Python in GIS

Group 08

Analysing 'Flight Distances' of White Fronted Goose

PROJECT REPORT

24 July 2020

Authors:

Anu Shrestha, Christian Carranza, Edwige Mukundane & Mateen Mahmood

1 INTRODUCTION

Bird's migration is a common phenomenon generally observed twice a year. Mostly it is the travel between liveable habitat and a reproduction environment (Drent et al., 2003). Geographical extent and temporal occurrence of these migrations vary based on many aspects and is mostly in relation with the photoperiod (van Wijk et al., 2012). Range of influencing factors include season, hunting sites, plant growth and many more (Kölzsch et al., 2016).

White fronted geese hereafter 'geese' depicts a similar pattern of migration, however being 'capital breeders' by their nature (meaning they acquire energy required for the breeding during migration) is influenced by 'plant growth' as a critical factor (Spaans et al., 1999). Plant growth itself relies on aspects like season, temperature, rainfall and snow melt, which are all indirect factors influencing geese migration (van Wijk et al., 2012).

Apart from various influencing factors, 'distance' and 'time' are the core components of migration studies. Teitelbaum et al., (2015) suggests that distance, timing and period of migration give a richer understanding of migrational patterns, direction of movement, routes taken and physiological adaptations of birds. Continuing research has explored these traits in detail, where Morbey & Hedenström, (2020) considered departure time, flight distance and travel speed to explore migration schedules; Moore, (1996) studied departure time from a stopover site and related it to estimated flight ranges and Kölzsch et al. (2018) formalised a migration network with stopover (nodes) that are linked by flight distances (links).

In this study, we focused on exploring distance and time, excluding all other factors, to realize how these two core components are related to each other. As geese rest at multiple stopover sites to consume energy en route to the breeding site, adding the duration information at these stopovers can also assist in identification of its relation with the following flight distance (van Wijk et al., 2012)

Migration distance is the total distance from the current habitat to the place where the geese rest for a prolonged duration (many months); however, 'Flight Distance' is the distance between potential stopovers during migration (Ojaste et al., 2019). This study focuses on whether these flight distances are affected with the time geese leave the stopover, and whether the duration of rest at the stopover site influences the travelled distance of the following flight.

Research aims to explore the relation between Departing Time and Flight Distances and further intends to answer the question, "Do white geese leaving early morning fly more distant?" and a subquestion, "Do they fly more distant after a good rest period?"

2 METHODOLOGY

Input dataset is acquired from Movebank data repository (Kölzsch et al., 2018)(Kruckenberg et al., 2018), for 7 geese; recorded between the years 2007 to 2008. Geese are recorded from their wintering grounds around the Netherlands, migrating towards their breeding habitat in different parts of Russia. Total of 7363 recordings are available in the form of points hereafter 'nodes' with time of recording and the gps location, fulfilling the requisites of spatio temporal analysis.

METHODS

Answering the research question requires identification of potential stopovers and respective departing nodes for each identified stopover; whose departing time and following (flight's) distance are the variables to be discussed. Similarly, once a stopover is delineated, its stopover duration can also be computed which will be the difference of time from last node to first node of the same stopover. See Figure-1 which describes this conceptual workflow:

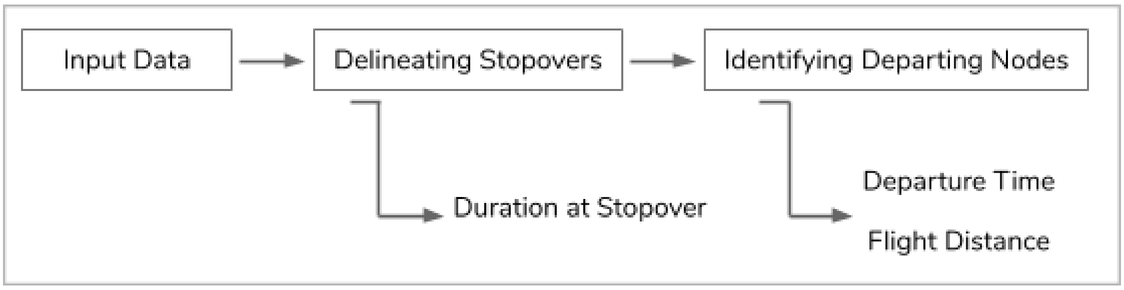

Figure 1 - Conceptual Workflow

van Wijk et al., (2012) and Kölzsch et al., (2016), both, suggest delineation of stopovers as a two step process where a stopover will consist of all those successive nodes where geese did not displace more than 30 kms between pairs of positions, and secondly geese stayed for at least 48 hours.

