

Nonparametric Analysis of US Dairy Production and Consumption

Spatial GAM

Teo Bucci* Filippo Cipriani† Gabriele Corbo‡ Andrea Puricelli§

2023-02-17

Contents

1 Load libraries and data	1
2 Model 1: PREDICT ON INTERACTION BETWEEN COORDINATES	2
3 Model 2: PREDICT ON INTERACTION BETWEEN COORDINATES AND POPULATION	5
4 Prediction	11

1 Load libraries and data

```
library(ISLR2)
library(car)
library(sp)
library(mgcv)
library(rgl)
library(splines)
library(pbapply)
library(devtools)
library(visreg)
library(ggplot2)
library(mgcViz)
```

```
source("05-Prepare-data-spatial.R", echo=F)
```

*teo.bucci@mail.polimi.it

†filippo.cipriani@mail.polimi.it

‡gabriele.corbo@mail.polimi.it

§andrea3.puricelli@mail.polimi.it

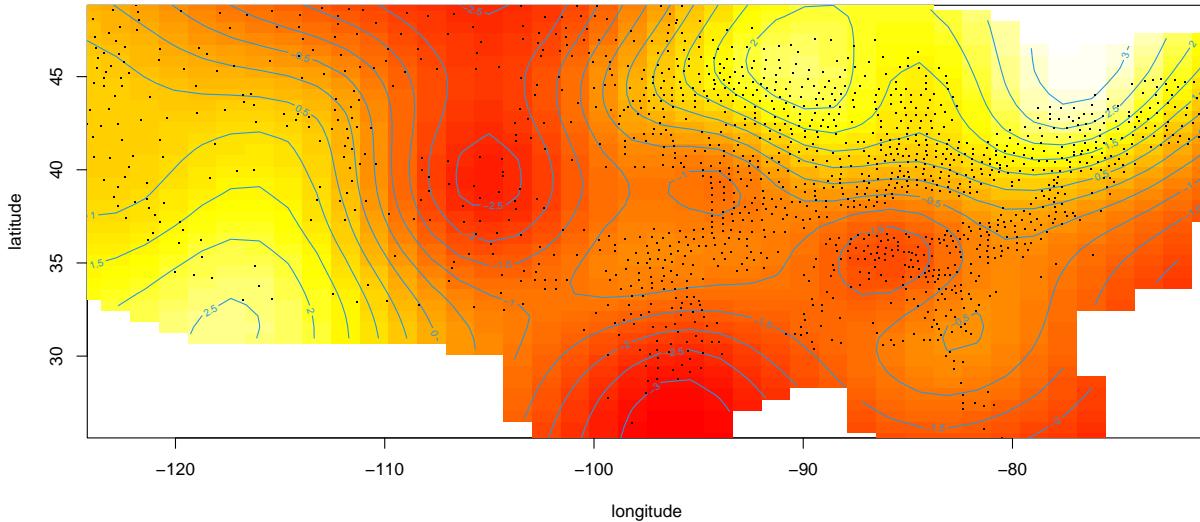
2 Model 1: PREDICT ON INTERACTION BETWEEN COORDINATES

```
mod2d <- mgcv:::gam(log(Value) ~ s(y,x), data = all_sales_county_2007, method = "REML")
summary(mod2d)
```

