Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography A Journal of Macroecology

METHOD OPEN ACCESS

tinyVAST: R Package With an Expressive Interface to Specify Lagged and Simultaneous Effects in Multivariate Spatio-Temporal Models

James T. Thorson[1] | Sean C. Anderson[2] | Pamela Goddard[3] | Christopher N. Rooper[2]

[1]Resource Ecology and Fisheries Management, Alaska Fisheries Science Center, Seattle, Washington, USA | [2]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada | [3]Lynker Technologies, LLC, Under Contract to Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

Correspondence: James T. Thorson (james.thorson@noaa.gov)

Received: 29 January 2024 | **Revised:** 10 March 2025 | **Accepted:** 22 March 2025

Handling Editor: Joseph Maina

Funding: The authors received no specific funding for this work.

Keywords: generalised additive model | habitat associations | joint species distribution model | spatio-temporal model | structural causal model | structural equation model | vector autoregressive model

ABSTRACT

Aim: Multivariate spatio-temporal models are widely applicable, but specifying their structure is complicated and may inhibit wider use. We introduce the R package *tinyVAST* from two viewpoints: the software user and the statistician.

Innovation: From the user viewpoint, *tinyVAST* adapts a widely used formula interface to specify generalised additive models and combines this with arguments to specify spatial and spatio-temporal interactions among variables. These interactions are specified using arrow notation (from structural equation models) or an extended arrow-and-lag notation that allows simultaneous, lagged and recursive dependencies among variables over time. The user also specifies a spatial domain for areal (gridded), continuous (point-count) or stream-network data. From the statistician viewpoint, *tinyVAST* constructs sparse precision matrices representing multivariate spatio-temporal variation, and parameters are estimated by specifying a generalised linear mixed model (GLMM). This expressive interface encompasses vector autoregressive, empirical orthogonal functions, spatial factor analysis and ARIMA models.

Main Conclusion: To demonstrate, we fit to data from two survey platforms sampling corals, sponges, rockfishes and flatfishes in the Gulf of Alaska and Aleutian Islands. We then compare eight alternative model structures using different assumptions about habitat drivers and survey detectability. Model selection suggests that towed-camera and bottom trawl gears have spatial variation in detectability but sample the same underlying density of flatfishes and rockfishes and that rockfishes are positively associated with sponges while flatfishes are negatively associated with corals. We conclude that *tinyVAST* can be used to test complicated dependencies representing alternative structural hypotheses for research and real-world policy evaluation.

Global Ecology and Biogeography, 2025; 34:e70035
https://doi.org/10.1111/geb.70035

1 | Introduction

Multivariate spatio-temporal models are applicable to many questions throughout ecology, human health, econometrics and geosciences (Thorson and Kristensen 2024). For example, these models are used to study:

1. *Community assembly* when applying joint species distribution models to estimate how species traits and local environmental conditions explain variation in density for multiple species across space and time (Ovaskainen et al. 2017);
2. *Species and/or temporal ordination* when identifying a reduced set of variables that can collectively explain biophysical variation for many processes across space and time. When simplifying dynamics over time, estimated indices (and spatial responses) are often called 'empirical orthogonal functions' (Wikle and Cressie 1999), although ordination can instead (or simultaneously) be applied to identify a small number of 'species archetypes'. where density for each species is estimated as a weighted average of different archetypes (Latimer et al. 2009);
3. *Species interactions*, that is, estimating how species density for one species affects local productivity for other species. Species interactions can be approximated by fitting a vector-autoregressive model to log-densities (Wootton and Emmerson 2005), and the estimated 'community matrix' then represents many important properties about community stability (Ives et al. 2003).

In the following, we will use the term 'multivariate spatio-temporal models' for this large family of analyses.

In the following, we introduce a new package, *tinyVAST*, for the R statistical environment (R Core Team 2023) that provides a simple and user-friendly interface for multivariate spatio-temporal models and which builds upon a wide range of existing models and software (Table 1). We explain the model from two contrasting viewpoints: the software user (to view the user-interface) and the statistician (to view the statistical machinery). We then introduce a case-study example involving data from two sampling programs for seafloor-associated fishes as well as biogenic habitat (sponges and corals). We outline eight alternative hypotheses regarding how these sampling programs might relate to one another, fit all models using *tinyVAST* and use conventional model selection to identify a parsimonious description of habitat associations for these species.

2 | Viewpoint #1: User-Interface and Feature Set

We aim to develop software that involves a minimal set of user-level inputs, but which can still encompass a wide range of spatio-temporal model configurations. We specifically want a user interface that resembles existing regression software in R, e.g., *glmmTMB* for hierarchical models (Brooks et al. 2017) and *mgcv* for generalised additive models (Wood 2017). Similar to these well-known regression packages, *tinyVAST* requires a data frame 'data' that contains any combination of continuous and categorical variables, and then an argument 'formula' that identifies the response variables, some transformation of predictor variables and an offset variable. This regression interface allows new users to quickly fit a model and then interrogate model components using some combination of S3-generic functions, e.g., 'predict', 'summary', 'simulate', 'residuals', etc.

By itself, however, the well-known 'formula' interface for regression packages is not sufficiently expressive, e.g., to represent dependencies among predictor variables that are relevant during mediation, instrumental variable, or attribution analysis (Pearl 2009). We, therefore, predict the mean of a response variable as the sum of four different terms:

1. *Spatial interactions among variables*: The user can specify interactions among variables at a given site (or for spatially correlated latent variables) using 'arrow' notation derived from path analysis (Wright 1934), based on the interface from R package *sem* (Fox et al. 2020);
2. *Temporal interaction among variables*: The user can specify simultaneous and lagged interactions among variables over time using an expanded 'arrow-and-lag' notation that is derived from the R package *dsem* (Thorson et al. 2024), where these interactions construct the annual intercept for a given variable therefore apply uniformly for all locations.
3. *Spatio-temporal interactions among variables*: The user can specify simultaneous and lagged interactions among variables over time, where these interactions occur on a site-by-site basis.
4. *Generalised additive model*: The user specifies a formula for a generalised additive model (GAM) that is interpreted by the R package *mgcv*. If other model components are missing, *tinyVAST* estimates parameters that are similar to *mgcv*, with small differences resulting from different methods for parameter estimation;

Combining these four terms results in a flexible and generic ('expressive') model structure, and each term is useful individually (or in combination) depending upon study goals.

2.1 | Arrow Notation for Interactions Among Variables

Specifying spatial interactions among variables allows users to estimate one or more spatial variables that are constant over time, e.g., representing the long-term utilisation distribution that results from species interactions. To represent interactions among variables over time, we draw upon the existing notation for structural equation models and path analysis (Wright 1934). In its simplest form, this 'arrow notation' specifies a set of linear dependencies among variables and is written using multiple lines of text, where each line specifies a dependency (i.e., coefficient linking two variables or exogenous variance). Each line then includes three arguments separated by commas. The first argument specifies which variables are involved, where a one-headed arrow indicates a slope parameter and a two-headed arrow indicates a variance or covariance parameter. The second argument then defines a name for the parameter, and the (optional) third defines a starting value. This interface allows users to define multiple parameters with the same value (by using the same parameter name in multiple lines) or fix

TABLE 1 | List of model types and software functionality (rows) and R packages (columns), indicating what functionality is supported by each R package.

	R package					
Model type	***sem***	***mgcv***	***dsem***	***sdmTMB***	***VAST***	***tinyVAST***
Available on CRAN	X	X	X	X		X
No spatial or temporal correlations						
Instrumental variables	X					X
Time-series correlations						
Vector autoregressive model			X			X
Dynamic factor analysis			X			X
Spatio-temporal correlations						
Spatial vector autoregressive model					X	X
Empirical orthogonal function analysis					X	X
Spatial factor analysis					X	X
Univariate species distribution model		X		X	X	X
Multivariate species distribution model		X		X	X	X
Smoothing spline response for covariates		X		X		X
Multiple distribution/link functions for different partitions of data					X	X

parameters at a hypothesised value a priori (by naming the parameter NA and fixing it at the starting value).

