Update

John Best

October 13, 2020

Model review

The proposed model uses a zero-inflated log-normal observation likelihood with a Poisson link function. The linear predictors for numbers density, n, and weight per group, w, are

$$\log \vec{n} = \vec{X}_n \vec{\beta}_n + \vec{A}_\omega \vec{\omega}_n + \vec{A}_\epsilon \vec{\epsilon}_n + \vec{Q}_n \vec{\lambda}_n + \vec{A}_\phi \vec{\phi}_n + \vec{A}_\psi \vec{\psi}_n$$
$$\log \vec{w} = \vec{X}_w \vec{\beta}_w + \vec{A}_\omega \vec{\omega}_w + \vec{A}_\epsilon \vec{\epsilon}_w + \vec{Q}_w \vec{\lambda}_w + \vec{A}_\phi \vec{\phi}_w + \vec{A}_\psi \vec{\psi}_w$$

Where each linear predictor may include:

- $\vec{X}_. \vec{\beta}_.$: abundance fixed effects (e.g. year)
- $\vec{A}_\omega \vec{\omega}_.$: abundance spatial effects
- $\vec{A}_\epsilon \vec{\epsilon}_.$: abundance spatiotemporal effects
- $\vec{Q}_. \vec{\lambda}_.$: catchability fixed effects (e.g. vessel/fleet)
- $\vec{A}_\phi \vec{\phi}_.$: catchability spatial effects
- $\vec{A}_\psi \vec{\psi}_.$: catchability spatiotemporal effects

The spatial and spatiotemporal effects parameter vectors take multivariate normal distributions, each of which depends on independent parameters κ and τ. These parameters can be difficult to estimate. In two dimensions and fixing the Matérn smoothness parameter ν to 1,

$$\kappa = \frac{\sqrt{8}}{\rho}$$

where ρ is the distance at which correlation drops to roughly 0.1. This parameter cannot be consistently estimated under in-fill asymptotics (adding additional observations within a fixed domain). The parameter τ on the other hand *can* be consistently estimated under in-fill asymptotics[1]. For $\nu = 1$, τ is the ratio of the marginal standard deviation σ to the correlation range,

$$\tau = 4\sqrt{2\pi}\frac{\sigma}{\rho}.$$

The intuition here is that in a fixed domain, a field with short correlation range and small marginal variance will generate similar realizations to a field with long correlation range and large marginal variance. In a model that includes all the specified spatial and spatiotemporal effects, eight κ and eight τ parameters must be estimated.

[1] Zhang. 2008. *Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics*. J Am Stat Soc. https://doi.org/10/b6ttjp

