Using gradients in productivity to inform population structure for a managed species: a case study using Pacific cod in Alaska

# Appendix I

## Gaussian Markov random fields

Gaussian random fields (GRFs) are useful in ecology for representing naturally spatially-varying processes. These stochastic processes can be approximated using discretely indexed Gaussian Markov random fields (GMRFs) characterized by a multivariate Normal distribution with mean **μ** and a spatially-structured covariance matrix **Σ**. Any GRF with a neighbourhood structure of for a point *si* that satisfies is a GMRF. The Markovian property of the GMRF facilitates computational efficiency by ensuring that the precision matrix ***Q*** is sparse, or non-zero for neighbours and diagonal elements (Banerjee Carlin, & Gelfand 2004; Lasinio, Mastrantonio, & Pollice 2013).

Although there are many methods available to construct GMRFs, we used stochastic partial differential equations (SPDEs), or GRFs with Matérn covariance functions (Lindgren, Rue, & Lindstrӧm 2011), to reduce dense covariance matrices to sparse precision matrices. Consequently, the covariance matrix

Eq. 1

can be decomposed into the following two parts: the marginal variance,

, Eq. 2

and the Matérn spatial covariance function,

, Eq. 3

which depends on the Euclidean distance between locations *i* and *j*, (Cressie 1993; Cressie and Wikle 2011). Consequently, a GRF with a Matérn covariance function provides an exact solution to the SPDE,

, Eq. 4

which can be approximated using a finite element representation (Eq. 5). is the modified Bessel function of second kind and order λ (Abramowitz & Stegun 1972), measuring the degree of smoothness. In practice, λ is poorly identified, and, therefore, we used . The range, , is the distance at which the spatial correlation is approximately 0.1 and κ is a scaling parameter (Lindgren, Rue, & Lindstrӧm 2011). All simulations used a range value of 500 km, or ~1/6 the Longitudinal extent of the domain, and exploratory simulations verified that scaled with the true specified range (Fig. 1). is the Laplacian and τ controls the variance.

The solution to the SPDE (Eq. 4) was approximated using the finite element method, a numerical technique for finding approximate solutions to large problems, through basis functions. The spatial domain was subdivided into a set of non-intersecting triangles (i.e., “triangulation”), based on the spatial locations of the observed data and by heuristically adding triangles such that any two triangles met at most at a common edge or vertex. The total number of vertices in the triangulation *M* and the minimum interior angles of the triangles dictate tradeoffs between bias and precision of the approximation,

, Eq. 5

where is the set of basis functions and are the non-zero distributed weights. Basis functions were piecewise linear in each triangle and chosen such that was 1 at vertex *m* and 0 everywhere else (Lindgren, Rue, & Lindstrӧm 2011).

GRFs were approximated using functions available in the Integrated Nested Laplace Approximation (INLA) package (Rue *et al*. 2014; Illian, Sørbye, & Rue 2012) implemented in R. The precision matrix for the weight vector ,

, Eq. 6

is sparse and depends on the diagonal matrix of ***C***, , and the one of the sparse matrix ***G***, , where ∇ denotes a gradient (Lindgren, Rue, & Lindstrӧm 2011). INLA was used to approximate the multivariate normal distributions (i.e., provide values at each *mi* for ***C*** and ***G***). The marginal likelihood of the model parameters given the data could then be calculated using Template Model Builder (TMB; Kristensen *et al*. 2016), an R package for fitting random effect models to data. TMB uses the integrated nested Laplace approximation (Skaug & Fournier 2006) to calculate the marginal likelihood of the fixed-effect parameters integrated across all random fields, while also calculating the gradient of the marginal likelihood via automatic differentiation. Fixed effects could then be subsequently optimized using traditional tools (e.g., nlminb).

## Simulation variance

Spatial variation in the rate of population growth and spatial variation in process error were assumed to be equal in the simulation. When the former was increased, spatial locations in close proximity to each other exhibited less similar growth trajectories and when the latter was increased local populations at individual spatial locations had a higher probability of going extinct (Fig. 2).

