Fisheries and Oceans Canada | Pêches et Océans Canada

Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

**Canadian Science Advisory Secretariat (CSAS)**

**Research Document 2020/074**

**Central and Arctic Region**

# Lake Superior Pygmy Whitefish (*Prosopium coulterii*) population trends, habitat characteristics, and abundance

Adam S. van der Lee and Marten A. Koops

Fisheries and Oceans Canada
Great Lakes Laboratory for Fisheries and Aquatic Sciences
867 Lakeshore Rd.
Burlington ON L7S 1A1

December 2020

## Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

**Published by:**

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca


ISSN 1919-5044

**Correct citation for this publication:**

van der Lee, A.S., and Koops, M.A. 2020. Lake Superior Pygmy Whitefish (*Prosopium coulterii*) population trends, habitat characteristics, and abundance. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/074. iv + 18 p.

***Aussi disponible en français :***

*van der Lee, A.S., et Koops, M.A. 2020. Tendances de la population, caractéristiques de l'habitat et abondance du corégone pygmée (Prosopium coulterii) du lac Supérieur. Secr. can. de consult. sci. du MPO, Doc. de rech. 2020/074. iv +20p.*

# TABLE OF CONTENTS

## ABSTRACT

Pygmy Whitefish (PWF, *Prosopium coulterii*) in Lake Superior have been assessed as Threatened by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC). In support of a recovery potential assessment, long-term bottom trawl survey data are used to assess trends in density, identify important habitat characteristics, and estimate population size. This analysis makes use of a statistical approach, R-INLA, that can accommodate complex covariance structures in spatial-temporal data. Trends from the spatial-temporal model indicate that lake-wide PWF biomass (kg/ha) has been in decline since 2013 (~ 1 generation) and may be at the lowest mean biomass since the nearshore bottom trawl survey was expanded to include Canadian locations. Prior to 2013, PWF biomass appeared to follow periodic fluctuations suggesting the population could rebound following successful recruitment. Depth was the only important habitat characteristic, fitted as a quadratic term, in predicting occurrence or biomass. PWF were more likely to inhabit depths between ~ 50 and 110 m, preferring 75 to 90 m, with maximum biomass occurring between 80 and 95 m. Based on lake-wide spatial projections 2018 PWF biomass is estimated to be 68,707 kg (CI: 2,465 – 1,357,612) with 9,774 $km^2$ (CI: 712 – 26,014) of suitable habitat.

# INTRODUCTION

The Committee on the Status of Endangered Wildlife in Canada (COSEWIC) has assessed Pygmy Whitefish (PWF, *Prosopium coulterii*) in Lake Superior as Threatened on the basis of a 48% decline in abundance over the previous 3 generations (COSEWIC 2016). Fisheries and Oceans Canada has developed the recovery potential assessment (RPA; DFO 2007a, 2007b) as a means of proving information and science advice needed to meet the requirements of the *Species at Risk Act* (SARA). Included in a RPA is an assessment of recent population trajectory, identification of habitat characteristics and quantification of available habitat.

The United States Geological Survey (USGS) has conducted annual bottom trawl surveys in Lake Superior to assess the status and trends of the fish community (USGS 2018). These data are used to develop statistical models to determine the long-term trends in PWF biomass density in Lake Superior and identify important habitat characteristics. In addition, the statistical model is used to make whole lake projections to estimate population size and quantify available habitat.

# METHODS

## STUDY SPECIES

PWF is a small whitefish (maximum size ~ 150 mm) with perhaps the most discontinuous range of any freshwater fish in North America (COSEWIC 2016). PWF occupy Lake Superior and four additional lakes in northwestern Ontario and disjunct locations in northwest USA and western Canada.

PWF populations in Lake Superior are not well studied. Eschmeyer and Bailey (1955) described many life-history characteristics of Lake Superior PWF and identified the potential preferred depth range. The majority of PWF were captured between 18 and 90 m depths with maximum CPUE at 55 to 62 m depth (Eschmeyer and Bailey 1955). Few other studies have been devoted to PWF (Stewart et al. 2016); however, additional multi-species studies have reported similar depth preferences (Dryer 1966, Selegby and Hoff 1996).

## DATA

USGS conducts annual nearshore and offshore bottom trawl surveys in Lake Superior to assess the status and trends of the fish community (USGS 2018; Figure 1). Sampling of nearshore sites on both the American and Canadian sides of the lake began in 1989. Nearshore trawl locations extend around the perimeter of the lake. Trawling typically takes place in May and June (~ 3% of trawl took place in April, July, or August) during daylight hours. Trawls are conducted with a 12 m Yankee bottom trawl with either a chain or 6 inch rubber roller foot rope. The roller foot rope was used at sites with steeper, rockier bottoms to reduce snagging. A mean of 77 nearshore sites were sampled annually (range: 52 – 87).

Sampling of offshore sites began in 2011. Trawls are also conducted using a 12 m Yankee bottom trawl with 6 inch rubber roller foot rope. Trawling typically takes place in July with the exception of 2012 where sampling was conducted in August. A mean of 38 offshore sites were sampled annually (range: 30 – 53). The variability in the number of samples across years was due to years 2011 and 2016 where additional sites at shallower (~ 85 m) were sampled; the number of locations sampled in other years ranged from 30 – 36 sites.

Trawl collections are counted and weighed by species. Total lengths of up to 50 individuals are measured per species. Species-specific density (fish/ha) and biomass (kg/ha) were estimated

by dividing sample counts and weights by the area swept during each trawl. For each trawl depth and bottom temperature were recorded. Nearshore trawls were conducted down depth contours. The median increase in depth was 36.5 m. The mean of start and end depth was taken as a representation of depth in the statistical models. Offshore trawls were conducted on-contour.

In addition, water profile data have been collected since 2013 on temperature, specific conductivity, pH, dissolved oxygen, chlorophyll a, and photosynthetic active radiation (PAR) (USGS 2018). The data were represented by the mean of values at maximum depth at the start and end locations of the trawls. The long-term nearshore dataset (1989-2018) consisted of 2,314 trawls including 949 positive catches. The data set with all covariates (2013-2018) contained data from 653 trawls and 209 positive catches.

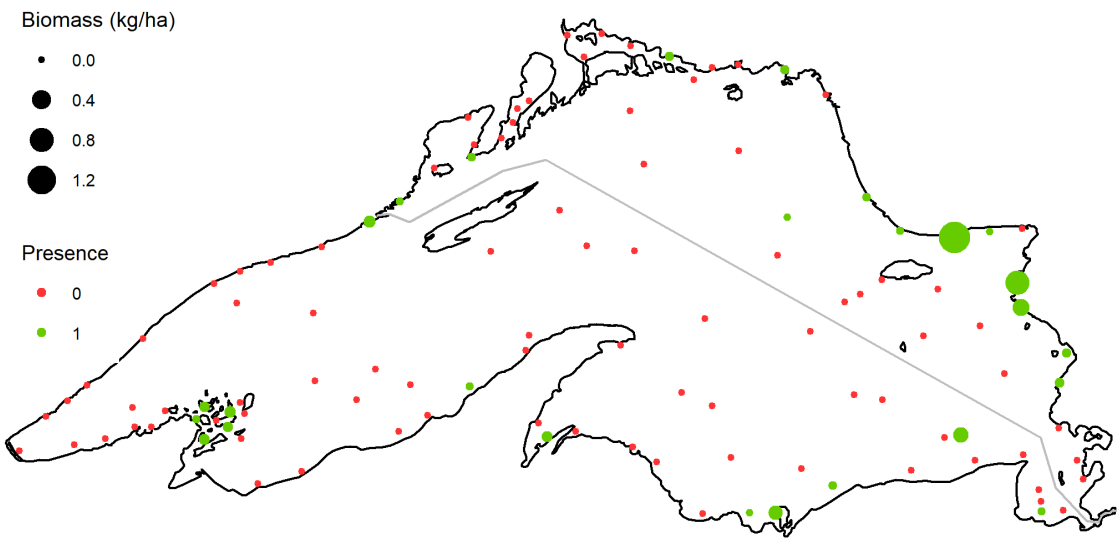


*Figure 1. Trawl location from nearshore and offshore bottom trawl survey for 2018 (USGS 2018). Red indicates trawls where PWF were not captured, green indicates trawls where PWF were captured with point size corresponding to PWF biomass (kg/ha).*

