Spatiotemporal models of variability

1 Introduction

Ecological and biological processes demonstrate variability in space and in time. Characterizing and understanding this variability is necessary to understand the generative processes. Sampling design tries to approach such issues by trying to balance information obtained vs costs of sampling. Strategies can range from completely random sampling in the absence of additional information, to some form of stratified random or even blocked designs that randomly chooses samples from strata constrained by factors believed to be pertinent or informative or at least one tries to abstract away as "background variability". A common one is of course areal stratification based upon some prior knowledge that is known or believed to be informative (e.g., depth, temperature or some oceanic feature), such that the variability within strata will be smaller than that between strata. The lower the variability within strata (relative to between-strata variability), the better the stratification of spatial areas (“design”) has captured local homogeneities in the process of interest (e.g., abundance of some organism); that is, each sample is thought to be more representative of the stratum that it represents.

The problem of course is that the size of these strata can shrink to unmanageable numbers as the number of informative factors increase and the kinds of processes also increase. Further, the locations of such strata can shift if they are based upon features that are not geographically fixed, such as with temperature, oxygen levels, abundance of prey or predators, light levels, etc. The fixed area approach, therefore, crudely “adjusts” for the influence of these “extraneous” factors by re-weighting of the total variance such that they can, thereafter, be ignored. These factors, are however, also highly informative and ignoring them for the sake of simplicity by “factoring them out” can lead to erroneous conclusions about the focal process(es) of interest, especially when they are dynamic.

There exist two main approaches attempting to incorporate such additional information: (1) a spatially continuous process and (2) spatially aggregated areal units. Both approaches decompose the spatial patterns into those that are associated with 1) informative factors; 2) structured spatial autocorrelation patterns; and 3) completely spatially unstructured errors. In the following, we will summarize the general background to the field, and focus upon the spatially continuous case, following closely Banerjee et al.' s (2004) exceptionally clear and thorough exposition. To assist in the context of stock assessment and general spatial and spatiotemporal modeling of potentially large areas, some of these methods have been formulated in an R-package, "**stmv**" (https://github.com/jae0/stmv). This document will also serve to document these methods. The spatially aggregated case will be treated in a separate followup document (or appended to this document in some future iteration).

2 Continuous representation

2.1 Spatial autocorrelation

To be precise, we focus upon any spatially referenced observation Y_s at locations s, measured in a coordinate space whose domain D has dimensionality d such that $\{s \in D \in \Re^d\}$. We focus upon the simple case of $d = 2$ spatial dimensions, such that for example, s = (northing, easting). The observations Y_s are assumed to be realizations of a **spatial stochastic process**, y, that is some latent unobservable but real, stochastic, generative **function** (i.e., a spatial random field) such that $y_s \rightarrow Y_s$ at $\{k = 1,\ldots,K\}$ spatial locations. The manner in which the variability of y_s changes as a function of distance, $h = \| s - s' \|$, is known as the spatial autocorrelation function. The $\| \cdot \|$ indicates a spatial norm which in $d = 2$ spatial dimensions is simply the Euclidean distance, $h = (\Delta\text{northing}^2 + \Delta\text{easting}^2)^{1/2}$.

The spatial model is expressed as a regression model of a stochastic process (Banerjee et al. 2004):

$$g(Y_s) = \boldsymbol{x}_s^T \boldsymbol{\beta} + \omega_s + \varepsilon_s, \tag{1}$$

where, the observations Y_s are realizations of some mean process $\boldsymbol{x}_s^T \boldsymbol{\beta}$ (sometimes referred to as "external drift" in the kriging literature), and a residual error process $(\omega_s + \varepsilon_s)$, operating potentially under the context of Generalized Linear Models via

the link function $g(\cdot)$. The x_s are spatially referenced predictors with associated parameters $\boldsymbol{\beta}$. The residual error process is decomposed into spatially structured ω_s and spatially unstructured ε_s components, both with mean of zero. The latter is also commonly called the "nugget" error in geostatistics and used to represent measurement and/or microscale variability/processes; it is usually assumed to have a Normal distribution and standard deviation σ_ε. The spatial error is assumed to follow a **Gaussian process** with mean 0 and a spatial covariance function $C(s,s';\theta)$ that describes form of the variance of the process as a function of distance between data, controlled by the parameters θ and spatially structured standard deviation σ_ω (see below). The full model specification is, therefore:

$$\begin{array}{rcl} g(Y_s) & = & \boldsymbol{x}_s^T\boldsymbol{\beta} + \omega_s + \varepsilon_s, \\ \varepsilon_s & \sim & \mathrm{N}(0,\sigma_\varepsilon^2), \\ \omega_s & \sim & \mathrm{GP}(\mathbf{0}, C(s,s';\theta)). \end{array} \tag{2}$$

