Appendix C. Test of VAST spatio-temporal analysis of SMBKC from NMFS bottom-trawl survey data

Overview

This is an example application of `VAST` for estimating single-species abundance indices specifically applied to a subset of NMFS/AFSC bottom trawl survey data. Further details can be found at the GitHub repo mainpage, wiki, and glossary. The R help files, e.g., `?Data_Fn` for explanation of data inputs, or `?Param_Fn` for explanation of parameters. VAST has involved many publications for developing individual features (see references section below).

The following loads in the main libraries.

```
library(TMB)
library(VAST)
Version <- "VAST_v2_0_0"
```

Spatial settings and model configuration

The following settings define the spatial resolution for the model, and whether to use a grid or mesh approximation as well as specific model settings.

```
Method <- "Mesh"
grid_size_km <- 25
n_x <- 50  # Number of stations
Kmeans_Config <- list(randomseed = 1, nstart = 100,
    iter.max = 1000)

FieldConfig <- c(Omega1 = 1, Epsilon1 = 1, Omega2 = 1,
    Epsilon2 = 1)
RhoConfig <- c(Beta1 = 0, Beta2 = 0, Epsilon1 = 0,
    Epsilon2 = 0)
OverdispersionConfig <- c(Vessel = 0, VesselYear = 0)
ObsModel <- c(2, 0)
Options <- c(SD_site_density = 0, SD_site_logdensity = 0,
    Calculate_Range = 1, Calculate_evenness = 0, Calculate_effective_area = 1,
    Calculate_Cov_SE = 0, Calculate_Synchrony = 0,
    Calculate_Coherence = 0)
strata.limits <- data.frame(STRATA = "All_areas")
VesselConfig <- c(Vessel = 0, VesselYear = 1)
```


This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the National Marine Fisheries Service and should not be construed to represent any agency determination or policy.

Data preparation

Data-frame for catch-rate data

The following extracts a subset of the data file downloaded from AKFIN.

```
# Read in header names
m.df <- data.frame(read.csv("male_ge90.csv", header = T,
    as.is = T))
hnames <- read.csv("hdr.csv", header = T)
names(m.df) <- names(hnames)
# Get into format for VASt
p.df <- transmute(m.df, yr = as.numeric(SURVEY_YEAR),
    loc = STRATUM_NAME, lat = as.numeric(MID_LATITUDE),
    long = as.numeric(MID_LONGITUDE), CrabN = as.numeric(CRAB_NUM),
    cpueN = as.numeric(CRAB_CPUENUM), cpueKG = as.numeric(CRAB_CPUEWGT_MT)/1000)
Data_Geostat <- p.df %>% mutate(Catch_KG = cpueKG,
    Year = yr, Vessel = "missing", AreaSwept_km2 = 1,
    Lat = lat, Lon = long, Pass = 0)

# Create a coverage of this specific are (St.
# Matthews Island)
posLL <- p.df %>% select(Lat = lat, Lon = long)
# Apply to create the extrapolation grid
Extrapolation_List <- SpatialDeltaGLMM::Prepare_Extrapolation_Data_Fn(Region = "Other",
    observations_LL = posLL, strata.limits = strata.limits)

## Derived objects for spatio-temporal estimation
Spatial_List <- SpatialDeltaGLMM::Spatial_Information_Fn(grid_size_km = grid_size_km,
    n_x = n_x, Method = Method, Lon = Data_Geostat[,
        "Lon"], Lat = Data_Geostat[, "Lat"], Extrapolation_List = Extrapolation_List,
    randomseed = Kmeans_Config[["randomseed"]], nstart = Kmeans_Config[["nstart"]],
    iter.max = Kmeans_Config[["iter.max"]], DirPath = DateFile,
    Save_Results = FALSE)

# Add knots to Data_Geostat
Data_Geostat <- cbind(Data_Geostat, knot_i = Spatial_List$knot_i)
```

Build and run model

To estimate parameters, first create a list of data-inputs used for parameter estimation. `Data_Fn` has some simple checks for buggy inputs, but also please read the help file `?Data_Fn`.

```
library(VAST)
TmbData <- Data_Fn(Version = Version, FieldConfig = FieldConfig,
    OverdispersionConfig = OverdispersionConfig, RhoConfig = RhoConfig,
    ObsModel = ObsModel, c_i = rep(0, nrow(Data_Geostat)),
    b_i = Data_Geostat[, "Catch_KG"], a_i = Data_Geostat[,
        "AreaSwept_km2"], v_i = as.numeric(Data_Geostat[,
        "Vessel"]) - 1, s_i = Data_Geostat[, "knot_i"] -
        1, t_i = Data_Geostat[, "Year"], a_xl = Spatial_List$a_xl,
    MeshList = Spatial_List$MeshList, GridList = Spatial_List$GridList,
```

```
    Method = Spatial_List$Method, Options = Options)

# We then build the TMB object.
TmbList <- Build_TMB_Fn(TmbData = TmbData, RunDir = DateFile,
    Version = Version, RhoConfig = RhoConfig, loc_x = Spatial_List$loc_x,
    Method = Method)
Obj <- TmbList[["Obj"]]

## Estimate fixed effects and predict random effects
## Next, we use a gradient-based nonlinear minimizer
## to identify maximum likelihood estimates for
## fixed-effects
Opt <- TMBhelper::Optimize(obj = Obj, lower = TmbList[["Lower"]],
    upper = TmbList[["Upper"]], getsd = TRUE, savedir = DateFile,
    bias.correct = FALSE)

# Store output
Report <- Obj$report()
```

