Lecture 6. Spatio-Temporal Data

Spatial Big Data Analysis with GIS

Korean Statistical Society, Winter School, February 24, 2023

Sea of Surface Temperatures (SST) Data

- `data_SST.mat`: contains SST data collected by satellite for the Agulhas and surrounding areas off the coast of South Africa from January 1 to November 26, 2004 (331 days)
- `SST.zone.period`: SSTs given in degrees Celsius, which is a 3D $72 \times 240 \times 331$ matrix (latitude, longitude, day)
- `lon_zone` and `lat_zone`: longitude and latitude values
- Spatial resolution: roughly 25 kilometers, though exact values depend on latitude
- Time resolution: 1 day

Note: As the data are in Matlab format, they must be converted for analysis in R. This can be done using the function `readMat` from the `R.matlab` package.

```r
### Loading the SST data
#library(R.matlab)

#Use readMat to locate the SST data in your working folder:
data.SST <- readMat("../Dataset/Lec6_tmp/data_SST.mat",
                    maxLength=NULL, fixNames=TRUE)
```

```
## Warning in readMat("../Dataset/Lec6_tmp/data_SST.mat", maxLength = NULL, :
## strings not representable in native encoding will be translated to UTF-8
```

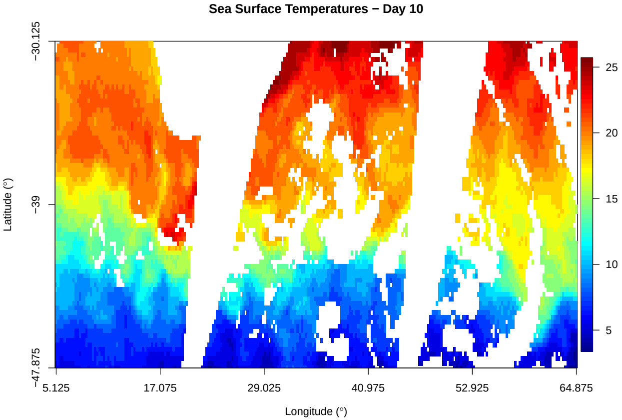
Exploratory Data Analysis

To familiarize ourselves with the geography of the dataset, we initially ignore the temporal component of the data set and examine the spatial distribution of temperatures on a single day.

The SST on Day 10

- `image.plot` in the `fields` package
- As a map, the plot uses Mercator projection, wherein lines of latitude and longitude form a regular grid.

- Pronounced temperature gradient: from highs of over 25°C in the north to a low of 3.5°C towards the southern boundary
- The mean temperature varies strongly with latitude and variance has some dependence on latitude.



Numerous gaps are present in the data, corresponding to three main causes:

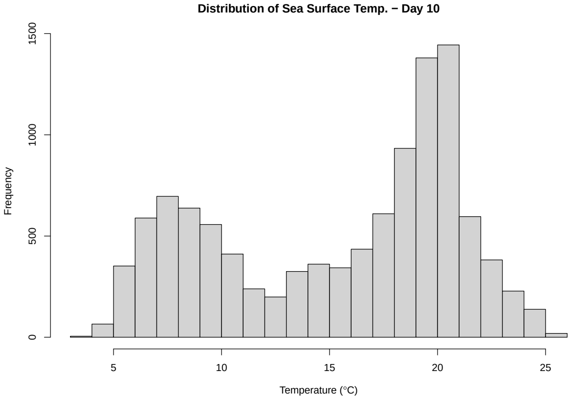
1. **land**: specifically South Africa and Lesotho, visible in the left-center of the top of the plot, as well as two small islands towards the southern boundary
2. **clipping**: two large-wedge-shaped voids cutting N-S across the picture resulting from the satellite's orbital path
3. **cloud cover**: all or most of the remaining swirls and dots present in the image