The above is equivalent to assuming a Multivariate Normal likelihood for the observations $\mathbf{Y} = (Y_{s_1},\ldots,Y_{s_K})^T$, with mean $\boldsymbol{\mu} = g(\mathbf{Y}) = \left[x_{s_i}^T\right]_{i=1}^K\boldsymbol{\beta}$ and a covariance matrix $\Sigma = \left[C(s_i,s_j;\theta)\right]_{i,j=1}^K + \tau^2 I_K$, such that $\mathbf{Y} \sim \mathrm{MVN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$; with I_K an identity matrix of size K. It is also computationally more efficient as fewer likelihood evaluations are conducted and fast and sparse implementations of the Multivariate Normal exist.

The spatial covariance function $C(h) = C(s,s';\theta)$ expresses the tendency of observations closer together to be more similar to each other than those further away. Commonly used forms include:

$$\begin{array}{rcl} C(h)_{\text{Spherical}} & = & \begin{cases} \sigma_s^2(1 - \frac{3}{2}h/\phi + \frac{1}{2}(h/\phi)^3); & 0 < h <= \phi \\ 0; & h > \phi, \end{cases} \\ C(h)_{\text{Exponential}} & = & \sigma_s^2 e^{-h/\phi}, \\ C(h)_{\text{Gaussian}} & = & \sigma_s^2 e^{-(h/\phi)^2}, \\ C(h)_{\text{Powered exponential}} & = & \sigma_s^2 e^{-|h/\phi|^p}, \\ C(h)_{\text{Matérn}} & = & \sigma_s^2 \frac{1}{2^{\nu-1}\Gamma(\nu)}(\sqrt{2\nu}h/\phi)^\nu\ K_\nu(\sqrt{2\nu}h/\phi). \end{array} \tag{3}$$

At zero distance, $C(0) = \mathrm{Cov}(Y_s,Y_s) = \mathrm{Var}(Y_s) = \sigma_\varepsilon^2 + \sigma_s^2$ (*i.e.*, global variance), where σ_ε is the nonspatial, unstructured error, σ_s is the spatially structured error, and $\theta = \{\phi,\nu,p,\ldots\}$ are function-specific parameters including ϕ the *range* parameter. $\Gamma(\cdot)$ is the Gamma function and $K_\nu(\cdot)$ is the Bessel function of the second kind

with smoothness ν. The Matérn covariance function is frequently used in the more recent literature as the shape of this function is more flexible (Figure 2.1).

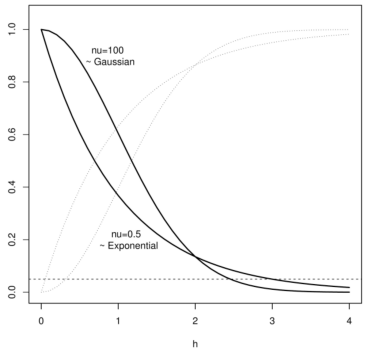

Figure 2.1: Matérn autocorrelation function, $\rho(h) = C(h)/C(0)$, the covariance function $C(h)$ scaled by the total variance $C(0)$, for two values of ν (dark lines). As ν increases ($\nu = 100$), it approaches the Gaussian curve (upper dark curve on the left side) while at smaller values ($\nu = 0.5$) the curve is exponential (lower dark curve on the left side). This flexibility has made it a popular choice in geostatistics. The associated semivariograms (scaled to unit variance) $\gamma(h)$ are shown in light stippled lines. Spatial scale is defined heuristically as the distance h at which the autocorrelation falls to 0.05 (dashed horizontal line)– in this example between 2.5 and 3 distance units, depending upon value of ν. The semivariance (also called "semivariogram") $\gamma(h)$, is more commonly used in the kriging literature, and is simply the covariance function $C(h)$ reflected on the horizontal axis of the global variance $C(0)$ such that $\gamma(h) = C(0) - C(h) = \frac{1}{2}\,\mathrm{Var}[Y_s - Y'_s] = \sigma^2_\omega[1 - \rho(h)]$.