This two step process can be achieved with the application of 'time parameterized line segment clustering method' which respects the order of nodes. This clustering method considers timestamps of nodes and segregates them into different clusters, even if they are spatially adjacent (as they are not successive) (Xu et al., 2013). Once the stopovers are identified, Tang et al., (2011) suggests to spatially generalize a stopover by using its mean (average) location as cluster representative.

Figure-2 illustrates this clustering process and extends the conceptual flow into an implementation workflow followed for the execution of the study.

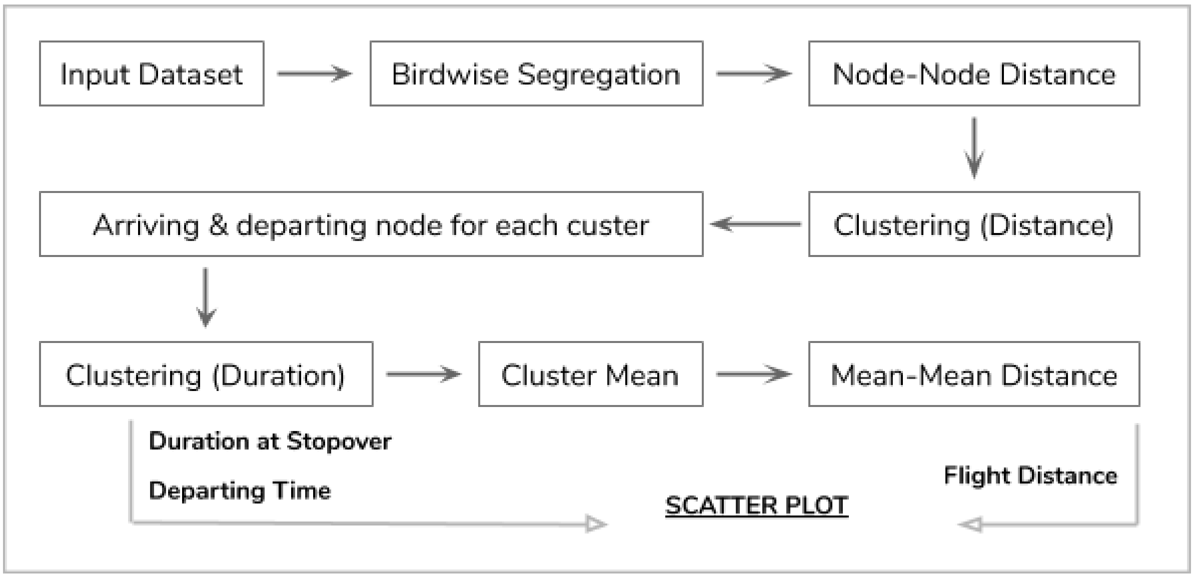

Figure 2 - Implementation Workflow

Explanation: Firstly, the input dataset was segregated birdwise. Time parameterized clustering can be manually achieved by applying both steps explicitly. Step 1 included computation of node to node distance, and identifying temporary clusters based on 30 kms condition. In step 2, for the identified temporary clusters, duration was computed for each cluster as the difference of timestamps between last node and first node. Afterwards, final clustering was assigned based on deduced stopover durations. As a representation of clusters, the mean of all clusters were identified. Last node of each cluster was marked as the departing node of that cluster which included the departing time from that cluster (stopover site). Last node also held the computed duration and the mean to mean distance from executed steps of clustering. Hence, all three variables of departing time, flight distance and duration of stopover were obtained from this implementation.

As data is recorded periodically and does not provide with the clear indication of when exactly the geese departed, (Smetzer & King, 2018) considered last observed time as departure time to next stop. Furthermore, as data is only recorded on hour marks (e.g 0700hrs, 0900hrs), and not with minute level details, hence departure time can be represented as 24 bins of each hour mark. Both these considerations were implemented in this study for the deliberation of departing time from stopover.