Progress

Bugs

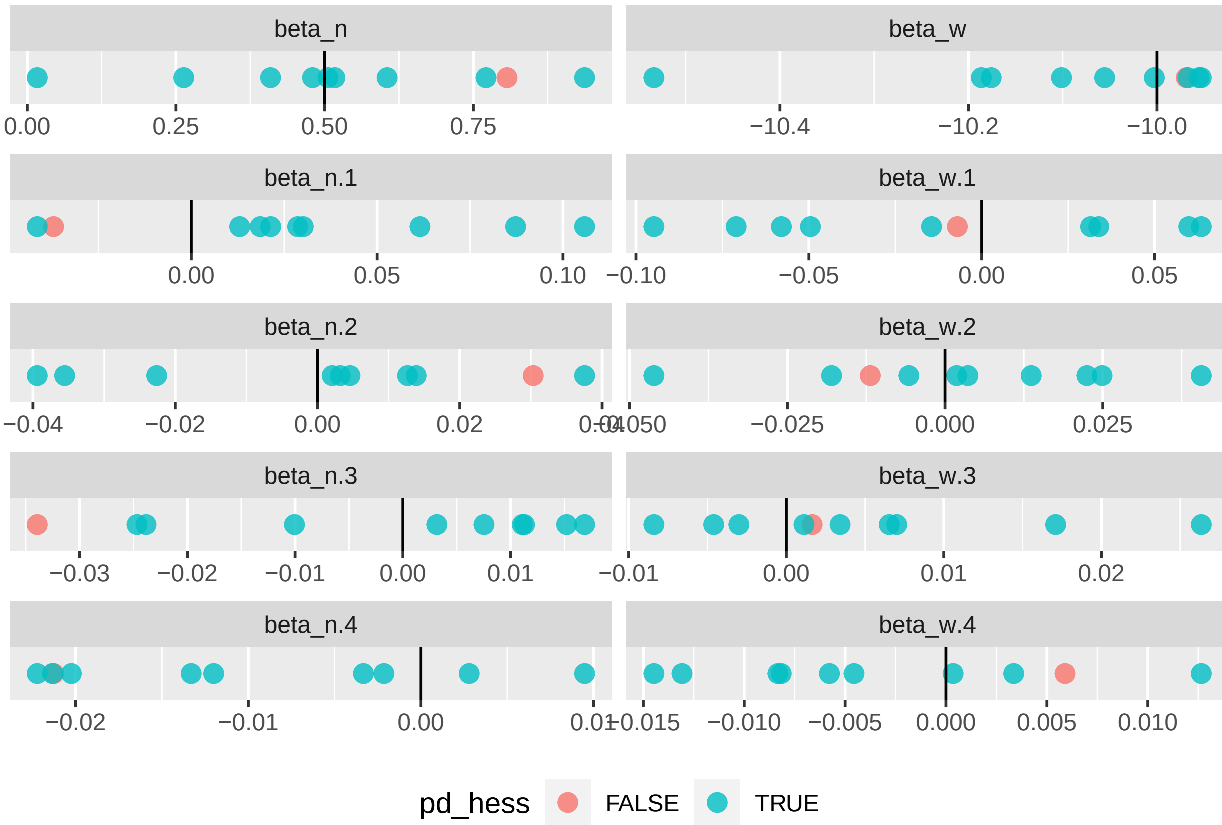
At the beginning of the summer I was struggling to fit models with spatial and spatiotemporal random fields for abundance, i.e. to be on par with the basic functionality of VAST. I discovered and fixed two major bugs in my TMB model. One covariance parameter was mis-indexed so that the parameter was shared by two random fields. Another parameter was not used by the model at all. This fix was straightforward once discovered.

The second was a misunderstanding of how the Eigen library can reshape a two-dimensional matrix into a vector. My spatiotemporal random fields were originally two-dimensional matrices, with mesh nodes as rows and years as columns. This was based on my understanding of the facilities in TMB for separable spatiotemporal random fields and made calculating the final indices of abundance simpler. It also simplified the indexing required for calculating the final index of abundance. However, I still needed a single vector in order to project the field to the observation locations and calculate the linear predictor. The method that I used to try to reshape the parameter matrix into a vector (`matrix.value()`) only returned the first value in the matrix rather than a vector. I rewrote this code to represent each spatiotemporal field as a vector.