Defining the spatial scale of a given observation or process is imperative for the development of any ecological assessment or monitoring. The **spatial autocorrelation function** is defined as the covariance function scaled by the global variance: $\rho(h) = C(h)/C(0)$. Heuristically, we define the **spatial autocorrelation scale** to be the distance at which the spatial autocorrelation decreases asymptotically to $\rho(x) \to 0.05$ (occasionally called the "practical" range in the literature). This spatial scale of an ecological process is informative in that when short-range processes dominate relative to the scale of the whole domain, such as when focusing upon

less mobile species, weakly dispersing, low currents, habitat heterogeneity, then monitoring these processes can be meaningful and fruitful in discriminating what is structuring an area of interest. If, however, long-ranging processes dominate relative to the scale of the whole domain, such as when focusing upon higher mobility species or dispersal processes/current, and stronger spatial connectivity, habitat heterogeneity, then there is a lower likelihood that monitoring such processes will provide insights to the internal structure of the area of interest.

This is perhaps clearest when spatial scale is studied in the context of specific organisms. For example, when a spatial feature (e.g., abundance distribution in space) demonstrates short characteristic spatial scales (i.e., a lot of spatial variability at smaller scales), sampling approaches must respect this and similarly operate at such shorter scales or even smaller if one is to be able to resolve the patterns and describe properly the subject of interest. Similarly, if a spatial feature is long-ranged and one wishes to resolve the patterns properly, then a sampling protocol must be similarly long-ranged to resolve the pattern. A sampling program much smaller than the characteristic spatial scale would be beneficial, but the accrued benefits relative to cost of sampling would diminish rapidly, in that time, effort and resources requirements generally increase more rapidly than any benefit (e.g., in the simplest case, if one is looking only naively at standard error as a measure of benefit, then it would increase asymptotically with increased effort with a power of $-1/2$).

3 Temporal autocorrelation

Ecological systems also exist in a temporal frame. As such, similar to the above spatial considerations, there also exists some characteristic temporal scale upon which the processes internal to an area of interest and time period of interest operate. The canonical example is how some quantity changes from one discrete-time period to another. This discrete-time notion of temporal autocorrelation is the slope parameter from a plot of a variable as a function of itself with an offset of one time unit:

$$v_{t+1} = \rho v_t + \eta_t,$$

with $\eta_t \sim N(0,\sigma_t^2)$ and a temporal (linear) autocorrelation parameter ρ. This is known as an AR(1) process, where the 1 indicates 1 unit time lag. More complex

models with moving averages and additional time-lags can also be specified. Collectively these are known as AR, ARMA, and ARIMA models. The difficulty with these autocorrelation timeseries formulations is the requirement of a complete data series without missing data.

The **cumulative periodogram** expresses the variance $f(\omega)$ as a function of temporal distance (wavelengths ω) and so is an analogue of the spatial semivariogram. It is a discrete sample estimate of the continuous concept of spectral density, $\gamma(t)$:

$$\gamma(t) = \int_{-1/2}^{1/2} e^{2\pi i\omega t} f(\omega) d\omega \quad \longleftrightarrow \quad f(\omega) = \sum_{h=-\infty}^{h=\infty} \gamma(t) e^{-2\pi i\omega t}.$$

Usefully, as the autocovariance and spectral density are Fourier transform pairs, a Fast Fourier Transform can be used to rapidly assess the power spectrum and determine the empirical form of the periodogram. This also of course in classical usage faces the same requirement of a complete data series without missing data. However, when used as the temporal analogue of an empirical spatial covariance generating **stochastic process**, it remains useful as a first order descriptive tool of the error processe. Indeed, any spatial autocorrelation function can be used to describe the empirical form of the temporal autocorrelation pattern and modeled in a manner completely analogous to the spatial case as a **temporal stochastic process**, y_t, that is, some latent, unobservable but real, stochastic, generative **function** such that $y_t \rightarrow Y_t$, where Y_t are any temporally referenced observation at some time t, measured in a coordinate space whose domain D has dimensionality 1 such that $\{t \in D \in \Re\}$ with $\{l = 1, \ldots, L\}$ temporal locations. The manner in which the variability of Y_t changes as a function of the norm (distance), $h = \| t - t' \|$, is known as the temporal autocorrelation function. The latter can take any form including the same as the spatial autocorrelation functions. The model formulation is identical to the spatial case:

$$\begin{array}{rcl} g(Y_t) & = & \boldsymbol{x}_t^T \boldsymbol{\beta} + \omega_t + \varepsilon_t, \\ \varepsilon_t & \sim & \mathrm{N}(0, \sigma_\varepsilon^2), \\ \omega_t & \sim & \mathrm{GP}(\mathbf{0}, C(t,t';\theta)). \end{array} \tag{4}$$