For example, an analyst might envision a linear model with an independent variable X and dependent variable Y, where variation in X causes variation in Y. This is specified by defining a slope parameter β_1 measuring how a change in X causes a change in Y (using a one-headed arrow in first line of Equation 1), and then two variance parameters σ_X and σ_Y (using two-headed arrows in the second and third lines):

$$\begin{gathered} X \rightarrow Y,\text{beta1},1 \\ X \leftrightarrow X,\text{sigmaX},1 \\ Y \leftrightarrow Y,\text{sigmaY},1 \end{gathered} \tag{1}$$

This arrow notation specifies a linear model:

$$\begin{gathered} X \sim \text{Normal}\left(\mu_X, \sigma_X^2\right) \\ Y \sim \text{Normal}\left(\mu_Y + \beta_1 X, \sigma_Y^2\right). \end{gathered} \tag{2}$$

However, arrow notation can also be used to specify complex and recursive dependencies among variables. For example:

$$\begin{gathered} X \rightarrow Y,\text{beta1} \\ Y \rightarrow Z,\text{beta2} \\ X \rightarrow Z,\text{beta3} \end{gathered} \tag{3}$$

specifies an impact of X on Y with slope β_1 (Equation 3 1st line) and an impact of Y on Z with slope β_2 (Equation 3 2nd line), such that X has an indirect impact on Z with magnitude $\beta_1\beta_2$. However, it also specifies a direct impact of X on Z with slope β_3 (Equation 3 3rd line), such that the total impact of X on Z is $\beta_1\beta_2 + \beta_3$. By default, the software also includes a variance parameter for each variable (in this case, σ_X, σ_Y and σ_Z), such that it results in a set of linear models:

$$\begin{gathered} X \sim \text{Normal}\left(\mu_X, \sigma_X^2\right) \\ Y \sim \text{Normal}\left(\mu_Y + \beta_1 X, \sigma_Y^2\right) \\ Z \sim \text{Normal}\left(\mu_Z + \beta_2 Y + \beta_3 X, \sigma_Z^2\right) \end{gathered} \tag{4}$$

Importantly, this model can give useful insight about an exogenous change in a mediator, where an exogenous change in Y is expected to result in a change in response Z. We use this 'arrow' notation to specify interactions among variables that are constant over time.

This structural model could be viewed using two interpretations:

1. *Weak form*: The analyst may use arrow notation as an expressive and general interface to specify the covariance among variables;
2. *Strong form*: The analyst may view the resulting model as a 'structural causal model' (Pearl 2009) and use it to make predictions about system responses to counterfactual scenarios.

We believe that the 'weak form' interpretation remains useful even in cases when the analyst has not validated (or otherwise believes) the strong-form interpretation.

2.2 | Arrow-and-Lag Notation for Interactions Among Variables Over Time

Specifying temporal or spatio-temporal interactions among variables over time allows *tinyVAST* to estimate short-term deviations from long-term average conditions. However, arrow notation does not include any concept of time and cannot distinguish between simultaneous versus lagged effects. We therefore follow Thorson et al. (2024) in defining an expanded 'arrow-and-lag' notation to allow users to specify interactions among variables over time using multiple lines of text. This text follows the same format as the 'arrow notation', except it adds an additional (second) argument representing the lag involved and now involves three or four arguments per line.

For example, the analyst might approximate species interactions by estimating a lagged impact of each species on per-capita productivity for every other species (Ives et al. 2003). This requires specifying a bivariate vector autoregressive (a.k.a., cross-lagged) model:

$$\begin{aligned} X &\rightarrow X, 1, \ \text{betaxx}, 0.1 \\ X &\rightarrow Y, 1, \ \text{betaxy}, 0.1 \\ Y &\rightarrow X, 1, \ \text{betayx}, 0.1 \\ Y &\rightarrow Y, 1, \ \text{betayy}, 0.1 \end{aligned} \tag{5}$$

where this then specifies:

$$\begin{aligned} X_t &\sim \text{Normal}\left(\mu_X + \beta_{xx} X_{t-1} + \beta_{yx} Y_{t-1}, \sigma_X^2\right) \\ Y_t &\sim \text{Normal}\left(\mu_Y + \beta_{xy} X_{t-1} + \beta_{yy} Y_{t-1}, \sigma_Y^2\right), \end{aligned} \tag{6}$$

where μ_X and μ_Y are intercepts for each species. This involves estimating four cross-lagged interactions and can instead be written as $\left(X_t, Y_t\right) = \mathbf{B}\left(X_{t-1}, Y_{t-1}\right) + \boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_t \sim \text{MVN}\left(\mathbf{0}, \text{diag}\left(\sigma_X^2, \sigma_Y^2\right)\right)$ and $\mathbf{B} = \begin{bmatrix} \beta_{xx} & \beta_{yx} \\ \beta_{xy} & \beta_{yy} \end{bmatrix}$. This arrow-and-lag notation then allows the user to specify (among other models) a vector-autoregressive (VAR), autoregressive integrated moving average (ARIMA), dynamic factor analysis (DFA) and difference-in-differences model (Table 2).

2.3 | Spatial objects

The user can specify an object supplied to the argument 'spatial_domain' that is used to define a spatial precision matrix, $\mathbf{Q}_{\text{domain}}$, where the matrix inverse $\mathbf{Q}_{\text{domain}}^{-1}$ represents the covariance among different sites within the spatial domain. Options include the following:

1. *Continuous spatial domain*: Using the R package *fmesher* (Lindgren 2023), the user can apply the stochastic partial differential equation (SPDE) approximation to a Matérn correlation function over a two-dimensional spatial domain (Lindgren et al. 2011). Package *tinyVAST* then uses *fmesher* to construct the projection matrices that apply bilinear interpolation to calculate random variables at any sampled or predicted location within the defined spatial domain.

TABLE 2 | Examples of common time-series models and their specification using 'arrow-and-lag' notation. We envision data are available for two variables X and Y for vector autoregressive models (VAR) and dynamic factor analysis (DFA), or one variable X for other models, and DFA and autoregressive integrated moving average (ARIMA) models involve estimating a latent variable F. All models are assumed to estimate an exogenous and independent variance for each variable by default. However, this default can be overridden by explicitly specifying a two-headed arrow where, for example, '$X \leftrightarrow X$, 0, NA, 0' specifies that exogenous variance for variable X is fixed at 0 a priori. Note that the DFA, ARIMA(1,1,0) and ARIMA(0,0,1) specifications are all reduced rank (arising from variable X specified to have zero variance).

Model and specification using 'arrow-and-lag' notation	Explanation for each line
Vector autoregressive model (VAR)	
$X \rightarrow X$, 1, b_xx	First-order autoregression for X (e.g., density dependence)
$X \rightarrow Y$, 1, b_xy	Lagged impact of X_{t-1} on Y_t
$Y \rightarrow X$, 1, b_yx	Lagged impact of Y_{t-1} on X_t
$Y \rightarrow Y$, 1, b_yy	First-order autoregression for Y (e.g., density dependence)
Dynamic factor analysis (DFA)	
$F \rightarrow F$, 1, NA, 1	Random walk for factor F
$F \rightarrow X$, 0, b_fx	Loading of X on factor F
$F \rightarrow Y$, 0, b_fy	Loading of Y on factor F
$F \leftrightarrow F$, 0, NA, 1	Unit variance for factor F
$X \leftrightarrow X$, 0, NA, 0	Turn off process error for X
$Y \leftrightarrow Y$, 0, NA, 0	Turn off process error for Y
ARIMA(1,1,0)	
$F \rightarrow F$, 1, rho	AR1 for factor F
$X \rightarrow X$, 1, NA, 1	Random walk for time-series X (integrated component)
$F \rightarrow X$, 0, NA, 1	Fixed unit loadings for factor F on time-series X
$X \leftrightarrow X$, 0, NA, 0	Turn off additional process error for time-series X
ARIMA(0,0,1)	
$F \rightarrow X$, 0, NA, 1	Unit loadings for factor F_t on X_t in time t
$F \rightarrow X$, 1, rho	Estimated loadings for factor F_{t-1} on X_t in time t
$X \leftrightarrow X$, 0, NA, 0	Turn off additional process error for time-series X

2. *Areal spatial domain*: Using the R package *igraph* (Csardi and Nepusz 2006), the user can instead construct a graph representing polygons or other areal data. Package *tinyVAST* constructs the adjacency matrix and applies a

simultaneous autoregressive (SAR) model to construct the spatial precision matrix (Ver Hoef et al. 2018). The user can specify an unconnected graph, which then eliminates any spatial correlation among modelled locations, resulting in a multivariate time-series model.