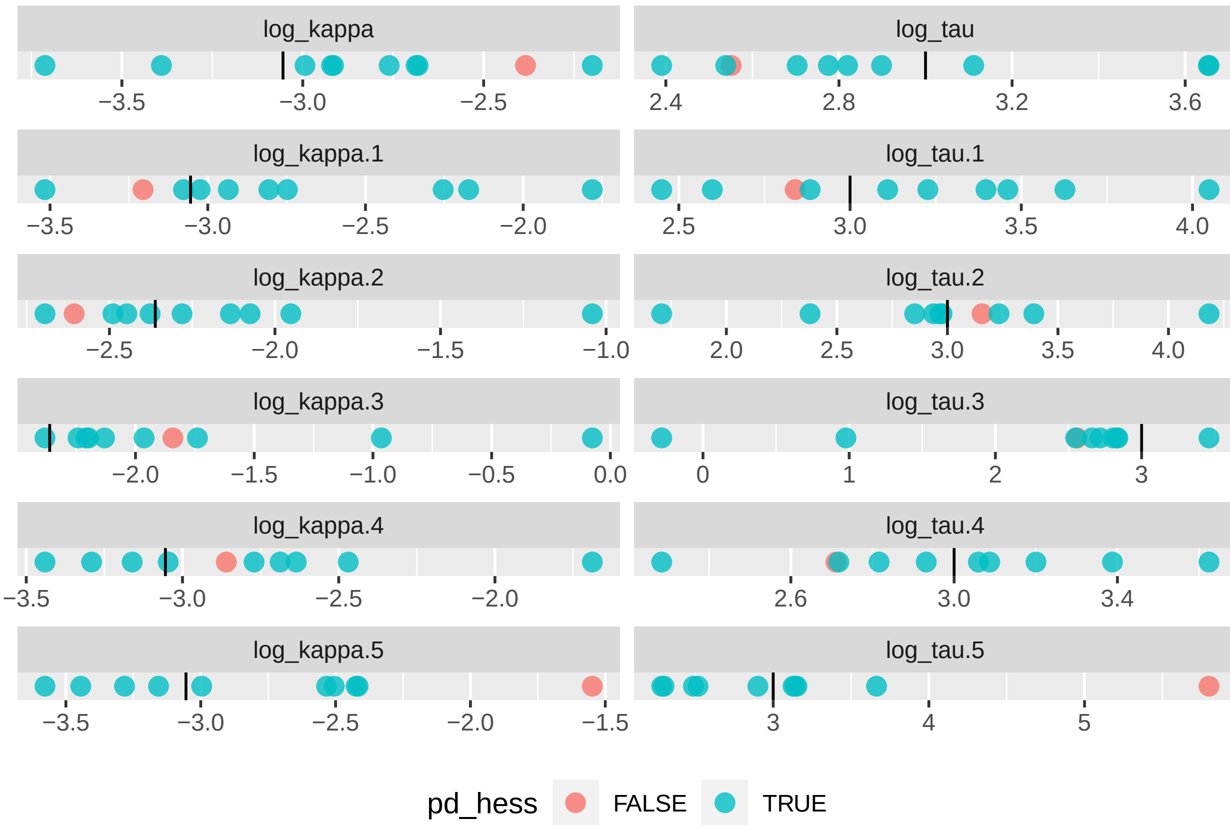
Fitting data generated by the estimation model

The TMB model includes the ability to simulate observations from the estimation model given a set of parameter values and observation locations. This simulated data was used to test the model's ability to recover generative parameters. The model did well recovering parameter values, with some important caveats. The primary issue is that the optimizer often gets stuck in areas of parameter space where the Hessian is not positive definite. This can indicate that the parameter values are at a saddle rather than a maximum, or be the result of numerical issues when calculating the eigenvalues of a poorly conditioned matrix. The optimizer consistently finds parameter values with a positive definite Hessian when spatial and spatiotemporal abundance and spatial catchability are estimated, along with their κ and τ parameters. Including spatiotemporal catchability usually results in a non-positive definite Hessian. It is also important to note that parameter values with positive definite Hessians are most often found when the initial parameter values are close to the final values. In the case of data generated from the estimation model, the generative values were used to initialize the optimization procedure. I take this as evidence that the major bugs have been resolved and the remaining challenges will mostly be related to finding appropriate starting values and ensuring the optimizer reaches a minimum with a positive definite Hessian.

To demonstrate this, ten data sets were generated from the estimation model. These were then fit using the generative values as initial values for the optimizer. Of the ten, nine reached parameter values with a positive definite Hessian. The generative values for the year effects are recovered well, without obvious bias.



Estimates of the parameters controlling the Gaussian random fields show a large amount of variation around the generative values. Pairs of rows below control spatial abundance, spatiotemporal abundance, and spatial catchability. Note that the fit with a non-positive definite Hessian includes values for `log_kappa` and `log_tau` far from the generative values.