The covariance function, for example, when expressed as an exponential decay model controlled by time range parameter ϕ_t is:

$$C(t,t';\theta_t) = \sigma_t^2 e^{-|h|/\phi_t}.$$

Similar to the case of spatial scales, temporal scales also have a simple implication in terms of monitoring and assessment. Short time-range variations require higher sampling effort to resolve/understand the issues and vice-versa. As temporal scale is an informative metric for monitoring and assessment of an ecological process, we must be precise in its definition. The cumulative distribution permits a rapid identification of the time scale at which correlation drops to some arbitrary level. To be approximately comparable to the spatial scale, we define the **temporal autocorrelation scale** as the time difference (wavelength) at which the temporal autocorrelation function (1 - Cumulative Power Spectral Density) decreases to 5% of the total variance. If resolving short-term processes is a study's goal, then sampling must also necessarily be more frequent. However, similar to spatial scale issues, there is a point where there will be diminishing returns for any increase in the resolution of a temporal signal.

4 Spatiotemporal autocorrelation

In reality, spatial and temporal patterns coexist and co-evolve. They are correlated processes and as such a challenge to model properly. This renders the independent treatment and estimation of autocorrelation in time and space problematic. Nonetheless, new developments in computational methods are bringing such models within range of use. This is primarily due to efficient methods associated with numerical modeling of Stochastic Partial Differential Equations (SPDEs), and the use of spectral (Fourier) methods.

Again, following Banerjee et al.'s (2004) development, spatiotemporal models can be seen as a simple extension of the spatial regression model. The observations, $Y_{s,t}$ are measured in a coordinate space $\{(s,t) \in D \in \Re^d \times \Re\}$ in the domain D of dimensionality $d+1$ with $\{k = 1,\ldots,K\}$ spatial and $\{l = 1,\ldots,L\}$ temporal locations. The space-time regression model can then be specified as:

$$g(Y_{s,t}) = \boldsymbol{x}_{s,t}^T \boldsymbol{\beta}_{s,t} + \omega_{s,t} + \varepsilon_{s,t},$$

where, $\boldsymbol{x}_{s,t}^T \boldsymbol{\beta}_{s,t}$ is the mean process (or "external drift" in the kriging literature) and the error process is decomposed into a spatiotemporally structured component ω and an unstructured component ε, operating again under a generalized linear model framework, through the action of the link function $g(\cdot)$. The parameters $\boldsymbol{\beta}_{s,t}$ of the spatially and temporally referenced predictors $\boldsymbol{x}_{s,t}$ can have variable forms:

- $\boldsymbol{\beta}$ – completely fixed with no variation in time and space;
- $\boldsymbol{\beta}_{-,t}$ – temporally varying and no spatial structure;
- $\boldsymbol{\beta}_{s,-}$– spatially varying and no temporal structure;
- $\beta_{s,-} \circledcirc \beta_{-,t}$ – space and time varying independently (separably, the "⊚" indicates additive or multiplicative);
- $\boldsymbol{\beta}_{s,t}$– varying in both time and space complex (nonseparable) and potentially hierarchically (nonsimply).

The *unstructured* error is usually assumed to be a Normal *iid* error process: $\varepsilon_{s,t} \sim N(0,\sigma^2_{\varepsilon})$. However, the manner in which the *spatiotemporally structured* error should be parameterized is not straight-forward. Some common approaches include:

- $\omega_{-,t}$ – temporal effects nested in sites (temporal autocorrelation at each site, no spatial autocorrelation);
- $\omega_{s,-}$ – spatial effects nested in time (spatial autocorrelation at each time slice, no temporal autocorrelation);
- $\omega_{s,-} \circledcirc \omega_{-,t}$ – *separable* (spatial and temporal autocorrelations are independent, the "⊚" indicates additive or multiplicative) with $\omega_{-,t} \sim \text{GP}(\mathbf{0}, C(\text{t},\text{t}';\theta_t))$ and $\omega_{s,-} \sim \text{GP}(\mathbf{0}, C(\mathbf{s},\mathbf{s}';\theta_s))$;
- $\omega_{s,t}$ – non-separable (both time and space structure evolve in a nonsimple manner).

If the spatial and temporal errors are assumed to be derived from a **Gaussian Process** with mean 0 and some covariance $C(\cdot,\cdot;\theta)$, then the spatial covariance can be modeled with a flexible form such as the Matérn: $C(\Delta s)_{\text{Matérn}} = \sigma_s^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} (\sqrt{2\nu}|\Delta s|/\phi)^{\nu} K_{\nu}(\sqrt{2\nu}|\Delta s|/\phi)$. Similarly, the temporal covariance can be formulated as any reasonable autocorrelation model such as for example the exponential: $C(\Delta t)_{\text{Exponential}} = \sigma_t^2 e^{-|\Delta t|/\phi_t}$.