Note: Various forms of interpolation can be used to fill those gaps by orbital clipping and cloud cover.

```
#First, some initial setup:
#library(fields)
lat <- (data.SST$lat.zone)
lon <- (data.SST$lon.zone)

#Next, isolate Day 10:
SST.10 <- (data.SST$SST.zone.period[1:72, 1:240, 10])

#Now plot Day 10 (with a legend):
par(mar=c(5,5,5,5))
image(t(SST.10), col=tim.colors(25), xaxt="n", yaxt="n", ylim=c(1, 0),
      main="Sea Surface Temperatures - Day 10",
      xlab=expression(paste("Longitude (",degree,")")),
      ylab=expression(paste("Latitude (",degree,")")))
axis(1, at = seq(0, 1, by = 1/5),
     labels=seq(min(lon), max(lon), (max(lon)-min(lon))/5))
axis(2, at = seq(0, 1, by = 1/2),
     labels=rev(seq(min(lat), max(lat), (max(lat)-min(lat))/2)))
image.plot(SST.10, legend.only=T)
```

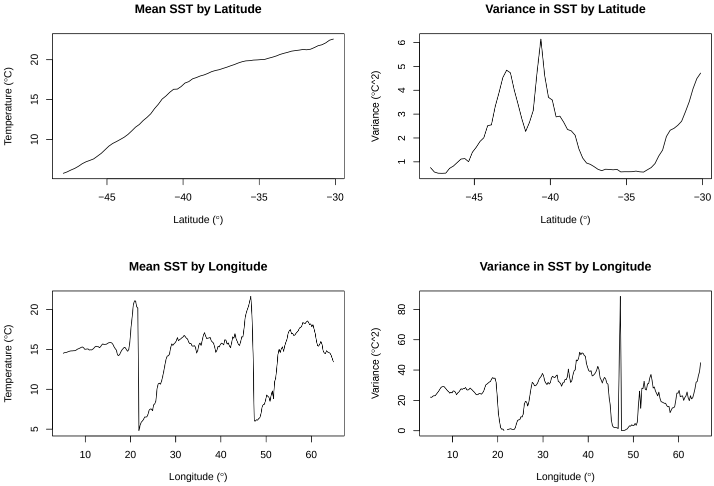
Day 10 Histogram

```
hist(SST.10, sqrt(240*2),
     main="Distribution of Sea Surface Temp. - Day 10",
     xlab=expression(paste("Temperature (",degree,"C)")))
```



- The temperatures have a bimodal distribution, with a warm peak at 20.5°C and a smaller cool peak at 7.5°C.
- The peaks correspond to the large regions of warm and cool waters at the north and south of the study area.
- A reflection of the relatively sharp transition from warm to cold in the zone mixing

Mean and Variance by Latitude and Longitude



- As latitude approach the equator, the mean temperature rises.
- Variances are highest in the zone of mixing between -44°N and -38°N.
- Both mean and variance are relatively constant across longitudes, with large deviation on both plots caused by the missing N-S bands of data.
- Variances by longitude are a order of magnitude higher because of the N-S temperature gradient.

Note: The plots of mean and variance across longitudes are rougher than those across latitudes.