## Simulation convergence

Simulations were replicated x times per scenario to ensure that parameter estimates reached a stable distribution and results were not overly influenced by random variation (Fig. 3). Simulations were able to replicate the true dynamics of the system and estimate variation in spatial productivity (**ω**) and spatially-explicit densities relatively accurately (Fig 4).

## Case study specification

Surveys conducted by the Alaska Fisheries Science Center (Fig. 5) occur on a yearly basis, but the entire study area (i.e., Gulf of Alaska and Aleutian Islands) was never sampled in a single year (Fig. 6). Prior to 1999, surveys in a given area were conducted once every three years, but more recently each area has been surveyed every other year.

A Constrained Refined Delaunay Triangulation “mesh” was generated for the study area based on the spatial extent and locations of sampled locations and specifications supplied to the INLA::inla.mesh.create function (Rue *et al*. 2014; Illian, Sørbye, & Rue 2012). A non-convex hull mesh was generated to avoid adding many small triangles at the boundary, and triangle sizes were controlled by specifying a minimum cutoff argument of 5% of the area of the spatial domain, which specifies the minimum distance between vertices. The resulting mesh (Fig. 7) had 191 vertices and larger triangles at the northern boundary (i.e., in the Bering Sea) where the Aleutian Islands and Gulf of Alaska surveys do not sample.

# References

Abramowitz, M., & Stegun, I.A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications, New York, NY. 470 pp.

Banerjee, S., Carlin, B.P., & Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Crc Press, Boca Raton.

Cressie, N.A.C. (1993) *Statistics for Spatial Data.* Wiley, NY.

Cressie, N., & Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data.* John Wiley & Sons, Inc., Hoboken.

Illian, J.B., Sørbye, S.H., & Rue, H. (2012) A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, **6**, 1499-1530.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. (2016) TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, **70**(5), 1-21.

Lasinio, G.J., Mastrantonio, G., & Pollice, A. (2013) Discussing the “big *n* problem”. *Statistical Methods and Applications*, **22**(1), 97-112.

Lindgren, F., Rue, H., & Lindstrӧm, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach [with discussion]. *Journal of the Royal Statistical Society B*, **73**(4), 423-498.

Rue, H., Martino, S., Lindgren, F., Simpson, D., & Riebler, A. (2014) INLA: functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation. R package version 0.0-1392038736.

Skaug, H., & Fournier, D. (2006) Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics & Data Analysis*, **51**, 699-709.

