Exploration for Data Filtering Speed

Anna Steel

August 12, 2016

Load and prepare dataset

- add UTM coordinates

```
 reddf <- readRDS("Maestros/AllFishPrimaryFilt.RData")

# Critical to order this properly for the ordered step analysis below (TypeI)
reddf = reddf[order(reddf$Id, reddf$Time),]
  reddf$IDcol = 1:nrow(reddf)

# # convert the Lat Long into UTMs using 'sp'
reddf.sp <- SpatialPointsDataFrame(coords = reddf[,c("Longitude","Latitude")],
              data = reddf, proj4string=CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"))
        # confirmed string with VEMCO; the XY coords in azimuthal equal area

options(digits=10)
reddf.utm <- spTransform(reddf.sp, CRS("+proj=utm +zone=10 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +
 reddf.utm@data[,c("east","north")] <- reddf.utm@coords

 reddf2 = reddf.utm@data
```

Use adehabitatLT to calculate descriptive metrics for each track after primary filtering (by HPEs only)

- irregular typeII track (time is recorded, not just order, and points are not evenly distributed)
- calculates distance between points, time lapsed between points, turning angles, and direct distance from start point – Use UTM so distances are calculated in meters (versus degrees) – Note: the distance & time calculated are for the SUBSEQUENT step (i to i+1)
- add column for speed over ground at each step
- calculate mean speed across all points, mean speed for each animal, and mean of the mean fish speeds

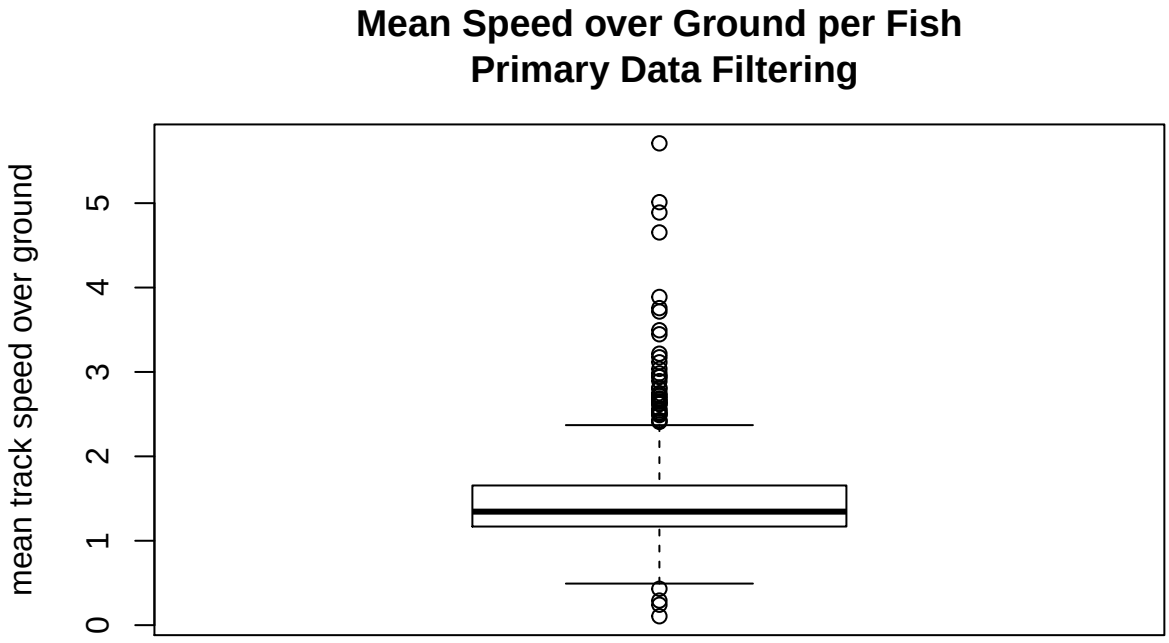
```
   red.ltraj = as.ltraj(xy=reddf2[,c("east","north")], date=reddf2$Time,
                        id=reddf2$Id, infolocs = reddf2[,c("Id","Hpes","east","north")])

   red2 = ld(red.ltraj)
   red2$spd_mps = red2$dist / red2$dt

   meanspds2 = data.frame(summarize(group_by(red2, Id), mean_mps=mean(spd_mps, na.rm=T), sd_mps=sd(spd_
```

```
## [1] "Primary data filtering: mean speed over ground (all posisions) = 1.4(SD=1.79) meters per second
```

