Generalized graphical mixed models connect ecological theory with widely used statistical models

James T. Thorson

Resource Ecology and Fisheries Management, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA

James.Thorson@noaa.gov

Abstract

Ecological dynamics are analyzed across multiple sites, times, and variables. Here, we introduce the family of generalized graphical mixed models (GGMMs) and show that it extends structural equation, generalized additive, and generalized linear mixed models. GGMMs represent ecological systems using a mathematical graph, where each analytic unit (node for each site-time-variable) has a direct effect on other units via specified linear interactions (edges). This graph is composed by combining elementary ecological relationships like ecological interactions, evolutionary trade-offs, time-lags, and spatial diffusion. GGMMs are then expressed using simultaneous equations, efficiently estimated using Gaussian Markov random fields, and used for prediction, inference, and causal analysis. We demonstrate GGMMs using three contrasting case studies: tracking cohorts in age-structured models; phylogenetic path analysis; and diffusion-enhanced spatio-temporal models. We conclude that GGMMs connect ecological

theory with statistical models that are applied for inference, prediction, and causal analysis throughout ecology.

Keywords: mathematical graph; generalized linear mixed model; generalized additive model; structural equation model; diffusion; species interactions; phylogenetic path analysis;

Introduction

Most ecological studies involve measuring, explaining, and predicting dynamics across multiple locations, times, and variables. For example, global conservation involves essential biodiversity variables (measurement of biomass across sites, times, and species), where statistical models are then used to fill in missing elements in this three-coordinate array (Jetz et al., 2019). Alternatively, macroevolution and paleo-ecology seeks to identify how species traits change over time among species subject to extinction and speciation events (Hautmann, 2020). Ecological subfields generally differ in how they discretize space, time, and variables, but these ordinates remain ubiquitous across ecology.

Ecologists then use these measurements across space, time, and variables for three distinct tasks:

1. Inference, i.e., measuring the functions and parameters that give rise to ecological dynamics in observational and experimental systems, so that these parameters can then be compared with ecological theory or other measured values across space and time;
2. Prediction, i.e., estimating the value for a system variable where it was not specifically measured, e.g., to allow biodiversity variables to be compared with target values;
3. Causal analysis, i.e., estimating the value for a system variable under some hypothetical change, allowing policy makers to compare the likely outcome of different potential management strategies and better understand mechanisms of the system.

These different tasks all involve some combination of experimental and observational studies, but an analysis suitable for one task may not be suitable for another (Levins, 1966). For example, a predictive model will often not be suitable for causal analysis (Arif & MacNeil, 2022), a model used to infer ecological parameters may explain a small proportion of predictive

variance, and a model with good in-sample prediction may not provide any transferable inference about system parameters.

Given these contrasting goals, ecologists draw upon a large toolbox of mechanistic and statistical models. However, we argue that most ecologists use some combination of generalized linear (mixed) models (GLMM), generalized additive models (GAM), and structural equation models (SEM) as their core toolbox for statistical analysis. For example, GAMs are widely used for estimating habitat utilization (Miller, 2025), GLMMs are used to control for pseudo-replication in the analysis of experimental designs (Bolker et al., 2009), and there is growing recognition that SEM (and variants like path analysis) have a distinct role for causal analysis (Grace, 2024). This large toolbox of models allows analysts to choose the most appropriate method for their problem. However, it also makes it harder for new researchers to identify what model is appropriate for their intended analysis, thereby raising "barriers to entry" for students and applied ecologists.

In this review, we therefore introduce a family of generalized graphical mixed models (GGMMs) that connects ecological theory with widely used statistical models. A GGMM starts by defining a mathematical "graph," where any combination of variables, times, and sites are defined as nodes (visualized as boxes), and nodes are then linked using edges (visualized as arrows) that represent linear structural interactions. The edges are represented using a path matrix in a simultaneous equation, and the GGMM composes this path matrix from elementary ecological processes including spatial diffusion, time-lags, ecological interactions among species, or evolutionary trade-offs among traits. The resulting GGM then includes GLMMs, GAMs, and SEM as nested submodels, and is useful for inference, prediction, and causal analysis. We specifically emphasize two insights:

1. Graphs can be expressed using a path matrix within a simultaneous equation, and efficiently fitted as a Gaussian Markov random field;
2. Graphs can be constructed from elementary ecological relationships such as time-lags, spatial diffusion, evolutionary relatedness, and unstructured interactions. The path matrix can then be constructed by summing across the statistical interaction (Kronecker product) of these different elementary relationships.

We demonstrate these insights using three examples drawn from population dynamics, macroevolution, and movement ecology.

Materials and Methods

Graphs, simultaneous equations, and random fields

We first introduce mathematical graphs including their construction, representation using simultaneous equations, and estimation using Gaussian Markov random fields.

As an illustrative example, we introduce the essential biogeographic variable **B** measuring biomass $b_{s,c,t}$ at each site $s \in \{1,2,\dots S\}$, species $c \in \{1,2,\dots,C\}$, and time $t \in \{1,2,\dots,T\}$. Assuming we have $S = 100$ sites, $C = 100$ species, and $T = 10$ times, we obtain an array **B** containing $STC = 100{,}000$ measurements of biomass. Ecologists are then interested in interactions among species, which might include nonlocal teleconnections (i.e., source-sink and predator avoidance behaviors) as well as both simultaneous and lagged effects (e.g., where species c_1 in time t_1 consumes juveniles and therefore has delayed impact on adult biomass for species c_2 in time $t_2 = t_1 + \Delta_t$). In the limit, nonlocal and lagged interactions can result in approximately $(SCT)^2/2 = 5 \times 10^9$ interactions (excluding interactions backwards in time) so we obtain a staggering number of potential interactions even using this relatively coarse discretization of space, time, and taxonomy. We therefore seek some conceptually clear,

computationally efficient, and statistically justified simplification for these nonlocal and lagged interactions.

A graphical model represents each element of response $\mathbf{B}$ as a node (visualized as a box), and represents linear relationships as edges (visualized as arrows). This can then be represented as a simultaneous equation:

$$\mathrm{vec}(\mathbf{B}) = \mathbf{P}\mathrm{vec}(\mathbf{B}) + \mathrm{vec}(\mathbf{E}) \quad (1)$$

$$\mathrm{vec}(\mathbf{E}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{L}^T\mathbf{L})$$

where $\mathbf{P}$ is the $K \times K$ path matrix where K is the length of $\mathrm{vec}(\mathbf{B})$, containing elements ρ_{k_2,k_1} that measure the effect of variable k_1 on k_2 (directed edges). Similarly, $\mathbf{E}$ is the array of exogenous variation representing processes that are not explicitly represented in the modeled interactions and $\mathbf{L}^T\mathbf{L}$ is the $K \times K$ covariance matrix for this unmodeled variation, represented using its square-root $\mathbf{L}$ containing λ_{k_2,k_1} (undirected edges). In its graphical representation, the path matrix $\mathbf{P}$ can then be visualized as one-headed arrows where ρ_{k_2,k_1} points from k_1 to k_2, and the exogenous variance $\mathbf{L}$ as two-headed arrows where λ_{k_2,k_1} connects k_1 and k_2. This "box and arrow" visualization is widely used in path analysis (Wright, 1921), and we discuss the connections to causal modelling in a later section.