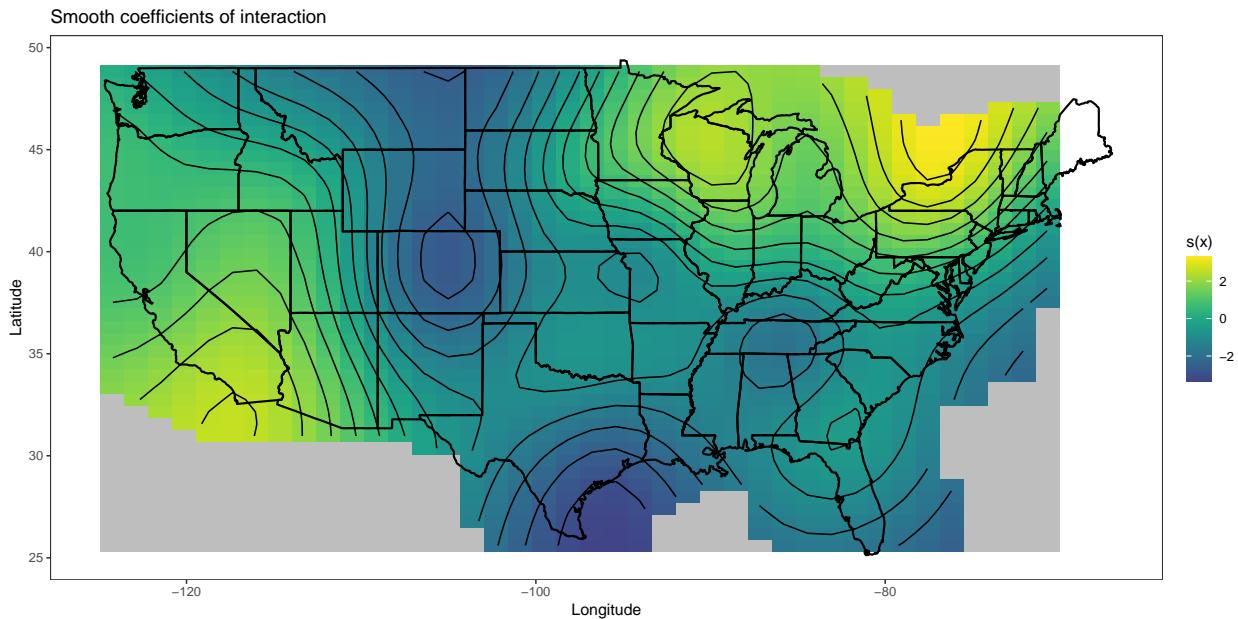
```
## 
## Family: gaussian
## Link function: identity
##
## Formula:
## log(Value) ~ s(y, x)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 14.6333   0.0525 278.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df F p-value    
## s(y,x) 25.84  28.39 22  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.303   Deviance explained = 31.5%
## -REML = 3135.2   Scale est. = 4.0302    n = 1462
```

```
par(mfrow=c(1,1))
plot(mod2d,scheme=2,pages=0,select = 1,
     main = "Smooth coefficients of interaction",
     xlab = "longitude",ylab = "latitude")
```

Smooth coefficients of interaction



```
b <- getViz(mod2d)
pl <- plot(sm(b, 1)) + l_fitRaster() + l_fitContour() +
  geom_polygon(data = map_data ("state"),
    aes(x=long, y = lat,group=group),
    fill=NA,color="black",lwd = 0.7,inherit.aes = F) +
  ggtitle("Smooth coefficients of interaction") +
  xlab("Longitude") + ylab("Latitude")
pl
```

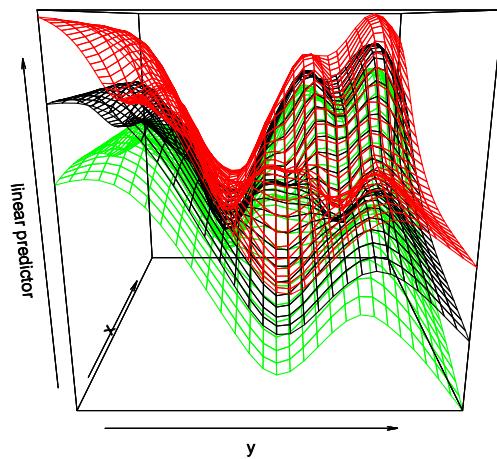


```

x_grid <- map_data ("state")$long
y_grid <- map_data ("state")$lat

# Make the perspective plot with error surfaces
vis.gam(mod2d, view = c("y", "x"),
        plot.type = "persp", se = 2)