Diagnostic plots

```
SpatialDeltaGLMM::Plot_data_and_knots(Extrapolation_List = Extrapolation_List,
    Spatial_List = Spatial_List, Data_Geostat = Data_Geostat,
    PlotDir = DateFile)
Region = "Other"
MapDetails_List <- SpatialDeltaGLMM::MapDetails_Fn(Region = Region,
    NN_Extrap = Spatial_List$PolygonList$NN_Extrap,
    Extrapolation_List = Extrapolation_List)
# Decide which years to plot
Year_Set <- seq(min(Data_Geostat[, "Year"]), max(Data_Geostat[,
    "Year"]))
Years2Include <- which(Year_Set %in% sort(unique(Data_Geostat[,
    "Year"])))
```

Convergence

Diagnostics generated during parameter estimation can confirm that parameter estimates are away from upper or lower bounds and that the final gradient for each fixed-effect is close to zero. For explanation of parameters, please see references (and specifically `?Data_Fn` in R).

[1] “”

Encounter-probability component

One can check to ensure that observed encounter frequencies for either low or high probability samples are within the 95% predictive interval for predicted encounter probability (Figure . Diagnostics for positive-catch-rate component was evaluated using a standard Q-Q plot. Qualitatively, the fits to SMBKC are reasonable but could stand some more evaluation for improvement as only one configuration was tested here (Figures and .

```
Enc_prob <- SpatialDeltaGLMM::Check_encounter_prob(Report = Report,
    Data_Geostat = Data_Geostat, DirName = DateFile)
```

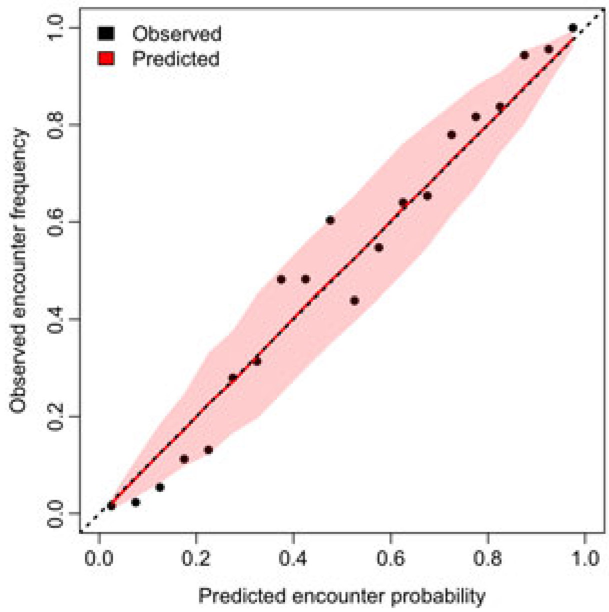

Figure 1: Observed encounter rates and predicted probabilities for SMBKC.

```
Q <- SpatialDeltaGLMM::QQ_Fn(TmbData = TmbData, Report = Report,
    FileName_PP = paste0(DateFile, "Posterior_Predictive.jpg"),
    FileName_Phist = paste0(DateFile, "Posterior_Predictive-Histogram.jpg"),
    FileName_QQ = paste0(DateFile, "Q-Q_plot.jpg"),
    FileName_Qhist = paste0(DateFile, "Q-Q_hist.jpg"))
```

Pearson residuals

Spatially the residual pattern can be evaluated over time. Results for SMBKC shows that consistent positive or negative residuals accross or within years is limited for the encounter probability component of the model and for the positive catch rate component (Figures 4 and 5, respectively). Some VAST plots for visualizing results can be seen by examining the direction of faster or slower spatial decorrelation (termed “geometric anisotropy”; Figure 6).

```
SpatialDeltaGLMM::plot_residuals(Lat_i = Data_Geostat[,
    "Lat"], Lon_i = Data_Geostat[, "Lon"], TmbData = TmbData,
    Report = Report, Q = Q, savedir = DateFile, MappingDetails = MapDetails_List[["MappingDetail
    PlotDF = MapDetails_List[["PlotDF"]], MapSizeRatio = MapDetails_List[["MapSizeRatio"]],
    Xlim = MapDetails_List[["Xlim"]], Ylim = MapDetails_List[["Ylim"]],
    FileName = DateFile, Year_Set = Year_Set, Years2Include = Years2Include,
    Rotate = MapDetails_List[["Rotate"]], Cex = MapDetails_List[["Cex"]],
    Legend = MapDetails_List[["Legend"]], zone = MapDetails_List[["Zone"]],
    mar = c(0, 0, 2, 0), oma = c(3.5, 3.5, 0, 0), cex = 1.8)
```

```
SpatialDeltaGLMM::PlotAniso_Fn(FileName = paste0(DateFile,
    "Aniso.png"), Report = Report, TmbData = TmbData)
```

