spatialSim: multi-species spatiotemporal size-structured operating model for management strategy evaluation

Christopher D. Nottingham and Russell B. Millar
Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

Corresponding author: Christopher Nottingham (e-mail: c.nottingham@auckland.ac.nz).
Abstract

Spatiotemporal processes have the potential to be one of the most influential factors governing how fisheries targeting sedentary species respond to harvesting. Despite this, management strategy evaluation often fails to account for space or does so at low resolutions due to compute constraints. In this paper, a multi-species spatiotemporal size-structured operating model for sedentary species is presented. The model combines a spatially continuous Gaussian Markov Random Field model of the population dynamics with an areal harvesting model that supports preferential targeting and site selection constraints (e.g., economic constraints). This approach is very compute efficient, which makes it feasible to simulate realistic fisher dynamics and catch data at true spatial scale (e.g., the swept area of a dredge). The New Zealand surfclam fishery was used as a case study to demonstrate the versatility of the operating model and to showcase the simulation of localized depletion, which was manifest in the generation of realistic catch-per-unit-effort data that were uncorrelated with the trends in population abundance. The model is available as part of the open-source R package spatialSim.

Keywords: operating model; management strategy evaluation; spatiotemporal; size-structured; multi-species

Introduction

Determining an optimal management strategy is a critical part of sustainably managing a fishery. Traditionally, management advice has been based on a ‘best assessment’ of the resource, which involves fitting a model to fisheries data and evaluating uncertainty using confidence intervals and sensitivity tests. Recommendations for management action are then determined by applying a harvest control rule to values derived from the model or from model-based projections that assume a constant catch or fishing mortality rate (Butterworth 2007).

Contemporary fisheries management has started moving away from the traditional approach in favor of a more holistic simulation procedure known as management strategy evaluation (MSE) (Haddon 2012). MSE has its origins in the field of operations research (Charles 1995), which is centered around problem-solving methodology that can be used in the pursuit of im-
proved decision-making and efficiency. The MSE process uses a feedback loop to compare the performance of alternate fisheries management strategies with respect to achieving pre-assigned management objectives. The loop contains a simulation model known as the operating model (OM) that represents the true state of nature and a management strategy, which is made up of an estimation method (e.g., a stock assessment model) and a harvest control rule. The sequence of events in the feedback loop are: (1) simulate the true state of nature from the OM for assessment period $t$, (2) simulate fisheries data (e.g., commercial catch and survey data) from the OM, (3) estimate the stock status from the assessment model using the simulated data, (4) apply the harvest control rule to determine management recommendations (e.g., quota, effort limit, size limit or time-area closure) for assessment period $t + 1$, and (5) subject the management recommendations to implementation error such as quota overages caused by unreported catch. The loop is repeated for a sufficient number of future assessment periods and performance statistics are calculated.

MSE can be a very valuable tool for fisheries management. However, it relies heavily on the specification of the OM(s) and, in particular, the range of assumptions and uncertainties that are included in the modelled processes. In the literature, there are a large number of cases where the adequacy of the assumptions are questionable. One widespread issue is a “home cooking” problem, where the OM used to test a management strategy or assessment model is developed by the same person and makes the same or similar assumptions as the assessment model. This approach is suboptimal as it is likely to result in an unfair test, particularly when evaluating assessment methods. Another issue that is also prevalent in many cases and pertinent in the context of this work is the inadequate representation of spatial processes, which are often ignored (Punt et al., 2016) or incorporated by modelling a stock across a relatively small number of discrete zones (e.g., Punt and Hobday, 2009; Fay et al., 2011). Generally, this means the effects of localized depletion are also ignored or poorly captured in the simulated system and catch data. Consequently, the performance of any assessed management strategy will likely be overstated. This issue has the potential to lead to very misleading management recommendations, particularly for sedentary species that exhibit a significant amount of spatially heterogeneity, density dependence, and limited population connectivity.
In this paper a novel multi-species OM with spatiotemporal size-structured dynamics is presented. It provides scientists with a tool that can be used to avoid the “home cooking” problem and has a bespoke formulation balancing simplicity and complexity for simulating high resolution spatiotemporal dynamics of correlated populations that exhibit spatially heterogeneity, density dependence, and limited population connectivity. The model provides users with large amounts of flexibility when specifying systems, and it was developed in a modular way to facilitate the further development of structures and dynamics that are currently not supported.

The model’s state dynamics are an extension of the single-species non-spatial size-structured dynamics described by [Punt et al.] (2013). They incorporate spatiotemporal variation in recruitment and natural mortality using Gaussian Markov Random Fields (GMRF), which results in the simulated populations varying continuously across space. The spatially continuous population dynamics are combined with an areal model that can be used to simulate scientific surveys and commercial fishing on a grid containing the projected numbers-at-size.

The harvest model supports the specification of different levels of preferential targeting (bias to harvest from areas with higher densities of individuals) and different levels of fishing intensity across space. In addition, it supports user specified site-selection probabilities that can be used to incorporate factors such as economic constraints and area closures. Combining the continuous population dynamics with the areal harvesting model is computationally very efficient, which makes it feasible to simulate catch data to scale (e.g., the swept area of a dredge or diver).

The model also supports density dependence in recruitment through the incorporation of a spatially explicit stock-recruit function, where the number of recruits at a given location depend on the density of sexually mature individuals within a user-set spatial range of the location. This specification makes the assumption that the simulated population(s) self-recruit with no input from outside the spatial domain. Therefore, it is only appropriate for species/populations that do not exhibit significant larval drift, for which, there are likely to be many [Hannan, 2014; Teske et al., 2016].

The multi-species aspect of the model provides users with the option of specifying a species correlation structure on recruitment and assumes that the natural mortality rates of each species
are proportionally related across spacetime. This allows fishery systems comprising species with correlated spatiotemporal population structures to be simulated, and facilitates a straightforward way of comparing multi-species assessment methods for fisheries with these characteristics. This part of the model was developed to be as simple as possible, while retaining the complexity required to answer a set of targeted questions and produce insightful results. This distinguishes this model from a lot of the multi-species alternatives (e.g., Christensen et al. 2005; Gray et al. 2006; Fulton et al. 2011), which are often overly complex, slow to run, and difficult to specify/understand.

The model is coded in the C++ template language TMB (Kristensen et al. 2016), which has a number of useful features. These include an elegant interface with R, simulation routines for GMRFs, multidimensional arrays, ragged arrays, sparse matrices and fast linear algebra routines through the Eigen library. In the presented work, automatic differentiation is not required and, therefore, double data types are used in place of CppAD data types. This enables a wider range of C++ functionality, which is taken advantage of through C++11 routines. This work is an open-source project and is available in the R (R Core Team 2020) package, spatialSim (https://github.com/cnotti/spatialSim).

The following sections of this paper first outline the structure of the operating model and different options users have when specifying each of the simulated processes. These include the growth of individuals, natural mortality, harvesting, recruitment, and model initialization. Following this, a simple case study of the New Zealand surfclam fishery is presented. The case study demonstrates the versatility of the operating model and the importance of incorporating spatial processes into a MSE. In addition, it provides insight into the computational cost of running the model by presenting execution times for a number of different sized spatial domains.