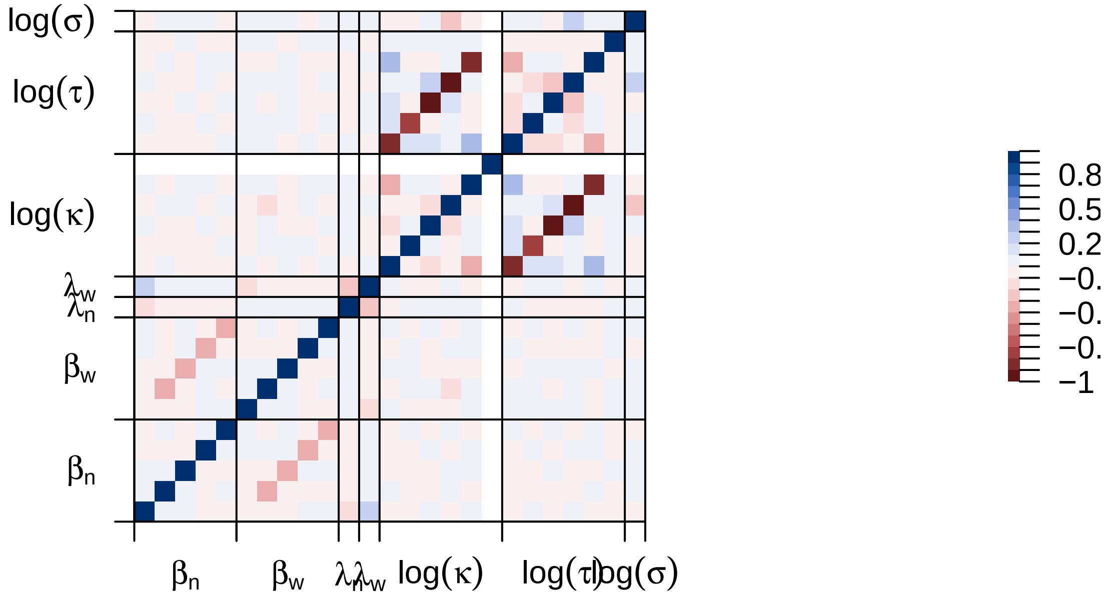


Where are the negative eigenvalues coming from?

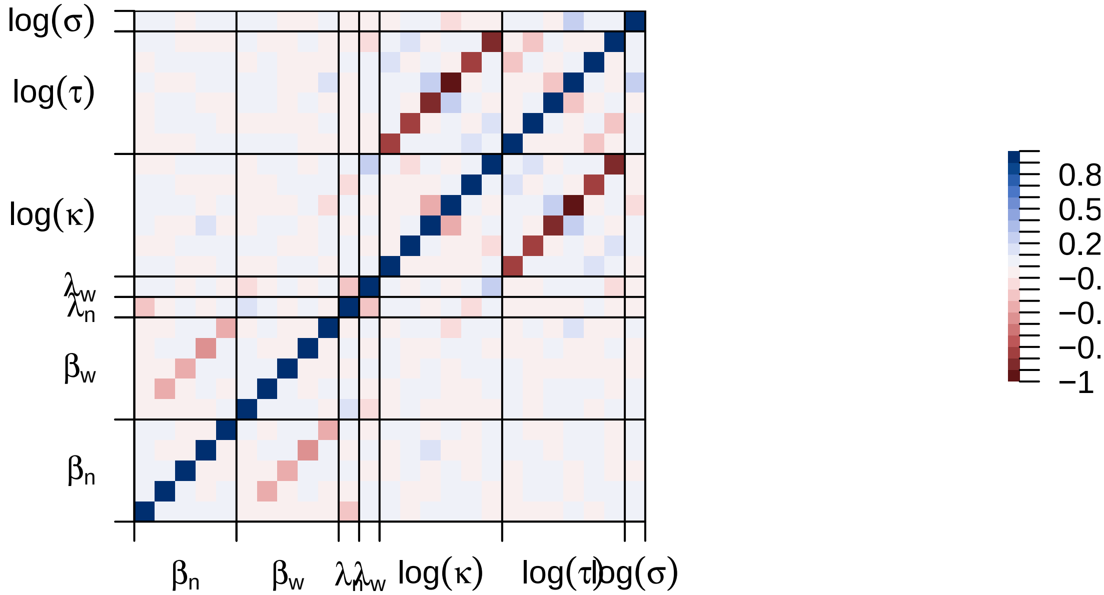
Looking more closely at the replicate with a non-positive definite Hessian, it is possible to see which parameters are most strongly associated with the small or negative eigenvalues by looking at the largest components of the eigenvectors associated with those eigenvalues. In this case one of the eigenvalues of this Hessian is negative. This eigenvalue is primarily associated with the parameters controlling the spatial catchability random field.

Parameter	Eigenvector loading
log_kappa.5	-0.9959629
log_tau.5	0.0897654
log_tau.4	0.0001548
log_tau.1	0.0001280
log_kappa.1	-0.0000889

This is consistent with the extreme values shown above for these parameters in this fit. It is also consistent with the very large off-diagonal correlations between each pair of κ and τ parameters in the parameter correlation matrix (with one missing row/column).