This simultaneous equation (and associated graphical representation) can be rearranged as a Gaussian Markov random field (GMRF), where $\mathrm{vec}(\mathbf{B})$ follows a multivariate normal distribution such that the inverse covariance ("precision") matrix can be constructed directly:

$$\mathrm{vec}(\mathbf{B}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q}^{-1}) \quad (2)$$

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P}^T)(\mathbf{L}^T\mathbf{L})^{-1}(\mathbf{I} - \mathbf{P})$$

The probability density in the first line can be computed efficiently as long as path matrix $\mathbf{P}$ and exogenous variation $\mathbf{L}$ are sparse, i.e., species primarily interact with a constrained set of other

118 species, within a localized neighborhood, and where interactions occur simultaneously or over a
119 reduced set of lags (see Supplementary Materials 1 for more details).
120 To identify parameters in **P** and **L**, we also require them to have constraints where e.g.,
121 species interactions might be stationary across space and/or time. These constraints show up
122 where element q_{k_2,k_1} of precision **Q** will be equal to q_{k_4,k_3} for other sites, times, or variables.
123 We summarize four common graphs here (see Fig. 1), where each corresponds to a simultaneous
124 equation and an associated path matrix **P** that is common in ecological analysis:
125 1. *Time-lagged dynamics*: Variable b_t is often predicted from its preceding value b_{t-1}, such
126 that $\rho_{t_2,t_1} = 1$ if $t_2 = t_1 + 1$ and 0 otherwise. This results in a sparse lag-1 matrix $\mathbf{P}_{\text{lag1}}$, and
127 a lag-2 matrix arises as $\mathbf{P}_{\text{lag2}} = \mathbf{P}_{\text{lag1}}{}^2$, and so on;
128 2. *Diffusive dynamics*: Ecological variables are often more similar when they are close together
129 (Tobler, 1970), and this spatial autocorrelation can arise as animals or their physical habitat
130 undergoes diffusion (Lindgren et al., 2011). In two-dimensions and discretizing space into
131 square grid cells, a location $\mathbf{s} = (x, y)$ affects its four neighbors $\{(x, y+1), (x+$
132 $1, y), (x, y-1), (x-1, y)\}$ such that each row of $\mathbf{P}_{\text{diffusion}}$ is nonzero for only four elements
133 and it represents the "weight matrix" in a simultaneous autoregressive spatial model (Ver
134 Hoef et al., 2018). Alternatively, metapopulation and metacommunity models often
135 discretize space into habitat patches (e.g., Hanski et al., 1994), and spatial adjacency and
136 diffusion (and resulting spatial autocorrelation) can also be defined in this context;
137 3. *Evolutionary dynamics*: Ecologists often study evolutionary dynamics along a lineage (a
138 phylogeny for species or a pedigree for individuals). A phylogeny is often represented as a
139 tree, wherein a parent taxon will split into two descendants and the evolutionary path matrix
140 $\mathbf{P}_{\text{phylogeny}}$ is nonzero for each pair of descendant (row) and ancestor (column). Genetic drift

occurring within a quadratic fitness landscape will result in a stabilizing selection towards the fitness peak (Lande, 1976);

4. *Interactive dynamics*: Finally, ecologists are often interested in structural linkages among variables. For example, a trophic cascade arises from two negative linkages $Predator \rightarrow Consumer$ and $Consumer \rightarrow Producer$, where the product of these two negative direct effects results in a positive indirect effect from predators to producers. Interactions can then be used to construct the path matrix $\mathbf{P}_{\text{interaction}}$ with whatever pattern is hypothesized;

These elementary relationships can then be combined to structure a larger multivariate model. For example, an analyst might specify interactive dynamics where predator X affects prey Y and prey Y affects consumer Z (Graph-4 above) and where all taxa exhibit spatial diffusion (Graph-2). This involves two interactions $\rho_{X\rightarrow Y}$ and $\rho_{Y\rightarrow Z}$ in 3×3 matrices $\mathbf{P}_{\text{X}\rightarrow\text{Y}}$ and $\mathbf{P}_{\text{Y}\rightarrow\text{Z}}$ where $\mathbf{P}_{\text{interaction}} = \mathbf{P}_{\text{X}\rightarrow\text{Y}} + \mathbf{P}_{\text{Y}\rightarrow\text{Z}}$, and an $S \times S$ matrix $\mathbf{P}_{\text{diffusion}}$, where the joint path matrix is:

$$\mathbf{P}_{\text{joint}} = \underbrace{\mathbf{P}_{\text{X}\rightarrow\text{Y}} \otimes \mathbf{P}_{\text{diffusion}}}_{\substack{\text{diffusive effect of} \\ \text{predator on consumer}}} + \underbrace{\mathbf{P}_{\text{Y}\rightarrow\text{Z}} \otimes \mathbf{P}_{\text{diffusion}}}_{\substack{\text{diffusive effect of} \\ \text{consumer on producer}}} \quad (3)$$

where $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of two matrices, such that resulting matrix $\mathbf{C}$ has dimensions $a_1 b_1 \times a_2 b_2$ when matrix $\mathbf{A}$ has dimension $a_1 \times a_2$ and matrix $\mathbf{B}$ has dimension $b_1 \times b_2$. Therefore, $\mathbf{P}_{\text{joint}}$ is the $3S \times 3S$ matrix arising from three parameters (two interactions and one diffusion rate), given that interaction matrix are 3×3 and $\mathbf{P}_{\text{diffusion}}$ has dimension $S \times S$.

To further illustrate, we next introduce how these simultaneous equations (and associated graphs) arise naturally in ecological analyses.

Case study 1: Tracking cohorts in age-structured demographics

As a first example, ecologists are often interested in predicting abundance at age $n_{a,t}$ for A ages and T years, which arises via survival from the preceding age and year $n_{a-1,t-1}$. However, $n_{a,t}$ might vary for all ages in a single year and therefore be predicted from $n_{a-1,t}$, or it might be affected by changes in survival availability for an age that is consistent across years (i.e., predictable from $n_{a,t-1}$). We therefore explore a model with three interactions, arising from a lag-1 process among years and a separate lag-1 process among ages which we call $\mathbf{G}_{\text{Year}}$ and $\mathbf{G}_{\text{Age}}$, respectively, to distinguish the two versions of the lag-1 matrix $\mathbf{P}_{\text{lag1}}$:

$$\mathbf{P}_{\text{joint}} = \underbrace{\rho_1\left(\mathbf{G}_{\text{Age}} \otimes \mathbf{I}_{\text{Year}}\right)}_{n_{a-1,t} \to n_a,t} + \underbrace{\rho_2\left(\mathbf{I}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}}\right)}_{n_{a,t-1} \to n_a,t} + \underbrace{\rho_3\left(\mathbf{G}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}}\right)}_{n_{a-1,t-1} \to n_a,t} \quad (4)$$

where $\mathbf{G}_{\text{Age}}$ is the $A \times A$ lag-1 matrix $\mathbf{P}_{\text{lag1}}$ among A ages, $\mathbf{I}_{\text{Age}}$ is the $A \times A$ identity matrix, and $\mathbf{G}_{\text{Year}}$ and $\mathbf{I}_{\text{Year}}$ are the corresponding $T \times T$ lag-1 and identity matrices across years (see Supplementary Materials 2 for more details). We fit this model to proportional abundance-at-age for rex sole in the Gulf of Alaska, which was sampled intermittently from 1992-2022 (McGilliard, 2024). We specify a log-linked Tweedie distribution for measurement errors, and fit the model using package *tinyVAST* (Thorson et al., 2025) in the R statistical environment. We then use 10-fold crossvalidation to compare parsimony among the eight models arising from estimating or fixing at zero the three parameters $\{\rho_1, \rho_2, \rho_3\}$ from Eq. 4, and representing the relative importance of within-cohort, within-year, and within-age drivers for observed abundance-at-age.

Case study 2: Phylogenetic trait imputation with varying stabilization rates

Ecologists also study how traits covary among natural populations, seeking to identify trade-offs that arise from adaptation to shared evolutionary constraints. Recent research has developed phylogenetic structural equation models from the Kronecker product of a single evolutionary

matrix that is shared across traits (Thorson et al., 2023), but this does not allow different traits to have different rates of stabilizing selection. We therefore present a novel extension, where we calculate the joint precision from a simultaneous equation that includes phylogenetic path matrix $\mathbf{P}_{\text{phylogeny}}$ and an interaction matrix $\mathbf{P}_{\text{interaction}}$. This results in joint precision:

$$\mathbf{Q}_{\text{joint}} = \left(\mathbf{I} - \mathbf{P}_{\text{joint}}^{T}\right)\mathbf{Q}_{\text{phylogeny}}\left(\mathbf{I} - \mathbf{P}_{\text{joint}}\right) \quad (5)$$

Where $\mathbf{P}_{\text{joint}} = \mathbf{I} \otimes \mathbf{P}_{\text{interaction}}$, $\mathbf{P}_{\text{interaction}}$ is the $C \times C$ matrix of interactions among traits, and $\mathbf{I}$ is the $S \times S$ identity matrix, and $\mathbf{Q}_{\text{phylogeny}}$ is the block-diagonal matrix of evolutionary precisions for each trait (see Supplementary Materials 3 for more details).