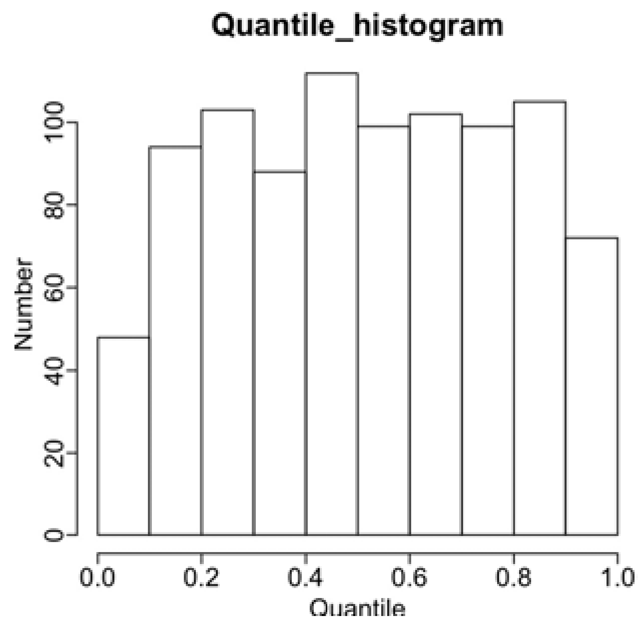

Figure 2: Plot indicating distribution of quantiles for "positive catch rate" component.

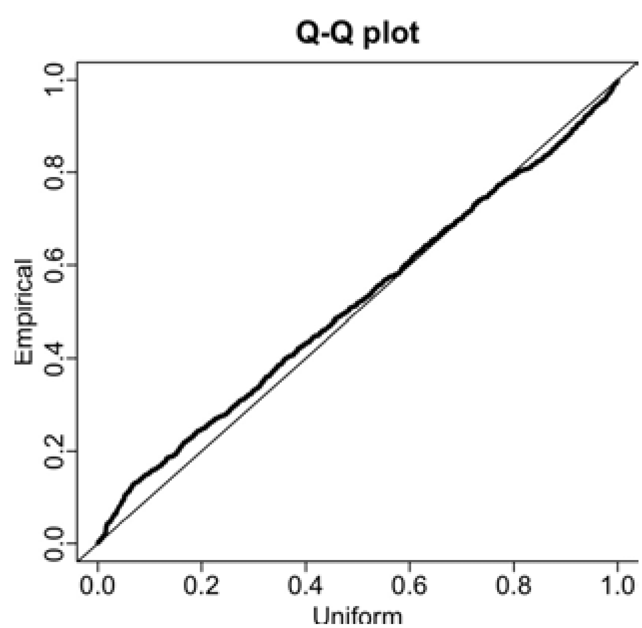

Figure 3: Quantile-quantile plot of residuals for "positive catch rate" component.

