Appendix 2. Spring phenology

Spatio-temporal population change of arctic-breeding waterbirds

We obtained data on the predicted start of the growing season (start of season time, SOST; day of the year) from vegetation indices derived from MODIS satellite imagery from 1992 to 2016 (Zhang et al. 2003, Didan et al. 2016). Data are available at approximately 2.5 km by 5 km spatial resolution across the Arctic Coastal Plain, Alaska (ACP). We summarized SOST by cell-year by calculating the mean, weighted by area covering each cell for cells with > 1 value. Some SOST values are not well predicted along the coast and particularly in river deltas subject to spring flooding. Therefore, we removed SOST values in cells that were < 122 (2 or 3 May) or > 228 (16 or 17 Aug) and imputed values for those cell-years by interpolating values from the 40 nearest neighbors in each year weighted by the inverse-weighted distance to the target cell. In most cases, this represented cells < 3 cell-widths or 17 km from the target cell boundary (i.e., on the diagonal). The days chosen as cutoffs represented natural break-points in the data. We used the knnimputation function in the R package DMwR (Torgo 2015, R Core Team 2018) to impute omitted values.

We evaluated trends in spring phenology through space and time in two ways. First, we evaluated a generalized additive model (GAM; Wood 2017) that incorporated spatial autocorrelation and an additive temporal trend to estimate changes in SOST through time. Second, we evaluated a space by year interaction to examine spatial patterns of temporal change in SOST. We implemented GAMs using the gam function in the R package mgcv (Wood 2018). We obtained confidence intervals around differences in SOST between 1992 and 2016 using Monte Carlo simulations. In this appendix, we provide R code and results to facilitate replication of our analyses.

RESULTS

Model selection results supported an interaction between space and time, and a smoothed spatial component (Table A2.1). The most supported model explained 72.6% of the deviance in the data. Predicted SOST occurred, on average, 9 days earlier in 2016 (day 158, SE = 0.065; 7 June) than in 1992 (day 167, SE = 0.065; 16 June) (Figure A2.1). The predicted difference in SOST from 1992 to 2016 varied through space (Figure S2), and all cells showed SOST decreasing through time (range: 5.32 to 15. 67 days earlier) with 95% confidence intervals (CI) that did not overlap zero (upper 95% CI range among cells: 4.37 to 14.49 days earlier).
Table A2.1. Model selection results from four generalized additive model structures evaluated that examined effects of space (cellX, cellY) and time (year) on the start of spring green up (SOST) across the Arctic Coastal Plain, Alaska, 1992-2016. Models varied in complexity and smooth parameters for the spatial component; either (cellX, cellY) was modeled using a smooth (s()) or a full tensor product (te()). The value with the lowest AIC value was used for subsequent model predictions.

<table>
<thead>
<tr>
<th>Model</th>
<th>ΔAIC</th>
<th>AIC</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOST~s(year)</td>
<td>51,393</td>
<td>299,122.4</td>
<td>6.00</td>
</tr>
<tr>
<td>SOST~ te(cellX, cellY) + s(year)</td>
<td>2,861.9</td>
<td>250,591.3</td>
<td>28.49</td>
</tr>
<tr>
<td>SOST~ s(cellX, cellY) + s(year)</td>
<td>1,549.2</td>
<td>249,278.6</td>
<td>34.70</td>
</tr>
<tr>
<td>SOST~ te(cellX, cellY) + s(year) + ti(cellX, cellY, year)</td>
<td>1,378</td>
<td>249,107.4</td>
<td>86.71</td>
</tr>
<tr>
<td>SOST~ s (cellX, cellY) + s(year) + ti(cellX, cellY, year)</td>
<td>0</td>
<td>247,729.4</td>
<td>92.13</td>
</tr>
</tbody>
</table>
Figure A2.1. Predicted start of the growing season (day) from 1992 to 2016 for 1,914 6 km by 6 km cells across the Arctic Coastal Plain, Alaska. We modeled the effect of space and time on start of growing season day using generalized additive models that apply a non-linear smooth to spatio-temporal effects. The gray shaded region denotes 95% confidence interval around predictions. The lower plot is identical, but includes semi-transparent data values across cells for each year (in blue).
Figure A2.2. The change (in days) in the start of spring growing season (day of year) between 1992 and 2016 within approximately 36 km² cells (6 km by 6 km) across the Arctic Coastal Plain, Alaska. We modeled the effect of space and time on start of growing season day using generalized additive models that apply a non-linear smooth to spatio-temporal effects, then subtracted cell-specific values in 1992 from 2016. Negative values denote a decreasing trend in the start of spring (i.e., start of the growing season occurred earlier in recent years).


