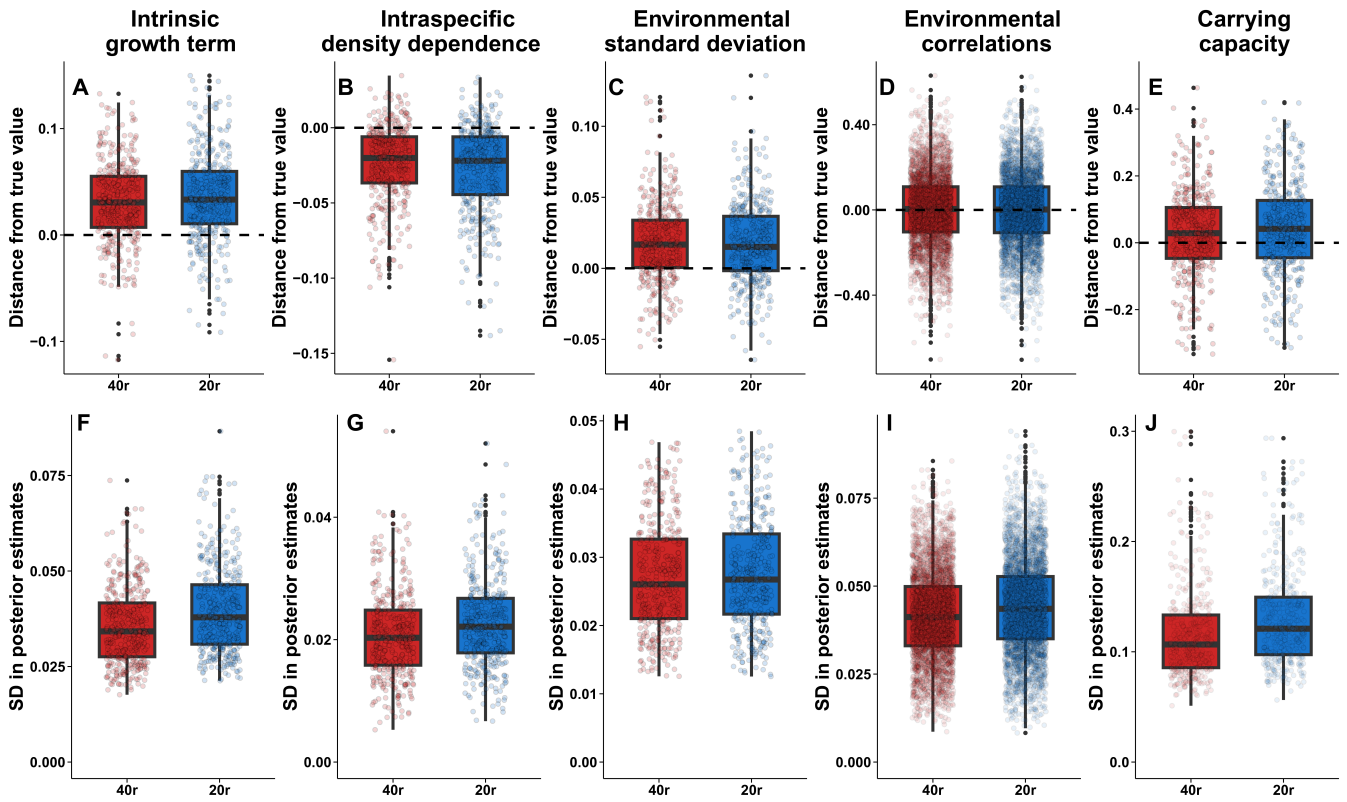

FIGURE 3 Same as Figure 2 but showing the accuracy and precision results of the *40 reefs fit* (40r) and *20 reefs fit* (20r) in red and blue, respectively. The *y*-axes were zoomed in for clarity and some outliers fell outside the range (see Appendix S1: Figure S8 for full *y*-axis range).

time series length, not observed trends, drove this shift in parameter estimation from the 11-year to the 25-year model. The environmental standard deviation and environmental correlation estimates were consistent across both model fits (Figure 4C,D), aligning with the simulation results. As with the simulation results, the precision in the posterior estimates was also higher for the model fitted to the whole 25-year period (Figure 4F–J).