3. *Stream network domain*: Using the R package *sfnetwork* (van der Meer et al. 2023), the user can construct a graph representing locations along a stream network (Hoef et al. 2006). Package *tinyVAST* then constructs the precision matrix representing an Ornstein-Uhlenbeck process (Charsley et al. 2023).

If the user does not specify the argument 'spatial_domain', then *tinyVAST* assumes that all observations are from a single site and, therefore, reverts to a univariate time-series model (see summary of arguments in Table 3). Alternatively, the user can specify that all sites are independent using the package *igraph*, and this then reverts to a multivariate time-series model. A fitted model can then be explored using standard S3-generic functions (Table 4).

2.4 | Formula Interface for Exogenous Covariates

To specify the GAM component of *tinyVAST*, we use the formula interface from package *mgcv*. This interface is appropriate for estimating how exogenous variables affect the specified response and allows the user to select from a wide range of nonlinear response functions. The *mgcv* formula includes univariate splines, multivariate splines, random intercept and random-slope models, as well as standard functionality for polynomial basis expansion of fixed effects. The most common options are supported within *tinyVAST*, and these options are well described elsewhere (Zuur et al. 2009; Wood 2017).

3 | Viewpoint #2: Statistical Model Structure

Package *tinyVAST* fits a generalised linear mixed model (GLMM) where random effects are specified using Gaussian Markov random fields (GMRFs). This GLMM involves combining four model components (numbered at the beginning of section Viewpoint #1: User-interface and feature set).

The first component is the 'spatial interaction among variables' represented using a spatial SEM. Given the arrow notation representing interactions among C variables, *tinyVAST* constructs a sparse precision matrix $\mathbf{Q}_{\text{space\_term}}$ with dimension $C \times C$:

$$\mathbf{Q}_{\text{space\_term}} = \left(\mathbf{I} - P^t\right)\mathbf{\Gamma}^{-t}\mathbf{\Gamma}^{-1}(\mathbf{I} - P), \tag{7}$$

where P is the sparse $C \times C$ matrix of path coefficients (specified using one-headed arrows in the arrow notation) and $\mathbf{\Gamma}$ is the sparse $C \times C$ matrix of exogenous covariance parameters (two-headed arrows in arrow notation). Package *tinyVAST* then constructs a separable precision for a $S \times C$ matrix $\mathbf{\Omega}$ of S sites and C variables:

$$\text{vec}(\mathbf{\Omega}) = \text{MVN}\left(\mathbf{0}, \mathbf{Q}^{-1}_{\text{space\_term}} \otimes \mathbf{Q}^{-1}_{\text{domain}}\right). \tag{8}$$

This matrix $\mathbf{\Omega}$ is then projected to the set of samples $i \in I$ by multiplication with a $I \times S$ projection matrix.

The second component is the 'temporal interaction among variables', represented using a dynamic SEM. The arrow-and-lag notation representing interactions among variables over time is used to construct a sparse $CT \times CT$ precision matrix $\mathbf{Q}_{\text{time\_term}}$. This is used to specify a nonseparable precision for a $C \times T$ two-dimensional matrix $\mathbf{D}$ across C variables and T times:

$$\text{vec}(\mathbf{D}) = \text{MVN}\left(\mathbf{0}, \mathbf{Q}^{-1}_{\text{time\_term}}\right). \tag{9}$$

This matrix $\mathbf{D}$ contains the annual intercept $d_{c,t}$ for each variable and year.

The third component is the 'spatio-temporal interaction among variables', represented using a separable process that includes a DSEM (representing site-by-site interactions among variables over time) and the spatial process. Arrow-and-lag notation again specifies interactions among variables over time to construct a sparse $CT \times CT$ precision matrix $\mathbf{Q}_{\text{spacetime\_term}}$. This is then used to specify a separable precision for a $S \times C \times T$ three-dimensional array $\mathbf{E}$ across S sites, C variables and T times:

$$\text{vec}(\mathbf{E}) = \text{MVN}\left(\mathbf{0}, \mathbf{Q}^{-1}_{\text{spacetime\_term}} \otimes \mathbf{Q}^{-1}_{\text{domain}}\right). \tag{10}$$

This array E is again projected to the set of samples $i \in I$ by multiplication with a $I \times S$ projection matrix.

However, some specifications using 'arrow notation' for the space-variable interaction $\mathbf{\Omega}$ or temporal interaction $\mathbf{D}$, or using 'arrow-and-lag notation' the space-variable-time interaction E will result in $\mathbf{Q}_{\text{space\_term}}$, $\mathbf{Q}_{\text{time\_term}}$ and/or $\mathbf{Q}_{\text{spacetime\_term}}$ not being full rank (see Table 2 for example). In these cases, the covariance $\mathbf{Q}^{-1}_{\text{space\_term}}$, $\mathbf{Q}^{-1}_{\text{time\_term}}$ and/or $\mathbf{Q}^{-1}_{dsem}$ does not exist. The user can instead specify an alternative parametrization:

$$\text{vec}\left(\widetilde{\mathbf{\Omega}}\right) = \text{MVN}\left(\mathbf{0}, \mathbf{I}_C \otimes \mathbf{Q}^{-1}_{spatial}\right)$$

$$\mathbf{\Omega} = \left(\mathbf{I} - P^t\right)^{-1}\mathbf{\Gamma}^t\widetilde{\mathbf{\Omega}} \tag{11}$$

where $\mathbf{I}_C$ is a $C \times C$ identity matrix such that $\mathbf{I}_C \otimes \mathbf{Q}^{-1}_{spatial}$ has full rank (and a similar reparametrization is done for $\mathbf{D}$ or E). This 'projection' parametrization ensures that, efficents $\widetilde{\mathbf{\Omega}}$, $\widetilde{\mathbf{D}}$ and $\widetilde{E}$ are full rank, while their projected values $\mathbf{\Omega}$, $\mathbf{D}$ and E have the reduced-rank structure specified by the user.

The fourth component is the 'generalised additive model', representing additional nonlinear effects and interactions. The GAM term includes fixed effects $\boldsymbol{\alpha}$ and random effects $\boldsymbol{\gamma}$, with design matrix $\mathbf{X}$ and $\mathbf{Z}$, respectively. The formula interface is used to construct the $I \times J$ design matrix $\mathbf{X}$ for fixed effects, and the $I \times K$ design matrix $\mathbf{Z}$ for random effects, where random effects represent spline coefficients or other shrinkage priors. The design matrix for fixed effects can include an offset term (i.e., a column of $\mathbf{X}$ for which the slope α is fixed at 1 a priori), and can

TABLE 3 | List of all user-defined arguments passed to function 'tinyVAST' (the core function in package '*tinyVAST*'), including different user options when specifying these.