While conceptually coherent and elegant, the evaluation of the likelihoods in theses models requires the repeated computation of the inverse of the covariance matrix $\Sigma_{n\times n}$ of size n, an operation that scales with $O(n^3)$ operations. This has been a bottleneck to further development and use of these covariance-based methods in

large scaled problems of space and space-time. Approximations have been suggested to overcome this computational limit: modeling the spatial process ω with a lower dimensional process via kernel convolutions, moving averages, low rank splines/basis functions and predictive processes (projection of spatial process onto a smaller subset; Sølna and Switzer 1996, Wikle and Cressie 1999, Hung et al. 2004, Xu et al. 2005, Banerjee et al. 2004); approximating the spatial process as a Markov random field with Laplace and SPDE Approximations (Lindgren and Rue 2015); and approximating the likelihood of the spatial-temporal SPDE process with a spectral domain process (Sigrist et al. 2012).

In the spatiotemporal setting, separable models are almost always used for the sake of computational speed as this treats space and time independently, reducing the problems crudely from $O((KL)^3)$ to $O(K^3) + O(L^3)$ operations; where K is the number of spatial locations and L the number of time slices. In reality, however, such separable models are usually inappropriate unless the study area is homogeneous and truly first and second order constant (i.e., constant mean, variance) across both time and space, a fact that is seldom true in most ecological systems (see below).

A central assumption of all spatial and spatiotemporal models is that the form and magnitude of the autocorrelation in space and usually also in time are second order stationary (constant mean and variance). This can be forced to be the case by modeling the mean effects and operating upon a residual error that is stationary. However, in practice, there is spatial heterogeneity of variance as well which cannot be easily modeled in a simple regression context. This is notoriously the case with biology where aggregation and behavior is highly clustered and context (location and time) dependent (nonlinear).

5 Spatiotemporal models of variability (stmv)

In **stmv**, we address this **nonstationarity** and **nonseparability** of spatial and temporal structure and associated issues of computational speed and complexity by formulating a simple, operational approach to the overall spatiotemporal problem. This is done by reducing the problem into small manageable subdomains ("support") where assumptions of stationary are valid and modeling of spatiotemporal processes become computationally feasible. There is, therefore, some conceptual similarity of this approach to “geographically weighted regression” (e.g., Fotheringham et al. 2002) in that each subdomain can have their own model parameters,

$\beta_{s,t}$. However, while this latter approach permit only the model parameters $\beta_{s,t}$ to vary, **stmv** permits both the model parameters $\beta_{s,t}$ and the spatiotemporal errors $\varphi_{s,t}$ to vary.

To be more precise, in the spatiotemporal domain D, where $\{(s,t) \in D \in \Re^d \times \Re | d = 2\}$ defines the coordinate space, we define statistical nodes $\{N_{m=(1,...,M)} | m \in \Re^d\}$ in a spatial lattice (or conceivably as centroids of a mesh, though this is not yet implemented). The norm (distance) of data from each node is $h_m = ||s_m, s_Y||$. A local subdomain of a given node m is $\{S_{m=(1,...,M)} \in D | h_m < h_u\}$ or more briefly as S_m which represents all locations within some distance to the statistical node $\{h_u | C(h_u)_{\text{Matérn}} = 0.05\}$; that is, the distance at which the local spatial autocorrelation drops to a negligible value (arbitrarily taken as $p < 0.05$) and associated parameter values are "supported". The data found within the subdomain m is $\{Y_{s,t} | (s,t) \in D | h_m < h_u\}$ is notationally abbreviated as $Y_{s,t|m}$.

Operating upon all components of the regression model is computationally prohibitive. Even with very simplistic Generalized Additive Model (GAM) or Generalized Additive Mixed effects Model (GAMM) parameterizations of spatial and temporal structure, the solutions take many days on fast machines (5 GHz CPU, 64GB RAM in 2016), depending of course upon the amount of data and resolution and model complexity. As a compromise between model complexity and computational speed, **stmv** uses a global covariate model $F(\cdot) \equiv \boldsymbol{x}_{s,t}^T \boldsymbol{\beta}_{s,t}$ parameterized using a linear, generalized linear or generalized additive model. Here, $F(\cdot)$ represents some potential penalized basis splines of low order (3 knots or less seem biologically plausible when modality can be expected) of the covariate predictors and potentially some function $g(\cdot)$ that represents a link function such that the residual error in the link-space can be assumed to be Normal with mean zero and standard deviation σ_φ, the latter accounting for the residual error process $\varphi_{s,t}$:

$$\begin{aligned} g(Y_{s,t}) &= F(\cdot) + \varphi_{s,t}, \\ \varphi_{s,t} &\sim \text{Normal}(0, \sigma_\varphi^2). \end{aligned}$$

The spatiotemporal structure is decomposed from this residual error process and so the approach is in fact quite similar to “regression kriging” and (universal) “kriging with external drift” (Hengl et al. 2004).