DISCUSSION

According to theoretical studies (Auger-Méthé et al., 2021; Dennis et al., 2006; Newman et al., 2023), the state-space approach can, in principle, account for missing data. However, the impact of missing data on the accuracy and precision of these models has been largely unexplored, especially in species-rich time series. In earlier work applying the multivariate Gompertz model to this system (Ruiz-Moreno et al., 2024), we used the complete-case 11-year dataset to evaluate the importance of heterogeneity in demographic rates, species-to-species interactions, and species-to-environment interactions in driving reef fish community dynamics. In contrast, here, we examined the effects of time series length and missing data on estimates of community dynamics parameters, comparing that original 11-year dataset to a longer time series with missing data, and using simulations to assess how the distribution of sampling effort (longer time series with missing years, shorter time series without missing years, or longer time series with less spatial replication) influences the bias and precision of parameter estimates. We find that the dynamics of reef fish communities on the GBR are better captured by analyzing the whole 25-year time series with missing data rather than the 11-year time series subset. The model fitted to the 25-year time series had comparable or higher accuracy and higher precision in all parameters than the model fitted to the 11-year time series, and it was consistent with the simulation results. Similarly, our simulation studies showed a small improvement in the accuracy and precision of the parameter estimates with higher spatial replication, even when the census frequency was changed from annual to biannual sampling midstream rather than halving the number of sampling sites and maintaining annual surveys. Indeed, previous studies emphasize the importance of maintaining adequate spatial replication to capture spatial heterogeneity and accurately detect trends in marine monitoring programs (Edgar & Barrett, 2012; Hewitt et al., 2001). Therefore,