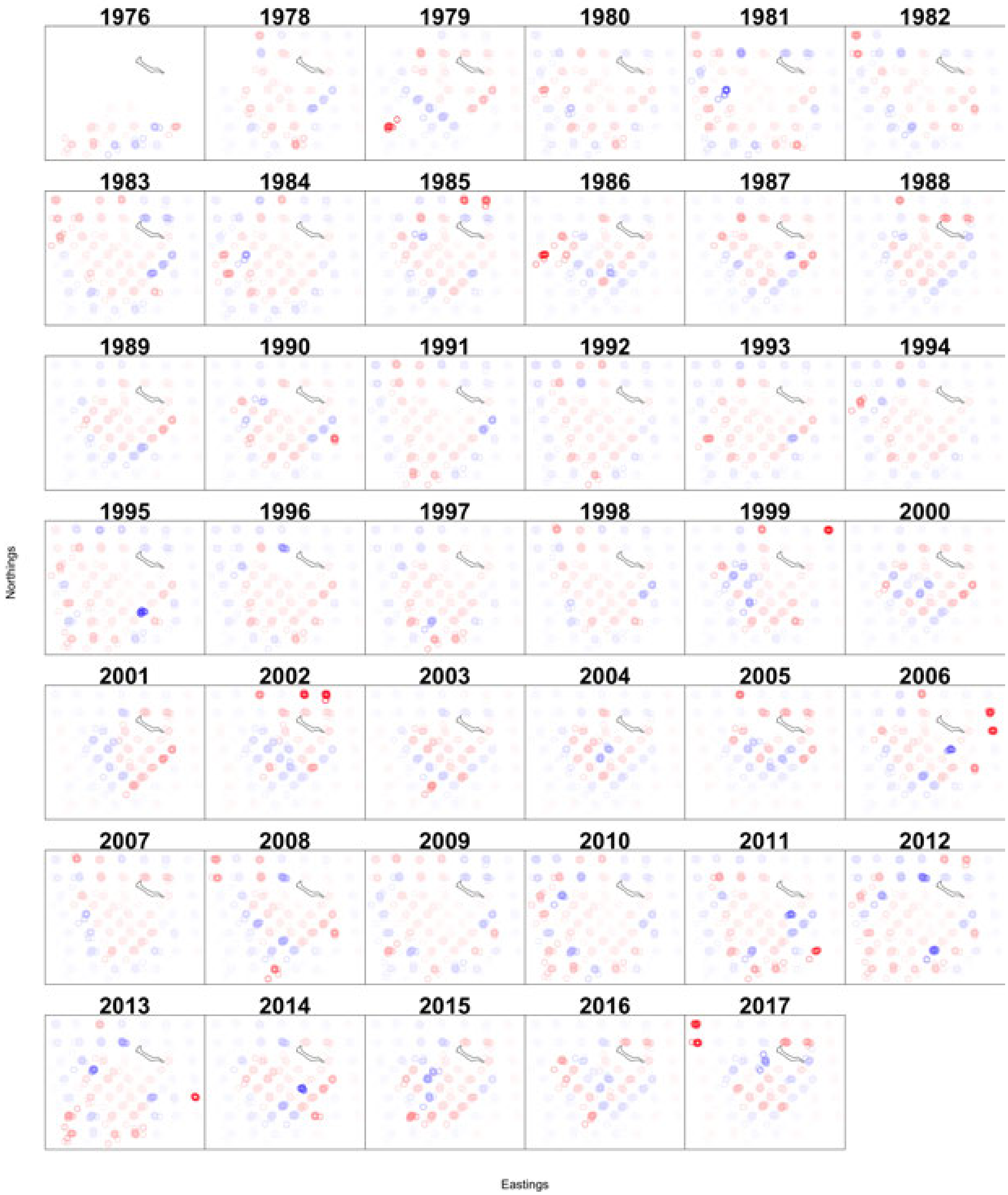

Figure 4: Pearson residuals of the encounter probability component at SMBKC stations, 1976-2017.

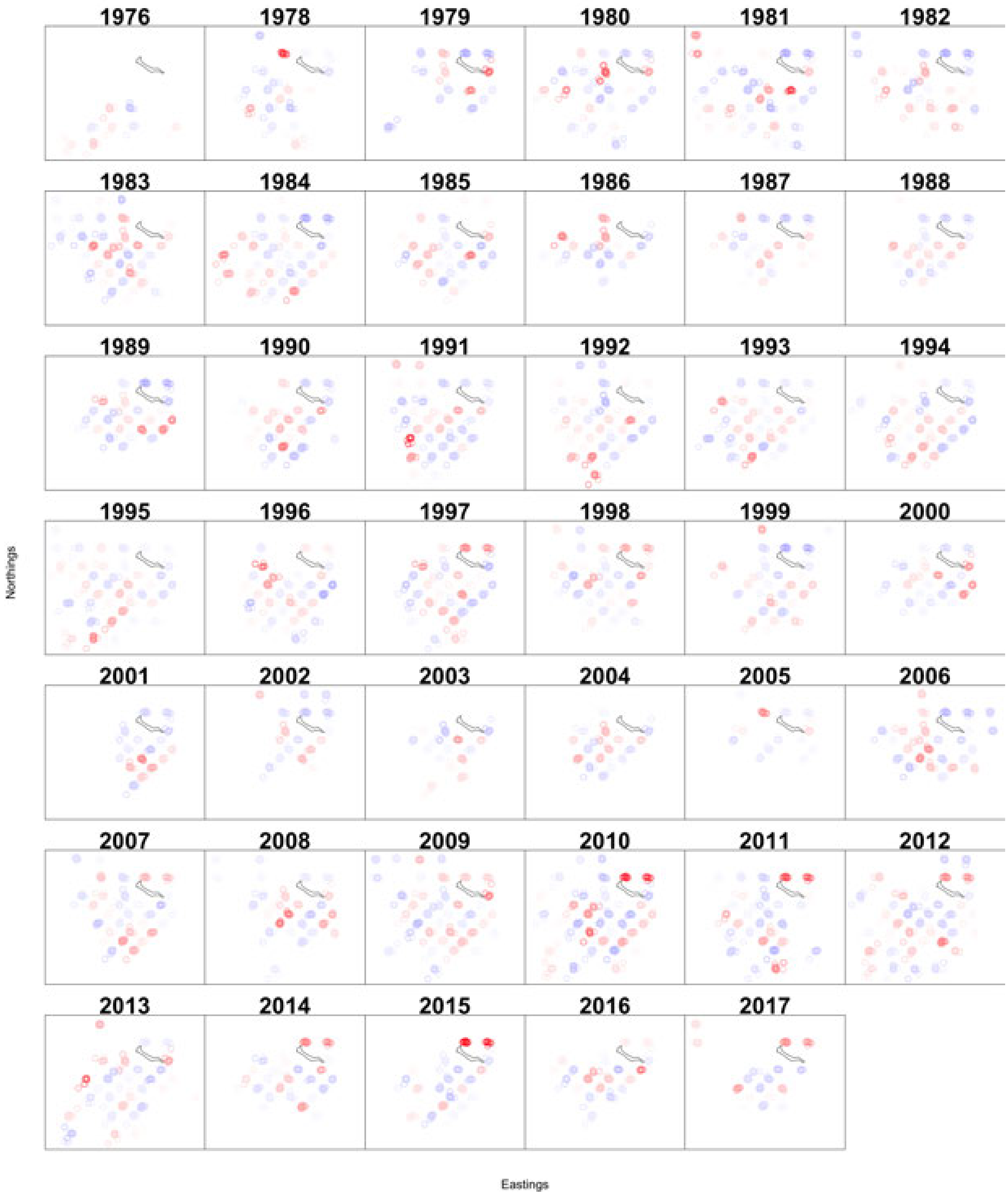

Figure 5: Pearson residuals of the positive catch rate component for SMBKC stations, 1976-2017.