**Model overview**

The model is made up of two phases, which include an initialization phase, during which the populations are initialized, and a harvesting phase that represents the period in which harvesting
occurs. Both phases are centered around a state equation that describes how the numbers-at-size in the simulated populations evolve through spacetime. Processes that are part of the state dynamics include the growth of individuals, natural mortality, harvesting, and recruitment. The dimensions of the different structures of the model are controlled by a number of user-set parameters. These are presented in Table 1.

**State dynamics**

The state equation is loosely based on the size-structured dynamics described by Punt et al. (2013), with the addition of multi-species processes that describe spatial and temporal variation in recruitment and natural mortality. The spatial processes are modelled using the stochastic partial differential equation (SPDE) approach of Lindgren et al. (2011). Hence, the numbers-at-size vary continuously across the spatial domain of the fishery \( \Omega \), which is defined by an isometric triangulation with \( v_S \) nodes, where each node represents a location referenced by \( s \),

\[
n_{c,s,t+1,j} = \sum_k G_{c,j,k} (n_{c,s,t,k} - x_{c,s,t,k} + r_{c,s,t,k}) e^{-M_{c,s,t,k}}.
\]

Here, \( n_{c,s,t,k} \) is the number of individuals belonging to species \( c \) that have size in the interval \( Q_{c,k} = (l_{c,k-1}, l_{c,k}) \) at time \( t \) and location \( s \), \( G_{c,j,k} \) is the probability of an individual belonging to species \( c \) growing from size interval \( Q_{c,k} \) into size interval \( Q_{c,j} \), \( M_{c,s,t,k} \) is the instantaneous rate of natural mortality for individuals belonging to species \( c \) with size \( l \in Q_{c,k} \) at location \( s \) between time \( t \) and \( t + 1 \), \( x_{c,s,t,k} \) is the number of harvested individuals belonging to species \( c \) with size \( l \in Q_{c,k} \) at location \( s \) between time \( t \) and \( t + 1 \), and \( r_{c,s,t,k} \) is the number of individuals belonging to species \( c \) with size \( l \in Q_{c,k} \) that are recruited to the population at location \( s \) between time \( t \) and \( t + 1 \).
| Parameter | Description |
|-----------|-------------|
| $v_B$     | Number of time periods in the initialization phase |
| $v_Y$     | Number of years in the harvesting phase |
| $v_P$     | Number of annual periods |
| $v_C$     | Number of species |
| $v_S$     | Number of triangulation nodes in spatial domain $\Omega$ |
| $v_L$     | Number of size intervals |
| $v_F$     | Number of cells in the harvest grid |

Table 1: Parameters specifying the dimensions of model structure.

**Growth**

The probability of an individual transitioning from size interval $Q_{c,k}$ to $Q_{c,j}$ depends on the distribution of sizes within interval $Q_{c,k}$ and the distribution of growth conditional on those sizes, denoted $f_{Uc}(u)$ and $f_{\Delta c}(\delta; u)$, respectively. It follows that

$$G_{c,j,k} = \int_{l_{c,j-1}}^{l_{c,j}} \int_{l_{c,k-1}}^{l_{c,k}} f_{\Delta c}(y - u; u) f_{Uc}(u) \, du \, dy,$$

where $\delta = y - u$ is the growth of an individual between time $t$ and $t + 1$, given initial size $u$ at time $t$. The two-dimensional integral in (2) is approximated using the method described by Millar and Nottingham (2019). This approximation assumes that the distribution of growth conditional on an individual’s initial size is lognormally distributed and the sizes of individuals are uniformly distributed within each interval, but, for tractability, that individuals with initial sizes in the interval $Q_{c,k}$ have the same distribution for the expected amount of growth, $\delta_{c,k}$. The specification of $G$ depends on a user-defined growth function and a parameter controlling the variability of $\delta_{c,k}$. Additionally, users must specify a function that defines how an individual’s weight changes with growth. A detailed overview of the user-set growth parameters and functions is presented in Table 2.
Table 2: Parameters and functions that specify the growth of individuals.

Natural Mortality

Natural mortality is specified using a parameter representing the expected instantaneous mortality rate for each species, $\alpha_M$, and two stochastic processes that include a spatiotemporal field $\eta$ and a temporally correlated process $\iota$. In total, it depends on eight user-specified parameters, which are presented in Table 3. Explicitly, the instantaneous natural mortality rate on individuals belonging to species $c$ with size $l \in Q_{c,k}$ at location $s$ between time $t$ and $t + 1$ is defined as,

$$\Delta t M_{c,s,t,k} = \alpha_{M_c} \frac{\text{logistic}(\iota_t) \exp(\eta_{s,t,k})}{\nu_M},$$

where $\alpha_{M_c}$ is the expected instantaneous mortality rate for species $c$, $\nu_M$ is a normalizing constant that standardizes the fraction to have a mean of one, the exponential term ensures that spatiotemporal variation in natural mortality is on the interval $(0, 1)$, the logistic term provides additional flexibility by allowing temporally auto-correlated scaling of the spatiotemporal field (e.g., cyclical scaling), while retaining stability by bounding the term to be on the interval $(0, 1)$, $\iota_t \sim N(\mu_t (1 - \rho_t) + \rho_t \iota_{t-1}, \sigma_t)$ is a first order autoregressive (AR1) temporal process, and $\eta$ is a Gaussian stochastic process with correlation between space, time, and the sizes of individuals, $\text{vec}(\eta) \sim \text{MVN}(\vec{\mu}_\eta, \sigma_\eta S_\eta \otimes T_\eta \otimes L_\eta)$. Here, $\text{vec}$ denotes the vectorization operator, which stacks the elements of an object over its dimensions to form a vector, $\otimes$ is the Kronecker product, $\vec{1}$ is a column vector of ones, $\mu_\eta$ is the marginal mean of $\eta$, $\sigma_\eta$ is the marginal sd of $\eta$, and $S_\eta$, $T_\eta$, and $L_\eta$ are the correlation matrices for space, time, and size, respectively. The spatial, temporal, and size components of $\eta$ are separable, such that, if
time and size are marginalized, $\eta \sim \text{MVN}(1\mu_\eta, \sigma_\eta S_\eta)$ is a Matèrn field, if space and size are marginalized, $\eta_t \sim N(\mu_\eta(1 - \rho_\eta) + \rho_\eta \eta_{t-1}, \sigma_\eta)$ is an AR1 process, and if time and space are marginalized, $\eta_k \sim N(\mu_\eta(1 - \psi_\eta) + \psi_\eta \eta_{k-1}, \sigma_\eta)$ is an AR1 process. The marginal standard deviation of $\eta$ is $\sigma_\eta = (\tau_\eta \kappa_\eta \sqrt{4\pi})^{-1}$, where $\tau_\eta$ and $\kappa_\eta$ are the scale and range parameters from the SPDE approach of [Lindgren et al. (2011)], which is used to approximate the Matèrn field. The range parameter $\kappa_\eta$ determines the rate of spatial decorrelation (e.g., the distance for which correlations decline to 10% is approximately $\sqrt{8\kappa_\eta}$) and both $\kappa_\eta$ and $\tau_\eta$ scale the marginal sd.

| Parameter | Description |
|-----------|-------------|
| $\mu_\eta$ | Expected value of $\eta$ |
| $\psi_\eta$ | Size autocorrelation parameter of $\eta$ |
| $\rho_\eta$ | Temporal autocorrelation parameter of $\eta$ |
| $\kappa_\eta$ | SPDE range parameter of $\eta$ |
| $\tau_\eta$ | SPDE scale parameter of $\eta$ |
| $\mu_\iota$ | Expected value of $\iota$ |
| $\rho_\iota$ | Temporal autocorrelation parameter of $\iota$ |
| $\sigma_\iota$ | Marginal standard deviation of $\iota$ |
| $\alpha_{Mc}$ | Expected rates of instantaneous natural mortality for each species $c \in \{1, \ldots, v_C\}$ |

Table 3: Parameters that specify natural mortality.

