README for Geese Migration Analysis Workflow

This guide explains how to process geese migration data using R and Python scripts across multiple folders. Users will work with the `3_Autumn`, `4_Spring`, and `5_Clustering` folders to analyse and interpret the results. Preprocessing and annotation steps are already completed, and the final datasets are available for download.

Folder Structure Overview

`1_Preprocessing` (preprocessing done by the author):

Removed GPS fixes with insufficient satellite data, filtered low-quality tracks, and aggregated datasets.

Excluded dependent individuals and retained only active migratory data.

`2_Annotation` (annotation done by the author):

Annotated data with environmental variables (e.g., wind, elevation, geomagnetic data).

Final annotated datasets are available for use.

`3_Autumn` and `4_Spring`:

Process seasonal and diurnal migratory datasets for Autumn and Spring.

`5_Clustering`:

Python scripts for clustering analysis:

`1_Validation_Indices.py`: Calculates clustering validation indices (Silhouette and Calinski-Harabasz).

`2_AHC_Clustering.py`: Performs Agglomerative Hierarchical Clustering and generates dendrograms and cluster IDs.

Workflow Guide

Step 1: Starting with 3_Autumn and 4_Spring

1. Download the final datasets.
2. Navigate to the `3_Autumn` and `4_Spring` folders.
3. Run the R scripts **sequentially**, starting with:

- o `1_Annotated_Track_Process.R` → Preprocess annotated GPS data.
- o `2_Day_Night_Classification.R` → Classify data points as day or night.
- o `3_Wind_Parameters.R` → Calculate wind support and crosswind values.
- o `4_Magnetic_Parameters.R` → Calculate magnetic headings and apparent inclination.
- o `5_Clustering_Features.R` → Calculate features for clustering analysis.
- o `6_Deviations.R` → Compute angular deviations.
- o `7_Before_Clustering.R` → Filter and scale data for clustering analysis.

4. Outputs from **`7_Before_Clustering.R`**:
   - o **3_Autumn**:
     - `Autumn_Day_CF.csv`
     - `Autumn_Night_CF.csv`
   - o **4_Spring**:
     - `Spring_Day_CF.csv`
     - `Spring_Night_CF.csv`

Copy these files to the `5_Clustering` folder for clustering analysis.

Step 2: Clustering in 5_Clustering

1. Navigate to the `5_Clustering` folder.

2. Run the Python scripts in order:

   - o **`1_Validation_Indices.py`**:
     - Calculates Silhouette and Calinski-Harabasz validation indices for clustering quality.
     - Command: `python 1_Validation_Indices.py`
     - Output validation plots for each dataset.

   - o **`2_AHC_Clustering.py`**:
     - Performs clustering using Agglomerative Hierarchical Clustering (AHC).
     - Generates cluster IDs and dendrograms.
     - Command:
     - `python 2_AHC_Clustering.py`
     - Outputs:
       - Clustered CSV files:
         - `Autumn_Day_with_ClusterID.csv`
         - `Autumn_Night_with_ClusterID.csv`
         - `Spring_Day_with_ClusterID.csv`
         - `Spring_Night_with_ClusterID.csv`
       - Dendrogram visualizations:
         - `Dendrogram_Autumn_Day.png`
         - `Dendrogram_Autumn_Night.png`
         - `Dendrogram_Spring_Day.png`
         - `Dendrogram_Spring_Night.png`

3. Move the clustering results back to their respective seasonal folders:
   - o **Autumn** (`3_Autumn/8_After_Clustering`):
     - `Autumn_Day_with_ClusterID.csv`

- Autumn_Night_with_ClusterID.csv
  - **Spring** (4_Spring/8_After_Clustering):
    - Spring_Day_with_ClusterID.csv
    - Spring_Night_with_ClusterID.csv

Step 3: Post-Clustering Analysis

1. Return to the 3_Autumn and 4_Spring folders.

2. Continue running scripts sequentially from 8_After_Clustering.R onwards:
   - 8_After_Clustering.R: Join clustering IDs with other variables.
   - 9_Density_Plots.R: Visualize key variables across clusters.
   - 10_chi_squared_test.R: Test for independence of variables across clusters.
   - 11_Boxplot_MP_Dev.R: Analyse movement parameters and deviations.
   - 12_Deviation_Analysis.R: Perform statistical analysis of deviations.
   - 13_CircularHistogram.R: Generate circular histograms of headings.

Key Notes

- **Preprocessed Data**:
  - Preprocessing and annotation are completed; final datasets are available for download.
  - Users start directly from 3_Autumn and 4_Spring.
- **Clustering Outputs**:
  - Results from the 5_Clustering folder must be copied back to the respective seasonal folders before continuing with interpretation.
- **Dependencies**:
  - **R**: Ensure required libraries (maptools, dplyr, etc.) are installed.
  - **Python**: Install required packages using:
  - pip install numpy pandas scipy matplotlib scikit-learn

For questions, contact am636@st-andrews.ac.uk.