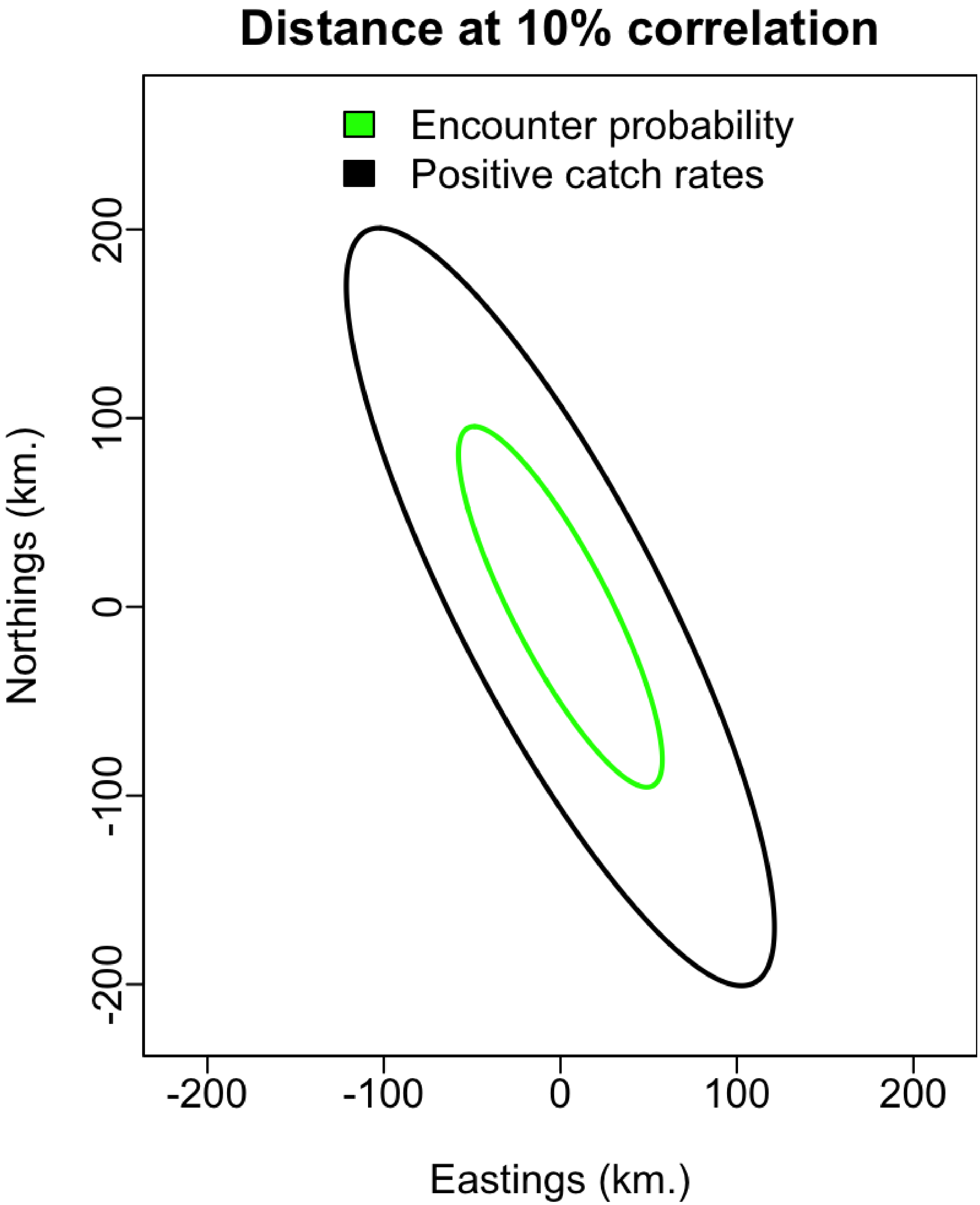

Figure 6: Directional decorrelation for SMBKC stations, 1978-2017.

```
SpatialDeltaGLMM::PlotResultsOnMap_Fn(plot_set = c(3),
    MappingDetails = MapDetails_List[["MappingDetails"]],
    Report = Report, Sdreport = Opt$SD, PlotDF = MapDetails_List[["PlotDF"]],
    MapSizeRatio = MapDetails_List[["MapSizeRatio"]],
    Xlim = MapDetails_List[["Xlim"]], Ylim = MapDetails_List[["Ylim"]],
    FileName = DateFile, Year_Set = Year_Set, Years2Include = Years2Include,
    Rotate = MapDetails_List[["Rotate"]], Cex = MapDetails_List[["Cex"]],
    Legend = MapDetails_List[["Legend"]], zone = MapDetails_List[["Zone"]],
    mar = c(0, 0, 2, 0), oma = c(3.5, 3.5, 0, 0), cex = 1.8,
    plot_legend_fig = FALSE)
```

Densities and biomass estimates A heatmap of the relative densities over

time shows a consistent pattern in the relative biomass of males >89mm (Figure 7). For the application to SMBKC, the biomass index was scaled to have the same mean as that from the design-based estimate (5,763 t) of abundance is generally most useful for stock assessment models (Table 2).