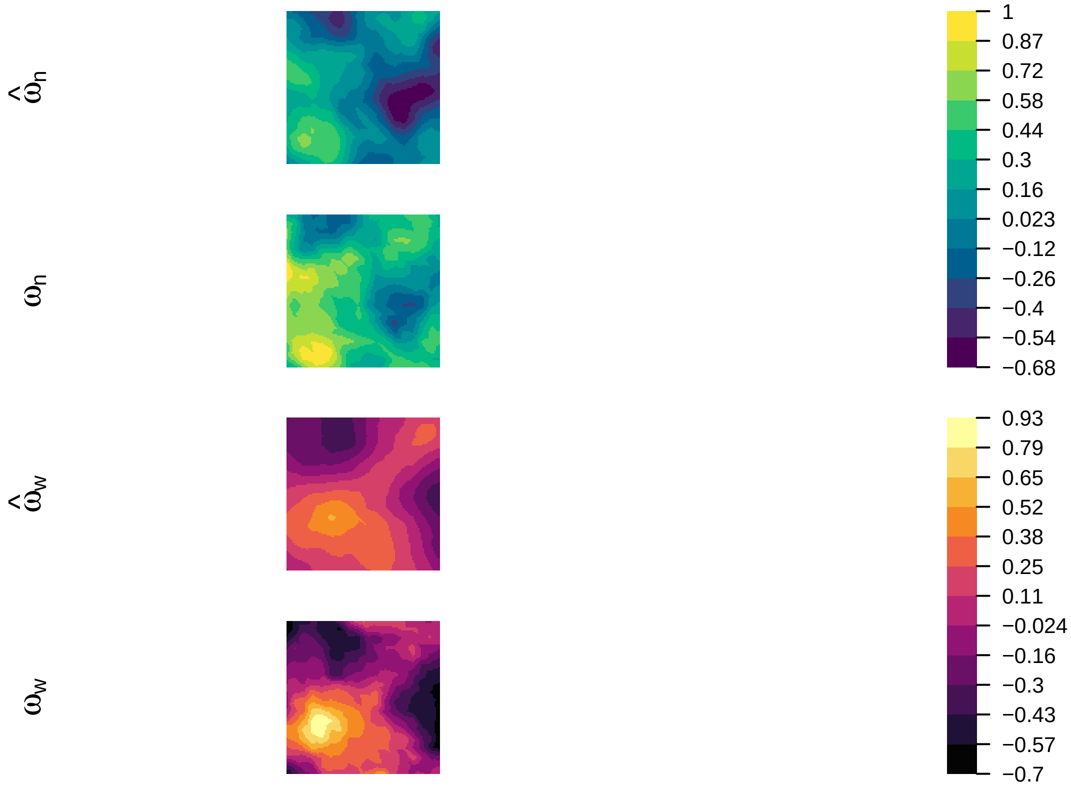
This is also present in fits where the Hessian is positive definite.