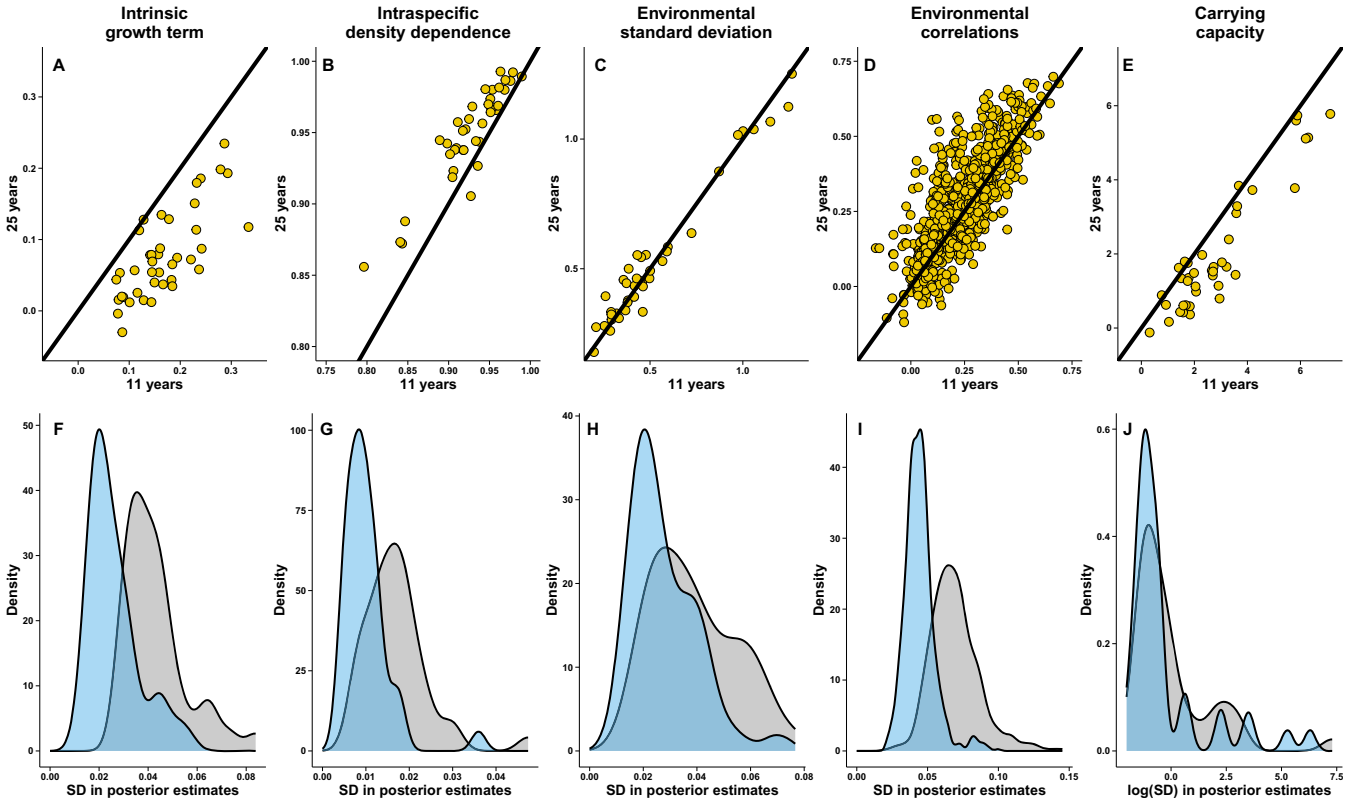

FIGURE 4 Top panels (A–E) compare parameter estimates from the model fit to the 40 species using 11 consecutive years of the Long Term Monitoring Program data (1995 to 2005) on the *x* axis versus 25 years with missing data (1995–2019) on the *y*-axis. Parameters shown include intrinsic population growth term (A), intraspecific density dependence (B), environmental standard deviation (C), environmental correlations (D), and carrying capacity (E). For panels A–C and E, each yellow point represents a species' estimate from both models, while in panel (D), each point is the correlation between a pair of species in their response to environmental fluctuations. Black lines on panels (A–E) represent the unity line. Proximity to the line indicates similar estimates from both models. Bottom panels (F–J) show the precision estimates for each parameter: gray for the 11-year model fit and blue for the 25-year model fit.

keeping the same number of sampled sites or increasing them may be beneficial, even if that requires increasing the time between censuses.

Our analyses revealed that the multivariate Gompertz model exhibits consistent biases in the intraspecific density dependence, the intrinsic population growth parameter and carrying capacity. Although different combinations of intrinsic growth and intraspecific density dependence can yield the same carrying capacity (i.e., $K_i = a_i/(1-b_{ii})$), our results with the reef fish data showed that carrying capacity was overestimated in the shorter time series and this was not a consequence of changes in mean abundances of the species from the annually sampled portion of the time series versus the biannually sample portion (Appendix S1: Figure S5). Furthermore, the simulation study showed a similar trend of overestimation of carrying capacity by the shorter time series, even though true carrying capacity remained constant through time in those simulations.

Studies using the Gompertz model have frequently highlighted biases concerning the estimation of density dependence. For example, single-population studies indicate that the Gompertz model estimates weaken density dependence when observation error is considered instead of ignored (Knape & de Valpine, 2012), although this effect is diminished when a hierarchical approach is employed in which multiple population time series mutually inform one another (Thibaut & Connolly, 2020). Moreover, lack of spatial heterogeneity in the Gompertz model may underestimate the strength of density dependence or falsely show over-compensatory dynamics when not present (Thorson et al., 2015). However, reef fishes on the GBR show no evidence of spatial variability in density dependence (nor intrinsic growth term) (Ruiz-Moreno et al., 2024). Furthermore, we found that most species show no spatial autocorrelation in their process error estimates in our model fitted to the 25-year LTMP data (see Appendix S4). Our community-wide average estimate of the strength of intraspecific density dependence aligns with single-population analyses from insects, fishes, birds and mammals (Thibaut & Connolly, 2020), as well as our previous analysis of the same reef fish assemblages, allowing for interspecific interactions over an 11-year period (Ruiz-Moreno et al., 2024) (Appendix S1: Figure S6).

Notably, analyzing the longer 25-year time series with missing data produced weaker negative density dependence compared to the 11-year time series. These biases, also observed in simulated and empirical single-population studies (Humbert et al., 2009; Thibaut & Connolly, 2020), highlight the need to be cautious when interpreting results derived from short-term time series.

In line with the consistently negligible estimates of interspecific interactions from Ruiz-Moreno et al. (2024), cross-validation results in our study provided stronger support for a model excluding interspecific interactions for the 11-year time series (Appendix S1: Table S1), highlighting the robustness of the pattern of intraspecific density dependence being stronger than interspecific interactions. This pattern has indeed been documented in other ecosystems for trees, plankton, vertebrate and invertebrate communities (Adler et al., 2018; Griffiths et al., 2016; Mutshinda et al., 2009; Ovaskainen et al., 2017; Sandal et al., 2022), suggesting that community dynamics are often highly individualistic, with mainly intraspecific density regulation, where interspecific interactions are either confined to small spatial and temporal scales that do not scale up to affect population-level dynamics, or so diffuse that their effects remain negligible.