References

Please cite 2016 (ICES J. Mar. Sci. J. Cons.) if using the package; 2016 (Glob. Ecol. Biogeogr) if exploring factor decomposition of spatio-temporal variation; 2015 (ICES J. Mar. Sci. J. Cons.) if calculating an index of abundance; 2016 (Methods Ecol. Evol.) if using the center-of-gravity metric; 2016 (Fish. Res.) if using the bias-correction feature; 2016 (Proc R Soc B) if using the effective-area-occupied metric.

Thorson, J.T., and Barnett, L.A.K. In press. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. J. Cons

Thorson, J.T., Ianelli, J.N., Larsen, E., Ries, L., Scheuerell, M.D., Szuwalski, C., and Zipkin, E. 2016. Joint dynamic species distribution models: a tool for community ordination and spatiotemporal monitoring. Glob. Ecol. Biogeogr. 25(9): 1144-1158. doi:10.1111/geb.12464. url: http://onlinelibrary.wiley.com/doi/10.1111/geb.12464/abstract

Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. J. Cons. 72(5), 1297-1310. doi:10.1093/icesjms/fsu243. URL: http://icesjms.oxfordjournals.org/content/72/5/1297

Thorson, J.T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fish. Res. 175: 66-74. doi:10.1016/j.fishres.2015.11.016. url: http://www.sciencedirect.com/science/article/pii/S0165783615301399

Thorson, J.T., Pinsky, M.L., Ward, E.J., 2016. Model-based inference for estimating shifts in species distribution, area occupied, and center of gravity. Methods Ecol. Evol. 7(8), 990-1008. doi:10.1111/2041-210X.12567. URL: http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12567/full

Thorson, J.T., Rindorf, A., Gao, J., Hanselman, D.H., and Winker, H. 2016. Density-dependent changes in effective area occupied for sea-bottom-associated marine fishes. Proc R Soc B 283(1840): 20161853. doi:10.1098/rspb.2016.1853. URL: http://rspb.royalsocietypublishing.org/content/283/1840/20161853.

To see these entries in BibTeX format, use 'print(, bibtex=TRUE)', 'toBibtex(.)', or set 'options(citation.bibtex.max=999)'.