## STATISTICAL ANALYSIS

Hurdle models were used to identify the long-term trends in occurrence and lake-wide biomass and identified preferred habitat characteristics. A hurdle model is a two part model, the first part a Bernoulli process used to analyse presence-absence data, and the second a continuous process, here the gamma distribution, used to analyse positive values.

Two hurdle models were fit to the data. The first using only the nearshore trawl data (from 1989) to investigate long-term trends in the PWF biomass. The second model used both the nearshore and offshore trawl data as well as the water profile data to identify preferred habitat by including all potential covariates.

To estimate the statistical models a Bayesian inference approach, Integrated Nested Laplace Approximation (INLA; Rue et al. 2009, Lindgren and Rue 2015), was used and estimated with the R-INLA package within the statistical program R 3.5.0 (R Core Team 2018). INLA uses deterministic approximations rather than stochastic simulations to make Bayesian inferences and, as a result, is much faster than Markov chain Monte Carlo (MCMC). In addition, when combined with the stochastic partial differential equation approach (SPDE; Lindgren et al. 2011)

INLA can estimate Gaussian Markovian random fields (GMRF) to account for complex covariance structures typical in spatial-temporal data. For example, one might expect PWF biomass from two trawls close together to be more related to each other than trawls from opposite sides of the lake.

Presence-absence of PWF was modelled in the hurdle model with a logistic regression and biomass (kg/ha) was modelled with a generalized linear model (GLM) with biomass assumed to follow a gamma distribution which was related to the linear predictors with the log link function. In addition to linear covariates, the model incorporated a random effect for year to allow for non-linear trends and account for dependence between successive years and a spatial random effect using Gaussian Markov random field (GMRF). Presence-absence was modelled as:

$$y_{i,t}^{01} = Bernoulli(\pi_{i,t}),$$

$$E(y_{i,t}^{01}) = \pi_{i,t} \; and \; var(y_{i,t}^{01}) = \pi_{i,t} \times (1 - \pi_{i,t}),$$

$$logit(\pi_{i,t}) = Z_{i,t} + u_t + v_i,$$

and positive biomass was modelled as:

$$y_{i,t}^{>0} = Gamma(\mu_{i,t}, r),$$

$$E(y_{i,t}^{0>01}) = \mu_{i,t} \; and \; var(y_{i,t}^{>0}) = {\mu_{i,t}}^2/r,$$

$$log(\mu_{i,t}) = Z_{i,t} + u_t + v_i,$$

with random effects:

$$u_t = u_{t-1} + w_t, where \; w_t \sim N(0, \sigma_w^2),$$

$$v_{i,k} = \Phi v_{i,k-1} + s_{i,k}.$$

Presence-absence was modeled as a Bernoulli process with mean $\pi$ and biomass as a gamma distribution with mean $\mu$ and shape parameter $r$. $Z_{i,t}$ represents all fixed effects (e.g. intercept or intercept and covariates). The trend in time, $u_t$, was modelled as a random walk function with noise term $w_t$. The spatial effect is represented by $v_i$ which is a GMRF with mean 0 and covariance matrix $\Sigma$. The long-term model represents 30 years of data and to allow for the possibility that the spatial distribution of PWF has changed over that time period multiple spatial fields were estimated. Rather than estimate a spatial field for each year of data, which is computationally intensive, four spatial fields were estimated at evenly distribution knots, $k$, in the time series (1992, 2000, 2008, and 2016). At each knot the spatial field was estimated as the weighted average of data from the surrounding years. Correlation between spatial fields was assumed following an AR1 process with correlation $\Phi$; therefore the spatial field was a function of the previous field and the new spatial effect $s_{i,k}$. The covariate model represents only 6 years of data and therefore a single spatial field was estimated for the model implying that PWF distribution did not change over that time period.

The covariance matrix, $\Sigma$, was estimated by the Matérn correlation function using SPDE (Lindgren et al. 2011). In INLA the Matérn correlation function is defined by an estimated range parameter, *R*, which describes the distance at which the correlation between observations drops to 0.1. The GMRF is estimated over a mesh consisting of non-overlapping triangles constructed with built-in INLA functions (Figure A.1).

To fit the covariate model, Zuur et al.'s (2017) recommended procedures were followed. Covariates were examined for outliers (likely errors) and for collinearity resulting in conductivity being excluded from analysis. All covariates were scaled and standardized. Preliminarily analysis revealed likely non-linear trends in both occurrence and biomass with depth. Depth, as a result, was modelled as up to a third degree polynomial. Diffuse priors were used for all fixed and random parameters. Final model selection was done with step-wise backward selection based on ΔWAIC (Watanabe 2010) changes < 3.

# RESULTS

## LONG-TERM MODEL

The spatial model (Table 1) produced a better fit to the long-term nearshore bottom trawl data than the GLM model based on WAIC for both the presence-absence and biomass models (ΔWAIC = 1463 and 1437 respectively).

*Table 1. Spatial hurdle model results for long-term nearshore bottom trawl data. $\sigma_e$ represents the error associated with the gamma observations, $\sigma_w$ represents the error associated with the year effect, $\sigma_s$ represents the error associate with the spatial field, range represents the distance (km) between observations where the correlation is ≥ 10%, and $\Phi$ represents the correlation between spatial fields.*

| **Presence-absence model** | | | | **Biomass model** | | | |
|---|---|---|---|---|---|---|---|
| **Fixed effects** | median | LCI | UCI | **Fixed effects** | median | LCI | UCI |
| Intercept | -0.94 | -4.93 | 2.86 | Intercept | -3.62 | -5.62 | -1.68 |
| **Hyper-parameters** | median | LCI | UCI | **Hyper-parameters** | median | LCI | UCI |
| $\sigma_e$ | NA | NA | NA | $\sigma_e$ | 0.98 | 0.94 | 1.02 |
| $\sigma_w$ | 0.007 | 0.002 | 0.019 | $\sigma_w$ | 0.21 | 0.13 | 0.35 |
| $\sigma_s$ | 3.36 | 2.52 | 4.56 | $\sigma_s$ | 1.67 | 1.39 | 2.01 |
| Range | 103.9 | 69.4 | 161.5 | Range | 53.5 | 37.8 | 75.4 |
| $\Phi$ | 0.97 | 0.92 | 0.99 | $\Phi$ | 0.86 | 0.77 | 0.92 |
| WAIC | 1671 | - | - | WAIC | -3483 | - | - |

There was no evidence of an overall trend in capture probability (Figure 2) or much inter-annual fluctuations ($\sigma_w$ was very small with broad credibility intervals; Table 1). Generally, the spatially independent estimate of occurrence was ~ 30% which was about 10% less than the GLM model that did not incorporate spatial dependency. The range value indicates the distance apart that samples were at least 10% correlated. This is a representation of the distance that presence in one location could influence presence in another. The spatial model estimated that PWF presence influences occurrence up to 104 km away in all directions (Table 1, FigureA.2). The median linear distance between nearest trawl locations was ~ 14.5 km across years. The spatial fields (Figure A.2) were highly correlated (Φ = 0.97) indicating that the distribution of PWF occurrence did not change much over the time series.

