# Post-Processing GEC Data

The standardization of the GEC gave a mostly representative and consistent picture of the distribution of African savannah elephants. However, not all survey regions such as Kruger National Park did deliver data at all or some did follow different guidelines. Other datasets were performed as a total count or block count instead of a transect flight. Therefore, several datasets could not be included into this study.

As Survey teams used different observation codes, geographic systems, time formats etc. – harmonization was required, in order to combine the whole dataset. Furthermore some spreadsheets had double column names and similar. Those issues have not been solved in the raw data but within R so the post-processing was strictly performed without manual corrections within the spreadsheets. Every subset of this chapter is represented by a single R-Script.

## Harmonization (harmonize\_RSOD.R)

The RSOD spreadsheet revealed at least some deviation from the standard protocol and some loops checked for different errors such as:

* Some data sets have redundant column names, a loop checks all datasets for double column names and adjusts the integer for the skip parameter in read.csv().
* COD\_VIR has two columns for lon and lat respectively. A general check looks for cases where lon and lat are doubled and merges the respective columns
* survey code names were missing in some cases and were deduced from the file names
* UTM coordinates of COD\_GAR were converted to EPSG 4326, if there are more datasets with UTM coordinates this approach has to be adjusted, as the UTM coordinate systems are diverging.

Different observations codes were summarized. All observation codes could be separated into the categories “mh”, “bh”, “car1” – “car4” and “elephant unkown”. The last category was not necessary as all observations could be separated into mixed herds und bull herds (mh/bh).

Some survey did not photo corrected count or did at least not mention it in the respective column. Sometimes photo corrected counts but no observed counts were given. In one case (…?) counts were in a different column.   
The output of this step was a SpatialPointsDataFrame with all observations combined with information about lon/lat, observation code, counts and the like.

## Segmentization (CmbTransects\_segmentize.R)

In order to create comparable areas, the observation fields left and right to the transects was divided into equally spaced segments with 2500m in length. All transects were converted to a metric coordinate system (EPSG: 102024). The end points of the transects were extended by

Currently this mathematical approach is replaced by a numerical approximation that minimizes eq. 4) with a fixed . A sequence from to with a step length of 2500m returned the start and end points of the single segments in lengthwise direction. The width of the fields of observation should be calculated from the height above ground level. However, this was given only in rare cases so a target height of 180m was assumed when calculating the area the of observation. The width of the segments left and right to the transect lines was therefore set to 150m. The segments were reprojected to the geographic coordinate system EPSG 4326.

The output of this step were equally spaced segments (EPSG 4326) that are aligned to all given transects. Only the spatiall information is of interest at this point.

## Datetime (datesfromgpx.R)

Datetimes could only be assigned to segments that could in turn be connected to an observation. Therefore .gpx files were used to connect spatial points with a datetime attribute to transects, by calculating the median of the datetime attributes of .gpx-file-points that are closest to the individual centroids of the segments of the transects.

This might lead to a small error if a single transect was surveyed on separate flights/dates. As some .gpx-files did not contain dates (BWA\_NOR, KEN\_LAM, XWA\_TBC), datetimes of the FSOD- (BWA\_NOR) or RSOD-spreadsheet (KEN\_LAM, XWA\_TBC) were used instead. In some cases (AGO\_LUI, COD\_GAR, COD\_VIR, KEN\_TSV) the gpx.files did contain dates but they were not in a common format and had to be adjusted.

# Extracting Covariates / Response

## Preprocessing

GEE has been used to extract covariates. As not all the products have been available in the archive, they were pre-processed in R.

### Shapefile of Africa (createafrica\_shape.R)

A combined .shp of Africa is created using GADM level = 0 files and combining them via. ISO3 country codes to a shapefile of Africa. The shapefile was used to clip the following covariates.

### Protected areas (??)

Using the WDPA data set, a binary raster file of the distribution of African protected areas was created and uploaded to GEE.

### Gridded Livestock of the world (GLW; preprocessing\_GLW\_creation.R)

Using the GLW data set a one-band raster file representing the density of livestock as Tropical Livestock Units (TLU). Maps of the density of cattle, sheep and goats were weighted by their individual coefficient and the sum of those 3 layers was uploaded to GEE as the final product.

## Extraction

### In R

Herd type & count data (extract\_COUNT-HT.R)  
Observations of the RSO spreadsheet submitted by the single GEC survey teams were assigned to the ID of previously created segments. If all segments have been further apart than 2000m, the observation has been dismissed (about 100 data points). Furthermore the temporal distance between RSO observations and .gpx files was calculated (s. Datetime chapter). If the difference was greater than 14 days the observation has also been dismissed (additional 100 data points were lost). The same goes for count data.   
As there is the possibility of more than observations being connected to one segments, mixed herds and bull herds could both be observed. In such cases the segments has been assigned the observation code “mh” for mixed herd. The sum of the count data has been calculated

### Using GEE

#### Poacher Cost

#### Protected Areas

#### Livestock Density

#### Daily Cycle

#### Seasonal Cycle

#### Agricultural Intensity

#### Human Density

#### Tree Coverage

#### Vegetation Density

#### Water availability

#### Natural Barriers

#### Below Ground Soil moisture

#### Ruggedness

# Modelling

## Collect covariates (accessGoogleDrive.R)

The exports by GEE are stored in the GEC Google Driver folder and can now be accessed using the googledrive-package within R. As previous exports are not overwritten, only the newest version of the covariates are downloaded into a single folder and read into R. Some of the data