```

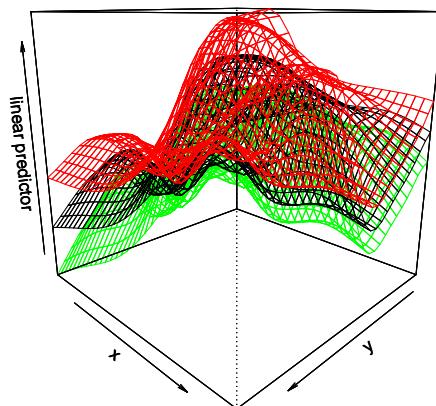


red/green are ± 2 s.e.

```

# Rotate the same plot
vis.gam(mod2d, view = c("y", "x"),
        plot.type = "persp", se = 2, theta = 135)

```

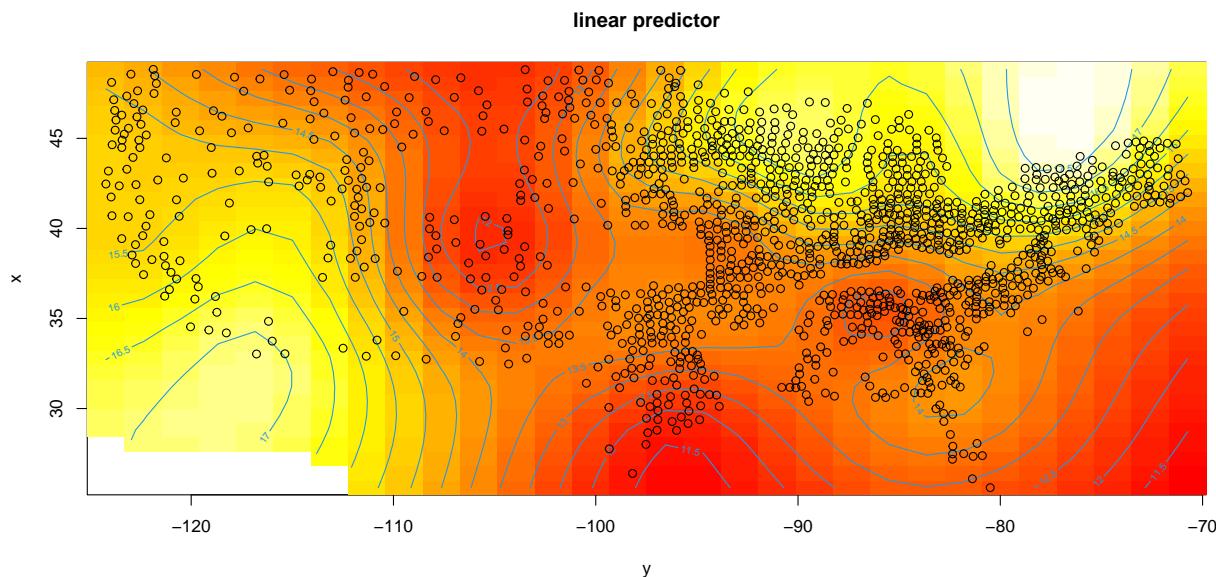


red/green are ± 2 s.e.

```

# Make plot with 10% extrapolation
vis.gam(mod2d, view = c("y", "x"),
         plot.type = "contour", too.far = 0.25)
dat1 <- all_sales_county_2007
coordinates(dat1) <- c("y", "x")
points(dat1)

```



3 Model 2: PREDICT ON INTERACTION BETWEEN COORDINATES AND POPULATION

```

tensor_mod2 <- gam(log(Value) ~ s(y, x)+s(log(Population)),
                     data = all_sales_county_2007, method = "REML")

```

```

# Summarize and plot
summary(tensor_mod2)

```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## log(Value) ~ s(y, x) + s(log(Population))
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.63331   0.05115 286.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