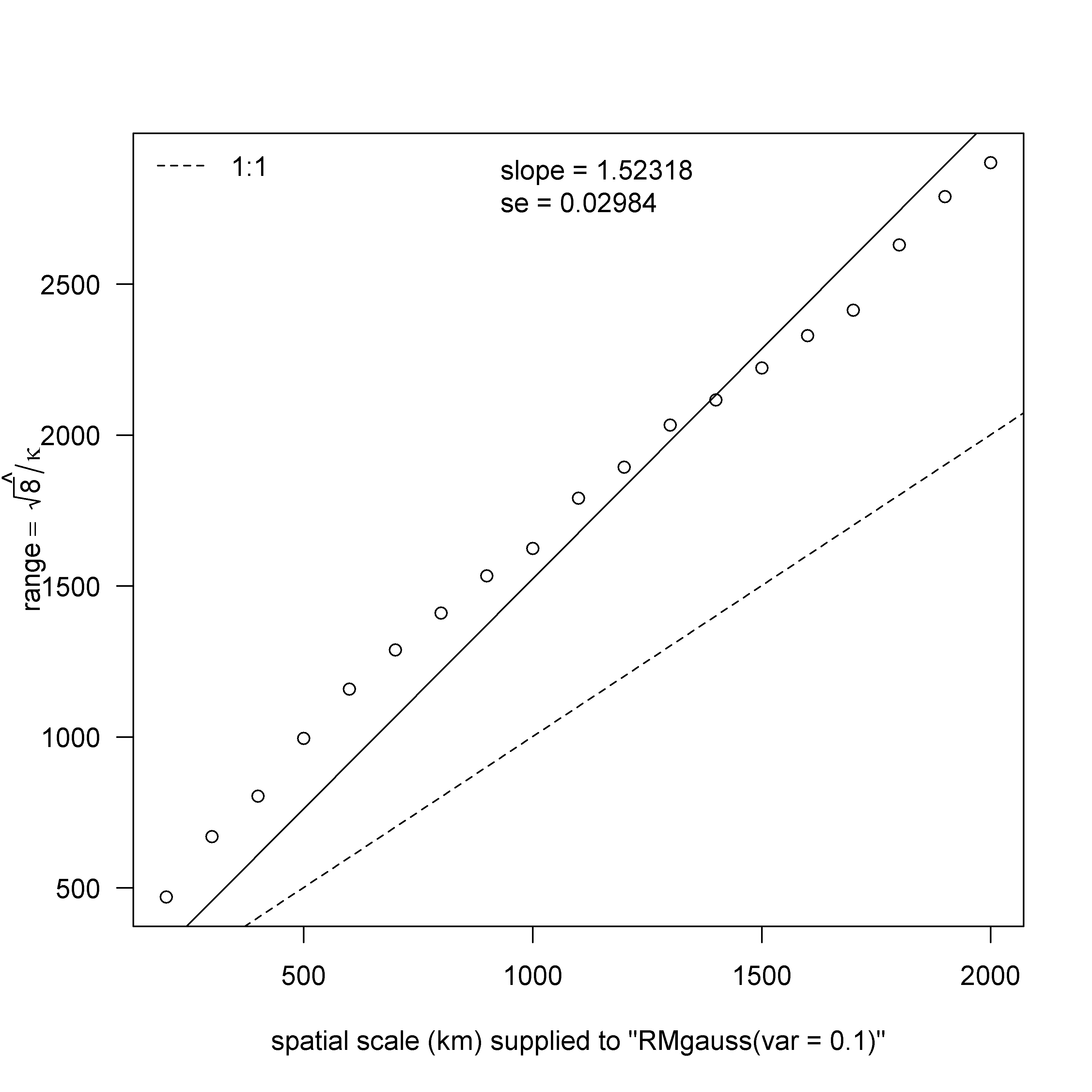


Fig. 1. Estimates of range (i.e., distance at which the spatial correlation is approximately 0.1) versus scale, an input of the RandomFields::RMgauss function, which is used to simulate stationary isotropic covariance models. Here, range is equal to , where κ is a scaling parameter. The solid black line represents the linear fit, and the slope and standard error (se) of the fit are reported in the lower right corner. The fitted line is force through the origin, and the dashed line represents a one to one ratio.