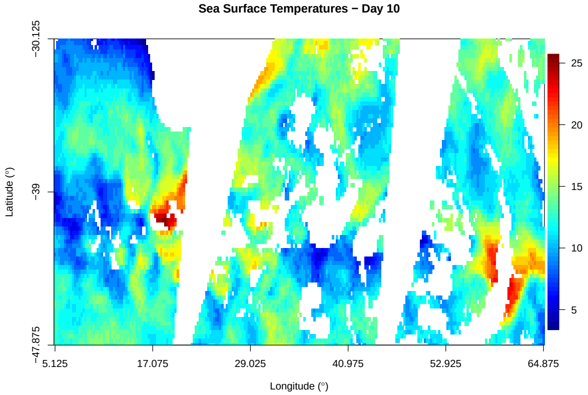
This could simply be due to the fact that with only 72 or fewer samples at any given longitude vs 240 at any latitude, there is more room for random variations to affect the longitudinal values.

```r
#To put all four figures in one plot, one can use the par command:
#library(graphics)
par(mfrow=c(2,2))

#Mean vs. Latitude:
latMean = rowMeans(SST.10, na.rm=TRUE)
plot(lat, latMean, type="n", main="Mean SST by Latitude",
     ylab=expression(paste("Temperature (",degree,"C)")),
     xlab=expression(paste("Latitude (",degree,")")))
lines(lat, latMean)

#Variance vs. Latitude:
latVar <- NULL
for (i in 1:72) {
  latVar[i] <- var(SST.10[i,1:240], na.rm=TRUE)
}
plot(lat, latVar, type="n", main="Variance in SST by Latitude",
     ylab=expression(paste("Variance (",degree,"C^2)")),
     xlab=expression(paste("Latitude (",degree,")")))
lines(lat, latVar)
```

```r
#Mean vs. Longitude:
lonMean = colMeans(SST.10, na.rm=TRUE)
plot(lon, lonMean, type="n", main="Mean SST by Longitude",
     ylab=expression(paste("Temperature (",degree,"C)")),
     xlab=expression(paste("Longitude (",degree,")")))
lines(lon, lonMean)

#Variance vs. Longitude:
lonVar <- NULL
for (i in 1:240) {
  lonVar[i] <- var(SST.10[1:72,i], na.rm=TRUE)
}
plot(lon, lonVar, type="n", main="Variance in SST by Longitude",
     ylab=expression(paste("Variance (",degree,"C^2)")),
     xlab=expression(paste("Longitude (",degree,")")))
lines(lon, lonVar)
```

Removing Latitudinal Means

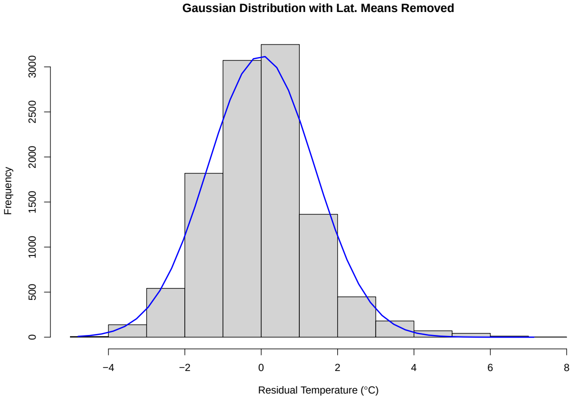
- Clear trend between mean temperature and latitude
- We subtract the latitudinal means from the Day 10 data.



```r
x <- rep(lon,each=72); y <- rep(lat,times=240); z <- c(SST.10)
XYZ <- cbind(x,y,z)

#The process for removing each latitudinal mean directly
M <- rep(latMean,times=240)
z.M <- z-M
XYZM <- cbind(x,y,z.M)
SST.10.M <- matrix(z.M, nrow = 72, ncol = 240)

#Plot results:
par(mar=c(5,5,5,5))
image(t(SST.10.M), col=tim.colors(25), xaxt="n", yaxt="n", ylim=c(1, 0),
      main="Sea Surface Temperatures - Day 10",
      xlab=expression(paste("Longitude (",degree,")")),
      ylab=expression(paste("Latitude (",degree,")")))
axis(1, at = seq(0, 1, by = 1/5),
     labels=seq(min(lon), max(lon), (max(lon)-min(lon))/5))
axis(2, at = seq(0, 1, by = 1/2),
     labels=rev(seq(min(lat), max(lat), (max(lat)-min(lat))/2)))
image.plot(SST.10, legend.only=T)
```