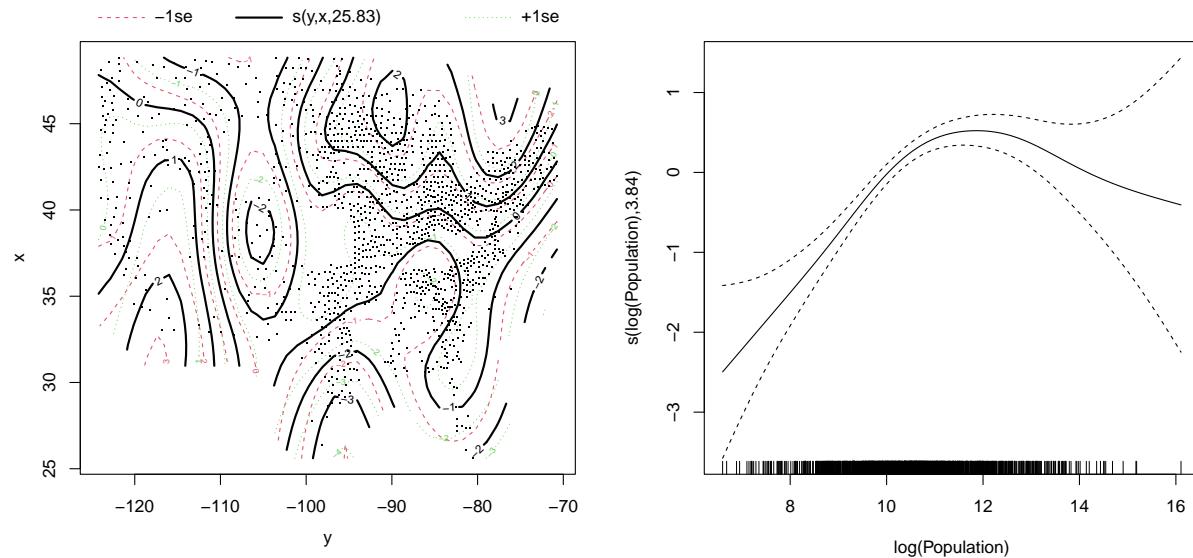
```

```

## 
## Approximate significance of smooth terms:
##          edf Ref.df      F p-value    
## s(y,x)      25.830 28.383 19.13 <2e-16 ***
## s(log(Population)) 3.841  4.871 16.06 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## R-sq.(adj) =  0.338   Deviance explained = 35.2%
## -REML = 3101.5   Scale est. = 3.8258    n = 1462

```

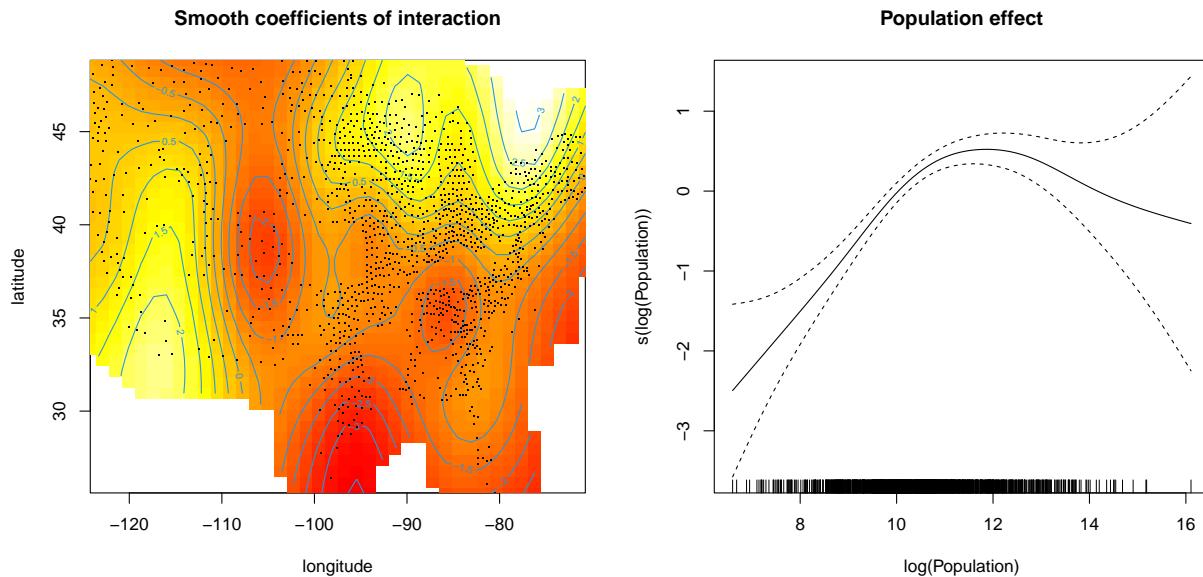
```
plot(tensor_mod2, pages=1)
```



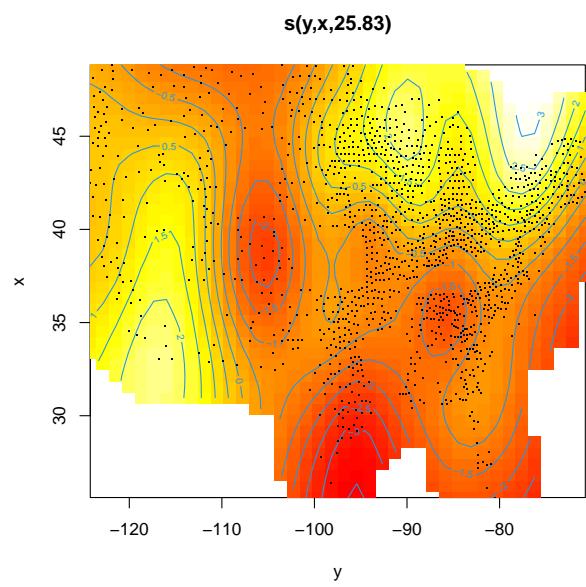
```