### Fishing Mortality

Harvesting is simulated on a grid with $v_F$ square cells that cover the domain of the fishery, $\Omega$. Each grid cell represents a potential fishing site and has area equal to the area fished by one unit of effort, $a_F$ (e.g., the area-swept by the tow of a dredge). The number of individuals with size $l \in Q_{c,k}$ at grid cell $f$ and time $t$ is interpolated from the values at the three closest triangulation nodes using barycentric coordinates, which calculate weightings of the vertex values to allot to any point within a triangle when interpolating the triangular lattice field into continuous space,

$$n_{c,f,t,k} = \frac{a_F}{a_S} \sum_{s \in \hat{N}_f} A_{f,s} n_{c,s,t,k}.$$  \hspace{1cm} (4)

Here, $A_{f,s}$ is the weight attributed to location $f$ from the node at location $s$ in the barycentric coordinate system, $a_S$ is the area associated with each of the triangulation nodes (i.e., the total area of $\Omega$ divided by the number of triangulation nodes within $\Omega$), and $\hat{N}_f$ is the neighborhood
of location $f$.

The number of harvested individuals from the $i$th haul that belong to species $c$ with size $l \in Q_{c,k}$ is $x_{i,c,k} = n_{c,f,t,k} \zeta_{c,k}$. Here, $\zeta_{c,k}$ is the probability that an individual belonging to species $c$ with size $l \in Q_{c,k}$ is retained by the dredge. The harvest of these individuals is apportioned to the triangulation nodes $\in N_f$ with proportion relative to $A_{f,s}$.

The total annual catch for species $c$ can be divided into catches representing scientific surveys and catches representing commercial harvesting. The number of survey sites and their distribution are user-specified. The distribution of commercial catch is determined by Algorithm A.4 (Appendix A), which supports the specification of preferential targeting, area closures, and other site selection constraints (e.g., economic constraints can be incorporated through the specification of lower site selection probabilities for areas further away from ports).

The harvest algorithm has two distinct configurations that are specified using different combinations of the parameters $F^{\text{int}}$, $p^{\text{targ}}$, and $p^{\text{loc}}$. The first configuration simulates vessels systematically fishing across $\Omega$, with harvesting focused on areas of higher abundance. The level of fishing intensity across space is controlled by the parameters, $F^{\text{int}}$ and $p^{\text{targ}}$, where the former determines the size of the sampling region for a given time step and the latter is the probability of sampling from a site where the biomass is above the median. It is worth noting that this setup does not use the third parameter, $p^{\text{loc}}$. The second configuration uses the entire domain as the sampling region for each time step (i.e., $F^{\text{int}} \approx 0$) and specifies the site selection probabilities as the product of the preferential targeting probability (i.e., $p^{\text{targ}}$ for a site where the biomass is above the median and $1 - p^{\text{targ}}$ otherwise), and $p^{\text{loc}}$, which represents fixed site-selection probabilities specific to each cell in the harvesting grid. This specification allows users to account for economic considerations such as a site’s distance from a port and other factors (e.g., depth, substrate type and other accessibility constraints) that may make areas more or less likely to be targeted.

An example of how the harvest algorithm can be specified to simulate different fishing patterns is presented in Figure 1. In this example, commercial harvesting is simulated from January to June and the parameters $p^{\text{targ}}$, $F^{\text{int}}$, and $p^{\text{loc}}$ are varied to illustrate their influence. In addition, to demonstrate how an area-closure could work, part of the domain (shaded in green)
is excluded from the commercial sampling zone. A detailed overview of the user-set harvest parameters and functions is presented in Table 4.

Figure 1: Plot (a) shows a spatial density map of the distribution of pre-harvest biomass (kg m\(^{-2}\)) for a fishery simulated using the spatialSim OM, and plots (b) and (c) show spatial density maps of the distribution of fishing effort resulting from different configurations of the harvest algorithm [Algorithm A.4] when applied to this fishery. In the former, each element of \(p_{\text{loc}}\) was 1 and harvesting was specified to systematically move across the domain with each time step for four scenarios that were specified using different values of \(p_{\text{targ}}\) and \(F_{\text{int}}\). Contrastingly, in plot (c) the probability of sampling commercial catch from a site given its geographic location, \(p_{\text{loc}}\), was specified to exponentially decrease as the target location got further away from the domain’s centroid and the other two parameters were specified as \(p_{\text{targ}} = 1\) and \(F_{\text{int}} = 0.1\). In all scenarios, the green shaded region within the dotted lines represents an area closed to commercial fishing.
Recruitment

The number of species $c$ individuals that are recruited at location $s$ between time $t$ and $t + 1$ with size $l \in Q_{c,k}$ depends on: a scaled Beverton-Holt stock-recruit function $SR(\cdot)$ (Beverton and Holt [1957]) of the biomass of species $c$ individuals that are sexually mature at time $t$ averaged across the nodes within $r_c^{\text{range}}$ km of node $s$, giving the overall term $r_{0c}SR(\bar{SSB}_{c,s,t})$; the proportion of recruitment allotted to size interval $Q_{c,k}$ given $\Delta t$, $v_k$; the probability of a species $c$ individual recruiting conditional on the environmental conditions at location $s$ between time $t$ and $t + 1$, $\xi_{c,s,t}$; a recruitment cutoff threshold for species $c$, $r_{c}^{\text{cut}}$; and random spatiotemporal variability $\epsilon_{c,s,t}$.

\begin{equation}
    r_{c,s,t,k} = \begin{cases} 
        r_{0c}SR(\bar{SSB}_{c,s,t}; SSBO_c, h_c)\xi_{c,s,t}\logistic(\epsilon_{c,s,t})v_k, & \frac{\bar{SSB}_{c,s,t}}{SSBO_c} \geq r_{c}^{\text{cut}} \\
        0, & \frac{\bar{SSB}_{c,s,t}}{SSBO_c} < r_{c}^{\text{cut}}.
    \end{cases}
\end{equation}

The stock-recruit function $SR(\cdot)$ is parameterized in terms of the steepness parameter, $h_c$, which is the ratio of recruitment when $\bar{SSB}_{c,s,t}$ is 20% of $SSBO_c$ to recruitment when $\bar{SSB}_{c,s,t}$ is equal $SSBO_c$ (Francis [1992] (Figure 2), where $SSBO_c$ is the expected unfished biomass of
individuals belonging to species $c$ that are sexually mature at a location with above average environmental conditions for recruitment (see the section describing initialization for further details). The biomass of species $c$ individuals that are sexually mature at time $t$ averaged across the spatial nodes that are within $r_c$ range km of node $s$ is mathematically defined as,

$$
SSB_{c,s,t} = \frac{\sum_{s^*} I_{s,s^*} SSB_{c,s^*,t}}{\sum_{s^*} I_{s,s^*}}.
$$

where $I_{s,s^*}$ is an indicator variable that is equal to one if location $s^*$ is within $r_c$ range km of location $s$ and zero otherwise, and $SSB_{c,s,t} = \sum_k (n_{c,s,t,k} - x_{c,s,t,k})\tilde{m}_{c,k}\tilde{w}_{c,k}$ is the total biomass of species $c$ individuals that are sexually mature at location $s$ and time $t$ after accounting for harvesting. Here, $\tilde{m}_{c,k}$ is mean proportion of mature species $c$ individuals with size $l \in Q_{c,k}$ and $\tilde{w}_{c,k}$ is the expected weight of a species $c$ individual with size $l \in Q_{c,k}$ (Table 2).