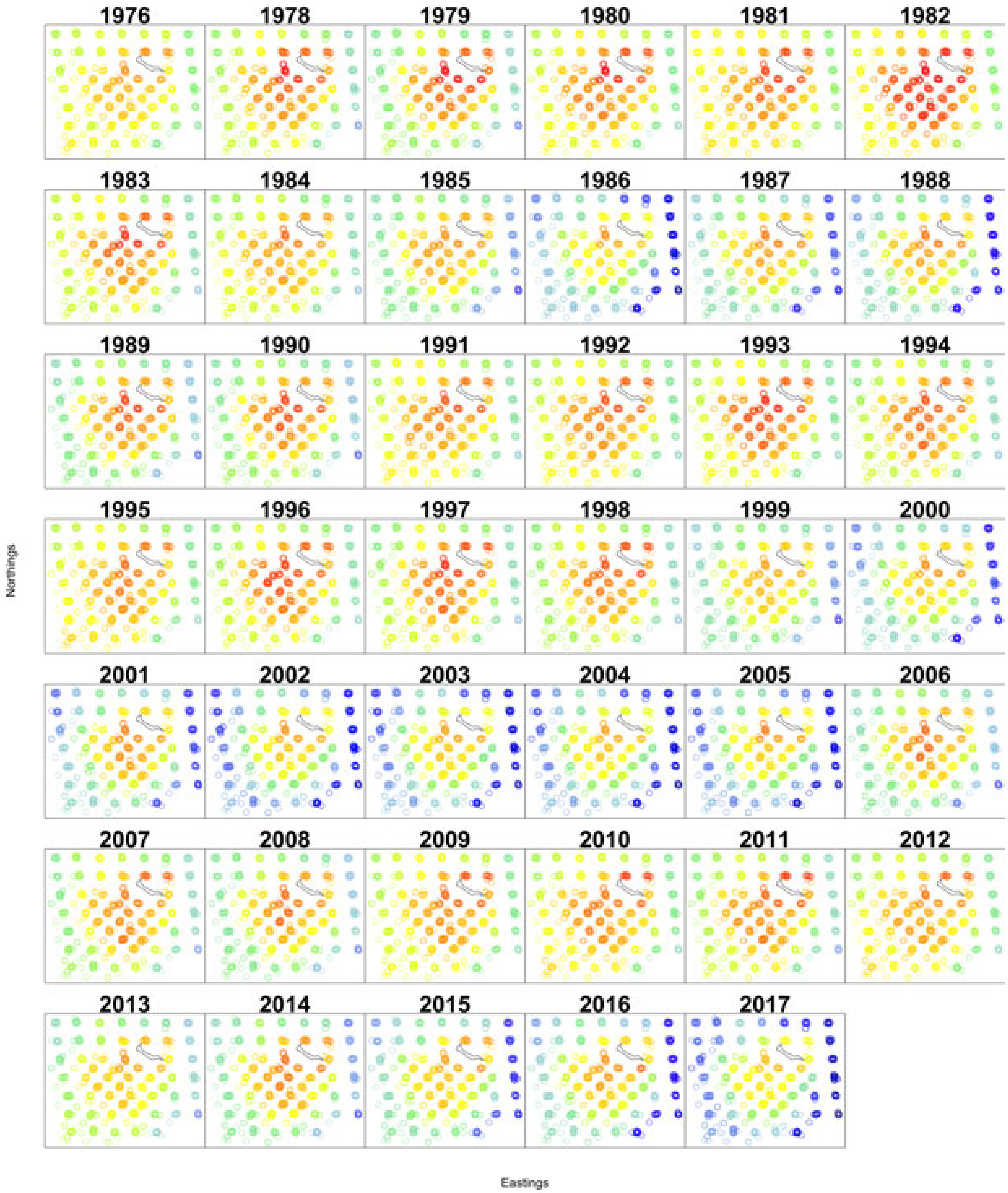

Figure 7: St. Matthews Island blue king crab (males >89mm) density maps as predicted using the VAST model approach, 1976-2017.

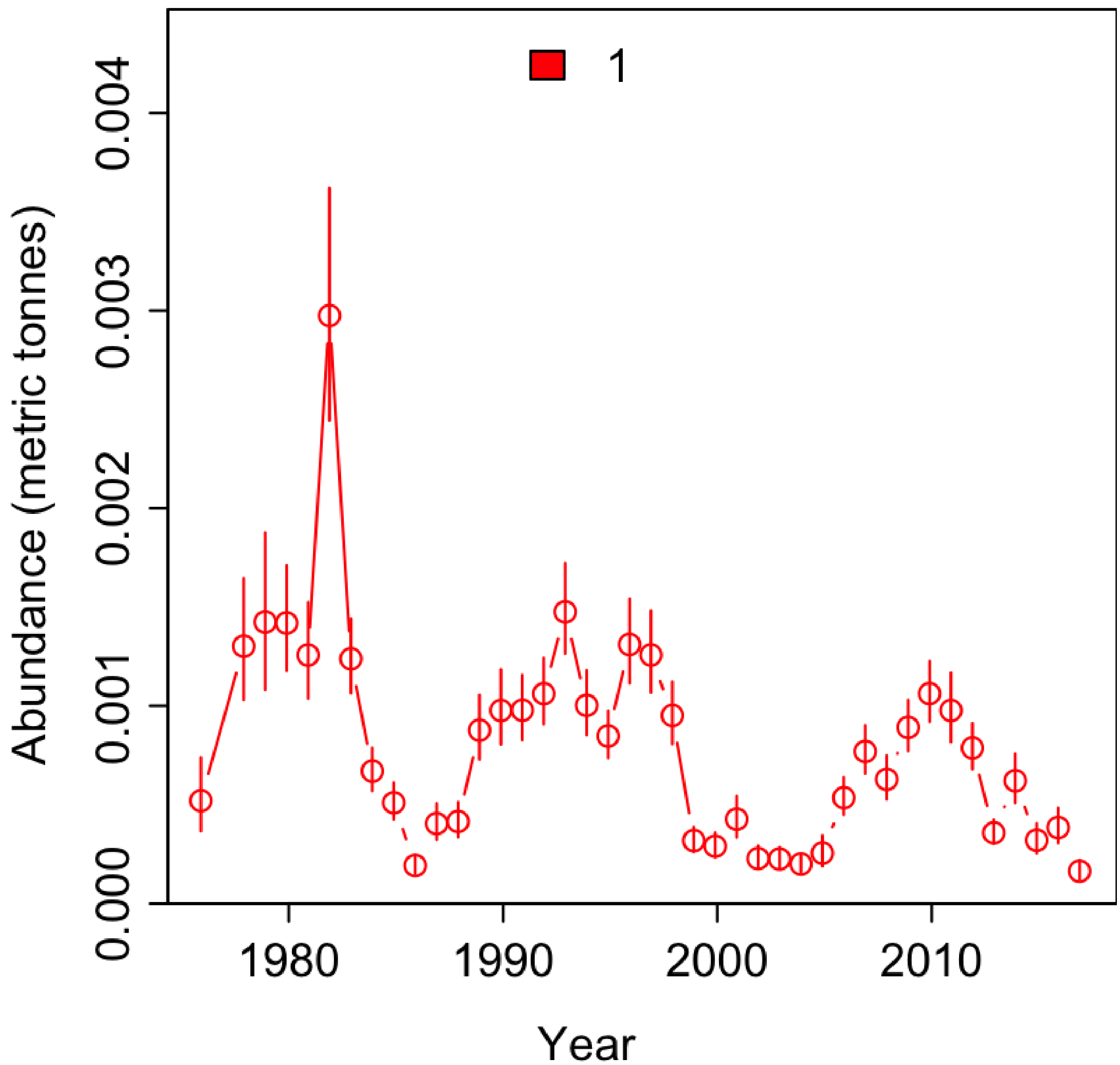

Figure 8: St. Matthews Island blue king crab (males >89mm) relative abundance as predicted using the VAST model approach.

Table 1: SMBKC parameter estimates, bounds, and final gradients as derived from the VAST modeling framework.