The local spatial autocorrelation scale is derived from a rapid (coarse grained) fit of the local residuals $\varphi_{s,t|m}$ to a Matérn autocorrelation function. To be symmetrical in time, one would also need to determine temporal nodes and define appropriate temporal autocorrelation scales. In practice, temporal data are often sparse

and limiting in survey data and so data from all time periods are used, essentially amounting to a temporally averaged spatial autocorrelation. Once the approximate bounds of the subdomain (support) are estimated, the $\varphi_{s,t|m}$ are modeled as some function $f_m(\cdot) \equiv \boldsymbol{\varphi}_{s,t|m}^T \boldsymbol{\beta}_{s,t|m}$ of a Fourier series with two harmonics, one interannual and one subannual (seasonal): f_m(interannual, seasonal). In other words, a full temporal autocorrelation (covariance) model is not used but rather one that uses only a subset of the components at fixed wavelengths:

$$\begin{aligned} \varphi_{s,t|m} &= f_m(\cdot) + \zeta_{s,t|m}, \\ \zeta_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\zeta|m}). \end{aligned}$$

Data are (optionally) weighted by the inverse squared distance h_m^{-2} from the coordinates of each statistical node m to make data closer to the area of interest and prediction more influential. The temporal autocorrelation is, therefore, carried by the individual temporal processes at each spatial datum and the temporally structured error $\sigma_{t|m}$ is the variance component of the model $f_m(\cdot)$, that is, $\sigma_{t|m} = \text{Var}[\varphi_{s,t|m}] - \sigma^2_{\zeta|m}$.

The spatial autocorrelation function is parameterized as being derived from the subdomain mean Gaussian process with a Matérn covariance function with parameters $\theta_m = \{\phi_m, \nu_m\}$ and a time-varying spatially structured standard error $\sigma_{s|m}$. As the data used to estimate the spatial autocorrelation structure are often sparse, the data are augmented by temporal predictions of the residual error process at each spatial datum (and notationally designated by an asterisk). These temporally "augmented" residual processes are modeled spatially at each time slice $\varphi^*_{s,t|m}$ as the sum of a time-varying spatial **Gaussian process** $\omega_{s,t|m}$ parameterized as a Matérn spatial covariance function $\sigma^2_{s,t|m} \frac{1}{2^{\nu_{t|m}-1}\Gamma(\nu_{t|m})} (\sqrt{2\nu_{t|m}} h/\phi_{t|m})^{\nu_{t|m}} K_{\nu_{t|m}}(\sqrt{2\nu_{t|m}} h/\phi_{t|m})$ with a local spatial error $\sigma_{s,t|m}$; and a spatially and temporally unstructured error process assumed to be derived from a Normal error process with mean zero and error $\sigma_{\varepsilon|m}$:

$$\begin{aligned} \varphi^*_{s,t|m} &= \omega_{s,t|m} + \varepsilon_{s,t|m}, \\ \omega_{s,t|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_{t|m} = \{\nu_{t|m}, \phi_{t|m}, \sigma_{t|m}\})), \\ \varepsilon_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}). \end{aligned}$$

The current approach represents a practical balance between computational time and model complexity/realism. The difficulty with the current implementation of

stmv is the focus upon prediction using a mosaic of solutions in space. This complicates overall likelihood or AIC evaluation and therefore global model evaluation. At present, predictive success is the only means to evaluate utility. A fully Bayesian approach is being considered that removes the need to work with "external drift" and facilitates global model evaluation. However, this approach would increase computational time potentially to unreasonable levels, rendering it impractical in a stock assessment framework where time lines are generally limited. These and other more flexible and complex models can be defined in this framework and they will be expanded upon in future versions of this document.