To illustrate, we download three traits from PanTHERIA (Jones et al., 2009), representing specific metabolic rate (mL O_2 / g), adult body mass (g), and home range size (km^2). Body size has the highest proportion of data (3340 measurements), while other traits have fewer measurements (Supplementary Materials 5, Table S1). We specify a phylogenetic structural equation model with two interactions $\log(\text{size}) \rightarrow \log(\text{metabolism})$ and $\log(\text{size}) \rightarrow \log(\text{range})$. We also estimate the Ornstein-Uhlenbeck (OU) parameter θ_c for each trait, used to calculate:

$$\mathbf{Q}_{\text{phylogeny}} = \begin{bmatrix} \mathbf{Q}_1 & 0 & 0 \\ 0 & \mathbf{Q}_2 & 0 \\ 0 & 0 & \mathbf{Q}_3 \end{bmatrix} \quad (6)$$

Where $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{Q}_3$ are the evolutionary precisions matrices given OU parameters θ_1, θ_2, and θ_3 for log-size, log-metabolism, and log-range respectively. We fit this model using a dated phylogeny across 5,911 mammal species and 185 million years of evolutionary history (Upham et al., 2019), where $\mathbf{P}_{\text{phylogeny,j}}$ is the 11821×11821 matrix across all tips and ancestral nodes.

Example 3: Diffusion-enhanced spatio-temporal models

Finally, ecologists have a long-running interest in invasive species including the trajectory and rate for expanding range edges. Spatial statisticians have developed non-separable spatio-temporal models that incorporate diffusive dynamics (Lindgren et al., 2023), but these see little use in ecology to date. Here, we present a novel demonstration that graphical models can represent both separable and diffusion-enhanced spatio-temporal dynamics using an additive path matrix, resulting from a lag-1 matrix $\mathbf{P}_{\text{lag1}}$ in time and a spatial diffusion matrix $\mathbf{P}_{\text{diffusion}}$:

$$\mathbf{P}_{\text{joint}} = \underbrace{\rho_1(\mathbf{I}_{\text{time}} \otimes \mathbf{P}_{\text{diffusion}})}_{d_{s,t} \to d_{s+1,t}} + \underbrace{\rho_2(\mathbf{P}_{\text{lag1}} \otimes \mathbf{I}_{\text{space}})}_{d_{s,t} \to d_{s,t+1}} + \underbrace{\rho_3(\mathbf{P}_{\text{lag1}} \otimes \mathbf{P}_{\text{diffusion}})}_{d_{s,t} \to d_{s+1,t+1}} \quad (7)$$

We demonstrate this using a deterministic simulation by visualizing the density matrix $\mathbf{D}$ resulting from diffusive dynamics $\text{vec}(\mathbf{D}) = (\mathbf{I} - \mathbf{P}_{\text{joint}})^{-1}\text{vec}(\mathbf{E})$, where $\text{vec}(\mathbf{E})$ is an indicator vector such that $\mathbf{E}_{s_{\text{center}},1} = 1$ for location s_{center} at the center of the spatial domain in time $t = 1$ and $\mathbf{E}_{s,t} = 0$ elsewhere. We visualize this diffusive process over a square spatial domain discretized into 21 rows and 21 columns ($S = 441$ square grid cells) and $T = 3$ times, while fixing $\rho_1 = 0.8$ and $\rho_2 = 0.1$, and varying the value of ρ_3. Diffusive dynamics are expected to result in a linear increase in the mean-square displacement for the utilization distribution over time (see Supplementary Materials 4 for more details).

Results

In the GGMM fitting to proportional abundance-at-age for rex sole in the Gulf of Alaska, 10-fold cross-validation indicates that the model with interactions along cohorts ($\rho_1 = n_{a,t} \to n_{a+1,t+1}$) and along years ($\rho_2 = n_{a,t} \to n_{a,t+1}$) has lowest predictive error (Supplementary Materials 5, Table S2). The fitted model (Fig. 2) estimates that cohort effects ($\rho_1 = 0.73$) are substantially stronger than year effects ($\rho_2 = 0.3$), and strong cohorts are also visually apparent starting at age-7 around 2005 and again in 2011, and progress visually through subsequent ages

and years. Leave-year-out cross-validation (Supplementary Materials 5, Fig. S1) confirms that including these interactions can result in skillful predictions for years without direct measurements.

The GGMM estimating interactions among adult body size [$\ln(g)$], specific metabolic rate [$\ln(mL\frac{0_2}{g})$], and range size [$\ln(km^2)$] for 5,911 mammal species over 185 million years (Fig. 3) estimates a isometric ($\rho = 1.00$) scaling of range size with adult body size, and an allometric ($\rho = 0.69$) scaling of metabolic rate with body size. Additionally, it estimates weakest stabilizing selection for body size, with a nearly 20% correlation between two species separated by 150 million years of divergent evolution. By contrast, range size has strongest stabilizing selection, where 20% correlation arises at approximately 25 million years of divergence. Finally, specific metabolism has an intermediate strength for stabilizing selection (20% correlation at 60 million years).

Finally, the GGMM for diffusion-enhanced spatio-temporal dynamics (Fig. 4) shows that diffusion arises simply from interactions across time ($n_{s,t} \rightarrow n_{s,t+1}$ using ρ_{time}) and interactions across spatial neighbors ($n_{s,t} \rightarrow n_{s+1,t}$ using ρ_{space}) (Fig. 4 top row). In this scenario, mean-squared displacement (MSD) shows a close-to-linear increases over time with rate ρ_{space}, as expected given diffusive dynamics (where the departure from a linear increases arises from spatial boundary effects). The model then reverts to separable spatio-temporal dynamics when adding a parameter lagged spatial effect ($n_{s,t} \rightarrow n_{s+1,t+1}$ using $\rho_{\text{spacetime}}$) and fixing $\rho_{\text{spacetime}} = -\rho_{\text{space}}\rho_{\text{time}}$ (Fig. 4 bottom row). Finally, intermediate dynamics arise when $-\rho_{\text{space}}\rho_{\text{time}} < \rho_{\text{spacetime}} < 0$. For example, when $\rho_{\text{spacetime}} < 0 = -0.5\rho_{\text{space}}\rho_{\text{time}}$ the MSD starts at 0.1 but then increases 0.05 per time-interval (Fig. 4 middle row). Importantly,

$\rho_{\text{spacetime}}$ allows for a continuous bridge between two ecological hypotheses, where a hotspot remains stationary or diffuses outwards over time.

Discussion

We introduced the family of generalized graphical mixed models (GGMMs), which represent variables as nodes and interactions as edges and are efficiently fitted as a Gaussian Markov random field. In particular, ecological variables are often indexed by space, time, and category (e.g., species or age). Ecological interactions are specified by combining several elementary graphical structures representing time-lags, phylogenetic relatedness, spatial diffusion, ecological interactions among species, or evolutionary trade-offs among species traits. Using three varied case-studies, we specifically showed that interactions among elementary graphical structures can represent population dynamics (age structure), evolutionary dynamics (stabilizing selection among traits), and movement dynamics (diffusion-enhanced spatio-temporal variation). This then yielded novel insights, e.g., that stabilizing selection is stronger for adult home range than body size within the mammal lineage.

As shown in our third case study (diffusion-enhanced spatio-temporal models), GGMMs can be used as a spatial smoother similar to generalized additive models (GAMs). In particular, the diffusion-enhanced spatio-temporal model constructs a set of local basis functions from structural parameters representing spatial diffusion, temporal autocorrelation, and the space-time interaction (see Fig. 4). A basis function is constructed for every combination of space and time, and summing across the estimated response to these basis functions constructs a piecewise smooth function similar to a smoothing spline. Previous research has already discussed the connection between GAMs and spatial GLMMs (Miller, 2025), while other research has derived a spatial smoother from diffusive movement (Lindgren et al., 2011). However, GGMMs extend

this literature by deriving a spatio-temporal smoother from ecological interpretable parameters where, e.g., the space-time interaction parameter ρ_3 determines whether spatial hotspots propagate outwards in space over time.