R code for regression and plots.

```r
library(fields)
library(mgcv)
library(plyr)
library(rgdal)
library(sp)

# load dataset of SOST values by year
sst <- read.csv("~/acp_sost.csv",header=T) # data available from corresponding author, camundson@usgs.gov

# quick and dirty plot of SOST through space, averaged across years
quilt.plot(sst$cellX,sst$cellY,sst$sost)

# look at simple to more complex models
# limit k to 5 for year to avoid overfit - meaningless curves
m0 <- gam(sost~s(year,k=5),method="REML",data=sst) # year only
m1 <- gam(sost~te(cellX,cellY)+s(year,k=5),method="REML",data=sst) # additive year, full tensor spatial effect
m2 <- gam(sost~s(cellX,cellY)+s(year,k=5),method="REML",data=sst) # additive year, smoothed spatial effect
m3 <- gam(sost~s(cellX,cellY)+s(year,k=5)+ti(cellX,cellY,year),method="REML",data=sst) # interactive year, smoothed spatial effect
m4 <- gam(sost~te(cellX,cellY)+s(year,k=5)+ti(cellX,cellY,year),method="REML",data=sst) # interactive year, tensor spatial effect

# model selection
AIC(m0,m1,m2,m3, m4)

# Examine model results
summary(m4)
m4
plot(m4,scheme=2)

# m3 fits best
vis.gam(m3)
fitted <- predict(m3,sst,se=T)
str(fitted)
plot(as.factor(sst$year),fitted$fit)
newd <- data.frame(cellX=sst$cellX[1],cellY=sst$cellY[1],year=seq(1992,2016))

# Predictions averaged over space
preds <- predict(m3, newdata=newd,exclude=c("ti(cellX,cellY,year)","s(cellX,cellY)"),
  se.fit=TRUE)
str(preds)
```

fit<-preds$fit
fit.up95<-fit-1.96*preds$se.fit
fit.low95<-fit+1.96*preds$se.fit

plot(newd$year, fit, type="n", lwd=3, #ylim=c(145,193),
    main="Green-up Trend, 1992-2016",xlab="Year",
    ylab="Start of season day")

#Include data points or not, if so, including ylim argument above
#points(sst$year,sst$sost, pch=16, col=rgb(0, 0, 1, 0.25))

# For the confidence grey polygon
polygon(c(newd$year, rev(newd$year)),
    c(fit.low95,rev(fit.up95)), col="gray",
    border=NA)

lines(newd$year, fit, lwd=1)

# Monte Carlo variance around difference in SOST from 1992 to 2016 for each cell

res<-double()
nsim<-1000# pick a large number here, but TAKES AWHILE
system.time(for (n in 1:nsim){
    tmp=rnorm(length(sst$year),fitted$fit,fitted$se.fit)
    res=rbind(res,data.frame(cellX=sst$cellX,cellY=sst$cellY,year=sst$year,prd=tmp))
})

yr2016<-subset(res,year==2016)
yr1992<-subset(res,year==1992)
diff<-yr2016$prd-yr1992$prd
diff.df<-data.frame(cellX=yr2016$cellX,cellY=yr2016$cellY,diff)
diff.df$Char.loc<-as.factor(paste(diff.df$cellX,diff.df$cellY,sep="_"))
df<-ddply(diff.df,.(char.loc),summarize,mndiff=mean(diff),se=sd(diff),lci=quantile(diff,0.025),
cellX=mean(cellX),cellY=mean(cellY),uci=quantile(diff,0.975))

str(df)
summary(df)

quilt.plot(df$cellX,df$cellY,(df$mndiff*-1),nx=147,ny=34)
# convert into ESRI shapefile for mapping
proj<-("+proj=aea +lat_1=55 +lat_2=65 +lat_0=50 +lon_0=-154 +x_0=0 +y_0=0
+datum=NAD83 +units=m +no_defs +ellps=GRS80 +towgs84=0,0,0")
spat.sost<-SpatialPointsDataFrame(cbind(df$cellX,df$cellY),df,proj4string=CRS(proj))
writeOGR(obj=spat.sost, dsn="D:/acp_predshapefiles",layer="acp_SOSTpreddiff",
driver="ESRI Shapefile")