Histogram with Gaussian Curve

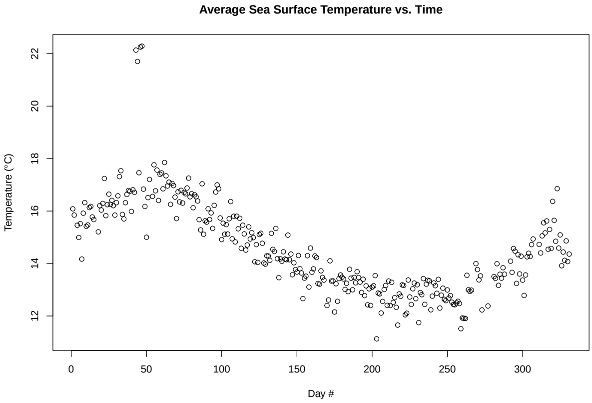
- The residuals have a nearly Gaussian distribution (with mean $= 0$ and variance $= 1.96$).



```r
#Histogram of residuals with Gaussian Curve:
h <- hist(z.M, main = "Gaussian Distribution with Lat. Means Removed",
          xlab = expression(paste("Residual Temperature (",degree,"C)")))

#Normal curve overlay:
xfit <- seq(min(z.M, na.rm=TRUE),max(z.M, na.rm=TRUE),length=40)
yfit <- dnorm(xfit,mean=mean(z.M, na.rm=TRUE),sd=sd(z.M, na.rm=TRUE))
yfit <- yfit*diff(h$mids[1:2])*(length(z.M)-6335)
lines(xfit, yfit, col="blue", lwd=2)
```

Exploratory Temporal Data Analysis

Average SST vs Time



A quick comparison of mean SSTs across the entire study area on each day

- Clear seasonal curve
- Temperatures highest around Day 60 in early March, and lowest around Day 240 in early September

Obvious outliers, with average temperatures around 22°C, occur just prior to Day 50.

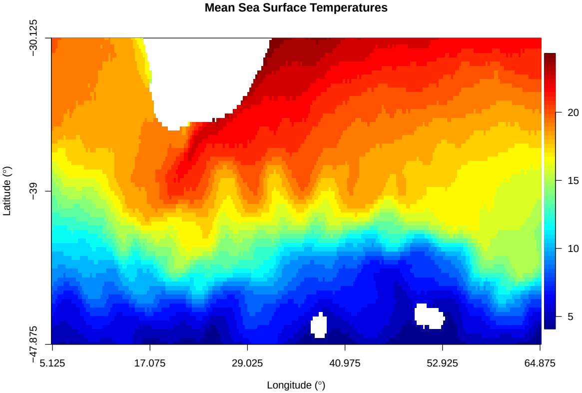
- They may result from some error in the data collection (values in the southern regions are missing completely).

Note: These outliers do not reflect the true nature of the dataset, and should be removed for any relevant model fitting or analysis.

```
#Average SST vs. Time:
Day <- 1:331
MeanT <- NULL
for (i in 1:331) {
  MeanT[i] <- mean(data.SST$SST.zone.period[1:72, 1:240,i], na.rm=TRUE)
}
plot(Day, MeanT, main="Average Sea Surface Temperature vs. Time",
     ylab=expression(paste("Temperature (",degree,"C)")), xlab="Day #")
```

Mean SSTs

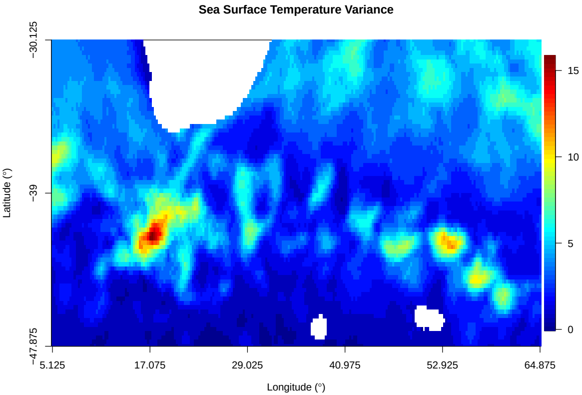
- The distribution of means deviates from a N-S gradient.



- The direction of the Antarctic Circumpolar current: west to east, as the Agulhas waters are pushed eastward when they move south.
- Cooler waters on the eastern edge of the figure hint at northward flow and thus the counter-clockwise direction of surface water circulation in the southern Indian Ocean.
- A large-scale feature known as the Indian Ocean Gyre

```r
#We can decipher land by taking the mean temperature
#throughout the 331 days at each point:
MeanT.Map <- matrix(0, 72, 240)
for (i in 1:72) {
  for (k in 1:240) {
    MeanT.Map[i,k] <- mean(data.SST$SST.zone.period[i, k, 1:331],
                           na.rm=TRUE)
  }
}
#Mean SSTs
par(mar=c(5,5,5,5))
image(t(MeanT.Map), col=tim.colors(25), xaxt="n", yaxt="n", ylim=c(1, 0),
      main="Mean Sea Surface Temperatures",
      xlab=expression(paste("Longitude (",degree,")")),
      ylab=expression(paste("Latitude (",degree,")")))
axis(1, at = seq(0, 1, by = 1/5),
     labels=seq(min(lon), max(lon), (max(lon)-min(lon))/5))
axis(2, at = seq(0, 1, by = 1/2),
     labels=rev(seq(min(lat), max(lat), (max(lat)-min(lat))/2)))
image.plot(MeanT.Map, legend.only=T)
```