Similarly, there is growing interest in using structural causal models to re-interpret a wide range of ecological analyses (Arif & MacNeil, 2022; Byrnes & Dee, 2025; Grace, 2024). Usefully, GGMMs predict the covariance across space, time, and categories by constructing the path matrix as the sum across structural interaction like spatial diffusion, time-lags, and interactions among species. Therefore, GGMMs can be used in the causal modelling workflow, i.e., developing a graph from scientific knowledge, validating it by determining whether it is consistent with available data, and subsequently using it to compute direct and indirect effects (summarizing Fig. 2 from Arif & MacNeil, 2023). Usefully, GGMMs allow us to compute causal effects that occur in some spatial and temporal neighborhood, e.g., how a change in density for one species affects other species at nearby sites or after a time-lag (Leibold et al., 2004). Spatial spillover and storage effects are important in determining species coexistence, and using GGMMs for causal inference to test modern coexistence theory seems like a fruitful direction for future research.

In summary, we see GGMMs as a useful avenue to integrate ecological theory (i.e., specific ecological interactions across space, time, and variables) with statistical estimation (hierarchical modelling tools). They derive the predicted covariance across coordinates from elementary structural relationships, and this structural model can then be interpreted as a structural causal model when appropriate (i.e., if assumptions are based on theory and consistent with available data). By providing a unified framework across ecological and evolutionary analyses, we hope

289 that GGMMs will allow researchers to move more easily between predictive, inferential, and
290 causal analyses.

291 **Acknowledgements**

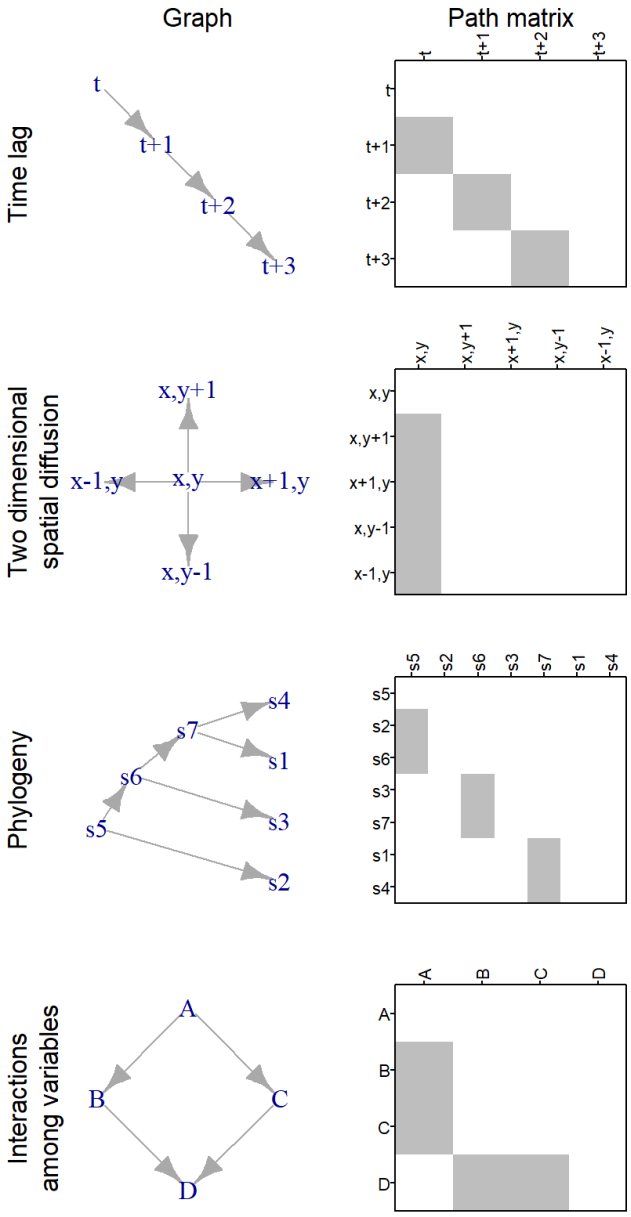
292 I thank H. Wu, who is continuing research related to the diffusion-enhanced model presented
293 here. I also thank C. McGilliard, who developed the rex sole proportional abundance-at-age
294 explored here. The manuscript was improved by comments from C. Monnahan and L.
295 Campbell.

296 **Data accessibility statement**

297 All code and data required to replicate analyses and figures are available on GitHub online
298 (https://github.com/James-Thorson/GGMM/) [To be made publicly accessible upon acceptance].
299 The mammal phylogeny was downloaded from VertLife (https://vertlife.org/phylosubsets/) and
300 was developed by Upham et al. (2019). The mammal traits were accessed from PanTHERIA
301 (Jones et al., 2009), available online from ESA archives
302 (https://esapubs.org/archive/ecol/E090/184/metadata.htm). The proportional abundance-at-age
303 data for rex sole in the Gulf of Alaska is publicly available (https://github.com/noaa-
304 afsc/goa_rex/blob/main/runs/2025_cie_review/2021_accepted_model_inputs/GOA_Rex_8_2021
305 .dat) from the 2024 stock assessment (McGilliard, 2024) and distributed for a Center for
306 Independent Experts 2025 review of the rex sole assessment.

307

Fig. 1 – Graphs (left column) representing common simplifying assumptions for ecological dynamics, and the path matrix **P** (right column) in simultaneous equation $\mathbf{y} = \mathbf{P}\mathbf{y} + \boldsymbol{\epsilon}$ that results from each graph (with grey box when $\rho_{i,j} \neq 0$ corresponding to graph arrows, and white boxes where $\rho_{i,j} = 0$), showing first-order autoregressive dynamics (top row), spatial diffusion from a central location (x, y) when using square boxes to discretize a spatial domain in two dimensions with four adjacent grid cells (2nd row), a dated phylogeny (3rd row) showing ancestral nodes $\{s5, s6, s7\}$ and extant species $\{s1, s2, s3, s4\}$, and interactions among four variables $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow D$, and $C \rightarrow D$ (4th row).


Graph
Path matrix
Time lag
t
t+1
t+2
t+3
Two dimensional spatial diffusion
x,y+1
x-1,y
x,y
x+1,y
x,y-1
Phylogeny
s4
s7
s1
s6
s3
s5
s2
Interactions among variables
A
B
C
D

316

317

Fig. 2 – The estimated interactions (top panel) when predicting proportional abundance at age $n_{a,t}$ for rex sole in the Gulf of Alaska showing the estimated effect of survival along a cohort ρ_3 and effects along a year ρ_2 (see Eq. 4), as well as the observed $n_{a,t}$ (middle panel) for each year (x-axis) and age (y-axis) showing low (purple) to high (yellow) values, and the estimated $n_{a,t}$ (bottom panel) including the estimated value for years with no direct sampling (white spaces in the middle panel).