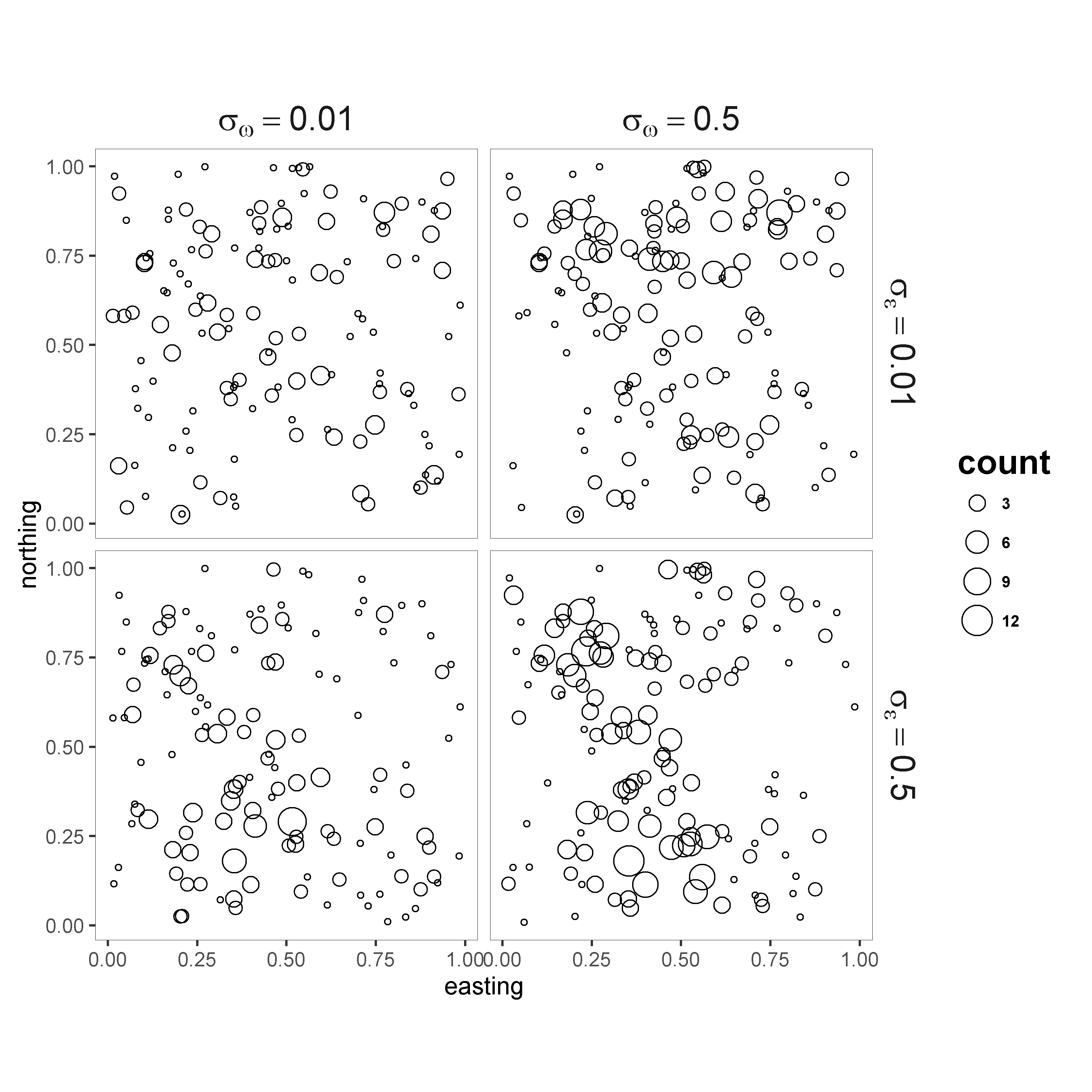


Fig. 2. Spatial variation in counts from the terminal year (i.e., year 26) of data simulated using a spatially-explicit Gompertz population dynamics model with a range of 0.25. Columns pertain to the standard deviation of spatially-explicit productivity and rows pertain to the standard deviation of spatially-explicit process error. Two hundred locations were simulated for each panel, and the size of the open circle plotted at each location is proportional to the count. Counts of zero were left off the figure.

Fig. 3. Median absolute relative error (MARE) in parameter estimates versus the number of replicates included in the analysis.

Fig. 4. True (left column) and estimated (right column) spatial variation in productivity (**ω**) and terminal year (*t* = 26) density (**n**t) for the first replicate of the baseline scenario with a single subpopulation (i.e., no gradient in mean productivity). The degree of density dependence ρ = 0.5, mean productivity α = 1, and expected log-ratio of median initial and equilibrium density ϕ = 0).