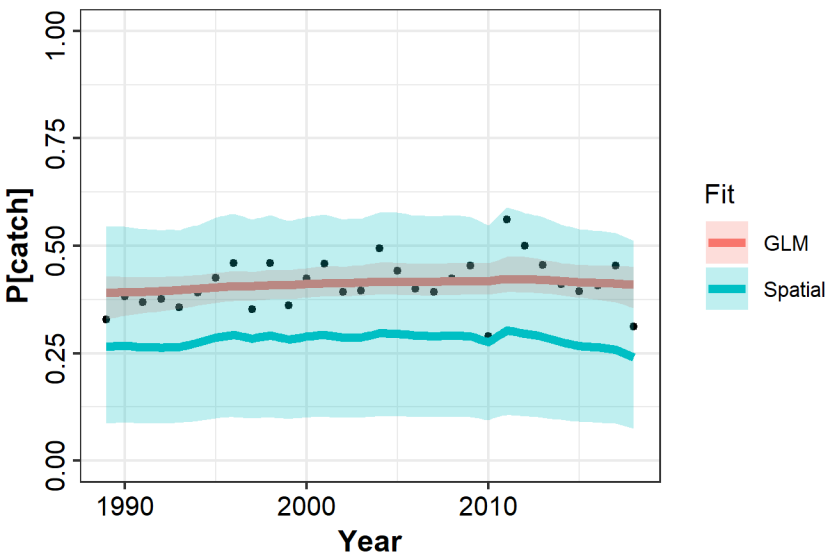


*Figure 2. Predicted lake-wide occurrence (P[catch], likelihood of capture in a trawl) through time. Relationships were fit to long-term near shore bottom trawl data using a spatial model and non-spatial GLM.*

There was a marked difference in the fitted pattern in mean biomass between the GLM and spatial models (Figure 3). The GLM estimated mean biomass to be an order of magnitude greater than the spatial model and that biomass has experienced an almost continuous decline since the 1990s. Conversely, the spatial model suggests mean biomass has followed periodic fluctuations with peak values at ~ 0.035 kg/ha occurring in the late 1997, 2006, and 2013. Biomass has declined since 2013 and in 2018 was at its lowest level in the 30 year time series.

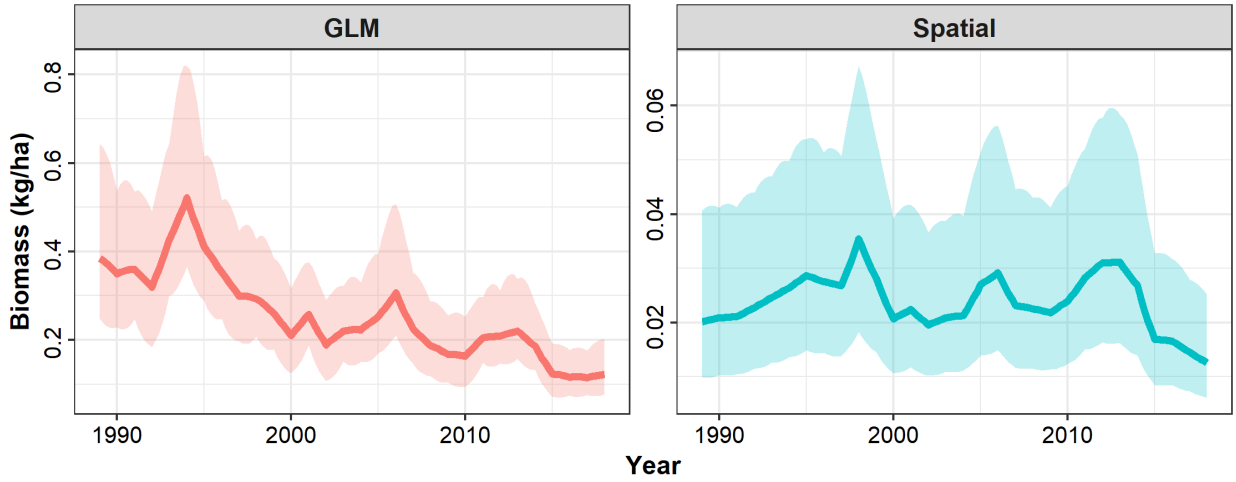


*Figure 3. Predicted lake-wide biomass (kg/ha) through time. Relationships were fit to long-term nearshore bottom trawl data using a spatial model and non-spatial GLM. NOTE: y-axis scales differ by an order of magnitude between panels.*

This difference between the model predictions is due to the impact of spatial dependency between sampling locations. The spatial model can account for this with spatially correlated random effects (Figure A.3) while the GLM treats all observation as independent. This results in the spatial model giving a better representation of lake-wide biomass while the GLM model is

biased towards the areas of highest catch. The GLM model's results could then be interpreted as showing a decrease in large catch values through time.

Over the 30 year time series there have been 17 trawls where the catch was > 3 kg/ha across six locations (Figure A.4). The largest catch in a single trawl was 13.5 kg/ha occurring in 1994 at site 450. In 2018 the largest catch was 1.6 kg/ha. The last instance of catch > 3 kg/ha was in 2014. Looking at the trends in catch at these six locations (Figure A.5) shows that two locations (450 and 466) are no longer sampled which may partially account for the decrease in frequency in large catches. The other locations seem to follow fluctuations similar to Figure 3 with most at low biomass in recent years.

## COVARIATE MODEL

Depth, as a second degree polynomial, was the only important predictor of PWF occurrence or biomass when using the spatial model (Table 2). All other predictors' credibility intervals overlapped with 0 and did not improve the model fit (based on WAIC). The spatial models were a significantly better fit to the data than the GLMs (ΔWAIC = 105 and 191). With the inclusion of offshore trawls the median estimate of range decreased in both the presence-absence and biomass models compared to the long-term models, 67 and 30 km respectively (Table 2, Appendix A.6 and A.7).