par(mfrow=c(1,2))
plot(tensor_mod2, scheme=2, pages=0, select = 1,
      main = "Smooth coefficients of interaction",
      xlab = "longitude", ylab = "latitude")
plot(tensor_mod2, scheme=2, pages=0, select = 2,
      main = "Population effect", ylab = "s(log(Population))")

```

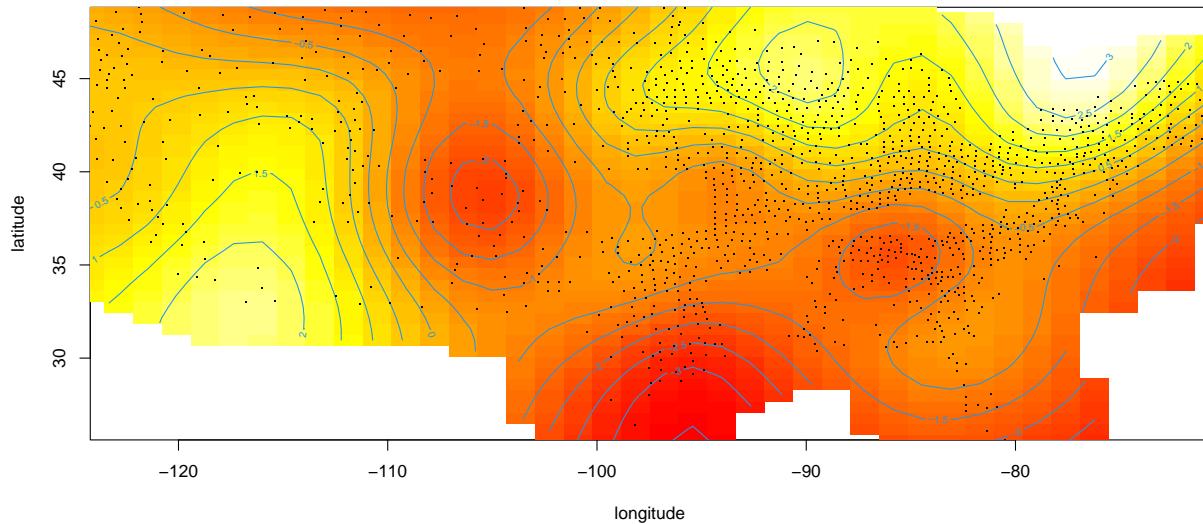