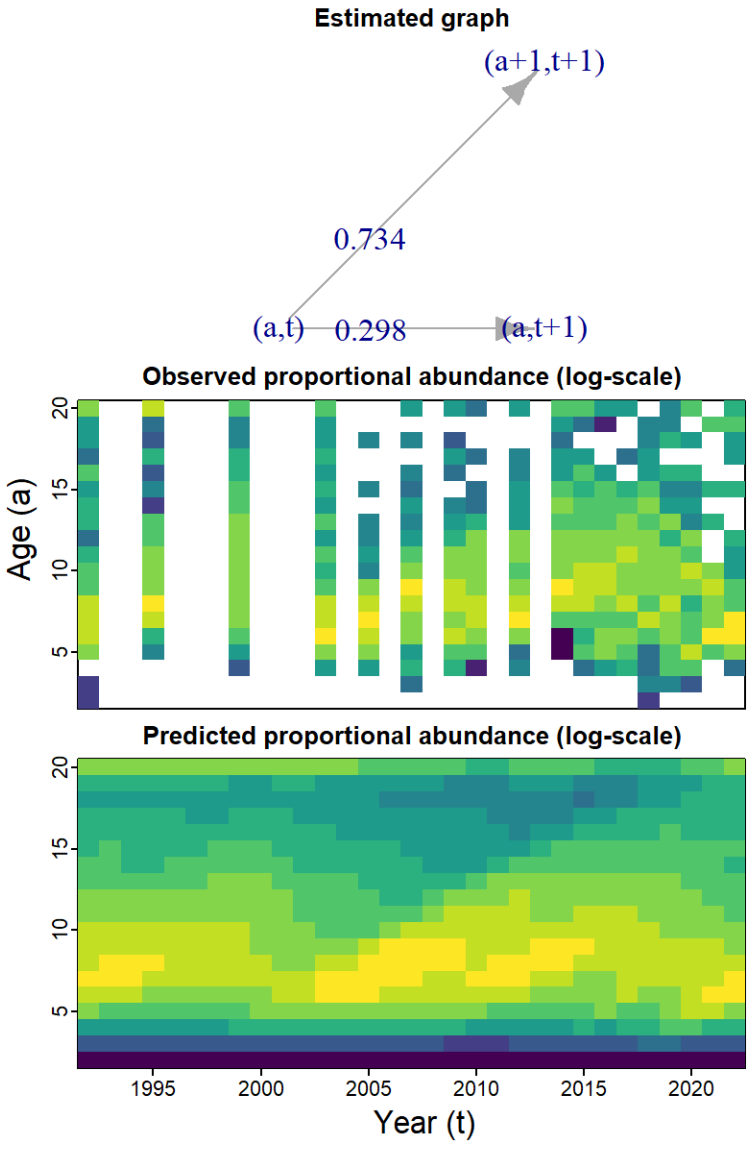


Fig. 3 – The estimated interaction (top panel) among three log-transformed traits (adult body mass [g], basal metabolic rate [mL O_2 / hour], and range size [km^2]) for 4999 mammal species (with 3340, 661, and 547 available measurements respectively), as well as the estimated correlation over time for residual patterns (bottom panel) showing the correlation (y-axis) over 185 million years (x-axis) of evolutionary history for mammals

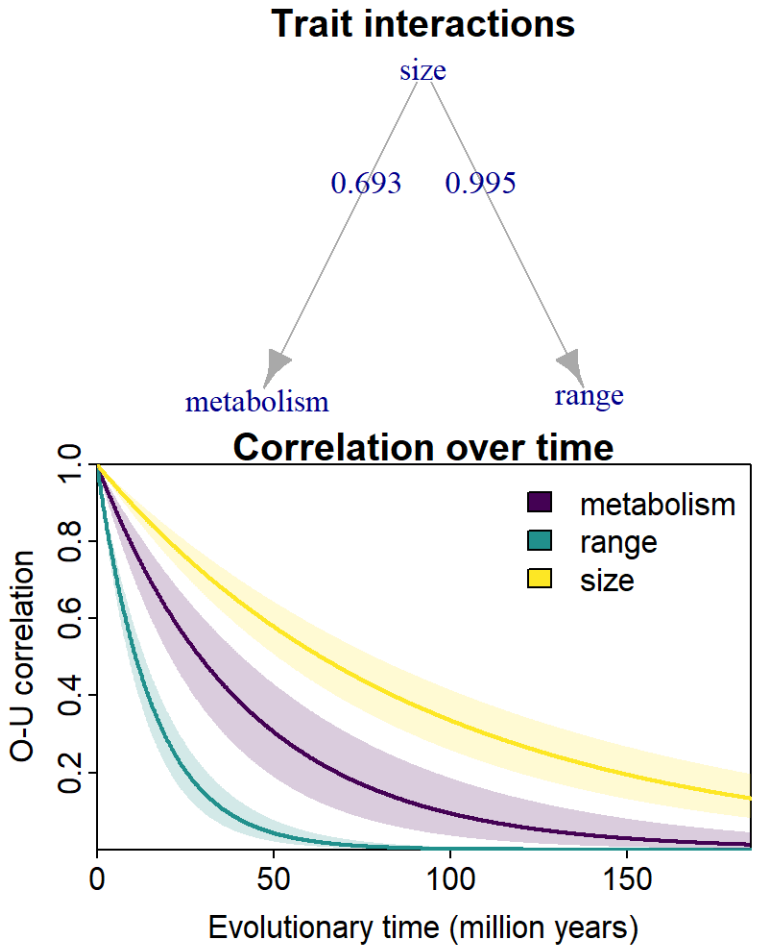
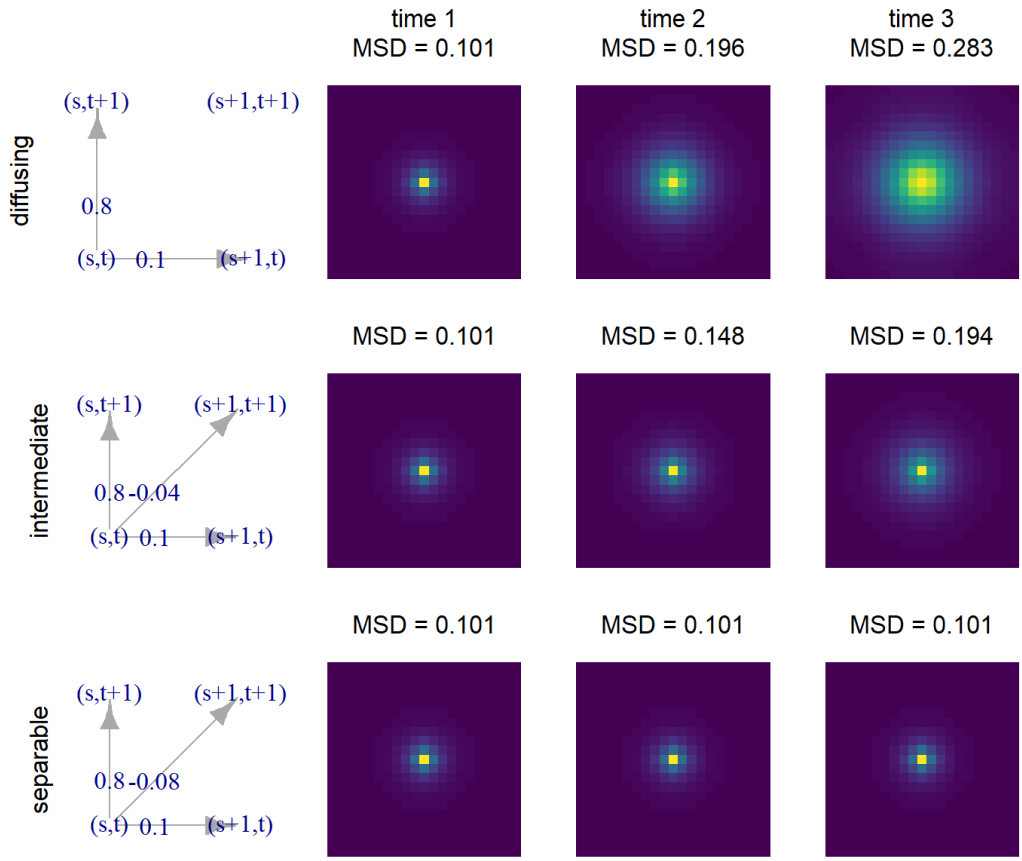


Fig. 4 – Visualizing diffusion-enhanced spatio-temporal dynamics (top-row), intermediate dynamics (middle row), and non-diffusive (separable) spatio-temporal dynamics (bottom row), including the graph (1st column) linking a focal cell (s,t) and adjacent cells in a given time $(s+1,t)$, the same cell in the next time $(s,t+1)$, or adjacent cells in the next time $(s+1,t+1)$. Nonseparable dynamics arise when $\rho_{s+1,t+1} = -\rho_{s,t+1}\rho_{s+1,t}$. We also visualize resulting dynamics from a concentrated density (purple is 0 density, yellow is high density) in time $t = 1$ (2nd column), and how this density evolves in times 2 (3rd column) and 3 (4th column). For each panel we also calculate the mean-squared displacement (MSD) as the variance of the density function. Diffusive dynamics results in MSD increasing linearly over time, although the diffusive MSD in time-3 is slightly lower due to boundary effects.



344

345

Bibliography

Arif, S., & MacNeil, M. A. (2022). Predictive models aren't for causal inference. *Ecology Letters*, *25*(8), 1741–1745. https://doi.org/10.1111/ele.14033

Arif, S., & MacNeil, M. A. (2023). Applying the structural causal model framework for observational causal inference in ecology. *Ecological Monographs*, *93*(1), e1554. https://doi.org/10.1002/ecm.1554

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135.