*Table 2. Final hurdle model for PWF presence-absence and biomass with covariates. $\sigma_e$ represents the error associated with the gamma observations, $\sigma_w$ represents the error associated with the year effect, $\sigma_s$ represents the error associate with the spatial field, and range represents the distance between observations where the correlation is ≥ 10%.*

| **Presence-absence model** | | | | **Biomass model** | | | |
|---|---|---|---|---|---|---|---|
| **Fixed effects** | median | LCI | UCI | **Fixed effects** | median | LCI | UCI |
| Intercept | 1.02 | -0.01 | 1.98 | Intercept | -2.30 | -2.99 | -1.64 |
| Depth | -0.52 | -1.45 | 0.29 | Depth | 0.44 | -0.41 | 1.33 |
| Depth$^2$ | -5.09 | -6.80 | -3.60 | Depth$^2$ | -3.90 | -5.26 | -2.46 |
| **Hyper-parameters** | median | LCL | UCI | **Hyper-parameters** | median | LCL | UCI |
| $\sigma_e$ | NA | NA | NA | $\sigma_e$ | 1.02 | 0.93 | 1.13 |
| $\sigma_w$ | 0.06 | 0.02 | 0.17 | $\sigma_w$ | 0.33 | 0.17 | 0.63 |
| $\sigma_s$ | 1.72 | 1.19 | 2.43 | $\sigma_s$ | 1.37 | 1.09 | 1.72 |
| Range | 66.8 | 37.1 | 124.1 | Range | 29.5 | 16.4 | 52.2 |
| WAIC | 495 | - | - | WAIC | -866 | - | - |

Depth had a significant non-linear effect on the probability of catching PWF (Figure 4). Preferred depth, occurrence probability ≥ 0.5, was predicted to be between ~ 50 and 110 m depth. Maximum median catch probability (~ 73%) occurred between 75 and 85 m. Beyond 160 m depth the model predicts a 0% catch probability. Depth also had a significant non-linear effect on PWF biomass (Figure 5). Maximum biomass was estimated to occur between ~ 80 and 95 m. Biomass was estimated to be ~ 0 kg at depths less than 20 m or over 150 m.

The effectiveness of the occurrence model was tested by examining the in-sample prediction accuracy comparing observed to fitted values; where predicted catch probability > 0.5 is taken as a prediction of positive occurrence. Overall the model correctly predicted presence-absence for 86% of trawls, 405/444 absences and 158/209 presences. The GLM was 77% accurate demonstrating that depth alone is a fairly good predictor of occurrence for PWF. The effectiveness of the biomass model was examined by calculating the correlation between fitted and observed values. The fitted values from the spatial model had a correlation value of 0.75

with observed catches. In contrast, the correlation value for the GLM biomass model was only 0.25.

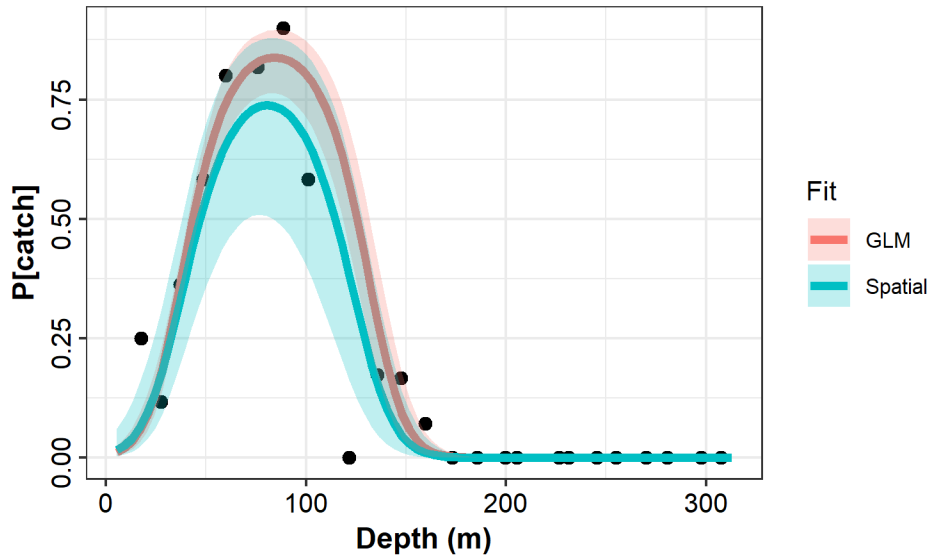


*Figure 4. The predicted relationship between depth and PWF probability of occurrence (P[catch]) using the GLM and spatial INLA models. Depth was fit as a second degree polynomial.*

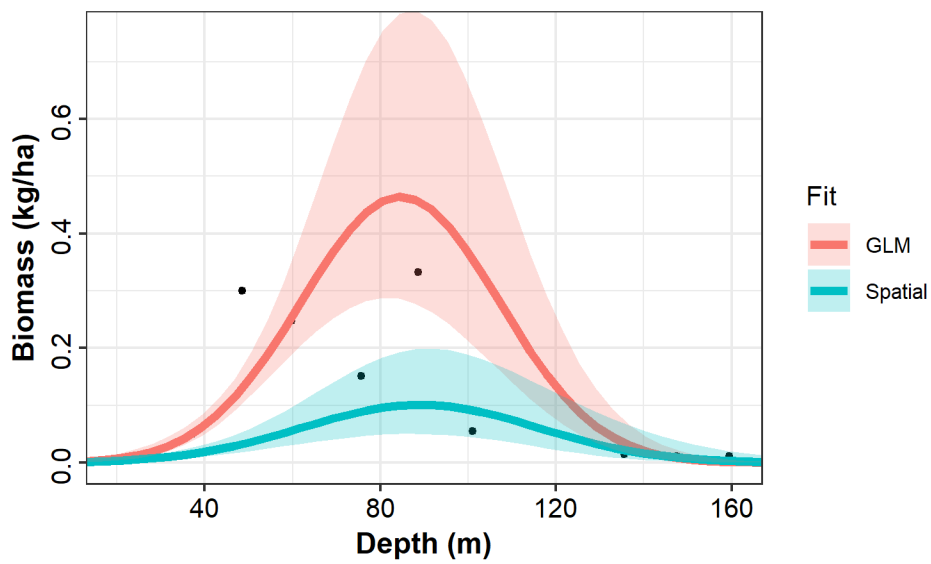


*Figure 5. The predicted relationship between depth and PWF biomass (kg/ha) using the GLM and spatial INLA models. Depth was fit as a second degree polynomial.*

The spatial models and Lake Superior bathymetry data were used to make lake-wide spatial biomass predictions for 2018 (Figure 6). This projection includes the occurrence and biomass spatial fields (Figure A.6 and A.7) and depth information to predict the median PWF biomass (kg/ha) in all areas of Lake Superior. The largest predicted PWF aggregations are around Michipicoten Island and the northeast shore of Lake Superior. Maximum predicted biomass in this area was 0.66 kg/ha which is equivalent to ~ 110 PWF/ha (based on the mean mass/fish from nearshore trawl captures of 6 g). The next largest predicted aggregation was east of the

Keweenaw Peninsula with densities up to 0.28 kg/ha. Additional areas with predicted densities > 0.1 kg/ha were east of the Slate Islands and around Grand Island.

The spatial biomass projection was used to estimate lake-wide PWF biomass. Biomass estimates were converted to units of kg/km$^2$ and 1 km$^2$ grid squares were summed. This gave a median lake-wide biomass estimate of 68,707 kg for 2018, although with very large credibility intervals (CI: 2,465 – 1,357,612) based on CI estimates from both the presence-absence and biomass models. Just over half of the population biomass, 35,944 kg, was represented by the aggregations in the northeast of the lake (Easting: 565-660, Northing: 5255-5330). The lake-wide median biomass density where PWF were predicted to occur (non-zero biomass) was 0.036 kg/ha (CI: 0.017 – 0.313). As well, the spatial projection was used to quantify available habitat. This was done with two methods; 1) summing all 1 km$^2$ grid squares predicted to contain positive biomass, and 2) summing all 1 km$^2$ grid squares predicted to have a > 50% likelihood of containing PWF. Method 1 estimated a median of 18,850 km$^2$ (CI: 1,468 – 43,319) of PWF habitat and method 2 estimated a median of 9,774 km$^2$ (CI: 712 – 26,014).