Argument name	Object passed	What it does
formula	Formula defining response, exogenous predictors and offset variables	Defines regression-style model, including nonlinear transformations
data	Data frame containing response, predictor and offset variables	Contains all sampling data
time_term	Options: 1. Character string parsed by 'make_dsem_ram(.)' internally Output from 'make_sem_ram(.)'	Defines interactions among variables over time (including lagged and simultaneous effects) that are constant across space
space_term	Options: 2. Character string parsed by 'make_sem_ram(.)' internally 3. Output from 'make_sem_ram(.)'	Defines interactions among variables that are constant over time. By default, *tinyVAST* uses 'arrow notation' parsed by 'make_sem_ram(.)'
spacetime_term	Options: 1. Character string parsed by 'make_dsem_ram(.)' 2. Output from 'make_dsem_ram(.)' 3. Output from 'make_eof_ram(.)' 4. Manually constructed 'reticular action model' representing associations among variables over time	Defines interactions among variables over time (including lagged and simultaneous effects). By default, *tinyVAST* uses 'arrow-and-lag notation' parsed by 'make_dsem_ram(.)'. However, other models can instead be specified using alternative constructors, for example, empirical orthogonal functions (EOF)
family	A standard R family object or a list with names matching elements of 'data' corresponding to user-specified error distributions, where each element specifies a family and link function for those distributions	Allows models to be specified that fit multiple data types or a single data type but with different dispersion parameters associated with different samples (e.g., different variance for each modelled variable)
space_columns	A character vector where 'data[,space_columns]' indicates the spatial coordinates for each sample	Matched against spatial coordinates or domain listed in 'spatial_domain'
spatial_domain	S3-object, where options include: 1. Topology used by SPDE method to define spatial correlations in 2D spatial domain, constructed using *fmesher* package (Lindgren 2023) 2. Topology used by Simultaneous Autoregressive (SAR) method to define spatial correlations for areal data (or independence for a multivariate time-series analysis), constructed using *igraph* package (Csardi and Nepusz 2006) 3. Topology used by autoregressive stream-network model for network data, constructed using *sfnetworks* package (van der Meer et al. 2023) 4. Missing, defining a time-series model (i.e., all samples are associated with a single site)	Defines spatial domain (extent and resolution) and resulting basis functions describing spatial correlations, where space-variable interactions arise as separable processes from 'spatial_domain' and 'space_term' and space–time-variable interactions arise as separable processes from 'spatial_domain' and 'spacetime_term'
variable_column	A string where 'data[,variable_column]' indicates the variable for each sample	Matched against variables listed in 'dsem' and 'sem'
variables	Character vector that defines the set of model variables, included as a column of 'data' and indicated by 'variable_column'	Defines the set of modelled variables (e.g., indicating species in a multispecies model).
time_column	A string where 'data[,time_column]' indicates the time for each sample	Matched against lags specified in 'dsem'
times	Integer vector that defines the set of model times, included as a column of 'data' and indicated by 'time_column'	Defines the temporal domain
control	Tagged list defining advanced features, constructed by 'tinyVASTcontrol(.)'	Allows detailed control over model specification and optimisation while avoiding argument clutter for casual users
delta_options	List with elements formula, space_term, time_term, spacetime_term, and spatial_varying	Allowing users to specify options when specifying a delta (hurdle) model via argument 'family'
weights	Vector of likelihood weights	Allows some data to be passed but not fitted
spatial_varying	Right-handed formula, listing variables for which a spatially varying coefficient (SVC) is estimated	Allows estimates of spatially nonstationary covariates, representing spatially varying detectability ratios or environmental responses

TABLE 4 | List of functions exported by *tinyVAST*, including their name and purpose.

Function name	Purpose
tinyVAST	Fit a *tinyVAST* model
make_sem_ram	Construct a reticular action model (RAM) using 'arrow notation' to construct the precision matrix for a structural equation model
make_dsem_ram	Construct a RAM using 'arrow-and-lag notation' to construct the precision matrix for a dynamic structural equation model
make_eof_ram	Construct a RAM for an empirical orthogonal function (EOF) analysis
summary	Summarise output for different model components
print	User-friendly output from a fitted model
logLik	The marginal log-likelihood, for example, to enable function AIC(.)
cAIC	Calculate conditional AIC (Zheng et al. 2024)
predict	Predictions in response or link-scale, or for link-scale for individual components
residuals	Compute deviance or response residuals from a fitted model
simulate	Simulate new data conditional upon fitted model, either with or without sampling from the predictive distribution of random effects (the latter being useful when assessing model residuals)
integrate_output	Compute plug-in or epsilon bias-corrected estimator (Thorson and Kristensen 2016) for a quantity calculated via Monte Carlo integration across a user-specified set of sampling locations (and associated covariates) for a single time. This can then compute an area-weighted abundance index and measure distribution shifts, range expansion and spatial overlap among variables

include any set of intercepts, slope and interactions. Each block of random effects is specified to follow a multivariate normal distribution, which is constructed by package *mgcv*:

$$\boldsymbol{\gamma}_z \sim \mathbf{MVN}\left(\mathbf{0}, \lambda_z^{-1}\mathbf{Q}_{\text{gam}}^{-1}\right), \tag{12}$$

where $\mathbf{Q}_{\text{gam}}$ is the precision matrix for the zth block of random effects $\boldsymbol{\gamma}_z$ and λ_k is the estimated wiggliness (inverse-variance) parameter for that block.

The expected value for sample i is then calculated by applying an inverse-link function to a linear predictor, which is calculated in turn by assembling all four components. Some distributions (e.g., delta models) require two linear predictors, so the user can provide a separate specification for linear predictor $p_{1,i}$ and $p_{2,i}$:

$$\begin{aligned} p_{1,i} &= \underbrace{\mathbf{X}_{1,i}\boldsymbol{\alpha}_1 + \mathbf{Z}_{1,i}\boldsymbol{\gamma}_1}_{GAM\#1} + \underbrace{\mathbf{A}_i\boldsymbol{\Omega}_{1,c[i]}}_{space\_term\#1} + \underbrace{\mathbf{D}_{1,c[i],t[i]}}_{time\_term\#1} + \underbrace{\mathbf{A_i}E_{1,c[i],t[i]}}_{spacetime\_term\#1} \\ p_{2,i} &= \underbrace{\mathbf{X}_{2,i}\boldsymbol{\alpha}_2 + \mathbf{Z}_{2,i}\boldsymbol{\gamma}_2}_{GAM\#2} + \underbrace{\mathbf{A}_i\boldsymbol{\Omega}_{2,c[i]}}_{space\_term\#2} + \underbrace{\mathbf{D}_{2,c[i],t[i]}}_{time\_term\#2} + \underbrace{\mathbf{A}_iE_{2,c[i],t[i]}}_{spacetime\_term\#2}, \end{aligned} \tag{13}$$

where $\mathbf{A}_i$ is a vector of length S that projects a spatial variable (e.g., the spatial term $\boldsymbol{\Omega}_{1,c[i]}$ for the 1st linear predictor) to its value at the location of sample i.

tinyVAST allows the user to specify a different link function and distribution for different partitions of data, where $e[i]$ indicates the (error) distribution for sample i, g_e is the bivariate link function for distribution e and f_e is the probability distribution function for that distribution:

$$Y_i \sim f_{e[i]}\left(g_{e[i]}^{-1}\left(p_{1,i}, p_{2,i}\right), \theta_{e[i]}\right), \tag{14A}$$

There are two use cases worth noting. For clarity, we introduce these cases assuming that there's a single error distribution for all samples, such that we can suppress notation $e[i]$ from the distribution f and link function g.