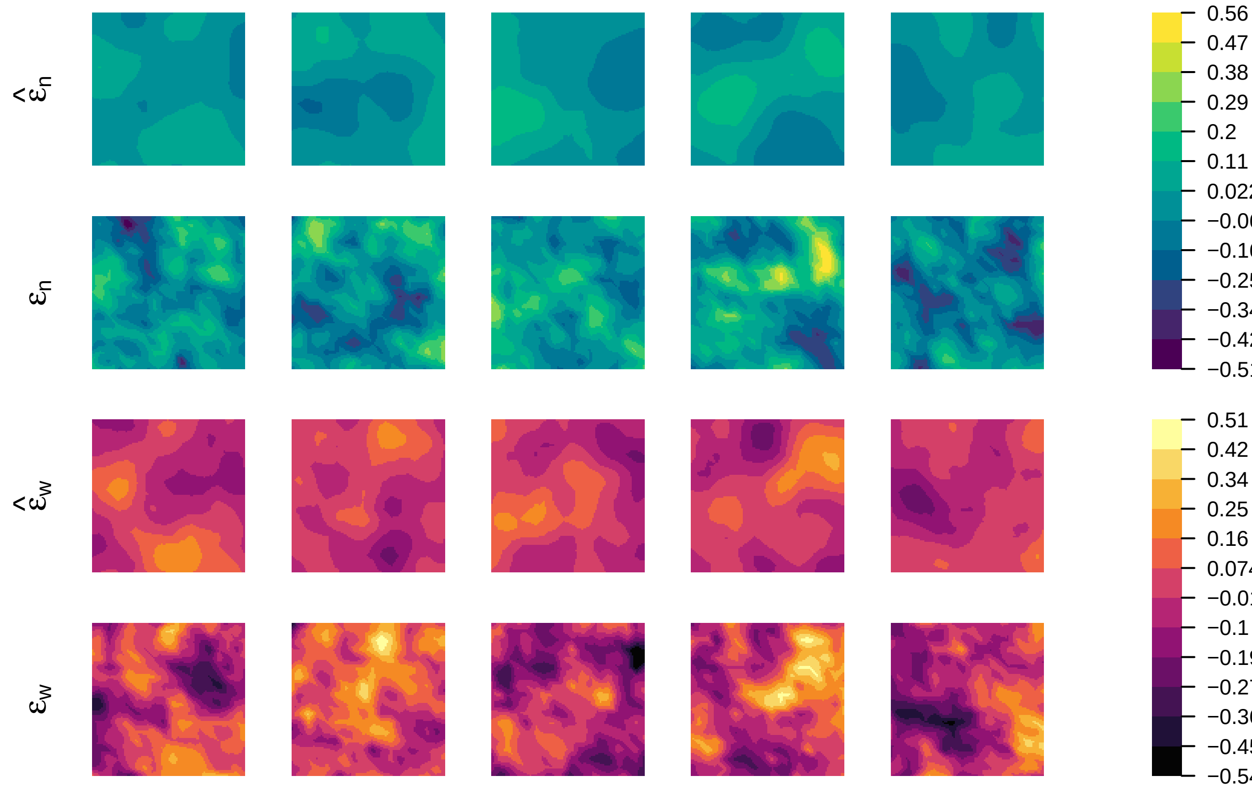
Note I have slightly reparameterized the model since these fits, with the result that the off-diagonal correlations between each pair of κ and τ are positive rather than negative, but of comparable magnitude.

Example random fields

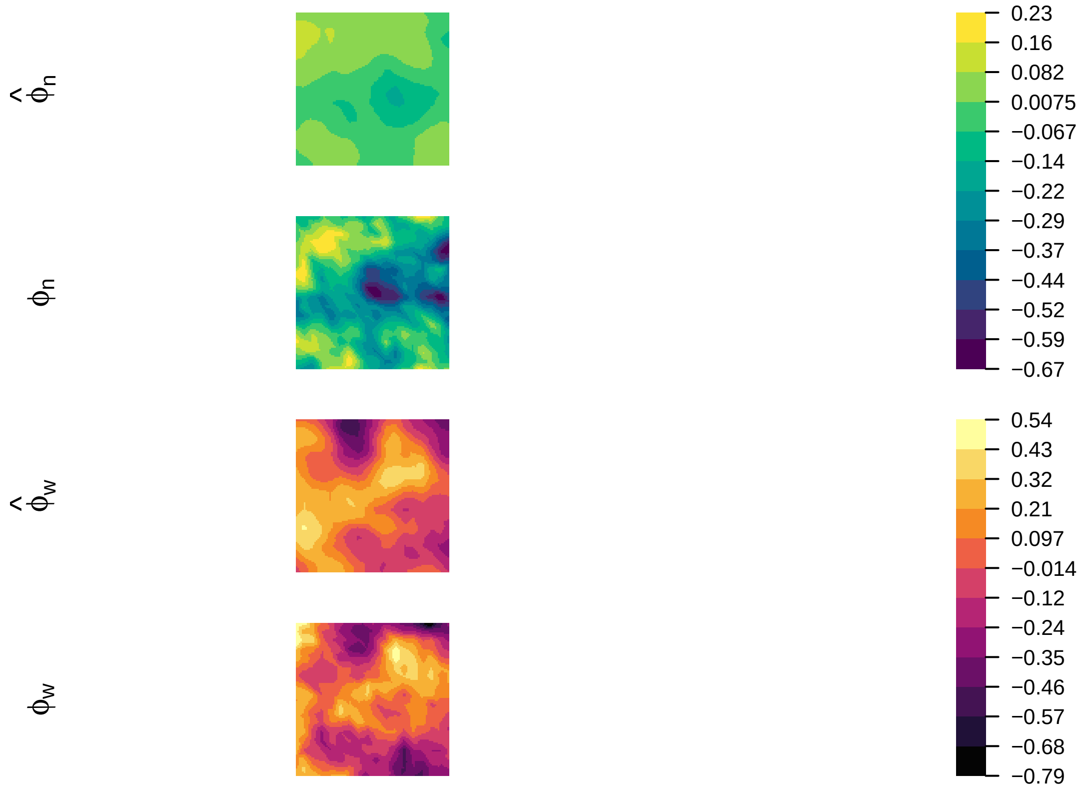
Using a fit with a positive definite Hessian, I can provide some examples of the empirical Bayes estimates of the random fields. The estimated spatial abundance process, $\vec{\tilde{\omega}}$, and the generative field $\vec{\omega}$ fitted