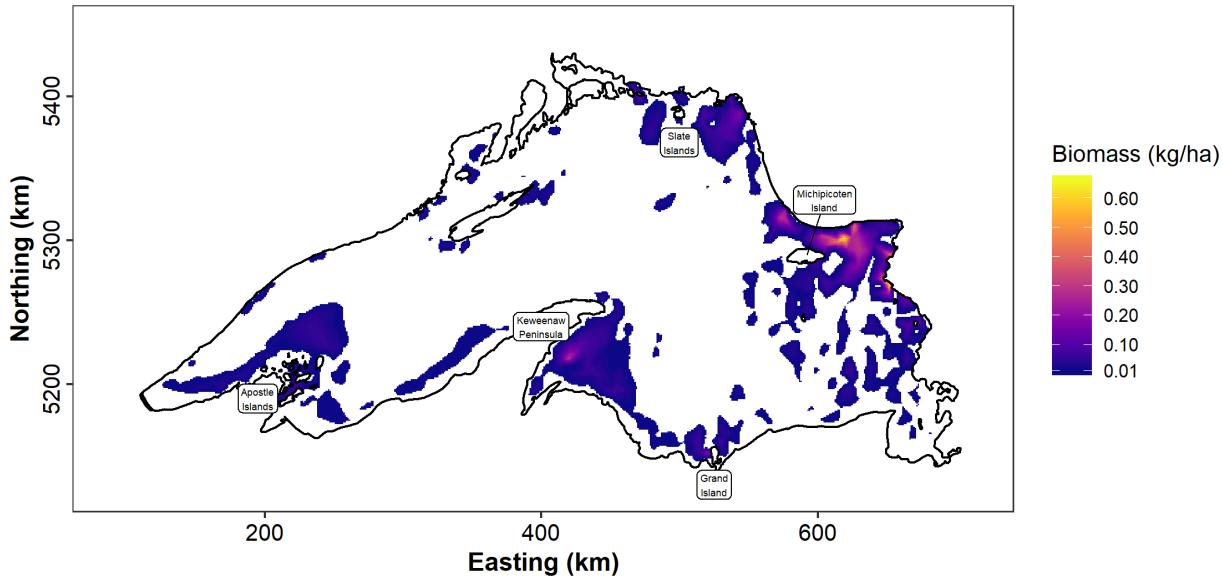


*Figure 6. Spatial prediction of PWF biomass (kg/ha) from combined spatial occurrence and biomass INLA models based on Lake Superior bathymetry. The x- and y-axes are Eastings and Northings (km) respectively and the z-axis is the average biomass (kg/ha) over ~ 1 km$^2$ grid squares.*

## DISCUSSION

The long-term bottom trawl data was analysed to determine trends in lake-wide PWF biomass, identify important habitat characteristics and quantify PWF population size. The statistical approach INLA was used, a method for making Bayesian inference using deterministic approximations rather than Markov-chain Monte Carlo allowing for more rapid estimation of model parameters. INLA allows for inclusion of complex spatial covariance structures often present in spatial-temporal data.

There was no noticeable trend in Lake Superior PWF occurrence over time (Figure 2) and the long-term trends in biomass appeared to follow periodic fluctuations (Figure 3). PWF have experienced recent (5 years; ~ 1 generation) declines in population size and may be at the lowest biomass in the past 30 years. Past trends in population trajectory would suggest the population should rebound over the next few years following strong recruitment (lake-wide

biomass peaked on 7-8 year cycles); however, if this does not occur there may be cause for concern. Other whitefishes in Lake Superior, such as Cisco (*Coregonus artedi*) and Bloater (*Coregonus hoyi*) have experienced declines in recruitment success since the 1990s (USGS 2018). Recruitment for PWF has not been monitored and the reproductive ecology of Lake Superior PWF is not well studied; however, if the requirements for successful recruitment for PWF are similar to that of other coregonids then PWF recruitment success may also be in decline and the population trajectory may not follow the past pattern.

The results of the spatial model contrast with the results of the GLM model and conclusions of the COSEWIC assessment (COSWEIC 2016). Both the GLM and COSEWIC assessment showed a decline in abundance/biomass through time. The likely reason for the difference between the GLM and spatial model is that GLM analysis is biased towards the largest catch values while the spatial model gives a better indication of lake-wide trends. Although the GLM result may not be representative of the lake-wide trend, the estimated pattern does represent a real phenomenon; that there has been a decline is the occurrence of large (> 3 kg) PWF catches in the annual bottom trawl survey since the 1990s. This result is potentially confounded due to changes in sampling effort where two of the six locations where large catches have occurred are no longer sampled; however, there is also the potential that there has been a noteworthy decline in what were the strongest PWF sub-populations.

Depth was the only covariate examined with an important impact on PWF occurrence (Table 2; Figure 4) or biomass (Figure 5). The occurrence and biomass relationships with depth both followed a non-linear parabolic relationship. PWF were likely (> 50%) to occur at depths between ~ 50 and 110 m and were most common between ~ 75 and 85 m. The impact of depth on biomass was less pronounced; however, greatest biomass was found between ~ 80 and 95 m. Previous analyses have found similar results at various locations in Lake Superior. PWF capture depths typically range from ~ 15 to 95 m (Eschmeyer and Bailey 1955, Dryer 1966, Selegby and Hoff 1996) and most commonly between 55 and 75 m (Dryer 1966). Maximum abundance has been reported to occur at depths slightly shallower than the model results, 55 to 62 m (Eschmeyer and Bailey 1955), 55 to 72 m (Dryer 1966), 60 m (Yule et al. 2008). However, these results were typically based on abundance rather than biomass.

Biomass and abundance estimates may differ due to potential shifts in depth preferences with body size or age. Younger PWF potentially occupy shallower depths (Eschmeyer and Bailey 1955, Gorman et al. 2012). Eschmeyer and Bailey (1955) and Yule et al. (2008) noted an increase in mean length of PWF with increased depth. As well, all PWF collected between ~ 18 and 26 m were young-of-year (YOY) (Eschmeyer and Bailey 1955). The proportion of YOY decreased to 7% beyond 73 m and 0% below 80 m depth (Eschmeyer and Bailey 1955). This analysis did not differentiate PWF by size or age, however, use of biomass as the response variable would down-weight the impact of young juveniles in biomass predictions. This potentially explains why the predicted depth for maximum biomass (80 – 95 m) was toward the deeper side of the preferred habitat estimate (50 – 110 m) and was deeper than previously reported depths of maximum abundance (Eschmeyer and Bailey 1955, Dryer 1966).

The depth values incorporated into model fits were based on the median depths of start and end locations of trawls. However, most near-shore trawls were conducted down depth contours. The median difference between start and end depth was 36.5 m (range: 0 – 129 m). The exact depth where PWF were captured during each trawl is not known. This adds error to the estimate of depth preference, likely broadening the interval of preferred depth, and leading to a potential overestimate of available habitat and population biomass. However, using the median depth rather than the start or end location allows the error to occur at both tails of the depth distribution preventing a consistent bias.

With depth as the only important covariate and detailed bathymetry data for Lake Superior available lake-wide, PWF biomass projections could be made and total biomass and available habitat could be estimated. Although the credibility intervals were wide, the model estimates that 2018 PWF biomass in Lake Superior was 68,707 kg occupying 18,850 $km^2$ of habitat including 9,774 $km^2$ of preferred habitat. This is the first quantification of PWF population size in Lake Superior.