![Figure 2: The effect of the steepness parameter, $h$, on the Beverton-Holt stock-recruit function.](image)

The random spatiotemporal variability, $\epsilon$, is a Gaussian stochastic process with correlation between species, time, and space, $\text{vec}(\epsilon) \sim \text{MVN}(\mu_\epsilon, \sigma_\epsilon C_\epsilon \otimes T_\epsilon \otimes S_\epsilon)$. Here, $\mu_\epsilon$, is the marginal mean of $\epsilon$, $\sigma_\epsilon$ is the marginal sd of $\epsilon$, $C_\epsilon$ is a user-defined species correlation matrix, and $T_\epsilon$ and $S_\epsilon$ are correlation matrices representing the temporal and spatial dimensions, respectively. If the temporal and species dimensions are marginalized, $\epsilon \sim \text{MVN}(\mu_\epsilon, \sigma_\epsilon S_\epsilon)$ is a Matérn field and if the species and spatial dimensions are marginalized, $\epsilon_t \sim N(\mu_\epsilon (1 - \rho_t) + \rho_t \epsilon_{t-1}, \sigma_\epsilon)$ is an AR1 process. The Matérn field is approximated using the same approach
that is used for the spatial field on natural mortality. Hence, the marginal standard deviation of $\epsilon$ is $\sigma_{\epsilon} = (\tau_{\epsilon}, \kappa_{\epsilon} \sqrt{4\pi})^{-1}$, where $\kappa_{\epsilon}$ and $\tau_{\epsilon}$ are the scale and range parameters from the SPDE approach.

The proportion of new recruits allotted to each size interval and the species specific probabilities of individuals recruiting conditional on the environmental conditions between each time period and at each location are defined externally by the user. An example of how $\xi$ and $\epsilon$ can be used to specify spatiotemporal variation in the recruitment of individuals is presented in Figure 3. In this example, $\xi$ includes a seasonal scaling effect and randomly shifting depth-zones that are optimal for recruitment. A detailed overview the parameters and functions that are used to specify recruitment is presented in Table 5.
Figure 3: An example of how $\xi$ and $\epsilon$ can be used to specify spatiotemporal variation in the recruitment of individuals of a given species.
Initialization

The state of the population at time $t = 0$ (i.e., the beginning of the first time period in the first year of harvesting) is calculated by executing the population dynamics (1) for $v_B$ prior time steps that act as an initialization period. At the start of the first time interval in the initialization period, the number of individuals belonging to each species is equal to zero. During the first two thirds of the $v_B$ steps, the SR relationship is equal to 1 for all locations and species: that is, $SSB_{c,s,b} = SSB0_c$ for $c \in \{1, \ldots, v_C\}$, $s \in \{1, \ldots, v_S\}$, and $b \in \{1, \ldots, 2v_B/3\}$. The expected unfished biomass of individuals belonging to species $c$ that are sexually mature at a location with above average conditions for recruitment is calculated from the biomass values obtained during the second third of the initialization period,

$$SSB0_c = \frac{3 \sum_b \sum_s SSB_{c,s,b}}{v_{\text{opt}}}, \quad \frac{v_B}{3} < b \leq \frac{2v_B}{3}, \quad (7)$$
where \( v_{S_{opt}} \) is the total number of spatial nodes where \( \xi_{c,s,b} \geq \sum_{s} \sum_{b} \xi_{c,s,b} v_S^{-1} v_B^{-1} \), for \( b \in \{v_B/3 + 1, \ldots, 2v_B/3\} \). During the final third of the \( v_B \) steps, recruitment is calculated using (5). This period provides time for the number of individuals at the sub-optimal recruitment locations to converge to lower values. To ensure \( SSB_0 \) is calculated over a number of full recruitment cycles, \( v_B/v_P \) must be an integer divisible by 3, where \( v_P \) is the number of time periods that make up an annual cycle.

**Case study: New Zealand surfclams**

**Model parameters**

In New Zealand, the total extent of suitable surfclam habitat exceeds 2300 km\(^2\) and densities are typically in excess of 200 tonnes per km\(^2\) (White et al., 2015), suggesting that a population biomass of 500 000 t or more is possible. However, total annual catch has not yet reached 800 t. To showcase some of the model’s features and demonstrate execution time, a scenario was specified representing a hypothetical fishery applied to two New Zealand surfclam species, *Spisula aequilatera* and *Mactra murchisoni*. Simulations were carried out for three scenarios, which differed in their specified level of commercial fishing intensity across space. The OM scenarios are denoted here as OM\(_{\text{high}}\), OM\(_{\text{med}}\), and OM\(_{\text{low}}\), where the subscript represents the associated level of fishing intensity. Each scenario was replicated 100 times and the specified parameters and functions were primarily based on published data for the Manawatu region. A detailed overview is presented in Appendix B.

The spatial domain was 56 km\(^2\), which is around the same size as the fishable area of the Manawatu fishery that the model’s parameters were based on. It comprised 500 triangulation nodes, 557 433 harvest cells and was divided into seven latitudinally extending bathymetrical depth bands ranging from 1 to 7 meters deep, with each representing a one meter change in depth. The fishery was simulated for a period of 15 years using monthly time increments. The first five years represented a period of unsustainable fishing and the ten years that followed were used to demonstrate the rate of recovery given no harvesting.

The growth of individuals from each species followed a von Bertalanffy curve (von Bert-
lanffy, 1938) with the parameterization described by Fabens (1965). The von Bertalanffy parameters were set to the average values estimated by Cranfield et al. (1993), Cranfield et al. (1996), and Cranfield and Michael (2001). The sd of the growth increment on the log scale was set to 0.15 for each species and the size distribution of each population was represented using 10 size intervals.

Harvesting was simulated on a grid where each cell represented a single dredge tow corresponding to an area-swept of approximately 100 m$^2$. Catch limits were set at 1000 t per annum and scientific biomass surveys were simulated at 50 random sites uniformly distributed across $\Omega$ per year. The commercial fishing target zones were defined as the entirety of $\Omega$ for both species. Selectivity was 0.95 for individuals in the first size-class and 1 otherwise. The probability of a commercial vessel harvesting from a cell with biomass above the median was set to $p_{\text{targ}} = 0.90$, which corresponds to fishers having very good knowledge about the distribution of individuals in the areas of harvesting.

The parameters of $\iota$ were set to $\sigma_\iota = 1$, $\rho_\iota = 0.75$, and $\mu_\iota = -2$. This combination of values produces a right skewed distribution with moderate variability and a high level of temporal autocorrelation, equating to relatively long periods of lower mortality rates followed by shorter periods with higher rates (Figure 4). The autocorrelation parameters of $\eta$ were set to $\psi_\eta = 0.75$ and $\rho_\eta = 0.75$, which corresponds to strongly correlated mortality rates for individuals of similar size and strongly correlated mortality rates between time periods, given $\iota$ (Figure 5).