1. *Univariate distribution*: The simplest and most widely used case involves a probability distribution function that uses a single linear predictor (e.g., a log-linked Poisson distribution). In this case, the user only specifies the first linear predictor $p_{1,i} = p_i$ and ignores the second linear predictor $p_{2,i} = 0$, such that mean $\mathbb{E}\left(Y_i\right) = \mu_i$ where $\mu_i = g^{-1}\left(p_i\right)$;
2. *Conventional delta models*: Alternatively, the GLMM might be applied to biomass samples that are continuous and positive but which also include a set of 0s (when a species was not encountered at all). In this case, the user specifies both linear predictors. The first linear predictor is typically transformed using a logistic function (i.e., logit-link), and the second linear predictor is typically transformed using an exponential function (i.e., log-link):

$$\mu_{1,i} = \frac{e^{p_{1,i}}}{1 + e^{p_{1,i}}}$$

$$\mu_{2,i} = e^{p_{2,i}} \tag{14B}$$

These are then used to compute the delta distribution:

$$\pi\left(Y_i \mid \mu_{1,i}, \mu_{2,i}, \theta\right) = \begin{cases} 1-\mu_{1,i} & \text{if } Y_i = 0 \\ \mu_{1,i} \times f\left(Y_i \mid \mu_{2,i}, \theta\right) & \text{if } Y_i > 0 \end{cases} \tag{14C}$$

where $f\left(Y_i \mid \mu_{2,i}, \theta\right)$ is the user-specified distribution for positive samples, and $\pi\left(Y_i\right)$ is the data likelihood including zeros and positive values for sample i.

3. *Poisson-linked delta model*: Finally, the user might specify a bivariate link function with two linear predictors representing areal samples of a marked point process (Thorson 2018). The first linear predictor is calculated using a complementary log–log link, while the second is defined such that $\mu_{1,i}\mu_{2,i} = e^{p_{1,i}} e^{p_{2,i}}$, where $e^{p_{1,i}}$ is represents group-densities for the point process and $e^{p_{2,i}}$ represents biomass per group:

$$\mu_{1,i} = 1 - e^{-e^{p_{1,i}}}$$

$$\mu_{2,i} = \frac{e^{p_{1,i}}}{\mu_{1,i}} e^{p_{2,i}} \tag{14D}$$

The data likelihood can then be computed similarly to the conventional delta model (Equation 14C).

These use-cases can be combined in various ways, e.g., to specify a different link function and distribution for different partitions of the data, where each sample i is associated with partition $e[i]$. For example, Grüss and Thorson (2019) combined presence-absence data using a cloglog link and a Bernoulli distribution, count data using a log link and Poisson distribution, and biomass data using a Poisson-linked delta model and a gamma distribution. In these data-integrated models, it is then often necessary to also include a factor representing survey in the GAM formula, to account for differences in detection probability relative to a reference survey.

Computing the joint log-likelihood for this GLMM is efficient given that $\mathbf{Q}_{\text{space\_term}}$, $\mathbf{Q}_{\text{time\_term}}$, $\mathbf{Q}_{\text{spacetime\_term}}$, $\mathbf{Q}_{\text{domain}}$ and $\mathbf{Q}_{\text{gam}}$ are all sparse. We use the R package *TMB* to apply the Laplace method to approximate the marginal likelihood and optimise the log-marginal likelihood using gradients computed using automatic differentiation (Kristensen et al. 2016). We then use a generalisation of the delta method to calculate standard errors (Kass and Steffey 1989), and the epsilon method to correct for skewness and retransformation bias of random effects (Thorson and Kristensen 2016).

4 | Case Study: Habitat Associations in Data-Integrated Species Distribution Model

To illustrate the wide range of models that can be expressed with *tinyVAST*, we introduce an integrated species distribution model (SDM) that combines samples from two surveys: a towed camera that measures densities of flatfishes (*Pleuronectiformes* spp.), rockfishes (*Sebastes* spp.), corals and sponges, and a separate bottom trawl that measures densities of flatfishes and rockfishes. Individually, historical studies looking at species associations have often found positive correlations between rockfishes and corals and sponges in both trawl and camera data (Yoklavich et al. 2000; Laman et al. 2015). However, these two gears typically sample different types of substrates since trawling is limited to soft-bottom substrates and at different spatial scales, since the area swept by an individual trawl haul is typically at least two orders of magnitude larger than a camera transect. So, it is unclear whether the strength of the positive impact on rockfish density is similar between the two gears. The opposite question is appropriate for flatfishes, which prefer soft-bottom sediments, are more likely to be effectively trawled, but are less likely to be observed with sponges and corals present. Previous research has typically fitted an SDM to coral and sponge densities separately, predicted densities at the location of bottom trawls, and used those predictions as fixed covariates to explain flatfish and rockfish associations with biogenic habitats (Laman et al. 2018). By contrast, we seek to answer the following questions:

1. Are the associations between fishes and biogenic habitats estimated to be the same in towed camera and bottom trawl sampling programmes?
2. Do the towed camera and bottom trawl measure similar variations in density for fishes, or does one or the other tend to measure more or less fluctuation in density between high and low-density areas due to density-dependent catchability for the bottom trawl (Kotwicki et al. 2014)?

The first set of data used in these analyses were from biennial (or triennial) bottom trawl surveys conducted from 1990 to 2019 in the Gulf of Alaska and the Aleutian Islands (Siple et al. 2024). The surveys are designed to assess the abundance of commercially important groundfishes and invertebrates in each ecosystem using a stratified random sampling design covering an area from Dixon Entrance in Southeast Alaska to Stalemate Bank in the Aleutian Islands at depths from 35 to 1000 m in the GOA and 35 to 500 m in the Aleutian Islands. Detailed sampling designs and protocols can be found in Stauffer (2004). In this analysis, catch and effort data from 14,877 research survey bottom trawl tows were used. The second set of data came from towed stereo-camera surveys conducted from 2012 to 2017 in the central and western Gulf of Alaska and Aleutian Islands from Yakutat, Alaska to Stalemate Bank (including Bowers Ridge and Bank). These data were collected using standardised sampling protocols of 15 min of on-bottom time, where all fishes and benthic invertebrates were identified and counted to the lowest possible taxonomic level. The Aleutian Islands data were from 216 transects chosen using a stratified random survey design covering depths of ~15 to 850 m (Rooper et al. 2016). The Gulf of Alaska data were from 227 transects chosen using a random or haphazard design (Sigler et al. 2023). The total number of each species of fish and invertebrate observed, as well as the area of the seafloor viewed, were calculated for each transect using the methods outlined in Rooper et al. (2016). However, in the analysis, both trawl survey and camera data used high-level classification of fish to avoid potentially erroneous species identifications in the camera data.

To answer these questions, we specify a log-linked Tweedie GLMM, which involves a single linear predictor p_i. We define a factor 'Group' with six levels: *S, C, F, R* (for sponges, corals, flatfishes and rockfishes in the drop camera), and *S2* and *R2* (for sponges and corals in the bottom trawl). We also include a factor 'Year' (indicating the year of sampling), numeric 'Area' (for the area swept by a given bottom trawl or drop camera sample) and 'Count' (for the count of a given sample). We then specify a GAM formula, 'Count ~0 + interaction(Group, Year) + offset(log(Area))'. This specifies a separate intercept for every combination of Group and Year that is present in the data set. We also specify a spatial SEM using arrow notation:

$$\begin{gathered} \mathrm{C} \rightarrow \mathrm{F,b1} \\ \mathrm{S} \rightarrow \mathrm{F,b2} \\ \mathrm{C} \rightarrow \mathrm{R,b3} \\ \mathrm{S} \rightarrow \mathrm{R,b4} \\ \mathrm{F} \rightarrow \mathrm{F2,a1} \\ \mathrm{R} \rightarrow \mathrm{R2,a2} \\ \mathrm{C} \rightarrow \mathrm{F2,d1} \\ \mathrm{S} \rightarrow \mathrm{F2,d2} \\ \mathrm{C} \rightarrow \mathrm{R2,d3} \\ \mathrm{S} \rightarrow \mathrm{R2,d4} \end{gathered} \tag{15}$$