Byrnes, J. E. K., & Dee, L. E. (2025). Causal Inference With Observational Data and Unobserved Confounding Variables. *Ecology Letters*, *28*(1), e70023. https://doi.org/10.1111/ele.70023

Grace, J. B. (2024). An integrative paradigm for building causal knowledge. *Ecological Monographs*, *94*(4), e1628. https://doi.org/10.1002/ecm.1628

Hanski, I., Kuussaari, M., & Nieminen, M. (1994). Metapopulation Structure and Migration in the Butterfly Melitaea Cinxia. *Ecology*, *75*(3), 747–762. https://doi.org/10.2307/1941732

Hautmann, M. (2020). What is macroevolution? *Palaeontology*, *63*(1), 1–11. https://doi.org/10.1111/pala.12465

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), 539–551. https://doi.org/10.1038/s41559-019-0826-1

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., … Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648–2648. https://doi.org/10.1890/08-1494.1

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 314–334.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 421–431.

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., & Rue, H. (2023). *A diffusion-based spatio-temporal extension of Gaussian Matérn fields* (No. arXiv:2006.04917). arXiv. https://doi.org/10.48550/arXiv.2006.04917

Lindgren, F., Rue, H., & Lindström, 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. https://doi.org/10.1111/j.1467-9868.2011.00777.x

McGilliard, C. R. (2024). *Assessment of the Rex Sole Stock in the Gulf of Alaska*. North Pacific Fishery Management Council. https://www.npfmc.org/wp-content/PDFdocuments/SAFE/2024/GOArex.pdf

Miller, D. L. (2025). Bayesian views of generalized additive modelling. *Methods in Ecology and Evolution*, *16*(3), 446–455. https://doi.org/10.1111/2041-210X.14498

Thorson, J. T., Anderson, S. C., Goddard, P., & Rooper, C. N. (2025). tinyVAST: R Package With an Expressive Interface to Specify Lagged and Simultaneous Effects in Multivariate Spatio-Temporal Models. *Global Ecology and Biogeography*, *34*(4), e70035. https://doi.org/10.1111/geb.70035

Thorson, J. T., Maureaud, A. A., Frelat, R., Mérigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. *Methods in Ecology and Evolution*, *14*(5), 1259–1275. https://doi.org/10.1111/2041-210X.14076

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, *17*(12), e3000494. https://doi.org/10.1371/journal.pbio.3000494

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85. https://doi.org/10.1016/j.spasta.2018.04.006

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Supplementary Materials 1: Derivation of GMRF for simultaneous equation

In the main text, we start with a simultaneous equation:

$$\mathbf{z} = \mathbf{P}\mathbf{z} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{L}^T\mathbf{L})$$

This can generalize a wide range of linear models including a linear regression. For example, consider a simple linear model with response Y and predictor X:

$$Y = \mu_\mathrm{Y} + \beta X + \epsilon_\mathrm{Y}$$

$$\epsilon_\mathrm{Y} \sim \mathrm{Normal}(0, \sigma_\mathrm{Y}^2)$$

We can then augment this with an expression for the predictor variable:

$$X = \mu_\mathrm{X} + \epsilon_\mathrm{X}$$

$$\epsilon_\mathrm{X} \sim \mathrm{Normal}(0, \sigma_\mathrm{X}^2)$$

We can the express this augmented linear model as a simultaneous equation, where $\mathbf{z} = (X - \mu_\mathrm{X}, Y - \mu_\mathrm{Y})$:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix}$$

And $\boldsymbol{\epsilon} = (\epsilon_\mathrm{X}, \epsilon_\mathrm{Y})$, where:

$$\mathbf{L} = \begin{bmatrix} \sigma_\mathrm{X} & 0 \\ 0 & \sigma_\mathrm{Y} \end{bmatrix}$$

It is easy to show that the simultaneous equation estimates identical parameters as the original linear model.

Extending this to a matrix response $\mathbf{B}$, we obtain a larger simultaneous equation:

$$\mathrm{vec}(\mathbf{B}) = \mathbf{P}\mathrm{vec}(\mathbf{B}) + \mathrm{vec}(\mathbf{E})$$

$$\mathrm{vec}(\mathbf{E}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{L}^T\mathbf{L})$$

We then subtract $\mathbf{P}\text{vec}(\mathbf{B})$ from both sides of S1A and re-arrange to obtain:

$$\text{vec}(\mathbf{B}) - \mathbf{P}\text{vec}(\mathbf{B}) = \text{vec}(\mathbf{E})$$

$$(\mathbf{I} - \mathbf{P})\text{vec}(\mathbf{B}) = \text{vec}(\mathbf{E})$$

We then multiply both sides by $(\mathbf{I} - \mathbf{P})^{-1}$ to obtain

$$\text{vec}(\mathbf{B}) = (\mathbf{I} - \mathbf{P})^{-1}\text{vec}(\mathbf{E})$$

We note the property that if $X \sim Normal(0, \sigma_X^2)$ then $aX \sim Normal(0, a^2\sigma_X^2)$. The same property holds for vectors, where $\mathbf{x} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma})$ implies that $\mathbf{Mx} \sim \text{MVN}(\mathbf{0}, \mathbf{M\Sigma M}^{\text{T}})$. Therefore, we can multiply $(\mathbf{I} - \mathbf{P})^{-1}$ into the covariance for the multivariate normal distribution for $\text{vec}(\mathbf{E}) \sim \text{MVN}(\mathbf{0}, \mathbf{L}^T\mathbf{L})$ to get:

$$\text{vec}(\mathbf{B}) \sim \text{MVN}(\mathbf{0}, (\mathbf{I} - \mathbf{P})^{-1}\mathbf{L}^T\mathbf{L}(\mathbf{I} - \mathbf{P}^T)^{-1})$$

Defining precision $\mathbf{Q} = \mathbf{\Sigma}^{-1}$, we then obtain:

$$\text{vec}(\mathbf{B}) \sim \text{MVN}(\mathbf{0}, \mathbf{Q}^{-1})$$

where:

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P}^T)(\mathbf{L}^T\mathbf{L})^{-1}(\mathbf{I} - \mathbf{P})$$

i.e, $\text{vec}(\mathbf{B})$ follows a multivariate normal distribution where we can construct the precision directly, and therefore is a Gaussian Markov random field. Furthermore, if $\mathbf{L}^T\mathbf{L}$ is diagonal, then $\mathbf{Q}$ will have the same sparsity pattern as $\mathbf{I} + \mathbf{P}^T\mathbf{P}$.

Expressing a simultaneous equation as GMRF is useful for two reasons:

1. *Evaluating the multivariate normal PDF*: Fitting a simultaneous equation requires evaluating the multivariate normal probability density function (MVN-PDF) for proposed values of parameters. The MVN-PDF is:

$$f(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{k/2}} |\mathbf{\Sigma}^{-1}|^{0.5} e^{-0.5(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x})}$$

Therefore, the covariance $\mathbf{\Sigma}$ only ever appears via its inverse $\mathbf{Q}$, and if we can construct the precision directly then we can avoid computing any matrix inversion.

2. *Approximating the log-marginal likelihood*: In a mixed model, we must calculate the log-likelihood of parameters $\boldsymbol{\theta}$ given data $\mathbf{y}$ while marginalizing across any random effects $\mathbf{z}$:

$$f(\boldsymbol{\theta}; \mathbf{y}) = \log \int_z \mathcal{L}(\boldsymbol{\theta}, \mathbf{z}; \mathbf{y}) dz$$

In the following, we approximate this using the Laplace approximation, which replaces the integral with a multivariate normal distribution with the same peak and curvature. Curvature is approximated using the matrix of second derivatives with respect to random effects (termed the Hessian matrix $\mathbf{H}$). If each datum y_i is calculated from fixed effects and (at most) a single random effect, then the Hessian matrix will have the same sparsity pattern as the precision $\mathbf{Q}$. This then allows computation to skip calculating elements of the Hessian matrix used in the Laplace approximation.