Contrary to the deterministic components of the Gompertz model, no study to our knowledge has explicitly examined biases in estimating components of the variance–covariance matrix (i.e., standard deviation: species' sensitivity to environmental fluctuations; the correlations: species correlated responses to environmental fluctuations) in the multivariate Gompertz model. Failing to properly model observation error can inflate the environmental variance in the univariate Gompertz model (Lele, 2006). A state-space approach seems to reduce this bias as it explicitly models both process and observation error (Dennis et al., 2006; Lele, 2006). Our study expands this state-space approach to the multivariate context, and contrary to the deterministic components of the Gompertz model, the components of the variance–covariance matrix showed no consistent bias in either the 11-year or the 25-year time series in reef fish community data. Furthermore, the simulation study revealed similar accuracy estimates for the model, regardless of time series length, or number of sampled reefs. In contrast to accuracy, analysis of both LTMP data and the simulation study showed an increase in the precision of the environmental standard deviation and environmental correlations when analyzing the longer time series.

One important assumption in our approach is that the Gompertz model was specified with constant carrying capacities (i.e., a and b parameters did not vary through time), so that abundances fluctuate around a fixed equilibrium without long-term trends. Yet many conservation scenarios involve persistent declines, recoveries, or invasions (Craigie et al., 2010; Green et al., 2012; Zerbini et al., 2019), which violate the fixed equilibrium assumption in our model. However, our approach can be modified to incorporate temporal trends in parameter values. In Appendix S5, we present a modified model, where the species' intrinsic population growth parameter is modeled as a linear function of time. Simulations showed that this "Changing a" model accurately captured both increasing and decreasing abundance trends, while our base model assuming constant a showed higher bias and uncertainty in the parameter estimates. These results highlight the flexibility of our framework and illustrate how it can be extended to account for conservation contexts where abundance trends are of particular interest.

Our findings have important implications for long-term community monitoring. First, extending time series length, even if this requires reducing census frequency from annual to biennial, significantly improves parameter estimation in our study. This suggests that monitoring programs should prioritize the duration of data collection over strict census frequency, consistent with evidence from other systems where long time series are critical to detecting population fluctuations and trends (Gorresen et al., 2016; Kiffner et al., 2017). Second, although longer time series provide more accurate and precise estimates, our simulations and empirical analyses showed that reliable estimates can be obtained from shorter (11-year) time series, although studies point out that the minimum required length varies among populations and species (Wauchope et al., 2019; White, 2019), with at least 10 years recommended to reliably detect population trends (Hatch, 2003; Rueda-Cediel et al., 2015; White, 2019). Monitoring programs should aim for at least a decade of monitoring, although certain strong trends can be detected with as little as 4–5 years of data (Wauchope et al., 2019). Third, high spatial replication improves the precision of parameter estimates, indicating that broad spatial coverage can enhance inference, even when time series are relatively short. This aligns with studies emphasizing the importance of accounting for both spatial and temporal variability in sampling designs to detect dynamics and trends over time (Emslie et al., 2015; Sale et al., 2005). These findings highlight that long-term ecological monitoring programs can flexibly adjust census frequency and spatial coverage without compromising inference.

Our multivariate process-based state-space framework reliably estimated fish community dynamics parameters even with missing data and abundance trends and is broadly applicable to other species-rich ecosystems facing similar challenges. Future studies should evaluate the performance of this modeling approach across different ecosystems to further investigate its robustness and potential use in time series data from adaptive monitoring programs.

Implementing this type of robust approach to high-diversity systems highlights the importance of flexible and adaptive long-term monitoring programs capable of effectively responding to changing environmental and management needs. Continued investment in sampling frameworks that consider both spatial and temporal dimensions can ensure accurate detection of community dynamics in species-rich systems and inform better conservation outcomes.

AUTHOR CONTRIBUTIONS

Alfonso Ruiz-Moreno conceived the study with supervisory support from Sean R. Connolly and Michael J. Emslie. Michael J. Emslie collected and provided the LTMP data. Alfonso Ruiz-Moreno performed the modeling work and analyzed the model output with support from Sean R. Connolly. Alfonso Ruiz-Moreno wrote the manuscript with substantial support from Sean R. Connolly and Michael J. Emslie.