where (b_1, b_2, b_3, b_4) is the effect of biogenic habitat (sponges and coral) on the portion of fish densities available to the drop camera, (a_1, a_2) is the relationship between fish densities available to the drop camera relative to the bottom trawl and (d_1, d_2, d_3, d_4) is the effect of biogenic habitat on the portion of fish densities in the bottom trawl after accounting for direct effects (a_1, a_2). This specification then results in the following structural equations:

$$\begin{aligned} \log(\mu_S) &= \alpha_S + \omega_S \\ \log(\mu_C) &= \alpha_C + \omega_C \\ \log(\mu_F) &= \alpha_F + b_1\omega_C + b_2\omega_S + \omega_F \\ \log(\mu_R) &= \alpha_R + b_3\omega_C + b_4\omega_S + \omega_R \\ \log(\mu_{F2}) &= \alpha_{F,t} + d_1\omega_C + d_2\omega_S + a_1\omega_F + \omega_{F2} \\ \log(\mu_{R2}) &= \alpha_{R,t} + d_3\omega_C + d_4\omega_S + a_2\omega_R + \omega_{R2}, \end{aligned} \tag{16}$$

where μ_S, μ_C, μ_F and μ_R are the modelled numerical density of sponges, corals, flatfishes and rockfishes in the towed camera, μ_{F2} and μ_{R2} are the modelled numerical densities of flatfishes

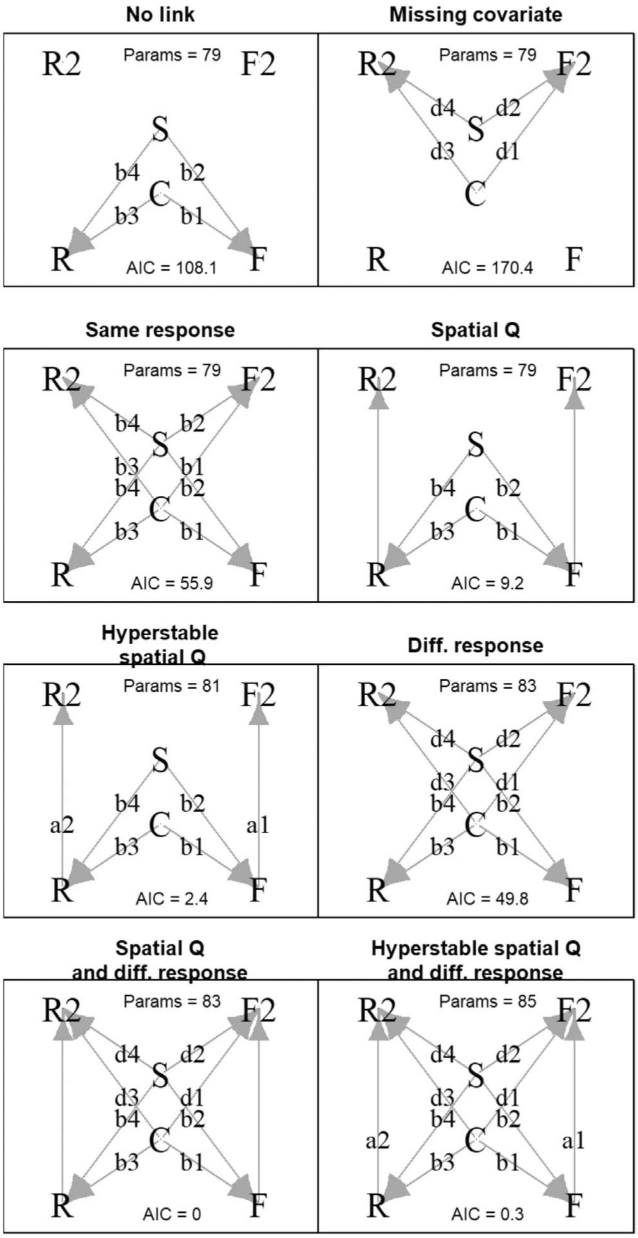

FIGURE 1 | Visualising the hypothesised relationship among six variables (Towed camera: *R* = rockfish; F = flatfish; C = corals; S = sponges; Bottom trawl: R2 = rockfish; F2 = flatfish) using 'tinyVAST::summary(.)' to summarise path coefficients, and functions 'graph_from_data_frame(.)' and 'plot.igraph(.)' in package *igraph* for plotting, where each arrows correspond to a specified path coefficient (*P*, see Equation 7). Estimated path coefficients list the parameter name, while path coefficients that are fixed a priori are shown without a name. We also show the marginal AIC relative to the most parsimonious model (bottom of each panel) and number of fixed effects (top of each panel), and all models include intercepts for each sampled combination of variable and year (56 parameters), two dispersion parameters per variable (2 × 6 = 12 parameters), one variance per species (6 parameters) and an estimated decorrelation rate (1 parameter) that are not shown here.

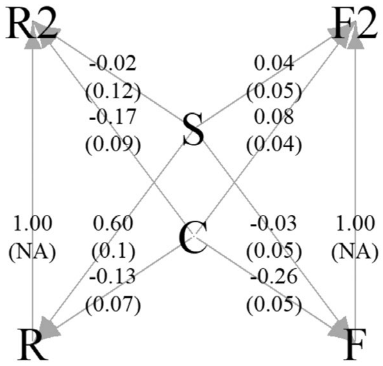

FIGURE 2 | Estimated path coefficients for the most parsimonious model (bottom-left of Figure 1), showing both maximum likelihood estimate and standard error (in parentheses), where standard error NA corresponds to parameters that are fixed a priori.

and rockfishes in the bottom trawl, $\alpha_{\blacksquare}$ represents the intercept for each variable (which varies among years t for the bottom trawl surveys but is constant for the drop-camera surveys) and $\omega_{\blacksquare}$ represents spatial variation in log-density for a given variable.

We then define a set of eight hypothesised relationships that arise as nested submodels. For example:

1. *Biogenic habitat as covariates*: We explore a model where sponges and corals are not linked to towed camera densities for fishes ($b_1 = b_2 = b_3 = b_4 = 0$). Instead, the sponges and corals are estimated as having a linear impact on bottom trawl estimates of fish densities (estimating d_1 through d_4), and other parameters are eliminated ($a_1 = a_2 = 0$). This then corresponds to a joint SDM that imputes the biogenic habitat densities and simultaneously estimates their use as covariates for predicting bottom trawl data.
2. *No link from drop camera and bottom trawl*: Alternatively, we explore a model where there is no link connecting drop camera and bottom trawl data ($d_1 = d_2 = d_3 = d_4 = a_1 = a_2 = 0$), while estimating habitat associations within the drop-camera data set. This measures the fine-scale association between habitat and fishes in the towed camera survey.