An added benefit of the lake-wide biomass projections is identification of areas of the lake that are not consistently sampled (Figure 1) but may contain PWF populations; for example east of the Keweenaw Peninsula, west of the Slate Islands, or north of the Apostle Islands (Figure 6). The area east of the Keweenaw Penisula was sampled twice as a part of the offshore survey, in 2011 and 2016 producing above average catches, 0.65 and 1.42 kg/ha respectively. The model predicts maximum biomass in the vicinity of these trawls for 2018 to be 0.28 kg/ha. Trawling near the area west of the Slate Islands occurred in 2004 at a mean depth 131 m (outside of preferred depth) producing a catch of 0.053 kg/ha. Our projection estimates a maximum of 0.05 kg/ha in this area. Finally, one trawl in 2011 (not part of the nearshore or offshore surveys) sampled close to the area north the Apostle Islands predicted to have above average biomass. The mean depth of the trawl was 61.8 m but only one PWF was captured with a biomass of 0.006 kg/ha. The model projects a maximum biomass in this area of 0.06 kg/ha. Further sampling would be required to determine if a population exists here.

In conclusion, a spatial Bayesian model was used to determine long-term trends in PWF biomass, identify and quantify preferred habitat, and estimate population size. These methods account for complex error structure of the spatial-temporal data and provide an increased understanding of PWF population status in Lake Superior.

## REFERENCES CITED


COSEWIC (Committee on the Status of Endangered Wildlife in Canada). 2016. COSEWIC assessment and status report on the Pygmy Whitefish *Prosopium coulterii,* Southwestern Yukon Beringian populations, Yukon River populations, Pacific populations, Western Arctic populations, Great Lakes – Upper St. Lawrence populations, Waterton Lake populations and Saskatchewan - Nelson Rivers populations in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. iv + 69 p.

Dryer, W.R. 1966. Bathymetric distribution of fish in the Apostle Island region, Lake Superior. Trans. Am. Fish. Soc. 95(3): 248–259.Eschmeyer, P.H., and Bailey, R.M. 1955. The Pygmy Whitefish, *Coregonus coulteri*, in Lake Superior. Trans. Am. Fish. Soc. 84(1): 161–199.

Gorman, O.T., Yule, D.L., Stockwell, J.D. 2012. Habitat use by fishes of Lake Superior. I. Diel patterns of habitat use in nearshore and offshore waters of the Apostle Islands region. Aquat. Ecosyt. Health Manag. 15(3): 332–353.

Lindgren, F., and Rue, H. 2015 . Bayesian Spatial Modelling with R-INLA. J. Stat. Softw. 63(19): 1–25.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. J. Royal Stat. Soc. B 73(4): 423–498.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. Royal Stat. Soc. B 71(2) 319–392.

Selegby, J.H., and Hoff, M.H. 1996. Seasonal bathymetric distributions of 16 fishes in Lake Superior, 1958-75. National Biological Service Biological Science Report 7. 14 p.

Stewart, T.R., Ogle, D.H., Gorman, O.T., and Vinson, M.R. 2016. Age, growth, and size of Lake Superior pygmy whitefish (*Prosopium coulterii*). Am. Midl. Nat. 175(1): 24–36.

USGS (U.S. Geological Survey). 2018. Compiled reports to the Great Lakes Fishery Commission of the annual bottom trawl and acoustics surveys for 2017 [online]. 80 p. (accessed 03-10-2019).

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11: 3571–3594.

Yule, D.L., Adams, J.V., Stockwell, J.D., and Gorman, O.T. 2008. Factors affecting bottom trawl catches: implication for monitoring of fish of Lake Superior. N. Am. J. Fish. Manag. 28(1): 109–122.

Zuur, A.F., Ieno, E.N., and Saveliev, A.A. 2017. Beginner's guide to spatial, temporal, and spatial-temporal ecological data analysis with R-INLA. Highland Statistics Ltd, Newburgh, UK. 362 p.