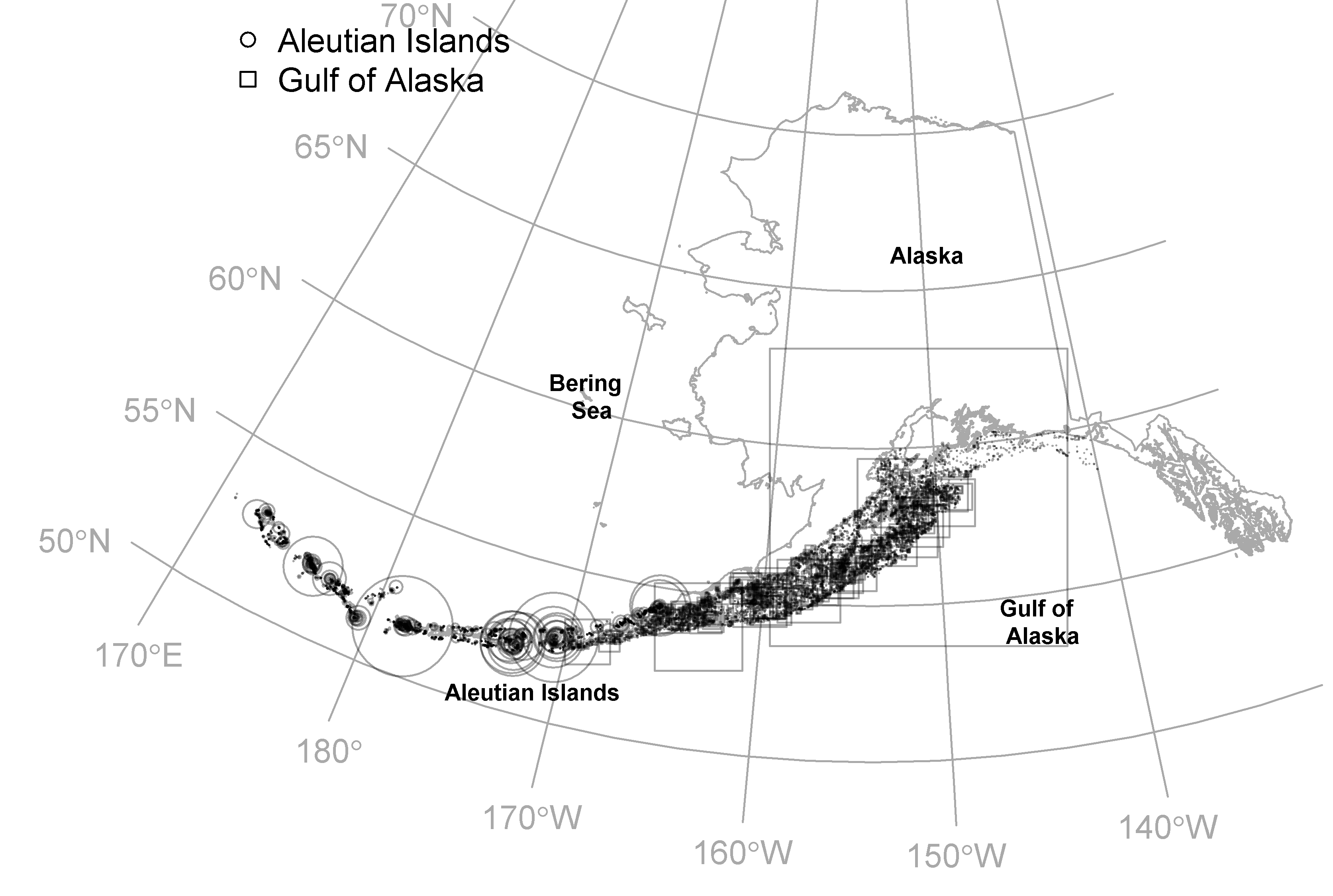


Fig. 5. Catch per unit effort (CPUE; kg ha-1) of Pacific cod for all years (1990-2016) sampled during the Aleutian Island (circles) and Gulf of Alaska (squares) bottom-trawl surveys conducted by the Alaska Fisheries Science Center of the National Oceanic and Atmospheric Administration. Point sizes are proportional to the observed CPUE and points are shown with transparency to depict areas with higher sampling effort.