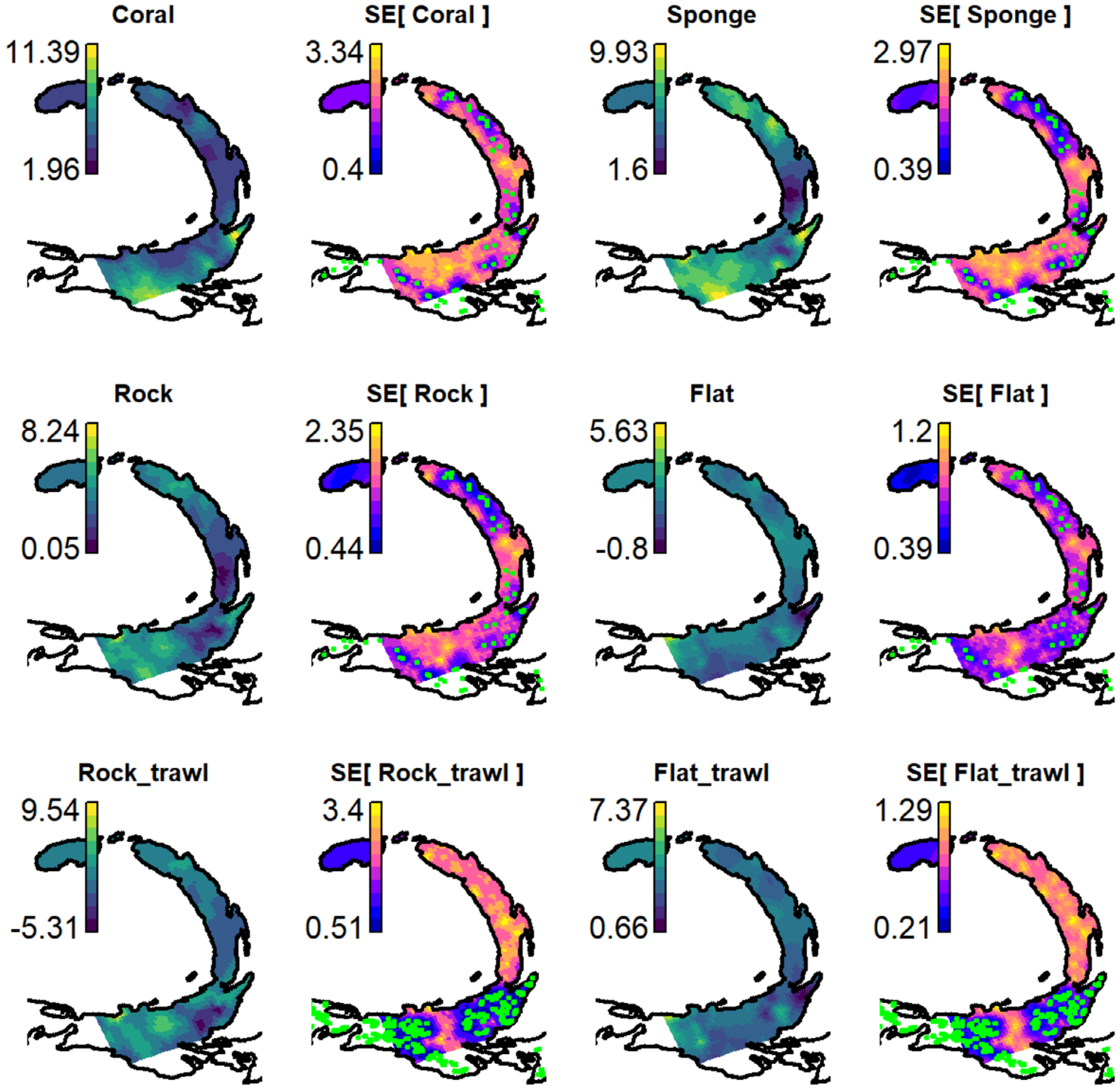

FIGURE 3 | Estimated log-densities (e.g., 'Coral' in top-left panel) and the standard error in log-density (e.g., 'SE[Coral]') for each of six modelled variables (Coral, S, Rock = rockfishes, and Flat = flatfishes in the drop camera, and Flat_trawl and Rock_trawl in the bottom trawl) showing the location of sampling data for that variable (green dots) on top of estimated standard errors. Here, we specifically show densities in Bower's Bank (a relatively shallow area north of the central Aleutian Islands that is not sampled by the bottom trawl survey) to show that drop camera samples can be predict densities from the bottom trawl.

3. *Linked and spatially varying catchability*: Finally, we explore a model that assumes a priori that bottom trawl and drop camera samples are both measuring the same underlying variation in densities (i.e., $a_1 = a_2 = 1$), while also estimating residual variation in the detectability ratio (i.e., spatial variation in ω_{F2} and ω_{R2}).

Other models are similarly formed via restrictions among parameters, which can be expressed using the 'arrow notation' for interactions among variables.

Results suggest that the 'Linked and spatially varying catchability' model is the most parsimonious of our hypothesised models (Figure 1), although the saturated model and another candidate also receive some support ($\Delta\text{AIC} < 3$). This parsimonious model estimates a significant impact of sponges on rockfishes ($b_4 = 0.59$, $SE = 0.10$; Figure 2), i.e., where a 10% increase in sponge densities is associated with a nearly 6% increase in rockfish density, and also estimates a significant and negative impact of corals on flatfishes ($b_1 = -0.26$, $SE = 0.05$). Model selection also indicates that it is parsimonious to specify that spatial variation in fish density in the towed camera is proportional to spatial variation observed using the bottom trawl ($a_1 = a_2 = 1$). Finally, the model estimates an 'indirect' impact of corals and sponges on bottom-trawl fish densities via their estimated impact on towed-camera densities (b_1 through b_4) and the proportional relationship between the towed camera and bottom trawl fishes ($a_1 = a_2 = 1$) but also a 'direct' impact of biogenic habitats on bottom trawl fishes (d_1 through d_4). These 'direct' impacts generally have a smaller magnitude than the indirect impacts, suggesting that the biogenic habitat has similar impacts for fish densities in both sampling programs. Finally, the model estimates spatial variation in the rockfish and flatfish densities calibrated to the bottom trawl (i.e., R2 and F2) and uses the drop camera to estimate variable densities in Bower's Bank, where the bottom trawl survey does not occur (Figure 3). The bottom trawl is used to index density across space (to define essential fish habitat) and time (as input for stock assessments), so the estimated bottom-trawl densities in Bower's Bank represent useful information for fisheries managers.

5 | Discussion

In this paper, we introduce the R package *tinyVAST* to provide an expressive (general and flexible) interface for multivariate

TABLE 5 | Vignettes included in package *tinyVAST*, including what model structures are specified and whether results are compared with alternative packages.

Vignette	What it demonstrates
tinyVAST model description	Summarises model notation, equations and structure in a way that software users can copy when describing their model
Dynamic structural equation models	Specifying a dynamic structural equation model (DSEM) and a comparison with package *dsem* (Thorson et al. 2024)
Comparison with mgcv	Specifying a generalised additive models (GAMs) and a comparison with package *mgcv* (Wood 2017)
Empirical orthogonal function analysis	Specifying an empirical orthogonal function (EOF) model, which cannot be represented using 'arrow-and-lag' notation and instead involves function 'make_eof_ram(.)'
Simultaneous autoregressive process	Specifying a simultaneous autoregressive (SAR) model and a comparison with a multivariate time-series model
Spatial modelling	Specifying a spatial model in two-dimensional coordinates and a comparison with package *mgcv*
Spatial factor analysis	Specifying a spatial factor analysis (SFA) model that estimates covariance among multiple variables
Stream network models	Specifying a spatial model on a stream network
Vector autoregressive spatio-temporal models	Specifying a univariate spatio-temporal model, including a comparison with *VAST* (Thorson 2019) and *sdmTMB* (Anderson et al. In press), and a bivariate vector-autoregressive model
Multiple data types	Specifying an integrated species distribution model using presence/absence, count and biomass sampling data
Age composition analysis	Specifying a multivariate SDM involving multiple ages, used to estimate age-composition using spatially unbalanced sampling
Condition and density	Specifying a marked point process involving numerical density and physiological condition and estimating the abundance-weighted average for the population mark

spatio-temporal models. This software includes functionality for a wide range of time-series and spatial models, including vector autoregressive, dynamic factor, empirical orthogonal function and joint species distribution models (see Table 1). This package is available via CRAN, and we envision it as a replacement for the package *VAST* (Thorson 2019) that is available via GitHub and is widely used in fisheries and marine sciences. It is 'tiny' compared with *VAST* in that it uses Gaussian Markov random fields to represent all spatio-temporal variation, and thereby achieves a more condensed code base and a more familiar user interface. It has a similar interface to the package *sdmTMB* (Anderson et al. In press) but includes additional options for multivariate analysis. It also includes package vignettes that demonstrate the user interface for a wide range of models (Table 5), and these vignettes compare results with alternative software, including *sdmTMB* (Anderson et al. In press), *VAST* (Thorson 2019), *mgcv* (Wood 2017) and *dsem* (Thorson et al. 2024). We hope that this simple yet expressive interface will expand access to multivariate spatio-temporal models among applied ecologists.