```
plot(tensor_mod2, scheme = 2, pages =0,select=1)  
par(mfrow=c(1,1))
```

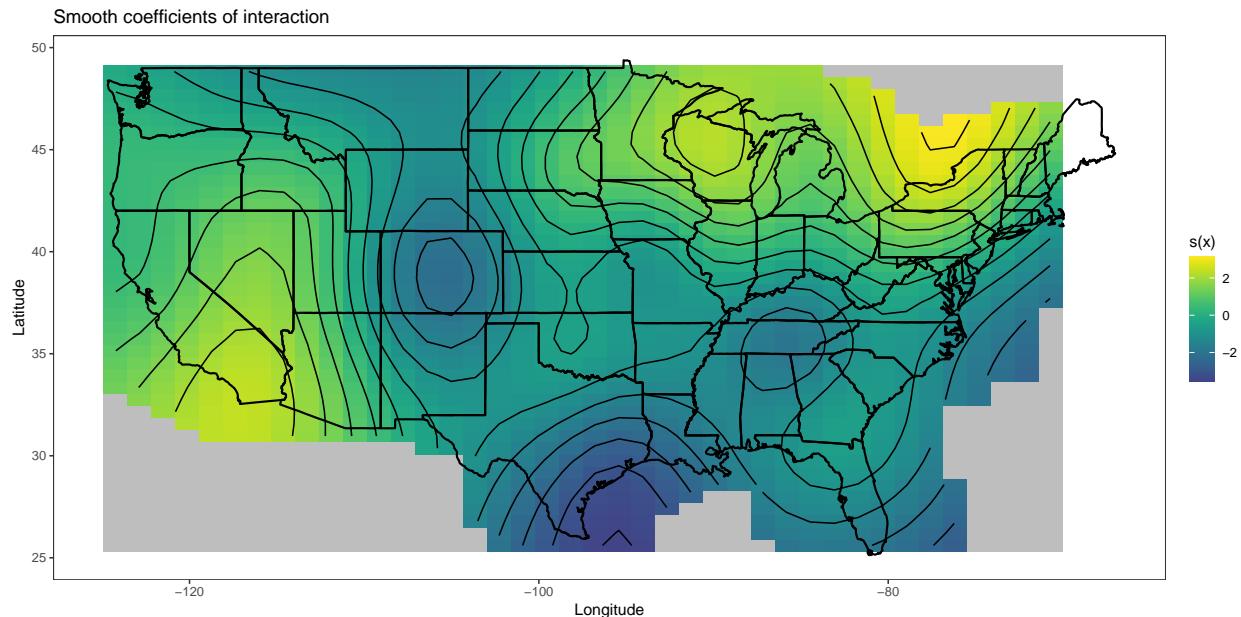


```
plot(tensor_mod2,scheme=2, pages=0, select = 1,  
      main = "Smooth coefficients of interaction",  
      xlab = "longitude", ylab = "latitude")
```

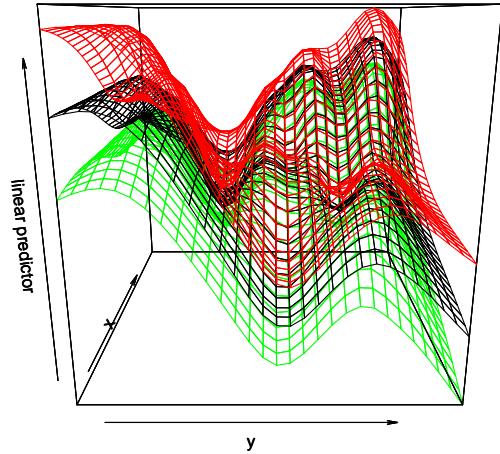
Smooth coefficients of interaction



```
b <- getViz(tensor_mod2)
pl <- plot(sm(b, 1)) + l_fitRaster() + l_fitContour() +
  geom_polygon(data = map_data ("state"),
    aes(x=long, y = lat,group=group),
    fill=NA,color="black",lwd = 0.7,inherit.aes = F) +
  ggtitle("Smooth coefficients of interaction") +
  xlab("Longitude") + ylab("Latitude")
pl
```

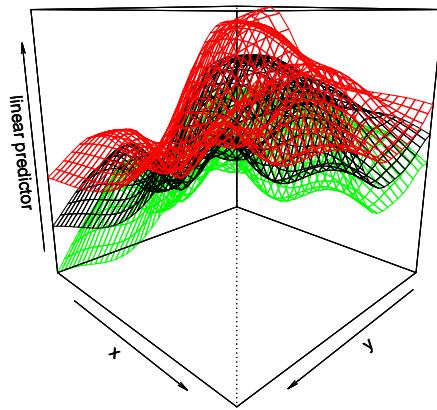