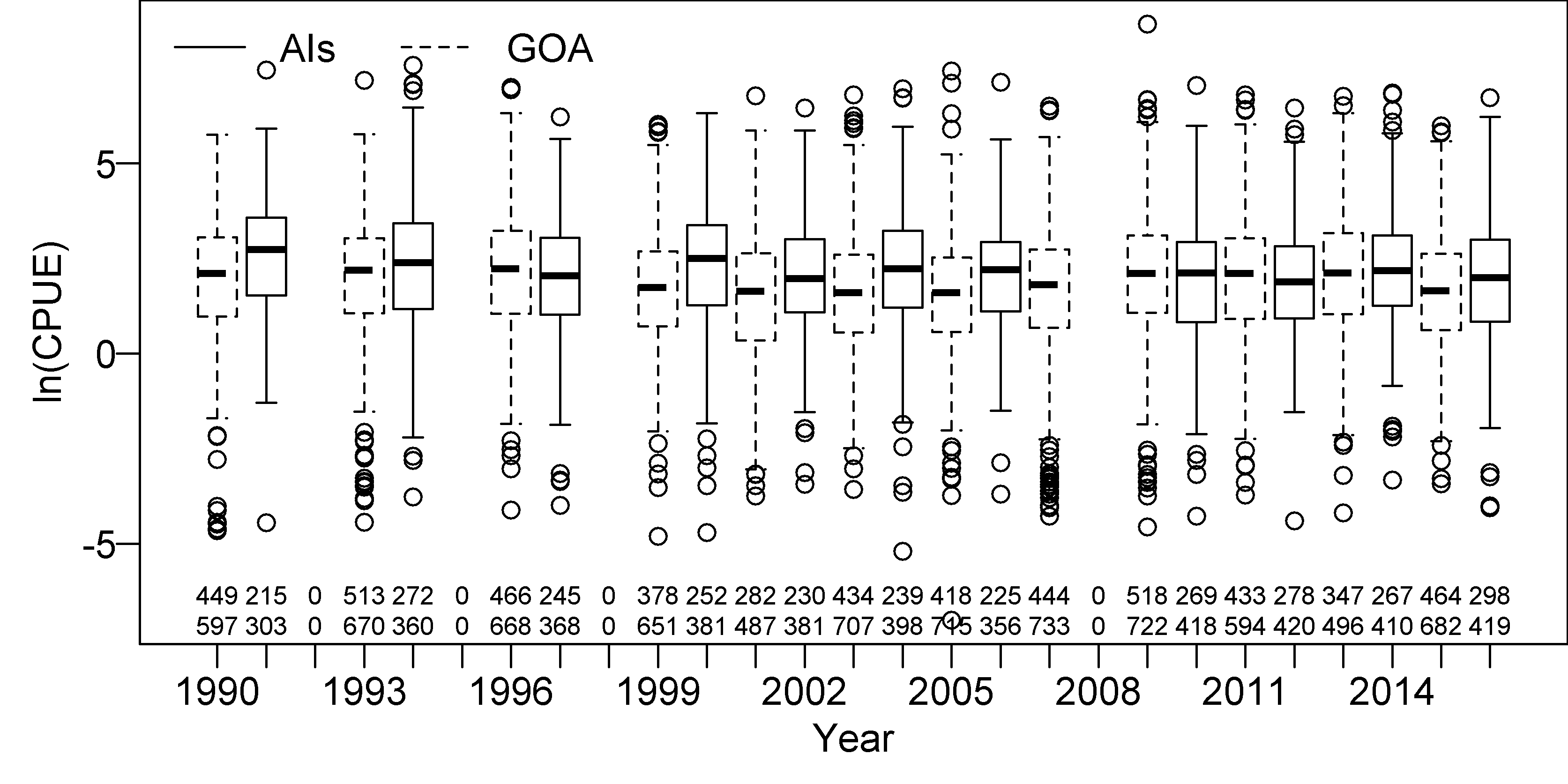


Fig. 6. Yearly variation in non-zero catch per unit effort (CPUE, kg ha-1) measurements on the natural log scale for Pacific cod from stations sampled during the Aleutian Islands (AIs; solid lines) and the Gulf of Alaska (GOA; dashed lines) bottom-trawl surveys. The number of tows per year with non-zero CPUE samples for Pacific cod (top) and total number tows per year (bottom) are provided below each box plot. Displayed ranges are the 25th and 75th percentiles (i.e., the lower and upper quartiles, respectively), and the band near the center of the boxplot is the 50th percentile (i.e., the median). Whiskers are either the minimum and maximum values or 50% and 150% of the quartiles.