Supplementary Materials 2: Model details for case study 1

In our first case study, we fit a generalized graphical mixed model (GGMM) to $A \times T$ matrix $\mathbf{N}$ of abundance at age $n_{a,t}$. We specify a generalized linear model:

$$n_{a,t} \sim \text{Tweedie}(\mu_{a,t}, \theta, \psi)$$

Where $\text{Var}(n_{a,t}) = \theta \mu_{a,t}^{\psi}$, i.e., θ is the dispersion and ψ controls the mean-variance relationship. The expected abundance-at-age $\mu_{a,t}$ arises from a log-linked linear predictor:

$$\log(\mu_{a,t}) = \beta_a + \omega_{a,t}$$

where β_a is an annually varying intercept that represents average survey availability and the net effect of average age-specific survival, and $\delta_{a,t}$ represents deviations around this long-term log-abundance at age.

As a graphical model, deviations $\omega_{a,t}$ follow a simultaneous equation:

$$\text{vec}(\boldsymbol{\Omega}) = \mathbf{P}\text{vec}(\boldsymbol{\Omega}) + \text{vec}(\mathbf{E})$$

$$\text{vec}(\mathbf{E}) \sim \text{MVN}(\mathbf{0}, \mathbf{L}^T\mathbf{L})$$

where $\mathbf{P}$ is the $AT \times AT$ path matrix and $\mathbf{E}$ is exogenous variation with covariance $\mathbf{L}^T\mathbf{L}$. As stated in the main text:

$$\mathbf{P}_{\text{joint}} = \underbrace{\rho_1(\mathbf{G}_{\text{Age}} \otimes \mathbf{I}_{\text{Year}})}_{n_{a-1,t} \to n_a,t} + \underbrace{\rho_2(\mathbf{I}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}})}_{n_{a,t-1} \to n_a,t} + \underbrace{\rho_2(\mathbf{G}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}})}_{n_{a-1,t-1} \to n_a,t}$$

Where $\mathbf{G}_{\text{Age}}$ is the $A \times A$ lag-1 matrix:

$$\mathbf{G}_{\text{Age}} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

i.e., a banded matrix with 1s immediately below the diagonal and zero elsewhere, and $\mathbf{G}_{\text{Year}}$ is the $T \times T$ lag-1 matrix with the same band-1 structure. Finally, $\mathbf{I}_{\text{Age}}$ is the $A \times A$ identity matrix and $\mathbf{I}_{\text{Year}}$ is the $T \times T$ identity matrix. We also specify the simplest structure for exogenous variation:

$$\mathbf{L} = \sigma\left(\mathbf{I}_{\text{Age}} \otimes \mathbf{I}_{\text{Year}}\right)$$

Such that exogenous variance $\mathbf{L}^T\mathbf{L} = \sigma^2\mathbf{I}$ is constant across ages and years.

To visualize this more concretely, we walk through the example with two ages ($A = 2$) and three times ($T = 3$). In this case:

$$\mathbf{G}_{\text{Age}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

and:

$$\mathbf{G}_{\text{Year}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Where deviation $\omega_{a,t}$ is modeled jointly as a 6 length vector $\text{vec}(\mathbf{\Omega}) = (\omega_{1,1}, \omega_{2,1}, \omega_{1,2}, \omega_{2,2}, \omega_{1,3}, \omega_{2,3})$. We can therefore write out each component of the 6×6 path matrix $\mathbf{P}_{\text{joint}}$ individually:

$$\underbrace{\rho_1\left(\mathbf{G}_{\text{Age}} \otimes \mathbf{I}_{\text{Year}}\right)}_{n_{a-1,t} \to n_a,t} = \rho_1 \begin{bmatrix} \mathbf{G}_{\text{Age}} & 0 & 0 \\ 0 & \mathbf{G}_{\text{Age}} & 0 \\ 0 & 0 & \mathbf{G}_{\text{Age}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \rho_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_1 & 0 \end{bmatrix}$$

and

$$\underbrace{\rho_2\left(\mathbf{I}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}}\right)}_{n_{a,t-1} \to n_a,t} = \rho_2 \begin{bmatrix} 0 & 0 & 0 \\ \mathbf{I}_{\text{Age}} & 0 & 0 \\ 0 & \mathbf{I}_{\text{Age}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \rho_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_2 & 0 & 0 \end{bmatrix}$$

and

$$\underbrace{\rho_3\left(\mathbf{G}_{\text{Age}} \otimes \mathbf{G}_{\text{Year}}\right)}_{n_{a-1,t-1} \to n_a,t} = \rho_3 \begin{bmatrix} 0 & 0 & 0 \\ \mathbf{G}_{\text{Age}} & 0 & 0 \\ 0 & \mathbf{G}_{\text{Age}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \rho_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_3 & 0 & 0 & 0 \end{bmatrix}$$

Such that $\mathbf{P}_{\text{joint}}$ has the following structure:

$$\mathbf{P}_{\text{joint}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \rho_1 & 0 & 0 & 0 & 0 & 0 \\ \rho_2 & 0 & 0 & 0 & 0 & 0 \\ \rho_3 & \rho_2 & \rho_1 & 0 & 0 & 0 \\ 0 & 0 & \rho_2 & 0 & 0 & 0 \\ 0 & 0 & \rho_3 & \rho_2 & \rho_1 & 0 \end{bmatrix}$$

Supplementary Materials 3: Model details for case study 2

In our second case study, we present a univariate Ornstein-Uhlenbeck process for evolution of a trait $\boldsymbol{\omega}$, with value ω_s each location s (tip or ancestral node) within a phylogenetic tree. The univariate O-U process with unit variance defines the value ω_{s_2} for a child taxon s_2 conditional upon the value ω_{s_1} for its parent s_1:

$$\omega_{s_2} \sim \text{Normal}\left(\rho_{s_2}\omega_{s_1}, \sigma_{s_2}^2\right)$$

with partial correlation calculated from the evolutionary distance l_{s_2} between parent s_1 and child s_2:

$$\rho_{s_2} = e^{-\theta l_{s_2}}$$

and residual variance:

$$\sigma_{s_2}^2 = \frac{1}{2\theta}\left(1 - e^{-2\theta l_{s_2}}\right)$$

Alternatively, this unit-variance O-U process can be expressed as a simultaneous equation:

$$\mathbf{D}\boldsymbol{\omega} = \mathbf{P}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}, \mathbf{D})$$

where the $S \times S$ path matrix $\mathbf{P}$ is nonzero for parent-child pairs:

$$\rho_{s_2,s_1} = 2\theta \frac{e^{-\theta l_{s_2}}}{1 - e^{-2\theta l_{s_2}}}$$

and the $S \times S$ matrix $\mathbf{D}$ is diagonal:

$$d_{s,s} = 2\theta \begin{cases} 1 & \text{if } s \text{ is the root} \\ 1 + \dfrac{e^{-\theta l_s}}{1 - e^{-2\theta l_s}} & \text{otherwise} \end{cases}$$

where the root starts at the marginal variance of the O-U process and therefore (in a loose sense) has distance $l_{s_2} \to \infty$ from its parent taxon (and a similar expression can be written for a random

walk ("Brownian motion") process. We can therefore specify the S length vector of traits in an O-U process as a Gaussian Markov random field:

$$\boldsymbol{\omega} \sim \mathrm{MVN}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1})$$

$$\mathbf{Q} = (\mathbf{D} - \mathbf{P}^T)\mathbf{D}^{-1}(\mathbf{D} - \mathbf{P})$$

Where $\mathbf{P}$, $\mathbf{D}$ and therefore $\mathbf{Q}$ depend upon the O-U parameter θ, and σ^2 is the variance of the O-U process given that $\mathbf{Q}$ is defined to have unit variance.