```
vis.gam(tensor_mod2, view = c("y", "x"),
        plot.type = "persp", se = 2)
```



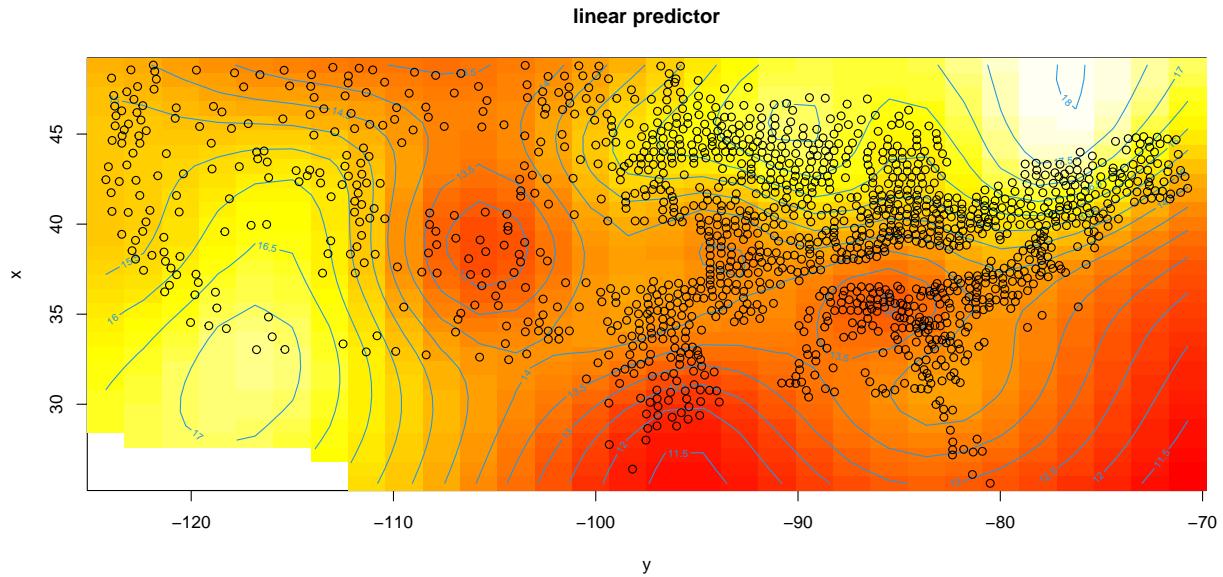
red/green are ± 2 s.e.