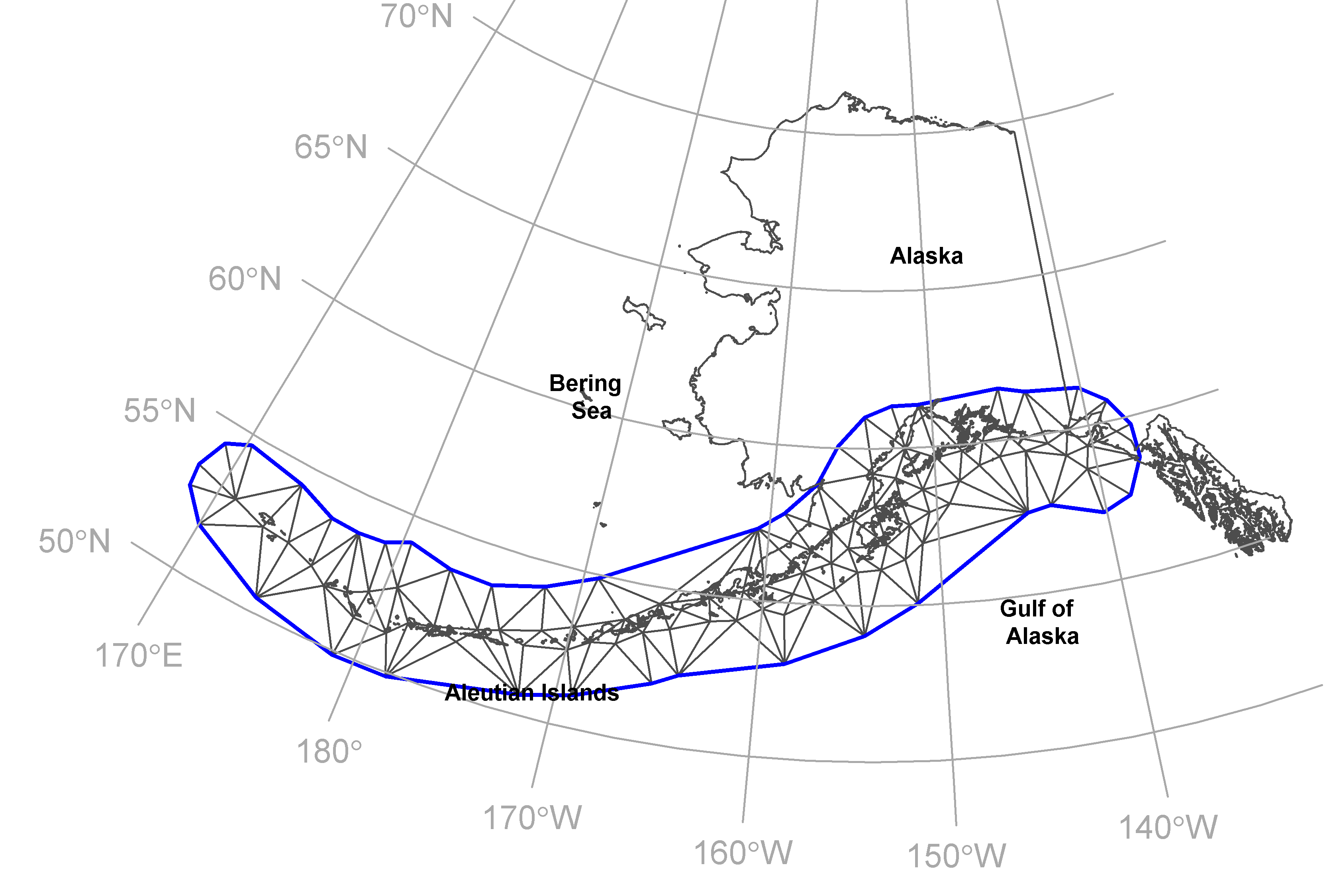


Fig. 7. Dark black lines display the triangulation mesh used to approximate a continuous process (i.e., a Gaussian random field) using a Gaussian Markov random field. The clustering algorithm was provided estimates of the spatial variation of productivity located at each vertex of the triangulation.