ACKNOWLEDGMENTS

The LTMP data analyzed in this study were collected from the traditional sea Country of the Gudang Yadhaykenu, Wuthathi, Kuuku Ya'u, Umpila, Dingaal, Thaanill-warra, Ngurruu- mungu, Eastern Kuku Yalanji, Yirrganydji, Gunggandji, Gunggandji-Mandingalbay yidinji, Mandubarra, Bandjin, Girramay, Gulngay, Djiru, Nywaigi, Manbarra, Wulgurukaba, Bindal, Ngaro, Yuwibara, Wiri, Yuwi-Bara, Koinjmal, Guninmal, and Darumbal peoples, and those represented by the Mackay Coastal Group and the Port Curtis Coral Coast Trust. We pay our respects to their Elders past, present, and emerging. We thank past and present members of the LTMP of the Australian Institute of Marine Science, and the crews of the research vessels Harry Messel, Sirus, Lady Basten, Cape Ferguson, and Solander. We thank the HPC facilities of James Cook University for computational resources. We thank Anthony Ives and an anonymous reviewer for their thoughtful comments on the thesis version of this manuscript, which substantially improved our study. We also thank the two anonymous reviewers and the editor for their constructive comments that helped us improve this manuscript. This research was supported by an AIMS@JCU scholarship to Alfonso Ruiz-Moreno, the Smithsonian Tropical Research Institute, and the Australian Research Council Centre of Excellence Program (grant number CE140100020).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code are available in Figshare as follows (Ruiz Moreno, 2025a, 2025b, 2025c, 2025d, respectively): LTMP data and the code to run the analysis, https://doi.org/10.6084/m9.figshare.28785908.v1; simulation code and simulated data for *"Simulation study comparing length of time series"*, https://doi.org/10.6084/m9.figshare.28783709.v1; simulation code and simulated data for *"Simulation study comparing number of sampled reefs"*, https://doi.org/10.6084/m9.figshare.28774304.v1; simulation code and simulated data for *"Simulation study with abundance trends"* (in Appendix S5), https://doi.org/10.6084/m9.figshare.30189733.v1.

ORCID

Alfonso Ruiz-Moreno https://orcid.org/0000-0002-5991-438X

REFERENCES

Adler, P. B., D. Smull, K. H. Beard, R. T. Choi, T. Furniss, A. Kulmatiski, J. M. Meiners, A. T. Tredennick, and K. E. Veblen. 2018. "Competition and Coexistence in Plant Communities: Intraspecific Competition Is Stronger than Interspecific Competition." *Ecology Letters* 21(9): 1319–29.

Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, et al. 2021. "A Guide to State–Space Modeling of Ecological Time Series." *Ecological Monographs* 91(4): e01470.

Blowes, S. A., S. R. Supp, L. H. Antão, A. E. Bates, H. Bruelheide, J. M. Chase, F. Moyes, et al. 2019. "The Geography of Biodiversity Change in Marine and Terrestrial Assemblages." *Science* 366(6463): 339–345.

Calder, C., M. Lavine, P. Müller, and J. S. Clark. 2003. "Incorporating Multiple Sources of Stochasticity into Dynamic Population Models." *Ecology* 84(6): 1395–1402.

Clark, J. S., and O. N. Bjørnstad. 2004. "Population Time Series: Process Variability, Observation Errors, Missing Values, Lags, and Hidden States." *Ecology* 85(11): 3140–50.

Craigie, I. D., J. E. M. Baillie, A. Balmford, C. Carbone, B. Collen, R. E. Green, and J. M. Hutton. 2010. "Large Mammal Population Declines in Africa's Protected Areas." *Biological Conservation* 143(9): 2221–28.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. "Estimating Density Dependence, Process Noise, and Observation Error." *Ecological Monographs* 76(3): 323–341.

Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. "Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss." *Science* 344(6181): 296–99.

Dornelas, M., A. E. Magurran, S. T. Buckland, A. Chao, R. L. Chazdon, R. K. Colwell, T. Curtis, et al. 2013. "Quantifying Temporal Change in Biodiversity: Challenges and Opportunities." *Proceedings of the Royal Society B: Biological Sciences* 280(1750): 20121931.