The scale and range parameters of $\eta$ were set to $\tau_\eta = 0.1$ and $\kappa_\eta = \exp(3)$, which correspond to relatively small amounts of spatial variability in $\eta$ with correlation decaying slowly over large distances (the distance for which correlation declines to 10% was approximately 11.8 km). The expected rates of instantaneous natural mortality for individuals belonging to each species were set to $\alpha_M = (0.79, 0.42)^T$, which are the averaged estimates of Cranfield et al. (1993).
The probability of a species individual recruiting conditional on the environmental conditions between time $t$ and $t + 1$ at location $s$, was specified using two independent components. These included (1) an annual seasonality term consistent with research on the gametogenic cy-
cles of surf clam species in New Zealand (Nuha, 2013; Nottingham and White, 2014) and (2) a time-invariant bathymetrical term based on the stratified biomass survey by White et al. (2012).

The marginal mean of $\epsilon$ was set to $-7$, which produced a very sparse and patchy recruitment pattern (Figure 6) that resulted in distinct beds of individuals. The temporal autocorrelation parameter of $\epsilon$ was set to $\rho_\epsilon = 0.75$, which corresponds to a high level of temporal autocorrelation in recruitment rates. The scale and range parameters of $\epsilon$ were set to $\tau_\epsilon = 0.0005$ and $\kappa_\epsilon = \exp(5)$, which correspond to relatively large amounts of spatial variability in $\epsilon$ with correlation decaying quickly over relatively short distances (the distance for which correlations decline to 10% is approximately 1.6 km). The species correlation matrix of $\epsilon$ was set as the identity matrix. This combination of parameters produces a patchiness in the spatial distribution of the populations that is consistent with what was reported by White et al. (2012).

![Figure 6](image.png)

Figure 6: The plot on the left (a) is a spatial density map of the simulated logistic transformed $\epsilon$ values for the first 6 months of the first year of the case study and the plot on the right (b) is a marginal histogram of all of the simulated logistic transformed $\epsilon$ values.

The spawning range and steepness parameters of the stock-recruit function were set to $r_{\text{range}} = 0.5 \text{ km}$ and $h_c = 0.9$ for each species. These values were not based on any real data, but were chosen to support the simulation of serial depletion over the five year harvesting...
period. The values of \( r_0 \) were chosen such that each species had biomass densities in the ranges indicated by Haddon et al. (1996), Triantafillos (2008), White et al. (2012), and White et al. (2015).

To provide insight into the computational cost of running the model, the first five years of the OM\textsubscript{high} scenario was repeated on 20 different domains ranging in size from 56 km\(^2\) to 224 km\(^2\) and the median execution times were recorded from ten replicates. The number of triangulation nodes and harvest cells changed in proportion with the area of \( \Omega \) and ranged from 500 to 2000 nodes and 557 433 to 2 244 792 cells, respectively. Execution times were recorded from an R session linked to the multithreaded Intel Math Kernel Library (Intel Corporation, 2009) on a desktop computer with an Intel 7700K processor and 32 gigabytes of memory with a clock frequency of 2667 MHz. The recorded times excluded the overheads associated with the initial setup process, which included creating a list specifying the model settings and simulating the random processes.
Figure 7: The population densities (kg m$^{-2}$) of the simulated surfclam populations for one of the replicates of the three OM variants specifying high, medium, and low levels of fishing intensity across space.
Figure 8: Plot (a) shows the biomass of each simulated surfclam population across time for each of the three operating model scenarios averaged across the 100 replicates, and plot (b) shows the biomass of each simulated surfclam population aggregated into 4 equally sized latitudinal regions across time for each of the three operating model scenarios averaged across the 100 replicates. Latitudinal region 1 is the southern most followed by regions 2, 3, and 4. The shaded regions represent the time period with no harvesting.
Figure 9: The mean population densities, commercial CPUE, and survey CPUE across time for one of the replicates of the three OM variants specifying high, medium, and low levels of fishing intensity across space.

Figure 10: Median execution times from 10 replicated sets of 20 model runs with different sized spatial domains that ranged from 56 km$^2$ to 224 km$^2$. 
Results

During the five year period of harvesting there were significant declines in the abundance of individuals and varying levels of spatial serial depletion among the different OM cases (Figure 7; Figure 8). During this period, the changes in abundance and spatial serial depletion were strongly correlated to the level of fishing intensity with the higher levels resulting in smaller decreases in population size and greater increases in the level of spatial serial depletion. At the highest level of fishing intensity, population levels in the southern half of the domain were almost completely wiped out, while population levels in the northern half of the domain were largely unaffected.

During the ten year period post-harvesting, the recovery rates in population abundance were limited for all scenarios. Overall, there was a slight increase in the abundance of *M. murchisoni* individuals, but no substantial recovery in the simulated *S. aequilatera* populations. Nevertheless, it is notable that the populations from the scenario with the lowest level of fishing intensity showed the strongest signs of a recovery, followed by the populations from the OM$_{high}$ and OM$_{med}$ scenarios. The reason for this is made clear in Figure 8, which shows the biomass trajectories across time aggregated into 4 equally sized latitudinal regions. The recovery rates associated with the OM$_{high}$ and OM$_{med}$ scenarios were primarily driven by increases in abundance in the northern parts of domain, which had seen less fishing. Alternatively, the recovery rates associated with the OM$_{low}$ scenario were driven by increases in abundance across the entire domain.

In all OM scenarios, temporal trends in stock depletion were captured reasonably well by the simulated survey CPUE data that represented 50 annual survey dredge tows per year (e.g., Figure 9). In contrast, the simulated commercial CPUE data from the models with more intensive fishing regimes (OM$_{high}$ and OM$_{med}$) were often extremely poor indicators of the states of the fisheries (e.g., Figure 9). This was because intensive harvesting was localized to one region each season. Consequently, in each subsequent fishing season harvesting was occurring in new areas with relatively high densities of individuals. In the case of OM$_{low}$, where harvesting was the least intensive, the simulated commercial CPUE was, on average, a
reasonable measure of the biomass of each species.

Model execution times increased exponentially with domain size. The median execution times for each of the 20 different sized domains in the 10 replicated sets ranged between 15.6 seconds for the case where the domain was 56 km$^2$ to 158.0 seconds for the case where the domain was 224 km$^2$ (Figure 10).

**Discussion**

There are a wide range of different spatial OMs that have been described in the literature. These range from very simplistic models that partition a population across a small number of discrete zones (e.g., Punt and Hobday [2009] Fay et al. [2011] Ives and Scandol [2013]), to extremely complex ecosystem models (e.g., Christensen et al. [2005] Gray et al. [2006] Fulton et al. [2011]) that, in practice, can only be run at low spatial resolutions due to computational constraints. Some authors have rightly inferred that rather than providing insight, the latter approach can lead to two things that are not understood, the real world system that was never properly understood to begin with, and a highly detailed model of it (Paola [2011] Needle [2015]).

This has prompted recent attempts to develop models that fall between the overly simplistic and overly complex ecosystem models. An example of this is the Honeycomb model (Needle [2015]), which incorporates discrete spatial variability and fleet dynamics, but deliberately avoids incorporating superfluous elements that make understanding the model difficult. The interpretable simplicity of these kinds of model provide a much more user friendly experience than the complex ecosystem alternatives. Parameter setting is straightforward and the computational benefits associated with a simpler model enables users to model the system at much higher resolutions.