```
# Rotate the same plot
vis.gam(tensor_mod2, view = c("y", "x"),
        plot.type = "persp", se = 2, theta = 135)
```

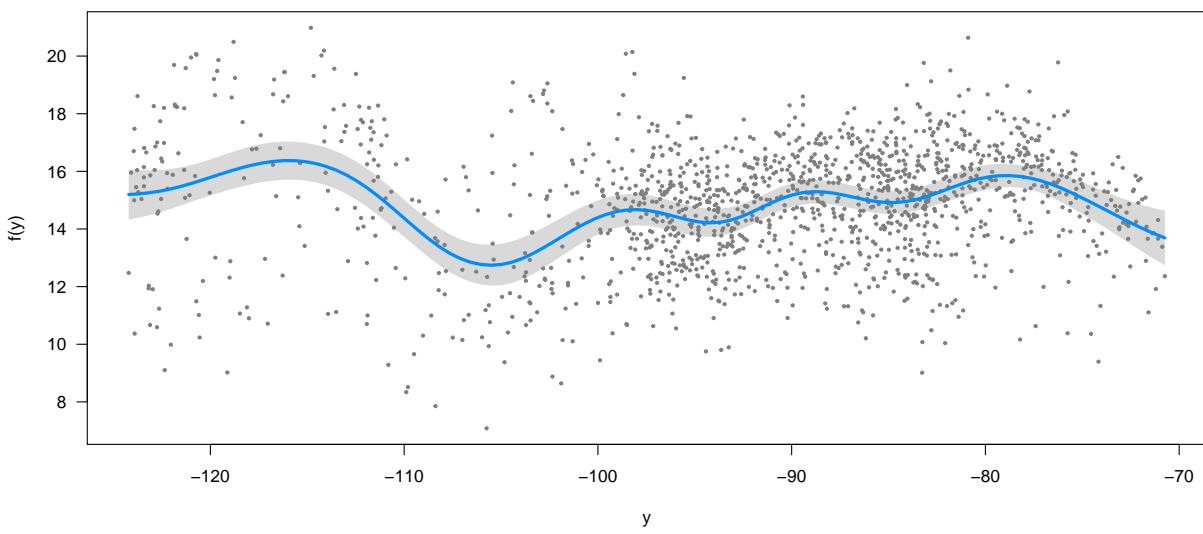


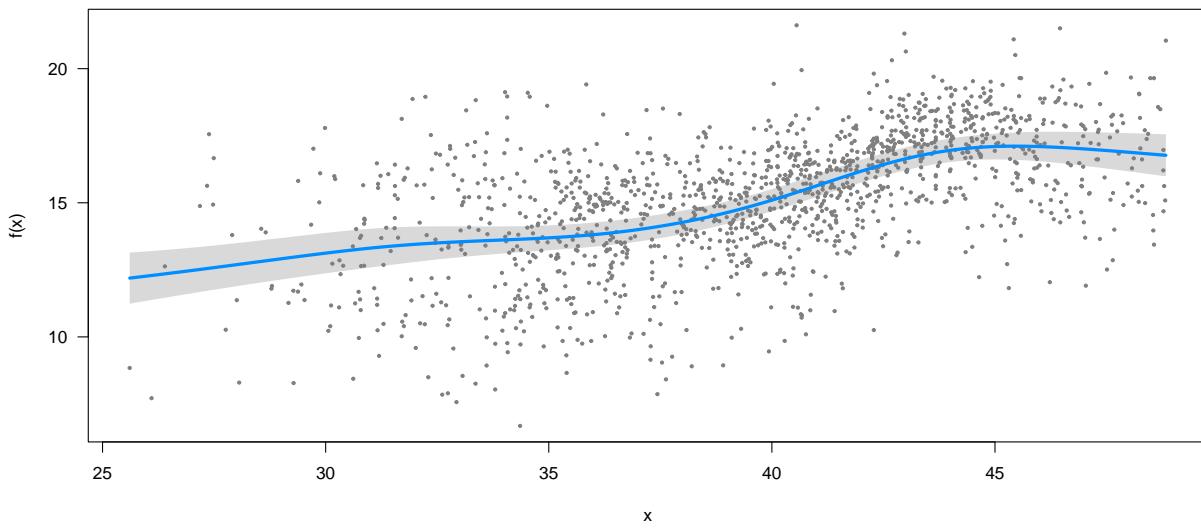
red/green are ± 2 s.e.

```
# Make plot with 10% extrapolation
vis.gam(tensor_mod2, view = c("y", "x"),
         plot.type = "contour", too.far = 0.25)
points(dat1)
```



```
visreg(tensor_mod2)
```





4 Prediction

```

x_grid <- seq(-118,-104, by = 0.01)
y_grid <- seq(36,43, by = 0.01)
grid <- expand.grid(x_grid,y_grid)
grid <- data.frame(x=grid$Var2, y=grid$Var1)
preds <- predict(mod2d,newdata=grid)
preds
df <- cbind(grid,"log(Sales)" =as.data.frame(preds))
ggplot(as.data.frame(df), aes(y,x, col=preds)) +
  geom_point(alpha=0.5) +
  scale_color_gradient(low="red", high="yellow") +
  coord_cartesian(xlim=c(-115.5, -108), ylim = c(37, 42))+ 
  geom_polygon( data=map_data("county"), aes(x=long, y=lat, group=group),
                color="black", fill="lightblue", alpha=0.1 ) +
  guides(color = guide_legend(title = "log(Sales)"))

```