6 The snow crab assessment

The snow crab estimation process interpolates a number of covariates to the same continuous spatial support of the snow crab assessment as they are surveyed using alternate survey designs. This is required to refine predictions of snow crab abundance and habitat while avoiding issues of bias due to aliasing (also known as "upscaling" and "downscaling" issues). Some of these covariates change on geological time scales relative to the timescale of the biological and ecological processes of interest and so can be considered functionally a “pure” spatial model (though of course they are not truly static). And others that are more biological in nature vary at similar or even shorter time scales and so require a temporal component. Here we detail some of these core data sources and their model assumptions in the context of the temporal autocorrelation scale of snow crab abundance in the Maritimes Region of Canada. The data handling methods and model parameterizations are encoded in the **aegis** R-package (http://github.com/jae0/aegis).

Bathymetry (depth; m) is a spatial covariate which is informative in that it determines ambient light levels, surface complexity/rugosity, hydrodynamic stability and overall environmental stability. Here, it is modeled as a Lognormal process:

$$\begin{aligned}
\log(Y_s) &= F(\text{constant offset}) + \varphi_s, \\
\varphi_s &\sim \text{Normal}(0, \sigma^2_\varphi), \\
\varphi_{s|m} &= \omega_{s|m} + \varepsilon_{s|m}, \\
\omega_{s|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_m = \{\nu_m, \phi_m, \sigma_m\})), \\
\varepsilon_{s|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}$$

As it is a pure space model, there is no need to temporally “augment” the data leaving a direct decomposition of the global residual error process $\varphi_{s|m}$ into a local

spatial process $\omega_{s|m}$ and a local unstructured error $\varepsilon_{s|m}$. An FFT-based Matérn convolution implementation is used to express the spatial process for computational speed improvements.

Similarly, substrate grain size (mm) is a pure space model which is a proxy measure of the type of substrate (mud, sand, gravel, rock, etc.) and so informative for benthic, demersal and infaunal habitat. Unfortunately, the only available data is an (over-)smoothed surface provided by Kostylev and Hannah (2007) and not the source data. Some data have been added from the snow crab surveys. It is also modeled as a Lognormal process and an FFT spatial process implementation of the Matérn covariance:

$$\begin{aligned}
\log(Y_s) &= F(\text{depth, slope, curvature}) + \varphi_s, \\
\varphi_s &\sim \text{Normal}(0, \sigma^2_\varphi), \\
\varphi_{s|m} &= \omega_{s|m} + \varepsilon_{s|m}, \\
\omega_{s|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_m = \{\nu_m, \phi_m, \sigma_m\})), \\
\varepsilon_{s|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}$$

Temperature is a fundamentally important to metabolism, growth, reproduction, predator and prey distribution and abundance, and disease incidence. Bottom temperatures, in particular, are the focus due to their relevance to benthic and demersal organisms and modeled as an hierarchical, spatiotemporal, "inseparable" spatiotemporal process. As their variations have high frequency variations, some additional complexity is required in modeling their spatiotemporal variations. Here, the temporal effects are nested in spatial subdomains S_m. The global covariate model is simply an intercept model with an identity link such that $\varphi_{s,t}$ are centered upon zero. Salinity or water density data can conceivably enter to delineate water masses and origins, however, this data does not exist at sufficient density and coverage to be informative enough to merit the additional computational load (at present). Instead, the residuals errors are modeled locally in each subdomain as a weighted timeseries with two Fourier harmonics in time (an interannual and a subannual/seasonal component). The weights are determined from the inverse squared distance from each statistical node h_m. Additional penalized thin-plate spline smooth terms for local depth and position are used to resolve local spatial trends and aliasing to third order or less (via shrinkage). Temporal predictions at each spatial datum are then used to "augment" the modeling of the spatial processes $\varphi^*_{s,t|m}$ which are treated independently for each time slice as a **Gaussian process**. The temporal autocorrelation is, therefore, carried only indirectly by the individual temporal processes centered at each spatial datum. For faster computations, a

Fast Fourier Transform (FFT) based convolution method is used to approximate the spatial Gaussian process. The model specification is, therefore:

$$
\begin{aligned}
Y_{s,t} &= F(\text{identity}) + \varphi_{s,t}, \\
\varphi_{s,t} &\sim \text{Normal}(0, \sigma^2_{\varphi}), \\
\varphi_{s,t|m} &= f_m(\text{interannual,seasonal,northing,easting,depth}) + \zeta_{s,t|m} \\
\zeta_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\zeta|m}), \\
\varphi^*_{s,t|m} &= \omega_{s,t|m} + \varepsilon_{s,t|m}, \\
\omega_{s,t|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_{t|m} = \{\nu_{t|m}, \phi_{t|m}, \sigma_{t|m}\})), \\
\varepsilon_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}
$$

The temperature data originate from a number of different sources, including the snow crab survey, groundfish survey, AZMP survey, FSRS survey, scallop survey and many other opportunistic samples maintained and kindly provided by Roger Petitpas (OSD, DFO).