## APPENDIX

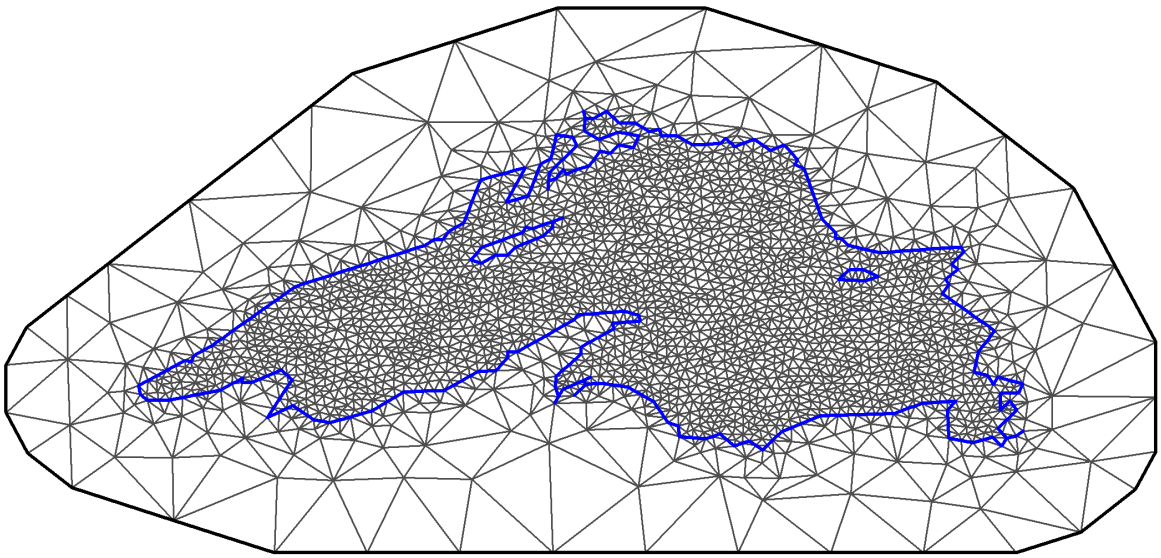

*Figure A.1. Mesh used to project the spatial field in the INLA model.*

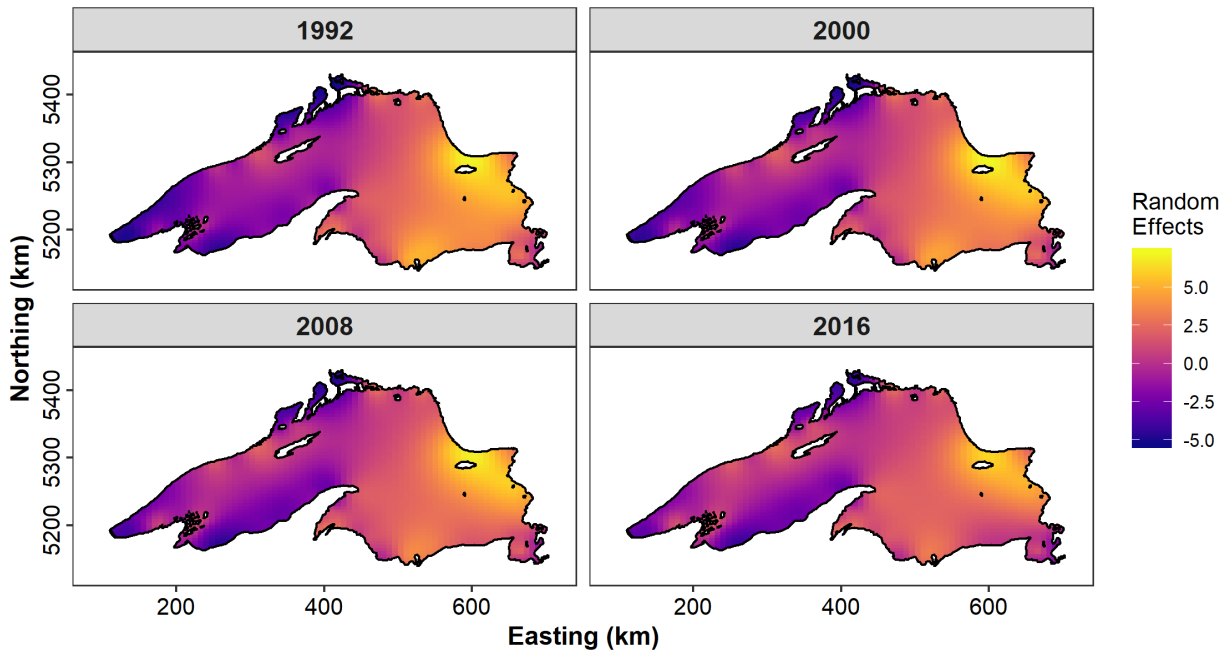


*Figure A.2. Long-term model occurrence spatial field. The x- and y-axis represent Mercator projections in km. The z-axis represents the random effects of occurrence in space on the logit scale, therefore brighter colours indicate areas of above average occurrence and darker colour indicate areas of below average occurrence.*

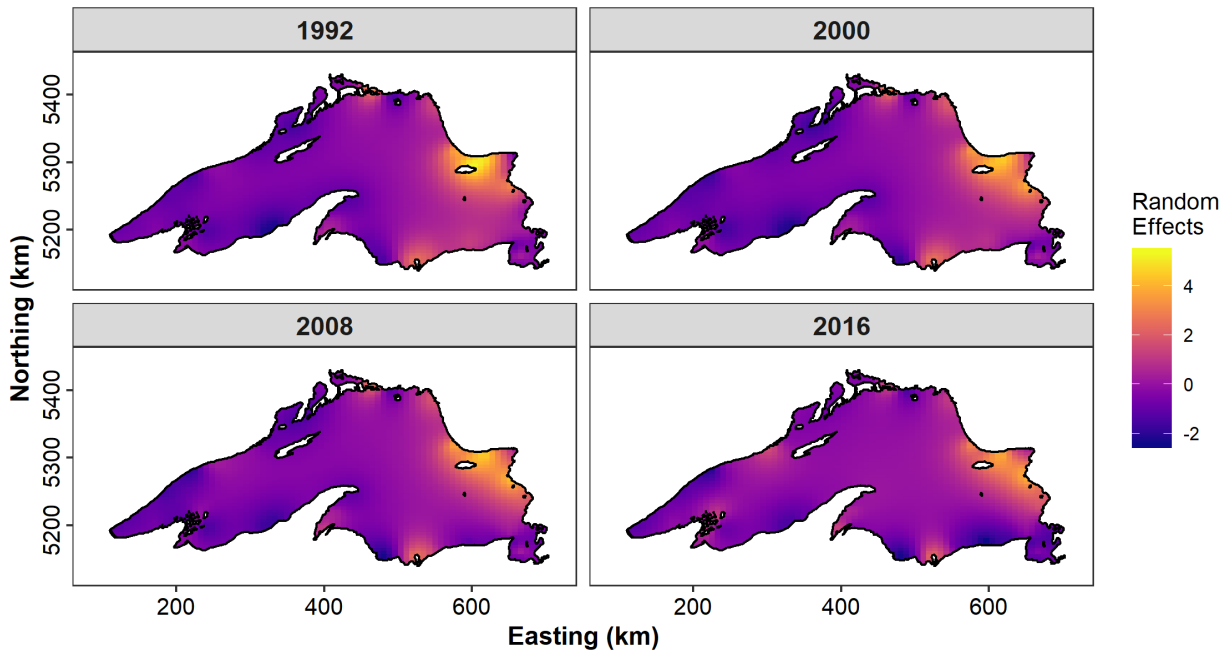


*Figure A.3. Long-term model biomass spatial field. The x- and y-axis represent Mercator projections in km. The z-axis represents the random effects of biomass in space on the $\log_e$ scale, therefore brighter colours indicate areas of above average biomass and darker colour indicate areas of below average biomass.*