Edgar, G. J., and N. S. Barrett. 2012. "An Assessment of Population Responses of Common Inshore Fishes and Invertebrates Following Declaration of Five Australian Marine Protected Areas." *Environmental Conservation* 39(3): 271–281.

Emslie, M. J., and A. J. Cheal. 2018. *Visual Census of Reef Fish*. Townsville: Australian Institute of Marine Science.

Emslie, M. J., M. Logan, D. H. Williamson, A. M. Ayling, M. A. MacNeil, D. Ceccarelli, A. J. Cheal, et al. 2015. "Expectations and Outcomes of Reserve Network Performance Following Re-Zoning of the Great Barrier Reef Marine Park." *Current Biology* 25(8): 983–992.

Ferguson, J. M., and J. M. Ponciano. 2014. "Predicting the Process of Extinction in Experimental Microcosms and Accounting for Interspecific Interactions in Single-Species Time Series." *Ecology Letters* 17(3): 251–59.

Gorresen, P. M., K. W. Brinck, R. J. Camp, C. Farmer, S. M. Plentovich, and P. C. Banko. 2016. "State-Space Modeling of Population Sizes and Trends in Nihoa Finch and Millerbird." *The Condor* 118(3): 542–557.

Green, S. J., J. L. Akins, A. Maljković, and I. M. Côté. 2012. "Invasive Lionfish Drive Atlantic Coral Reef Fish Declines." *PLoS One* 7(3): e32596.

Griffiths, J. R., S. Hajdu, A. S. Downing, O. Hjerne, U. Larsson, and M. Winder. 2016. "Phytoplankton Community Interactions and Environmental Sensitivity in Coastal and Offshore Habitats." *Oikos* 125(8): 1134–43.

Hampton, S. E., E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and E. J. Ward. 2013. "Quantifying Effects of Abiotic and Biotic Drivers on Community Dynamics with Multivariate Autoregressive (MAR) Models." *Ecology* 94(12): 2663–75.

Hatch, S. A. 2003. "Statistical Power for Detecting Trends with Applications to Seabirds Monitoring." *Biological Conservation* 111(3): 317–329.

Henrys, P. A., T. O. Mondain-Monval, and S. G. Jarvis. 2024. "Adaptive Sampling in Ecology: Key Challenges and Future Opportunities." *Methods in Ecology and Evolution* 15(9): 1483–96.

Hewitt, J. E., F. T. Simon, and J. C. Vonda. 2001. "Assessing Environmental Impacts: Effects of Spatial and Temporal Variability at Likely Impact Scales." *Ecological Applications* 11(5): 1502–15.

Holmes, E. E., E. J. Ward, and K. Wills. 2012. "MARSS: Multivariate Autoregressive State-Space Models for Analyzing Time-Series Data." *The R Journal* 4(1): 11–19.

Humbert, J.-Y., L. Scott Mills, J. S. Horne, and B. Dennis. 2009. "A Better Way to Estimate Population Trends." *Oikos* 118(12): 1940–46.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. "Estimating Community Stability and Ecological Interactions from Time-Series Data." *Ecological Monographs* 73(2): 301–330.

Kiffner, C., H. Rheault, E. Miller, T. Scheetz, V. Enriquez, R. Swafford, J. Kioko, and H. H. T. Prins. 2017. "Long-Term Population Dynamics in a Multi-Species Assemblage of Large Herbivores in East Africa." *Ecosphere* 8(8): e02027.

Knape, J., and P. de Valpine. 2012. "Are Patterns of Density Dependence in the Global Population Dynamics Database Driven by Uncertainty about Population Abundance?" *Ecology Letters* 15(1): 17–23.

Lele, S. R. 2006. "Sampling Variability and Estimates of Density Dependence: A Composite-Likelihood Approach." *Ecology* 87(1): 189–202.

Lindenmayer, D. B., and G. E. Likens. 2010. "The Science and Application of Ecological Monitoring." *Annual Review of Ecology, Evolution, and Systematics* 41: 369–392.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2009. "What Drives Community Dynamics?" *Proceedings of the Royal Society B: Biological Sciences* 276(1677): 2923–29.

Nakagawa, S. 2015. "Missing Data: Mechanisms, Methods, and Messages." In *Ecological Statistics: Contemporary Theory and Application*, edited by G. A. Fox, S. Negrete-Yankelevich, and V. J. Sosa. Oxford: Oxford University Press.