Additional covariates that express the ecosystem state at a given time and location ("indicators") are informative in delineating spatiotemporal processes that are structured from those that are random. Their model formulation is similar in that they follow a similar model structure with temporal effects nested in spatial subdomains and the use of link functions in a Generalized Linear Model/Generalized Additive Model setting where the covariates used to model these indicators rely upon spatial predictions of depth and substrate grain size and the spatial derivatives of the former (slope and curvature). The spatiotemporal error process is modeled locally in each subdomain as a space-time "inseparable" model, using time-varying covariates related to bottom temperature variations and associated statistics:

$$
\begin{aligned}
g(Y_{s,t}) &= F(\text{depth, slope, curvature, substrate grainsize}) + \varphi_{s,t} \\
\varphi_{s,t} &\sim \text{Normal}(0, \sigma^2_{\varphi}), \\
\varphi_{s,t|m} &= f_m(\text{interannual,seasonal,northing,easting,depth}) + \zeta_{s,t|m} \\
\zeta_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\zeta|m}), \\
\varphi^*_{s,t|m} &= \omega_{s,t|m} + \varepsilon_{s,t|m}, \\
\omega_{s,t|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_{t|m} = \{\nu_{t|m}, \phi_{t|m}, \sigma_{t|m}\})), \\
\varepsilon_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}
$$

For the estimation of habitat preferences and the creation of species distribution maps that rely upon presence-absence data. The data Y are assumed to come from a Bernoulli binomial process with a logit link function $g(\cdot)$:

$$
\begin{aligned}
\text{logit}(Y_{s,t}) &= F(\text{depth, slope, curvature, substrate grainsize}) + \varphi_{s,t} \\
\varphi_{s,t} &\sim \text{Normal}(0, \sigma^2_{\varphi}), \\
\varphi_{s,t|m} &= f_m(\text{ecosystem indicators}) + \zeta_{s,t|m} \\
\zeta_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\zeta|m}), \\
\varphi^*_{s,t|m} &= \omega_{s,t|m} + \varepsilon_{s,t|m}, \\
\omega_{s,t|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_{t|m} = \{\nu_{t|m}, \phi_{t|m}, \sigma_{t|m}\})), \\
\varepsilon_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}
$$

For the estimation of abundance, the positive valued data Y are assumed to come from a lognormal process:

$$
\begin{aligned}
\log(Y_{s,t}) &= F(\text{depth, slope, curvature, substrate grainsize}) + \varphi_{s,t} \\
\varphi_{s,t} &\sim \text{Normal}(0, \sigma^2_{\varphi}), \\
\varphi_{s,t|m} &= f_m(\text{ecosystem indicators}) + \zeta_{s,t|m} \\
\zeta_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\zeta|m}), \\
\varphi^*_{s,t|m} &= \omega_{s,t|m} + \varepsilon_{s,t|m}, \\
\omega_{s,t|m} &\sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \theta_{t|m} = \{\nu_{t|m}, \phi_{t|m}, \sigma_{t|m}\})), \\
\varepsilon_{s,t|m} &\sim \text{Normal}(0, \sigma^2_{\varepsilon|m}).
\end{aligned}
$$

Banerjee, S., Carlin, B. P., and Gelfand, A. E.. 2004. Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester.

Hengl, T., Heuvelink, G.B.M., and Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120: 75-93.

Huang, H.-C., and Hsu, N.-J. 2004. Modeling transport effects on ground-level ozone using a non-stationary space-time model. Environmetrics 15 (3), 251–268.

Kostylev, Vladimir and Hannah, Charles. 2007. Process-driven characterization and mapping of seabed habitats. Special Paper - Geological Association of Canada. 47: 171-184.

Lindgren, F., Rue, H., and Lindstrom, 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: 423–498.

Sigrist, F., Künsch, H. R., and Stahel, W. A. 2012. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. Ann. Appl. Statist. 6: 1452–1477.

Sølna, K., and Switzer, P. .1996. Time trend estimation for a geographic region. Journal of the American Statistical Association 91: 577–589.

Wikle, C.K., and Cressie, N. 1999. A dimension-reduced approach to space-time Kalman filtering. Biometrika 86: 815–829.

Xu, K., Wikle, C. K., and N. I. Fox. 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. Journal of the American Statistical Association 100: 1133–1144.