In the main text, we the fit a simultaneous equation for $S \times C$ trait matrix $\mathbf{X}$ across $C = 3$ traits, $\mathbf{x}_s = (size, metabolism, range)$, where the O-U parameter θ_c also varies among traits:

$$\mathrm{vec}(\mathbf{X}) = \mathbf{P}_{\mathrm{joint}}\mathrm{vec}(\mathbf{X}) + \mathrm{vec}(\mathbf{E})$$

$$\mathbf{E} \sim \mathrm{MVN}\left(0, \mathbf{Q}_{\mathrm{phylogeny}}^{-1}\right)$$

Where:

$$\mathbf{P}_{\mathrm{joint}} = \mathbf{I} \otimes \mathbf{P}_{\mathrm{interaction}}$$

$$\mathbf{P}_{\mathrm{interaction}} = \begin{bmatrix} 0 & 0 & 0 \\ \rho_1 & 0 & 0 \\ \rho_2 & 0 & 0 \end{bmatrix}$$

Where ρ_1 is the impact of log-size on log-metabolism, and ρ_2 is the impact of log-size on log-range, and:

$$\mathbf{Q}_{\mathrm{phylogeny}} = \begin{bmatrix} \sigma_1^{-2}\mathbf{Q}_1 & 0 & 0 \\ 0 & \sigma_2^{-2}\mathbf{Q}_2 & 0 \\ 0 & 0 & \sigma_2^{-2}\mathbf{Q}_3 \end{bmatrix}$$

And where $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{Q}_3$ are the O-U precisions given parameters θ_1, θ_2, and θ_3 for log-size, log-metabolism, and log-range, and σ_1^2, σ_2^2, and σ_3^2 are their estimated variances, respectively. This then results in joint precision:

$$\mathbf{Q}_{\mathrm{joint}} = \left(\mathbf{I} - \mathbf{P}_{\mathrm{joint}}^T\right)\mathbf{Q}_{\mathrm{phylogeny}}\left(\mathbf{I} - \mathbf{P}_{\mathrm{joint}}\right)$$

By defining $\widetilde{\mathbf{P}} = \mathbf{D}^{-1}\mathbf{P}$, the evolutionary precision for the univariate Ornstein-Uhlenbeck process can instead be written as:

$$\mathbf{Q} = \left(\mathbf{I} - \widetilde{\mathbf{P}}^{T}\right)\mathbf{D}\left(\mathbf{I} - \widetilde{\mathbf{P}}\right)$$

We can then redefine an additive path matrix $\widetilde{\mathbf{P}}_{\text{joint}}$ as:

$$\widetilde{\mathbf{P}}_{\text{joint}} = \rho_1\left(\left(\mathbf{I} - \widetilde{\mathbf{P}}_{\text{metabolism}}\right) \otimes \mathbf{P}_{\text{size}\rightarrow\text{metabolism}}\right) + \rho_2\left(\left(\mathbf{I} - \widetilde{\mathbf{P}}_{\text{range}}\right) \otimes \mathbf{P}_{\text{size}\rightarrow\text{range}}\right) +$$

$$\left(\widetilde{\mathbf{P}}_{\text{size}} \otimes \mathbf{I}_{\text{size}}\right) + \left(\widetilde{\mathbf{P}}_{\text{metabolism}} \otimes \mathbf{I}_{\text{metabolism}}\right) + \left(\widetilde{\mathbf{P}}_{\text{range}} \otimes \mathbf{I}_{\text{range}}\right)$$

And calculate the joint precision as:

$$\mathbf{Q}_{\text{joint}} = \left(\mathbf{I} - \widetilde{\mathbf{P}}_{\text{joint}}^{T}\right)\mathbf{D}_{\text{joint}}\left(\mathbf{I} - \widetilde{\mathbf{P}}_{\text{joint}}\right)$$

where:

$$\mathbf{D}_{\text{joint}} = \begin{bmatrix} \sigma_1^{-2}\mathbf{D}_1 & 0 & 0 \\ 0 & \sigma_1^{-2}\mathbf{D}_2 & 0 \\ 0 & 0 & \sigma_3^{-2}\mathbf{D}_3 \end{bmatrix}$$

We instead present Eq. 5 in the main text because we believe that it is more intuitive, but here we have showed that the joint precision could instead be presented as an additive path matrix.

Supplementary Materials 4: Model details for case study 3

In our third case study, we present a diffusion-enhanced spatio-temporal model using a square spatial domain that is discretized into S square grid cells. To do so, we first define $S \times S$ adjacency matrix $\mathbf{A}$ where $a_{s_2,s_1} = 1$ if cells s_1 and s_2 share an edge and zero otherwise (i.e., rook adjacency). We then rescale this adjacency matrix to calculate the diffusion path matrix:

$$\mathbf{P}_{\text{diffusion}} = \text{diag}(\mathbf{A1})^{-1}\mathbf{A} - \mathbf{I}$$

Where $\mathbf{1}$ is a S length vector of ones, such that $\mathbf{A1}$ is the sum across rows of the adjacency matrix (i.e., number of neighbors for each location s_1), and $\text{diag}(\mathbf{A1})^{-1}$ is a diagonal matrix that rescales $\mathbf{A}$ by the number of neighbors. In summary, $\mathbf{P}_{\text{diffusion}}$ is the adjacency matrix transformed such that each row sums to zero and has diagonal of -1, and it is analogous to the 1st-order lag matrix where lag-n can be calculated as $\mathbf{P}^n_{\text{diffusion}}$, i.e., the lag-0 matrix is $\mathbf{P}^0_{\text{diffusion}} = \mathbf{I}$ and the 2nd order diffusion matrix is $\mathbf{P}_{\text{diffusion}}\mathbf{P}_{\text{diffusion}}$.

We can then use this diffusive path matrix to define a diffusion-enhanced spatio-temporal process:

$$\mathbf{P}_{\text{joint}} = \underbrace{\rho_1(\mathbf{I}_{\text{time}} \otimes \mathbf{P}_{\text{diffusion}})}_{d_{s,t} \to d_{s+1,t}} + \underbrace{\rho_2\left(\mathbf{P}_{\text{lag1}} \otimes \mathbf{I}_{\text{space}}\right)}_{d_{s,t} \to d_{s,t+1}} + \underbrace{\rho_3\left(\mathbf{P}_{\text{lag1}} \otimes \mathbf{P}_{\text{diffusion}}\right)}_{d_{s,t} \to d_{s+1,t+1}}$$

In the main text, we define this process over three times $T = 3$, and calculate the $S \times T$ density matrix $\mathbf{D}$ containing density $d_{s,t}$ that results from an exogenous pulse experiment represented by an $S \times T$ matrix $\mathbf{E}$, where $\epsilon_{s,t}$ is one for the midpoint of the spatial domain when $t = 1$ and zero otherwise. We specifically visualize:

$$\text{vec}(\mathbf{D}) = \mathbf{P}_{\text{joint}}\text{vec}(\mathbf{E})$$

588 corresponding to the diffusion across space and time resulting from density starting in a single
589 midpoint cell in the first time.

590

Supplementary Materials 5: Additional figures and tables

Table S1 – Count of trait measurements available for each individual trait (along the diagonal) or any pair of traits (off-diagonal), from the 4999 species available in PanTHERIA that can be matched with the Vertlife phylogeny based on scientific binomial.

	ln_metabolism	ln_range	ln_size
ln_metabolism	547		
ln_range	233	661	
ln_size	547	658	3340

Table S2 – Results from a simple-random 10-fold crossvalidation experiment for each of eight models for proportional abundance-at-age for rex sole in the Gulf of Alaska, arising from every combination of estimating three potential interaction parameters or fixing them at zero. For each model, we list the interaction parameters included, the number of fixed effects, the crossvalidation root-mean-squared error, and the proportion of crossvalidation mean-squared error relative to the null model. Note that each model includes 22 parameters in addition to the interactions: an intercept for each age $a \in \{2,3,\ldots,20\}$; the variance for exogenous variation in the graphical model; and the dispersion and power parameters for the Tweedie distribution for residual variation.

Parameter included			Number of parameters	Cross-validation RMSE	Proportion of cross-validation variance explained
Bin	Year	Cohort			

$\rho_{a,t} \rightarrow \rho_{a+1,t}$	$\rho_{a,t} \rightarrow \rho_{a,t+1}$	$\rho_{a,t} \rightarrow \rho_{a+1,t+1}$			
			22	0.046	0.00
X			23	0.040	0.22
	X		23	0.044	0.08
		X	23	0.037	0.36
X	X		24	0.038	0.32
	X	X	24	0.034	0.46
X		X	24	0.037	0.35
X	X	X	25	0.037	0.35

609

610

Fig. S1 – Observed (black bullets) and predicted (lines) proportional abundance-at-age (y-axis) for ages 2-20 (x-axis) in each year 1992-2022 (panels) for rex sole in the Gulf of Alaska, showing the prediction using all data (black line) or using a leave-year-out crossvalidation design (red line). Note that the red and black lines are identical in years with no data (no black dots, e.g., 1993) because the leave-year-out crossvalidation still fits to all data for that year, and that the age-20 category includes all animals aged 20+.