The estimated spatiotemporal abundance fields $\vec{\hat{\epsilon}}$ and corresponding generative field $\vec{\epsilon}$.



And the estimated spatial catchability $\hat{\vec{\phi}}$ and generative $\vec{\phi}$ fields.



The empirical Bayes estimates of the random fields appear to capture the spatial structure of the generative field.

Next steps

At this point I would like to focus on fitting models with spatial and spatiotemporal abundance random fields and spatial catchability random fields. I may consider including spatiotemporal catchability random fields later, but I would like to push forward in the direction that seems most promising. Once I consistently find optima with positive definite Hessians, I will run a large number of replicates of the operating model and get initial metrics for index improvement when spatial catchability is included. As previously discussed, the metrics will address bias, error, and coverage of the estimated indices of abundance.

Finding appropriate initial values

When fitting data simulated directly from the estimation model, the generative parameter values can be used as initial values for optimization. Obviously this is not possible with real-life data, nor is it feasible for data generated from the operating model, which uses a different parameterization. One way to deal with this is to implement a phasing scheme, where model components are added sequentially. In this case I am experimenting with estimating fixed effects, then including spatial abundance, spatiotemporal abundance, and spatial catchability. Further work will establish whether this is sufficient for fitting data from the operating model, or if additional work should be done, e.g. taking Newton steps between phases as in Jim's `TMBhelper::fit_tmb` functions.

Regularizing random field parameter estimates

Including a penalty term to keep the correlation range and marginal standard deviation under control can stabilize the estimation of these parameters. This may be necessary here where it is not in e.g. VAST due to the number of random fields being estimated and the fact that the catchability fields are estimated using a strict subset of the data used for the abundance random fields. This is likely exacerbated by

The penalized complexity framework aims to specify prior distributions for Bayesian analysis that shrink models toward simpler "base" models[2]. In the case of the Gaussian random fields used here, penalized complexity priors can be used as penalty terms for the κ and τ parameters. This prior/penalty tends to shrink the Gaussian random field toward having an infinite correlation range and zero variance, i.e. a constant[3]. These priors are intuitively parameterized, so that for some correlation range ρ_0 and tail probability α_1,

$$\Pr(\rho < \rho_0) = \alpha_1.$$

Similarly, for a marginal standard deviation σ_0 and tail probability α_2,

$$\Pr(\sigma > \sigma_0 \mid \rho) = \alpha_2.$$

The joint prior can be calculated in terms of κ and τ, even when it is parameterized in terms of ρ and σ. Attempts to use this prior as a penalty term have been promising when fitting data from the operating model. Fits result in positive definite Hessians, but more work is needed to validate the approach using data generated from the estimation model, and to choose appropriate values of ρ_0 and σ_0 for each estimated Gaussian random field. It may also be possible to apply a penalty to only the κ parameters.

[2]Simpson et al., 2017. *Penalising model component complexity: a principled, practical approach to constructing priors.* J Am Stat Assoc. https://doi.org/10.1214/16-STS576

[3]Fuglstad et al., 2018. *Constructing priors that penalize the complexity of Gaussian random fields.* J Am Stat Assoc. https://doi.org/10/ggkqqb