The model presented in this paper was developed with the same philosophy of interpretable simplicity that drove the development of the Honeycomb model. However, it is unique in that it makes it computationally feasible to simulate a fishery system at the same spatial resolution that it would operate in real world situations (e.g., the average area-swept by the tow of dredge).
This efficiency is achieved by combining a spatially continuous GMRF model of the population dynamics with a harvest algorithm based on the highly efficient sampling routine described by Efraimidis and Spirakis (2006) that operates on an areal grid system containing the projected numbers-at-size.

The presented model was designed to help answer management questions for sedentary species, while allowing users to consider the implications of spatial heterogeneity, fleet dynamics, and spawning range. These are of fundamental importance to providing management advice that protects sedentary species against serial depletion. This was demonstrated in the case study, which showcased the simulation of three scenarios that each represented a fictional fishery comprising two species. The number of recruits at a given location was specified to depend on the average density of sexually mature individuals within a 500 m range of that location and the effect of three different levels of fishing intensity across space were compared.

The results demonstrated, that for the given scenarios, more intensive fishing across a spatial area can lead to lower levels of stock depletion in the short term, but in the longer term can potentially lead to higher levels of depletion due to reduced recruitment and slower recovery rates. Another notable finding was that the commercial CPUE data generated from the two scenarios with more intensive fishing were completely uncorrelated with the abundance of individuals in the populations. In the scenario where harvesting was the most spatially extensive, the simulated CPUE data appeared to be a reasonable proxy for estimating total abundance. This phenomenon resulted from the more intensive fishing regimes producing highly localized commercial catch data that were generated from areas that had not been fished in previous seasons. In addition to these findings, the case study also demonstrated how the operating model can be used to generate CPUE data representing scientific surveys. In all of the presented cases, samples from 50 random sites uniformly distributed across the fishery appeared to be a reasonably good proxy for estimating abundance.

The phenomenon that was observed in the case study, where the commercial CPUE data were uncorrelated with the trends in population abundance, is quite likely to occur in many of the worlds fisheries. Nevertheless, few existing stock assessment models are designed to model a fishery with data containing these features. In the fisheries literature there are three ap-
proaches that have been described to account for spatial effects (Punt, 2019). These include (1) pre-processing data to remove spatial effects (e.g., CPUE standardization), (2) approximating a spatial distribution by size and/or age using selectivity (e.g., Waterhouse et al., 2014), and (3) applying a model with spatially explicit population dynamics.

Recent research has primarily focused on categories (1) and (3), which has led to number of recent advancements, although, it is uncommon to see methods belonging to category (3) used in practice. In contrast, there has been a surge in the use of geostatistical CPUE standardization methods in stock assessments. This has been largely due to a geostatistical delta GLMM maximum likelihood estimator for abundance index standardization developed by Thorson et al. (2015b), which has also been extended to handle multiple-species/categories (Thorson, 2019). Some examples of recently developed models that incorporate spatially explicit population dynamics include a spatiotemporal size-structured model (Kristensen et al., 2014), a geostatistical delay-difference model (Thorson et al., 2015a), and a geostatistical surplus production model (Thorson et al., 2017).

As technology develops it is becoming more economically feasible to record commercial catch data at the per-haul level. The OM presented in this paper could be a useful tool to aid in the development of new spatiotemporal assessment models designed to leverage the additional information contained in these data (e.g., models that build on the work of Kristensen et al., 2014, Thorson et al., 2015a, Thorson et al., 2017, and Thorson, 2019) and, ultimately, to help quantify the costs and benefits of recording data in this way. In these applications, the spatialSim model’s ability to simulate scenarios with realistic catch data (e.g., CPUE data with bias caused by localized depletion) could facilitate a more meaningful assessment than a lot the other OMs that are currently available.

The current version of the spatialSim model has a lot of potential for further development and its utility could be greatly increased by simply extending the range of options for specifying the growth, natural mortality, and recruitment processes. This is because the current specification of these processes are only appropriate for cases where it is reasonable to assume that growth is spatially homogeneous, natural mortality is proportionally related among species, and the populations self-recruit. Thus, the model could benefit from future work that adds support
for spatially varying growth, different natural mortality correlation structures among species, and population connectivity (i.e., recruitment input from populations outside of the domain).

There is also a very wide scope for extending the harvest algorithm. A couple of good options could include incorporating species interactions in the site selection process (e.g., bias towards areas with lower or higher diversities of the simulated species) and adding support for site selection biases based on the sizes of individuals (e.g., bias towards areas with larger individuals). These features would greatly enhance the multi-species aspect of the model and increase the range of management options that could be explored (e.g., maximizing catch while avoiding by-catch constraints or small-fish protocols).

The model presented here provides a way of carrying out a spatial MSE where data can be simulated to scale with the fishery. It adds more complexity and realism than a lot of the currently available alternatives, yet retains enough simplicity to allow the model and its parameters to be easily understood by users. This simplicity means the model is limited in the range of scenarios it can simulate and questions it can answer. Nevertheless, it was developed in a modular way that is easily extendable. Thus, this open-source project has a lot of potential for further development, which would help facilitate its use in a much wider range of simulation studies.

Acknowledgments

I C. D. Nottingham thank the New Zealand Ministry for Primary Industries and the University of Auckland for the funding provided through the Ministry for Primary Industries Postgraduate Science Scholarship and the University of Auckland Doctoral Scholarship, which helped to make this project possible. We also thank R. Fewster and I. Tuck for their valuable input.
References

Beverton, R. J. H. and Holt, S. J. (1957). *On the dynamics of exploited fish populations*. Ministry of Agriculture and Fisheries, London.

Butterworth, D. S. (2007). Why a management procedure approach? *ICES Journal of Marine Science*, 64:613–617.

Charles, A. T. (1995). Fishery science: The study of fishery systems. *Aquatic Living Resources*, 8:233–239.

Christensen, V., Walters, C., and Pauly, D. (2005). *Ecopath with Ecosim: A user’s guide*. Fisheries Centre, University of British Columbia, Vancouver, Canada and ICLARM, Penang, Malaysia.

Cranfield, H. J. and Michael, K. P. (2001). Growth rates of five species of surf clams on a southern North Island beach, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 35(5):909–924.

Cranfield, H. J., Michael, K. P., and Francis, R. I. C. C. (1996). Growth rates of five species of surf clams on a beach in the South Island of New Zealand. *Marine and Freshwater Research*, 47:773–784.

Cranfield, H. J., Michael, K. P., and Stotter, D. (1993). Estimates of growth, mortality and yield per recruit for new zealand surf clams. New zealand fisheries assessment document 93/20., MAF,, New Zealand.

Efraimidis, P. S. and Spirakis, P. G. (2006). Weighted random sampling with a reservoir. *Information Processing Letters*, 97(5):181–185.

Fabens, A. J. (1965). Properties and fitting of the von Bertalanffy growth curve. *Growth*, 29:265–289.
Fay, G., Punt, A. E., and Smith, A. D. M. (2011). Impacts of spatial uncertainty on the performance of age structure-based harvest strategies for blue-eye trevalla (*Hyperoglyphe antarctica*). *Fisheries Research*, 110:391–407.

Francis, R. I. C. C. (1992). Use of risk analysis to assess fishery management strategies: A case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise New Zealand. *Canadian Journal of Fisheries and Aquatic Sciences*, 49:922–930.