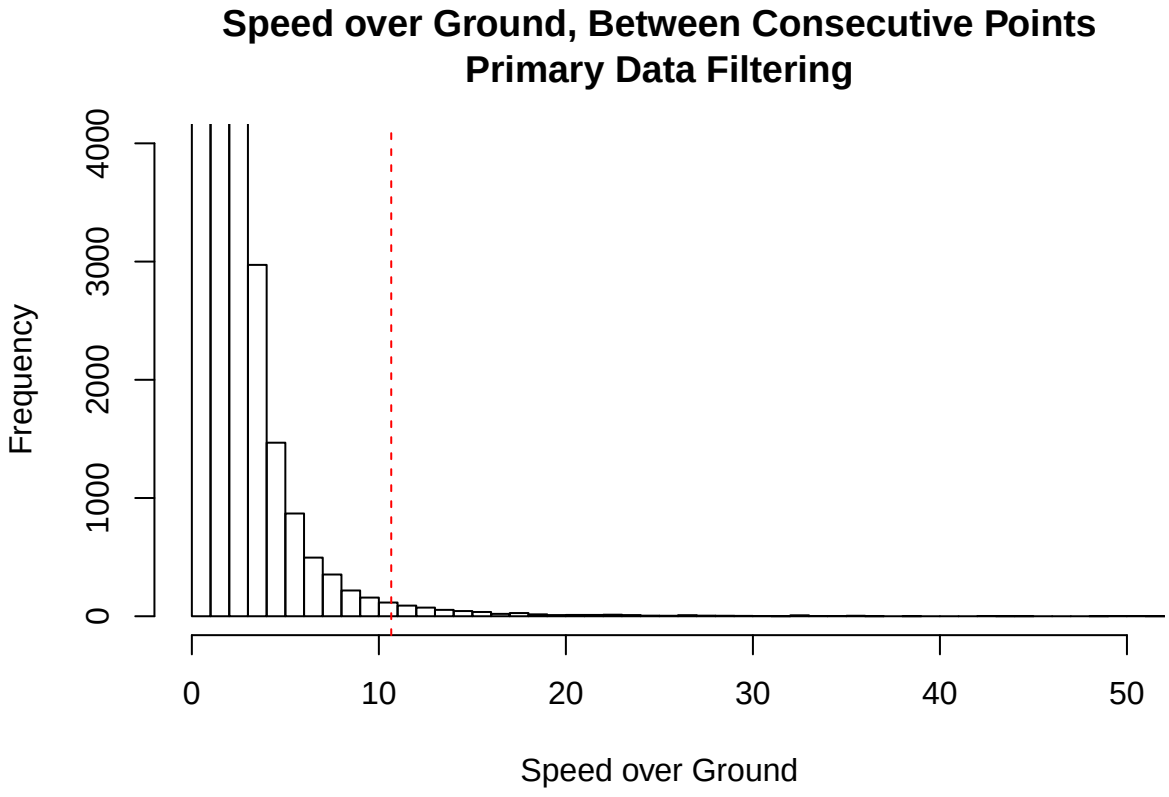
```
## [1] "Primary data filtering: mean speed over ground (mean of fish means) = 1.49(SD=1.24) meters per
```



Identify questionable detections based on speeds to and from a point

The concept that some points with low HPEs values are still erronious positions. If the speed over ground is excessive in both the prior (i-1) and subsequent (i) steps, it is likely that the position (i) has extordinarily high error and should be removed.

We will begin by considering any points as erronious with speeds greater than 7.7 (99% quantile) in both the prior and subsequent steps, then reassess the tracks to see if we need to set a lower cutoff