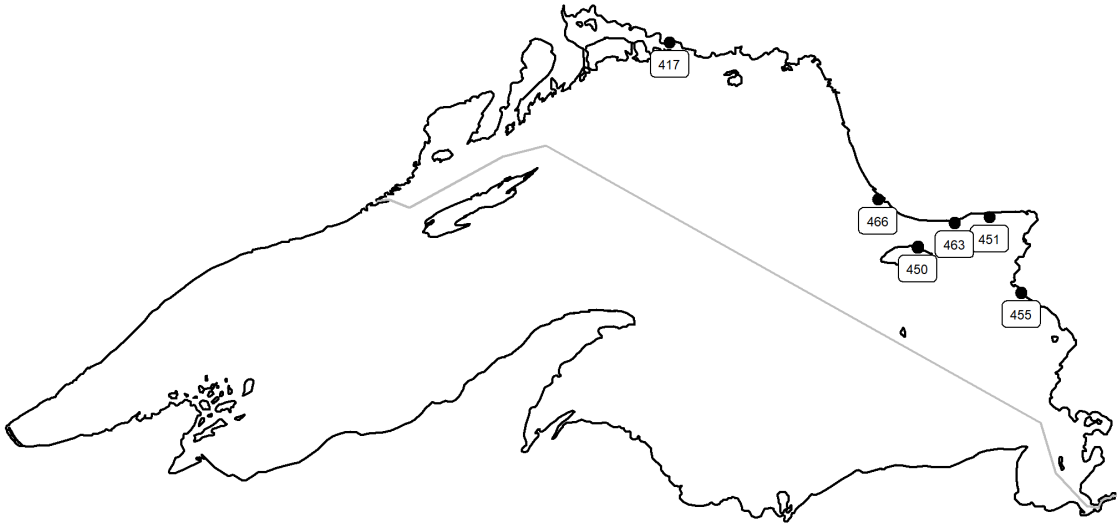


*Figure A.4. Trawl locations where catches have exceeded 3 kg/ha, 1989-2018.*

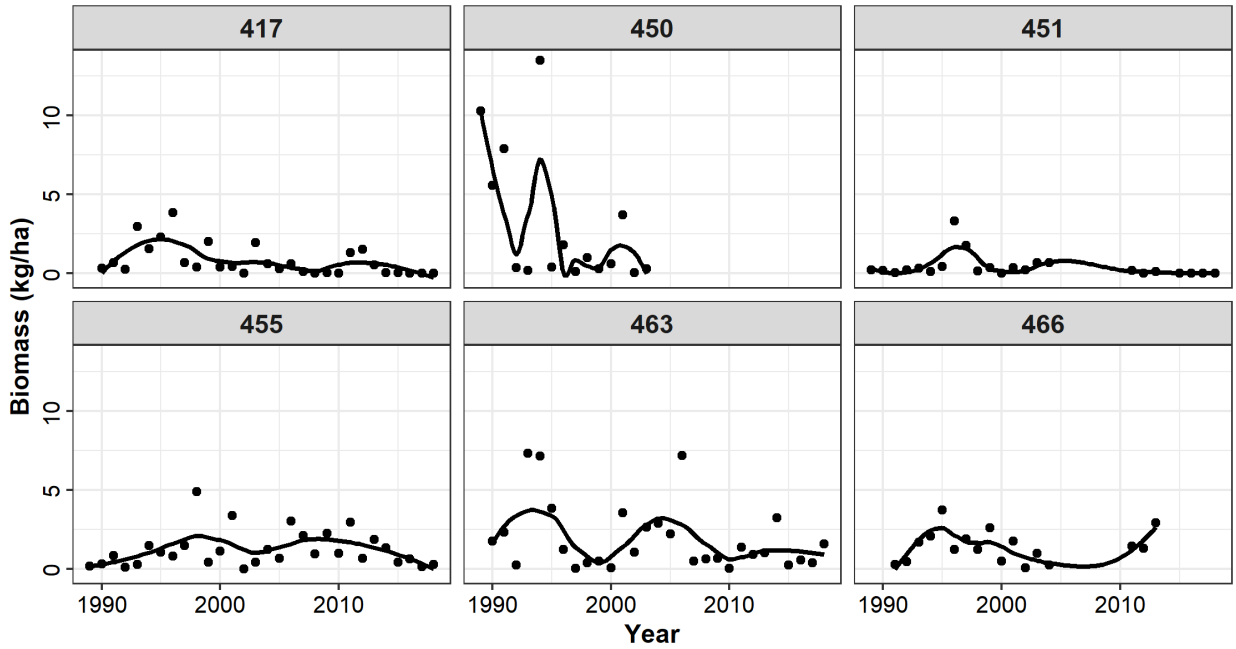


*Figure A.5. Time series of trawl captures at locations where captures have exceeded 3 kg/ha. The lines are loess curves fit to the data.*

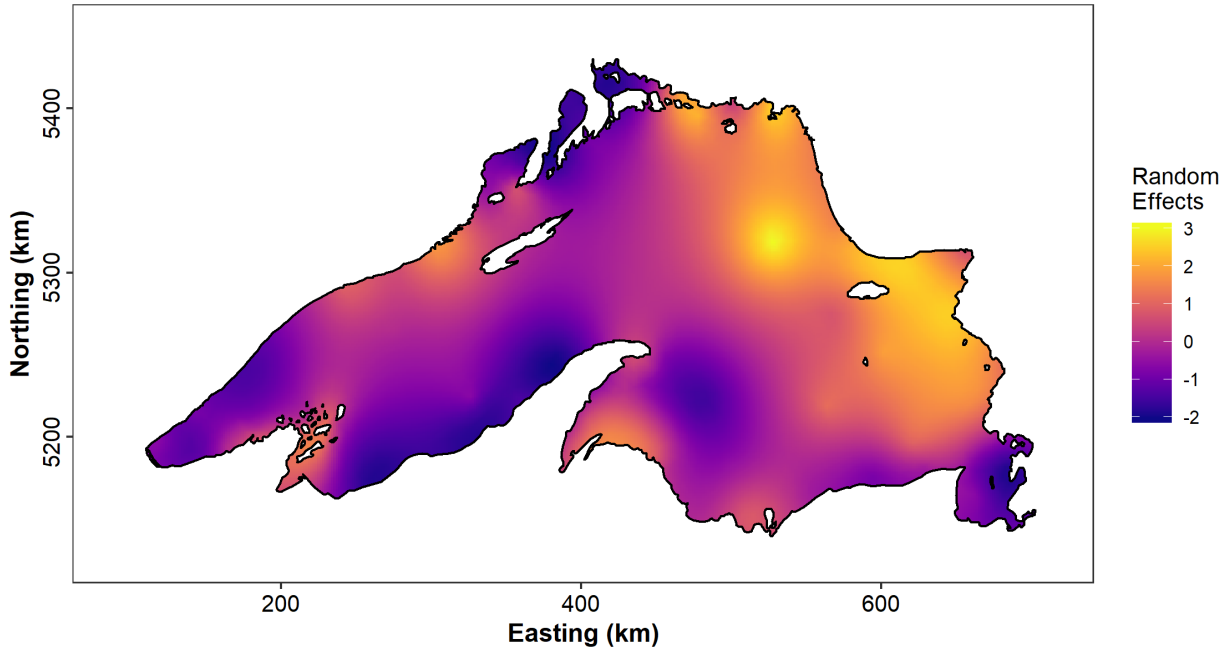


*Figure A.6. Covariate model occurrence spatial field. The x- and y-axis represent Mercator projections in km. The z-axis represents the random effects of occurrence in space on the logit scale; therefore, brighter colours indicate areas of above average occurrence and darker colour indicate areas of below average occurrence.*

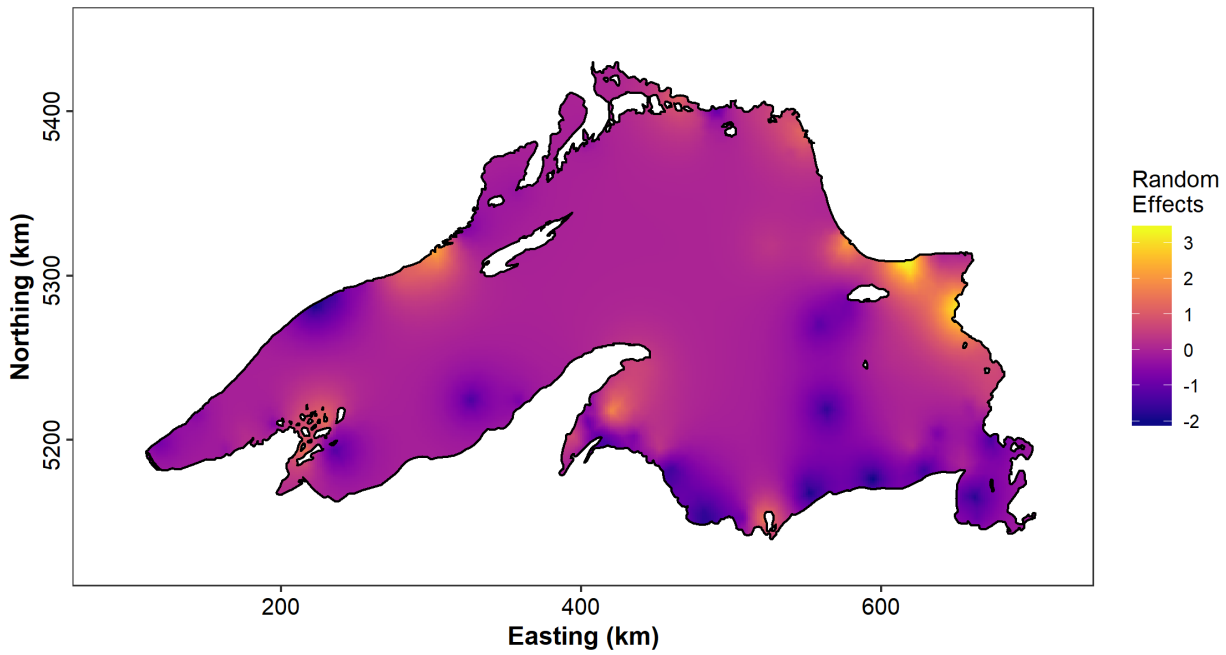


*Figure A.7. Covariate model biomass spatial field. The x- and y-axis represent Mercator projections in km. The z-axis represents the random effects of biomass in space on the $\log_e$ scale; therefore, brighter colours indicate areas of above average biomass and darker colour indicate areas of below average biomass.*