SST Variance

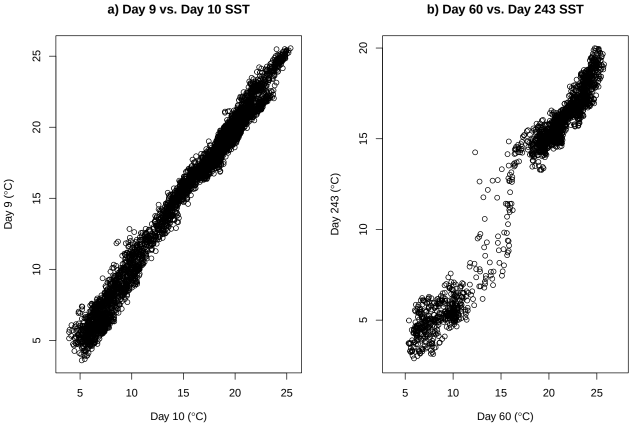


- The map shows that the Agulhas Current itself is fairly steady, forming only a light cyan line (variance ≈ 6) extending SW from South Africa.
- It shows zones where temperatures can fluctuate most dramatically due to mixing warm and cold waters, in the zone from 39°S to 44°S.

```r
##SST Variance
VarT.Map <- matrix(0, 72, 240)
for (i in 1:72) {
  for (k in 1:240) {
    VarT.Map[i,k] <- (sd(data.SST$SST.zone.period[i, k, 1:331],
                                 na.rm=TRUE))**2
    }
}
par(mar=c(5,5,5,5))
image(t(VarT.Map), col=tim.colors(25), xaxt="n", yaxt="n",
      ylim=c(1, 0), main="Sea Surface Temperature Variance",
      xlab=expression(paste("Longitude (",degree,")")),
      ylab=expression(paste("Latitude (",degree,")")))
axis(1, at = seq(0, 1, by = 1/5),
     labels=seq(min(lon), max(lon), (max(lon)-min(lon))/5))
axis(2, at = seq(0, 1, by = 1/2),
     labels=rev(seq(min(lat), max(lat), (max(lat)-min(lat))/2)))
image.plot(VarT.Map, legend.only=T)
```

SSTs on Different Dates

- A clear temporal correlation exists.



```r
#Compare sea surface temperatures on different dates
#Adjacent dates:
SST.9 <- (data.SST$SST.zone.period[1:72, 1:240, 9])
SST.11 <- (data.SST$SST.zone.period[1:72, 1:240, 11])

#Farthest Pair:
SST.60 <- (data.SST$SST.zone.period[1:72, 1:240, 60])
SST.243 <- (data.SST$SST.zone.period[1:72, 1:240, 243])

#Plot results:
par(mfrow=c(1,2))
z9 <- c(SST.9); z11 <- c(SST.11); z60 <- c(SST.60); z243 <- c(SST.243)
plot(z, z9, main="a) Day 9 vs. Day 10 SST",
     ylab=expression(paste("Day 9 (",degree,"C)")),
     xlab=expression(paste("Day 10 (",degree,"C)")))
plot(z60, z243, main="b) Day 60 vs. Day 243 SST",
     ylab=expression(paste("Day 243 (",degree,"C)")),
     xlab=expression(paste("Day 60 (",degree,"C)")))
```

Reference

- Berdahl, J. S. and Genton, M. G. (2016). Spatio-Temporal Analysis of Sea Surface Temperatures: A Tutorial.