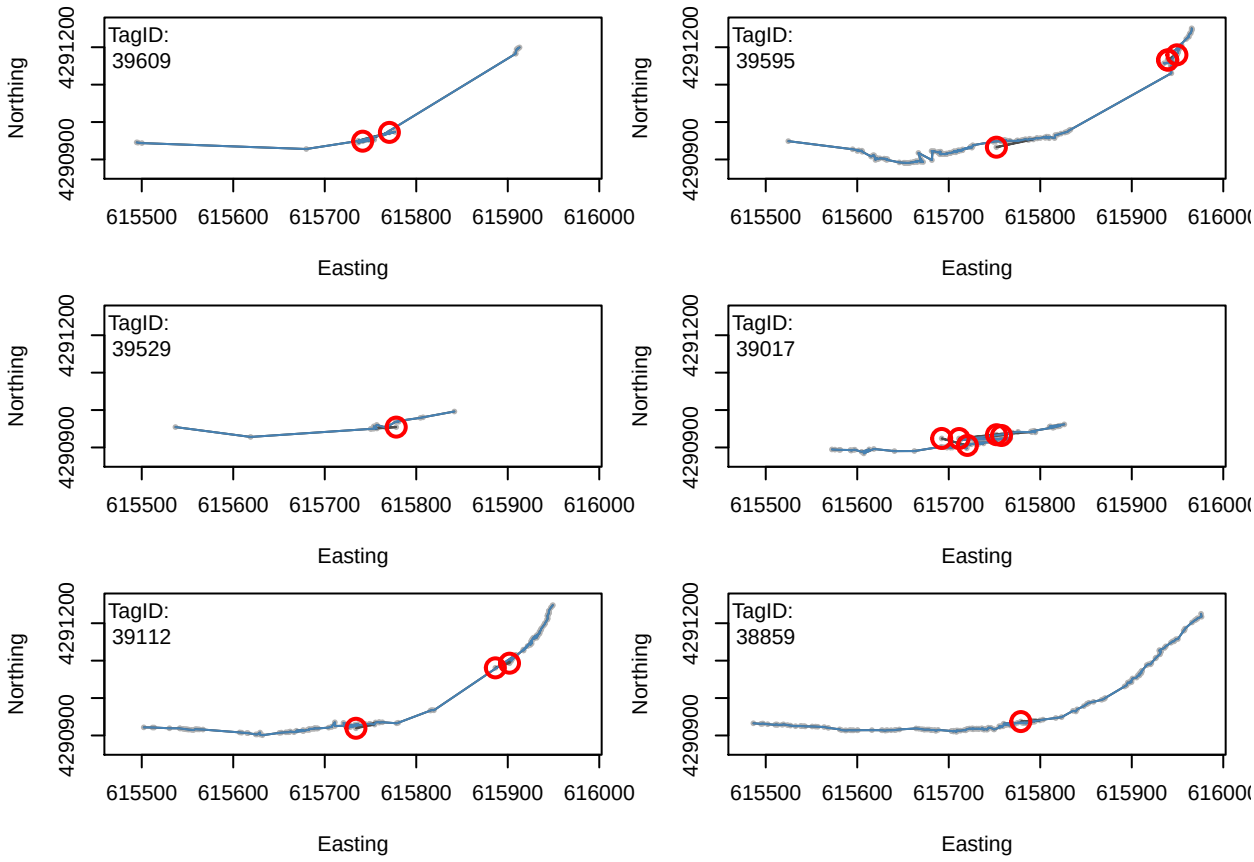



```
  red2$prevspd = lag(red2$dist)/lag(red2$dt)

  ## try filtering at the 99%ile (7.7mps)
  red2$badpos <- 0
   red2$badpos[red2$spd_mps>7.7 & red2$prevspd>7.7] <- 1
   table(red2$badpos)
```

```
##
##      0      1
## 104501    211
```

```
   # visualize a few sample tracks
   eg.badptID = unique(red2$Id[red2$badpos==1])
   par(mfrow=c(3,2), mar=c(4,4,1,1))
   for(i in sample(1:length(eg.badptID), 6))
    { egid = red2[red2$Id==eg.badptID[i],]
      plot(x=egid$east, y=egid$north, type="o", pch=16,
           xlim=range(red2$east), ylim=range(red2$north),
           xlab="Easting", ylab="Northing",
           col="grey70", cex=.6)
      text(615500, 4291200, paste0("TagID: \n",eg.badptID[i]))
      points(x=egid$east, y=egid$north, type="l",col="grey30")
      egid.f2 =egid[egid$badpos!=1,]
        points(x=egid.f2$east, y=egid.f2$north, type="l", pch=16, col="steelblue")
        points(x=egid[egid$badpos==1,]$east, y=egid[egid$badpos==1,]$north, col="red", cex=2, lwd=2)
    }
```



And remove questionable positions based on this threshold

- for the Fremont Weir analysis in 2015 we used a threshold of 5 mps based on a general estiamte of maximum biologically reasonable smolt movement speeds
- this year we've taken a more conservative approach (less extent of filtering interventions) by setting this threshold higher, and basing it upon quantiles in the data rather than expected movement rates

```
red3 = red2[red2$badpos==0,]
```

```
## [1] "Excessive-speed filter removed 211 positions, or 0.002% of recorded positions"
```

Recalculate movement speeds without these erronious positions

- expect to see very little change in central tendencies at this point because we only removed a tiny fraction of the overall positions; most of impact will be in std deviation
- at later stages of track smoothing we may see greater impacts on biologically relevant data

```
red3.ltraj = as.ltraj(xy=red3[,c("east","north")], date=red3$date,
                      id=red3$Id, infolocs = red3[,c("Id","Hpes","east","north")])

red4 = ld(red3.ltraj)
```

```
  red4$spd_mps = red4$dist / red4$dt

  meanspds4 = data.frame(summarize(group_by(red4, Id), mean_mps=mean(spd_mps, na.rm=T), sd_mps=sd(spd_
```

```
## [1] "Excessive-Speed Filtering: mean speed over ground (all posisions) = 1.35(SD=1.4) meters per sec
```

```
## [1] "Excessive-Speed Filtering: mean speed over ground (mean of fish means) = 1.42(SD=0.77) meters p
```