Fulton, E. A., Link, J. S., Kaplan, I. C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R. J., Smith, A. D. M., and Smith, D. C. (2011). Lessons in modelling and management of marine ecosystems: The Atlantis experience. *Fish and Fisheries*, 12(2):171–188.

Gray, R., Fulton, E. A., Little, L. R., and Scott, R. (2006). Operating model specification within an agent based framework. North west shelf joint environmental management study technical report no. 16., CSIRO., Hobart, Tasmania.

Haddon, M. (2012). *Reducing uncertainty in stock status: Harvest strategy testing, evaluation and development. General discussion and summary*. CSIRO Marine and Atmospheric Research.

Haddon, M., J., W. T., Wear, R. G., and C., A. V. (1996). Biomass and distribution of five species of surf clam off an exposed West Coast North Island beach, New Zealand. *Journal of Shellfish Research*, 15(2):331–339.

Hannan, D. A. (2014). *Population genetics and connectivity in Paphies subtriangulata and Paphies australis (Bivalvia: Mesodesmatidae)*. PhD thesis, School of Biological Sciences, Victoria University of Wellington, New Zealand.

Intel Corporation (2009). *Intel Math Kernel Library. Reference Manual*.

Ives, M. C. and Scandol, J. P. (2013). BIOMAS: A bio-economic modelling and assessment system for fisheries management strategy evaluation. *Ecological Modelling*, 249:42–49.
Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software, 70*(5):1–21.

Kristensen, K., Thygesen, U. H., Andersen, K. H., and Beyer, J. E. (2014). Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences, 71*:326–336.

Lindgren, F., Rue, H., and Lindstrøm, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(4):423–498.

Millar, R. B. and Nottingham, C. D. (2019). Improved approximations for estimation of size-transition probabilities within size-structured models. *Canadian Journal of Fisheries and Aquatic Sciences, 76*(8):1305–1313.

Needle, C. L. (2015). Honeycomb: A spatio-temporal simulation model to evaluate management strategies and assessment methods. *ICES Journal of Marine Science, 72*(1):151–163.

Nottingham, C. D. and White, W. L. (2014). Annual reproductive cycle and condition index of the New Zealand surf clam *Mactra murchisoni* Deshayes, 1854 (Bivalvia: Mactridae). *Journal of Shellfish Research, 34*(3):777–784.

Nuha, A. (2013). The reproductive biology of the surf clams triangle shell (*Spisula aequilatera*), ringed dosinia (*Dosinia anus*) and deep water tuatua (*Paphies donacina*) from the north east of the south island, New Zealand. Master’s thesis, School of Applied Sciences, Auckland University of Technology, New Zealand.

Paola, C. (2011). Simplicity versus complexity. *Nature, 469*:38–39.

Punt, A. E. (2019). Spatial stock assessment methods: A viewpoint on current issues and assumptions. *Fisheries Research, 213*:132–143.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., and Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries, 17*:303–304.
Punt, A. E. and Hobday, D. (2009). Management strategy evaluation for rock lobster, *Jasus edwardsii*, off Victoria, Australia: Accounting for uncertainty in stock structure. *New Zealand Journal of Marine and Freshwater Research*, 43(1):485–509.

Punt, A. E., Huang, T., and Maunder, M. N. (2013). Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. *ICES Journal of Marine Science*, 70(1):16–33.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Teske, P. R., Sandoval-Castillo, J., van Sebille, E., Waters, J., and Beheregaray, L. B. (2016). Oceanography promotes self-recruitment in a planktonic larval disperser. *Scientific Reports*, 6(1).

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210:143–161.

Thorson, J. T., Ianelli, J. N., Munch, S. B., Ono, K., and Spencer, P. D. (2015a). Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 72:1897–1951.

Thorson, J. T., Jannot, J., and Somers, K. (2017). Using spatio-temporal models of population growth and movement to monitor overlap between human impacts and fish populations. *Journal of Applied Ecology*, 54:577–587.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015b). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, 72(5):1297–1310.

Triantafillos, L. (2008). Survey of subtidal surf clams in Pegasus Bay, November–December 2007. Technical Report, NIWA, New Zealand.
von Bertalanffy, L. (1938). A quantitative theory of organic growth (Inquiries on growth laws. II). *Human Biology*, 10:181–213.

Waterhouse, L., Sampson, D. B., Maunder, M., and Semmens, B. X. (2014). Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. *Fisheries Research*, 158:15–25.

White, W., Millar, R., Breen, B., and Farrington, G. (2012). Survey of subtidal surf clams from the Manawatu Coast (FMA 8), October–November. Technical Report, MPI, Wellington, New Zealand.

White, W., Millar, R., Farrington, G., Breen, B., and Selvera, S. (2015). Stock assessment of surf clams from Cloudy Bay, NZ. Institute for Applied Ecology New Zealand Report 15/01, Applied Ecology New Zealand, Auckland University of Technology, Auckland, New Zealand.
Appendices

A Algorithms

The spatialSim harvest routine is presented in [Algorithm A.4]. Additionally, three sub-routines that are used within [Algorithm A.4] are presented. These include [Algorithm A.1] which returns the cumulative sums of a vector, [Algorithm A.2] which assigns each element of a vector as one of two values depending on whether the element of boolean input vector is TRUE or FALSE, and [Algorithm A.3] which takes a sample of size \( n \) without replacement from the elements of an input vector based on specified selection probabilities for each element of the input vector.

**Algorithm A.1:** cumsum(x)

\[
\begin{align*}
    n &= \text{the size of the vector } x \\
    y_i &= \sum_{k=1}^{i} x_k \text{ for all } i \in \{1, \ldots, n\} \\
    \text{return } y
\end{align*}
\]

**Algorithm A.2:** ifelse(x, a, b)

\[
\begin{align*}
    n &= \text{the size of the vector } x \\
    \text{for } i = 1 \text{ to } n \text{ do } \\
        \text{if } x_i = \text{TRUE} \text{ then } \\
        \quad y_i = a_i \\
        \text{else } \\
        \quad y_i = b_i \\
    \text{end} \\
    \text{return } y
\end{align*}
\]

**Algorithm A.3:** sample(x, n, p) [Efraimidis and Spirakis, 2006].

\[
\begin{align*}
    \text{input} & : x: \text{Vector of size } m \text{ from which samples will be drawn} \\
    & n: \text{Number of samples to draw} \\
    & p: \text{Vector containing the probabilities of drawing } x_i \text{ for } i \in \{1, \ldots, m\} \\
    r_i &= \text{simulate exponential}(1)/p_i \text{ for all } i \in \{1, \ldots, m\} \\
    y &= x \text{ sorted by } r \text{ in ascending order} \\
    \text{return} & \text{the the first } n \text{ elements of } y
\end{align*}
\]
Algorithm A.4: harvest\((b, n, F^{\text{lim}}, f^{\text{one}}, f^{\text{targ}}, f^{\text{surv}}, F^{\text{int}}, p^{\text{targ}})\)