Newman, K., R. King, V. Elvira, P. de Valpine, R. S. McCrea, and B. J. T. Morgan. 2023. "State-Space Models for Ecological Time-Series Data: Practical Model-Fitting." *Methods in Ecology and Evolution* 14(1): 26–42.

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B.-E. Sæther, and N. Abrego. 2017. "How Are Species Interactions Structured in Species-Rich Communities? A New Method for Analysing Time-Series Data." *Proceedings of the Royal Society B: Biological Sciences* 284(1857): 20170768.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rueda-Cediel, P., K. E. Anderson, T. J. Regan, J. Franklin, and H. M. Regan. 2015. "Combined Influences of Model Choice, Data Quality, and Data Quantity when Estimating Population Trends." *PLoS One* 10(6): e0132255.

Ruiz Moreno, A. 2025a. "LTMP Analysis 11-Year Versus 25-Year with Missing Data." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.28785908.v1

Ruiz Moreno, A. 2025b. "Simulation Study Comparing Length of Time Series." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.28783709.v1

Ruiz Moreno, A. 2025c. "Simulation Study Comparing Number of Sampled Reefs." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.28774304.v1

Ruiz Moreno, A. 2025d. "Simulation Study with Abundance Trends." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.30189733.v1

Ruiz-Moreno, A., M. J. Emslie, and S. R. Connolly. 2024. "High Response Diversity and Conspecific Density-Dependence, Not Species Interactions, Drive Dynamics of Coral Reef Fish Communities." *Ecology Letters* 27: e14424.

Sale, P. F., R. K. Cowen, B. S. Danilowicz, G. P. Jones, J. P. Kritzer, K. C. Lindeman, S. Planes, et al. 2005. "Critical Science Gaps Impede Use of No-Take Fishery Reserves." *Trends in Ecology & Evolution* 20(2): 74–80.

Sandal, L., V. Grøtan, B.-E. Sæther, R. P. Freckleton, D. G. Noble, and O. Ovaskainen. 2022. "Effects of Density, Species Interactions, and Environmental Stochasticity on the Dynamics of British Bird Communities." *Ecology* 103(6): e3674.

Stan Development Team. 2024. "Stan Modeling Language Users Guide and Reference Manual, 2.35." https://mc-stan.org.

Thibaut, L. M., and S. R. Connolly. 2020. "Hierarchical Modeling Strengthens Evidence for Density Dependence in Observational Time Series of Population Dynamics." *Ecology* 101(7): e02956.

Thibaut, L. M., S. R. Connolly, and H. P. A. Sweatman. 2012. "Diversity and Stability of Herbivorous Fishes on Coral Reefs." *Ecology* 93(4): 891–901.

Thorson, J. T., H. J. Skaug, K. Kristensen, A. O. Shelton, E. J. Ward, J. H. Harms, and J. A. Benante. 2015. "The Importance of Spatial Models for Estimating the Strength of Density Dependence." *Ecology* 96(5): 1202–12.

Vehtari, A., A. Gelman, and J. Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC." *Statistics and Computing* 27(5): 1413–32.

Wauchope, H. S., T. Amano, W. J. Sutherland, and A. Johnston. 2019. "When Can we Trust Population Trends? A Method for Quantifying the Effects of Sampling Interval and Duration." *Methods in Ecology and Evolution* 10(12): 2067–78.

White, E. R. 2019. "Minimum Time Required to Detect Population Trends: The Need for Long-Term Monitoring Programs." *BioScience* 69(1): 40–46.

White, E. R., and C. A. Bahlai. 2021. "Experimenting with the Past to Improve Environmental Monitoring." *Frontiers in Ecology and Evolution* 8: 572979.

Zerbini, A. N., G. Adams, J. Best, P. J. Clapham, J. A. Jackson, and A. E. Punt. 2019. "Assessing the Recovery of an Antarctic Predator from Historical Exploitation." *Royal Society Open Science* 6(10): 190368.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ruiz-Moreno, Alfonso, Michael J. Emslie, and Sean R. Connolly. 2025. "Longer Time Series with Missing Data Improve Parameter Estimation in a State-Space Model in Coral Reef Fish Communities." *Ecology* 106(12): e70270. https://doi.org/10.1002/ecy.70270