Acknowledgements

Thanks to researchers from the AFSC Groundfish Assessment Program for the collection of the historical bottom trawl survey data, and Rachel Wilborn, Kresimir Williams, Rick Towler and Darin Jones for their efforts in collecting and analysing the underwater camera survey data. We also thank Arnaud Grüss, Mallarie Yaeger, Nikolai Morokhovich and one anonymous reviewer for helpful comments on an earlier draft.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All code and data required to reproduce the case-study demonstration are available via Dryad 'Private for Peer Review' via confidential URL (http://datadryad.org/share/M2o70uUDE4al6sXLgWHvb4Z8sF-bw4wtQ5Fgaydrd4Y) using release 1.0.0 of R-package *tinyVAST* (https://zenodo.org/records/15001857). Data and code will be updated upon acceptance. This package is currently avaialble via CRAN.

References

Anderson, S. C., E. J. Ward, P. A. English, L. A. K. Barnett, and J. T. Thorson. In press. "sdmTMB: An R Package for Fast, Flexible, and User-Friendly Generalized Linear Mixed Effects Models With Spatial and Spatiotemporal Random Fields." Journal of Open Source Software: 2022.03.24.485545.

Brooks, M. E., K. Kristensen, K. J. van Benthem, et al. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9: 378–400.

Charsley, A. R., A. Grüss, J. T. Thorson, et al. 2023. "Catchment-Scale Stream Network Spatio-Temporal Models, Applied to the Freshwater Stages of a Diadromous Fish Species, Longfin Eel (*Anguilla dieffenbachii*)." *Fisheries Research* 259: 106583.

Csardi, G., and T. Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems* 1695: 1–9.

Fox, J., Z. Nie, and J. Byrnes. 2020. "Sem: Structural Equation Models." R package version 3.1–11.

Grüss, A., and J. T. Thorson. 2019. "Developing Spatio-Temporal Models Using Multiple Data Types for Evaluating Population Trends and Habitat Usage." *ICES Journal of Marine Science* 76: 1748–1761.

Hoef, J. M. V., E. Peterson, and D. Theobald. 2006. "Spatial Statistical Models That Use Flow and Stream Distance." *Environmental and Ecological Statistics* 13: 449–464.

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: 301–330.

Kass, R. E., and D. Steffey. 1989. "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)." *Journal of the American Statistical Association* 84: 717–726.

Kotwicki, S., J. N. Ianelli, and A. E. Punt. 2014. "Correcting Density-Dependent Effects in Abundance Estimates From Bottom-Trawl Surveys." *ICES Journal of Marine Science* 71: 1107–1116.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70: 1–21.

Laman, E. A., S. Kotwicki, and C. N. Rooper. 2015. "Correlating Environmental and Biogenic Factors With Abundance and Distribution of Pacific Ocean Perch (*Sebastes alutus*) in the Aleutian Islands, Alaska." *Fishery Bulletin* 113: 270–289.

Laman, E. A., C. N. Rooper, K. Turner, S. Rooney, D. W. Cooper, and M. Zimmermann. 2018. "Using Species Distribution Models to Describe Essential Fish Habitat in Alaska." *Canadian Journal of Fisheries and Aquatic Sciences* 75, no. 8: 1230–1255. https://doi.org/10.1139/cjfas-2017-0181.

Latimer, A. M., S. Banerjee, H. Sang Jr., E. S. Mosher, and J. A. Silander Jr. 2009. "Hierarchical Models Facilitate Spatial Analysis of Large Data Sets: A Case Study on Invasive Plant Species in the Northeastern United States." *Ecology Letters* 12: 144–154.

Lindgren, F. 2023. "fmesher: Triangle Meshes and Related Geometry Tools."

Lindgren, F., H. Rue, and J. Lindström. 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: 423–498.

Ovaskainen, O., G. Tikhonov, D. Dunson, et al. 2017. "How Are Species Interactions Structured in Species-Rich Communities? A New Method for Analysing Time-Series Data." *Proceedings of the Royal Society. Biological Sciences* 284: 20170768.

Pearl, J. 2009. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3: 96–146.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rooper, C., M. Sigler, P. Goddard, et al. 2016. "Validation and Improvement of Species Distribution Models for Structure-Forming Invertebrates in the Eastern Bering Sea With an Independent Survey." *Marine Ecology Progress Series* 551: 117–130. https://doi.org/10.3354/meps11703.

Sigler, M., C. Rooper, P. Goddard, R. Wilborn, and K. Williams. 2023. "Alaska Deep-Sea Coral and Sponge Assemblages Are Well-Defined and Mostly Predictable From Local Environmental Conditions." *Marine Ecology Progress Series* 712: 67–85. https://doi.org/10.3354/meps14317.

Siple, M. C., P. G. von Szalay, N. W. Raring, A. N. Dowlin, and B. C. Riggle. 2024. "Data Report: 2023 Gulf of Alaska Bottom Trawl Survey." 2024–09. Alaska Fisheries Science Center, National Marine Fisheries Service.

Stauffer, G. 2004. "NOAA Protocols for Groundfish Bottom Trawl Surveys of the Nation's Fishery Resources." NOAA Technical Memorandum. NMFS-SPO-65. Seattle, WA: National Oceanic and Atmospheric Administration (NOAA).

Thorson, J., and K. Kristensen. 2024. *Spatio-Temporal Models for Ecologists*. Chapman and Hall/CRC.

Thorson, J. T. 2018. “Three Problems With the Conventional Delta-Model for Biomass Sampling Data, and a Computationally Efficient Alternative.” *Canadian Journal of Fisheries and Aquatic Sciences* 75: 1369–1382.

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., A. G. Andrews, T. Essington, and S. Large. 2024. “Dynamic Structural Equation Models Synthesize Ecosystem Dynamics Constrained by Ecological Mechanisms.”

Thorson, J. T., and K. Kristensen. 2016. “Implementing a Generic Method for Bias Correction in Statistical Models Using Random Effects, With Spatial and Population Dynamics Examples.” *Fisheries Research* 175: 66–74.

van der Meer, L., L. Abad, A. Gilardi, and R. Lovelace. 2023. “sfnetworks: Tidy Geospatial Networks.”

Ver Hoef, J. M., E. M. Hanks, and M. B. Hooten. 2018. “On the Relationship Between Conditional (CAR) and Simultaneous (SAR) Autoregressive Models.” *Spatial Statistics* 25: 68–85.

Wikle, C. K., and N. Cressie. 1999. “A Dimension-Reduced Approach to Space-Time Kalman Filtering.” *Biometrika* 86, no. 4: 815–829. https://doi.org/10.1093/biomet/86.4.815.

Wood, S. N. 2017. *Generalized Additive Models: An Introduction With R*. CRC Press.

Wootton, J. T., and M. Emmerson. 2005. “Measurement of Interaction Strength in Nature.” *Annual Review of Ecology, Evolution, and Systematics* 36, no. 1: 419–444. https://doi.org/10.1146/annurev.ecolsys.36.091704.175535.

Wright, S. 1934. “The Method of Path Coefficients.” *Annals of Mathematical Statistics* 5: 161–215.

Yoklavich, M. M., H. G. Greene, G. M. Cailliet, D. E. Sullivan, R. N. Lea, and M. S. Love. 2000. “Habitat Associations of Deep-Water Rockfishes in a Submarine Canyon: An Example of a Natural Refuge.” *Fishery Bulletin* 98: 625.

Zheng, N., N. Cadigan, and J. T. Thorson. 2024. “A Note on Numerical Evaluation of Conditional Akaike Information for Nonlinear Mixed-Effects Models.” In press.

Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology With R*. Springer.