**input**: 
- \(b\): The selected biomass of individuals of a given species in each grid cell 
  \(f \in \{1, \ldots, v_F\}\)
- \(n_{f,k}\): The selected number of individuals of a given species in each grid cell 
  \(f \in \{1, \ldots, v_F\}\) with size in each interval \(k \in \{1, \ldots, v_L\}\)
- \(F^{\text{lim}}\): Harvest limit (biomass)
- \(f^{\text{one}}\): Grid cell to begin harvesting
- \(f^{\text{targ}}\): Vector containing indices to fishing grid cells that are in the area targeted by commercial fishers of a given species
- \(f^{\text{surv}}\): Vector containing indices to grid cells to be sampled for scientific surveys
- \(F^{\text{int}}\): Parameter controlling the spatial intensity of commercial harvesting
- \(p^{\text{targ}}\): Probability of fisher harvesting from a site with higher than the median biomass
- \(p^{\text{loc}}\): Vector containing the probabilities of sampling commercial catch from each grid cell given their geographic locations

**output**: 
- \(b^{\text{site}}\): elements of \(b\) with indices \(f^{\text{site}}\)
- \(p^{\text{site}}\): ifelse(\(b^{\text{site}} > \mu, p^{\text{targ}}p^{\text{loc}}, (1 - p^{\text{targ}})p^{\text{loc}}\))
- \(x^{\text{bio}}\): vector of size \(n^{\text{samp}}\) with elements equal to zero

**while** \(\sum(x^{\text{bio}}) < F^{\text{lim}}\) & \(n^{\text{samp}} < n^{\text{site}}\) **do**

- \(f^{\text{com}}\) = sample(\(f^{\text{site}}, n^{\text{samp}}, p^{\text{site}}\))
- \(x^{\text{bio}}\) = elements of \(b\) with indices \(f^{\text{com}}\)
- \(n^{\text{samp}}\) = index of the minimum value of \(|\text{cumsum}(x^{\text{bio}}) - F^{\text{lim}}|\)

**end**

**return** \(x^{\text{num}}, x^{\text{bio}}, f^{\text{com}}, f^{\text{one}}\)
**B Case study parameters and functions**

The parameters and functions that were used to specify the spatialSim case study scenarios are presented in the following pages that make up this appendix.

**Parameters**

| Process          | Parameter | Value (OM\textsubscript{high}) |
|------------------|-----------|-------------------------------|
| Growth           | $\sigma_\delta$ | $(0.15, 0.15)^T$ |
|                  | $\mu_\eta$ | 0 |
|                  | $\psi_\eta$ | 0.75 |
|                  | $\rho_\eta$ | 0.75 |
|                  | $\kappa_\eta$ | $\exp (3)$ |
|                  | $\tau_\eta$ | 0.1 |
|                  | $\mu_i$ | $-2$ |
|                  | $\rho_i$ | 0.75 |
|                  | $\sigma_i$ | 1 |
|                  | $\alpha_M$ | $(0.79, 0.42)^T$ |
| Natural Mortality| $F_{\text{lim}}^\text{int}$ | $(1e6, 1e6)^T$ |
|                  | $F_{\text{int}}^\text{arg}$ | 1.5 |
|                  | $p_{\text{arg}}$ | 0.90 |
|                  | $f_{\text{one}}$ | $(0, 0)^T$ |
| Fishing Mortality| $r_0$ | $(\exp(7.0), \exp(4.5))^T$ |
|                  | $h$ | $(0.9, 0.9)^T$ |
|                  | $r_{\text{range}}$ | $(0.5, 0.5)^T$ |
|                  | $r_{\text{cut}}$ | $(0.05, 0.05)^T$ |
| Recruitment      | $C_\epsilon$ | Identity matrix |
|                  | $\mu_\epsilon$ | $-7$ |
|                  | $\rho_\epsilon$ | 0.75 |
|                  | $\kappa_\epsilon$ | $\exp (5)$ |
|                  | $\tau_\epsilon$ | 0.0005 |
| Structure        | $v_B$ | 600 |
|                  | $v_Y$ | 10 |
|                  | $v_P$ | 12 |
|                  | $v_L$ | 10 |
|                  | $v_C$ | 2 |
|                  | $v_F$ | 557 433 |
|                  | $v_S$ | 500 |

Table B.1: Parameters for OM\textsubscript{high}. Note, these parameter values, with the exception of $F_{\text{int}}$, are the same for the other operating models. In the case of OM\textsubscript{med}, $F_{\text{int}} = 1.0$, and in the case of OM\textsubscript{low}, $F_{\text{int}} = 0.5$. 
Functions

Growth increment

The growth of an individual with size \( l \in Q_{c,k} \) was calculated from Faben’s von Bertalanffy curve (Fabens, 1965),

\[
F^\Delta(l; \beta^\Delta_c, l_{c,\infty}, \Delta t) = (1 - \exp(-\beta^\Delta_c \Delta t))(l_{c,\infty} - \bar{l}_{c,k}),
\]

where \( \beta = (1.09, 0.79)^T \) describes the growth rates of S. aequilatera and M. murchisoni individuals, \( l_{c,\infty} = (57.08, 77.70)^T \) are the asymptotic sizes S. aequilatera and M. murchisoni individuals can reach, and \( \bar{l}_{c,k} \) is the midpoint of size-interval \( Q_{c,k} \).

Size-weight relationship

The weight of an individual with size \( l \in Q_{c,k} \) was calculated from the following exponential growth curve,

\[
F^w(l; \beta^w_{1,c}, \beta^w_{2,c}, \beta^w_{3,c}) = \exp(\beta^w_{1,c} + \beta^w_{2,c} \bar{l}_{c,k} + \beta^w_{3,c} \bar{l}_{c,k}^2),
\]

where \( \beta^w_1 = (-8.586, -7.599)^T \), \( \beta^w_2 = (0.168, 0.118)^T \), and \( \beta^w_3 = (-0.001, 0.001)^T \).

Fishing selectivity

The probability of a species \( c \) individual being retained by fishing gear was 0.95 if they had size \( l \in Q_{c,1} \) and 1 otherwise,

\[
F^\zeta(l) = \begin{cases} 
0.95, & l \in Q_{c,1} \\
1.00, & l \notin Q_{c,1}.
\end{cases}
\]
Sexual maturity

The probability of a species $c$ individual with size $l \in Q_{c,k}$ being sexually mature was calculated from the logistic curve,

\[
\mathcal{F}^m(l; l_{50,c}, l_{95,c}) = \frac{1}{1 + \exp \left( - \log(19) \frac{l - l_{50,c}}{l_{95,c} - l_{50,c}} \right)}.
\]

Here, the size at which 50% of individuals were sexually mature was $l_{50} = 0.25l_e$ and the size at which 95% of individuals were sexually mature was $l_{95} = 0.5l_e$.

Probability of recruitment given environmental conditions

The probability of a species $c$ individual recruiting conditional on the environmental conditions at each time point and location was specified using two independent components that included (1) an annual seasonality term, $\xi_{seas}$, and (2) a time-invariant bathymetrical term, $\xi_{bath}$,

\[
\xi_{c,s,t} = \xi_{seas} \xi_{bath}
\]

The annual seasonality term, $\xi_{seas}^t$, was equal to 0 if $t$ represented one of the first six months of the year and $1/6$ otherwise. The bathymetrical term was calculated from a lognormal distribution,

\[
\xi_{bath}^{c,s} = \int_{e_s-1}^{e_s} \lognorm(x, \nu^e_c, \tau^e_c)dx,
\]

where $e_s$ is the depth at location $s$ to the nearest meter and $\lognorm(\cdot)$ is the lognormal density function. The mean and standard deviation parameters on the log scale for $S. aequilatera$ and $M. murchisoni$ individuals were $\nu^e = (1.32, 1.47)^\top$ and $\tau^e = (0.31, 0.27)^\top$, respectively.