Param	Lower	MLE	Upper	final_gradient
ln_H_input	-50.0	-0.157	50.0	0.00001
ln_H_input	-50.0	-0.637	50.0	-0.00006
beta1_ct	-50.0	1.068	50.0	0.00001
beta1_ct	-50.0	-1.381	50.0	0.00001
beta1_ct	-50.0	-2.306	50.0	-0.00002
beta1_ct	-50.0	-0.486	50.0	0.00001
beta1_ct	-50.0	0.556	50.0	0.00001
beta1_ct	-50.0	-0.774	50.0	0.00001
beta1_ct	-50.0	-0.643	50.0	-0.00004
beta1_ct	-50.0	-0.616	50.0	0.00000
beta1_ct	-50.0	-1.786	50.0	0.00000
beta1_ct	-50.0	-3.240	50.0	-0.00000
beta1_ct	-50.0	-2.464	50.0	0.00001
beta1_ct	-50.0	-2.955	50.0	0.00002
beta1_ct	-50.0	-2.080	50.0	0.00001
beta1_ct	-50.0	-1.924	50.0	-0.00001
beta1_ct	-50.0	-0.402	50.0	-0.00002
beta1_ct	-50.0	-0.534	50.0	-0.00001
beta1_ct	-50.0	-0.867	50.0	-0.00001
beta1_ct	-50.0	-1.032	50.0	-0.00001
beta1_ct	-50.0	0.265	50.0	-0.00002
beta1_ct	-50.0	-0.869	50.0	-0.00001
beta1_ct	-50.0	-1.201	50.0	-0.00001
beta1_ct	-50.0	-1.061	50.0	-0.00004
beta1_ct	-50.0	-1.742	50.0	0.00001
beta1_ct	-50.0	-2.691	50.0	-0.00001
beta1_ct	-50.0	-3.145	50.0	-0.00001
beta1_ct	-50.0	-3.401	50.0	-0.00004
beta1_ct	-50.0	-3.412	50.0	0.00002
beta1_ct	-50.0	-3.214	50.0	0.00002
beta1_ct	-50.0	-3.797	50.0	-0.00001
beta1_ct	-50.0	-1.776	50.0	0.00000
beta1_ct	-50.0	-1.032	50.0	-0.00002
beta1_ct	-50.0	-1.630	50.0	-0.00001
beta1_ct	-50.0	0.157	50.0	0.00001
beta1_ct	-50.0	0.141	50.0	0.00001
beta1_ct	-50.0	-1.206	50.0	-0.00003
beta1_ct	-50.0	0.143	50.0	0.00001
beta1_ct	-50.0	-0.956	50.0	0.00005
beta1_ct	-50.0	-2.236	50.0	0.00001
beta1_ct	-50.0	-2.546	50.0	-0.00001
beta1_ct	-50.0	-3.100	50.0	-0.00000
beta1_ct	-50.0	-3.756	50.0	0.00002
L_omega1_z	-50.0	2.282	50.0	0.00007
L_epsilon1_z	-50.0	0.683	50.0	-0.00009
logkappa1	-4.7	-3.695	-1.9	-0.00003
beta2_ct	-50.0	-8.669	50.0	0.00004
beta2_ct	-50.0	-7.498	50.0	0.00008
beta2_ct	-50.0	-7.295	50.0	0.00011
beta2_ct	-50.0	-7.582	50.0	0.00008
beta2_ct	-50.0	-7.801	50.0	-0.00014
beta2_ct	-50.0	-6.802	50.0	0.00000
beta2_ct	-50.0	-7.813	50.0	0.00013
beta2_ct	-50.0	-8.131	50.0	-0.00000
beta2_ct	-50.0	-8.362	50.0	-0.00010
beta2_ct	-50.0	-8.978	50.0	-0.00006
beta2_ct	-50.0	-8.486	50.0	0.00001

Table 2: SMBKC male >89mm biomass (t) estimates as derived from the VAST modeling framework.

Year	Estimate	CV
1977	3654.3	0.801
1978	9467.9	0.234
1979	10354.7	0.276
1980	10318.3	0.187
1981	9142.0	0.192
1982	21625.3	0.196
1983	9004.3	0.152
1984	4873.7	0.162
1985	3708.6	0.183
1986	1401.1	0.238
1987	2942.9	0.226
1988	3020.4	0.212
1989	6377.5	0.185
1990	7102.0	0.192
1991	7111.8	0.168
1992	7721.3	0.157
1993	10730.5	0.155
1994	7291.9	0.163
1995	6164.3	0.141
1996	9530.6	0.162
1997	9144.6	0.164
1998	6919.4	0.165
1999	2316.9	0.196
2000	2110.6	0.213
2001	3105.0	0.242
2002	1656.7	0.250
2003	1639.7	0.234
2004	1457.0	0.216
2005	1856.6	0.300
2006	3894.4	0.176
2007	5595.6	0.158
2008	4569.5	0.176
2009	6480.5	0.145
2010	7723.8	0.144
2011	7102.5	0.178
2012	5725.3	0.147
2013	2603.0	0.170
2014	4517.7	0.199
2015	2330.7	0.235
2016	2797.0	0.230
2017	1192.9	0.293