IMPLEMENTATION

Study was executed as QGIS based Script, as the implementation of discussed methodology. QGIS's processing functionalities were consumed for computation of node to node distances, identification of mean coordinates, and joining multiple tables. Selection by an expression was used while working with the subset of data, be it a specific cluster or a particular goose. Matplotlib was used for the production of results.

3 RESULTS

An intermediate result is of clustering, which can be observed spatially and can greatly help to understand the overall migration pattern, stopover delineation and flight distances. Figure-3 illustrates such clustering in form of a spatial plot for one goose (ID: 79694).

Result 01 - Spatial plot of single goose

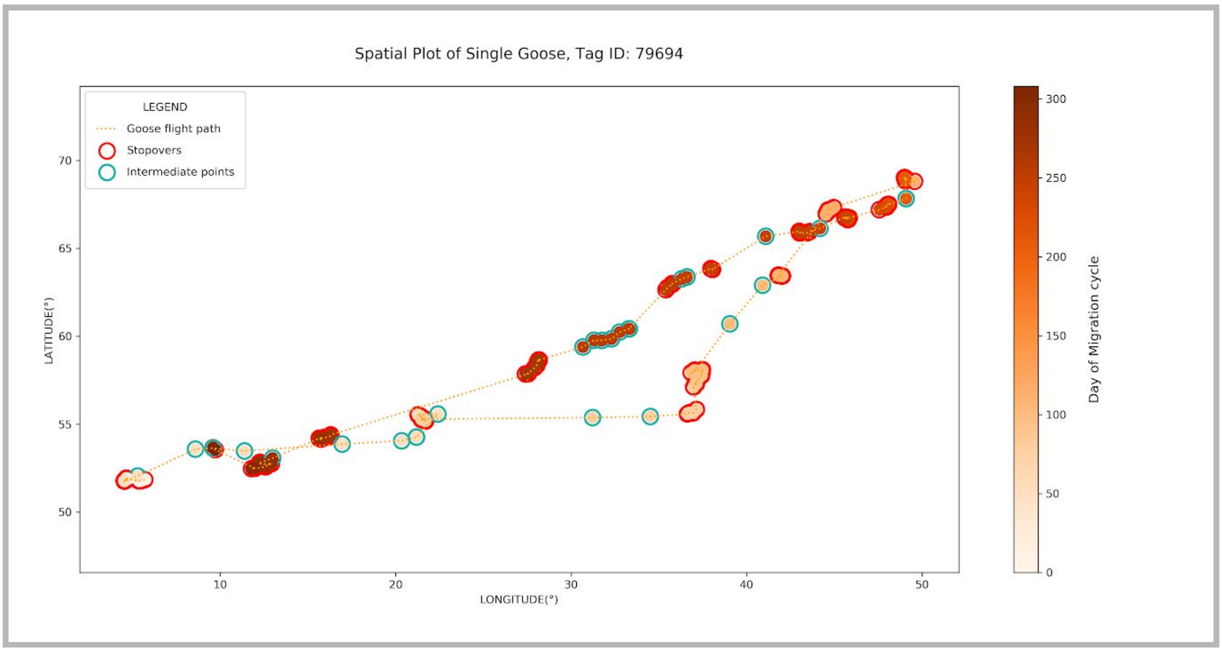

Figure 3 - Spatial plot representing stopovers, intermediate points, flight path and period of migration

Description: It is a plot of coordinates with longitudes on horizontal axes and latitudes on vertical axes. Stopovers are represented with red borders, whereas intermediate points are illustrated with teal (greenish) borders. Dashed line connects all the points in the form of trajectory highlighting the goose's flight path. Migration starts at the leftmost point (5.35°, 51.77°), with colors of all points based on the color ramp (on right) corresponding to the day of migration. Hence, as the goose moves northeast (direction of spring migration), the color of each point (cluster or intermediate) turns darker in shade.

Interpretation: Plot highlights that the goose (ID: 79694) travelled long distances between stopovers with an average of 360 ± 328 kms. Average duration at stopover of this goose is 243 hours which is relatable with the 270 hours average of all seven geese. Plot also identifies that generally the route for the spring migration and winter migration are over the same space, which relates with geese 'capital breeder' nature to fly over certain regions supporting their sustenance. Figure-4 is a similar spatial plot but for all geese to compare these among themselves.

Result 02 - Spatial plot of all seven geese

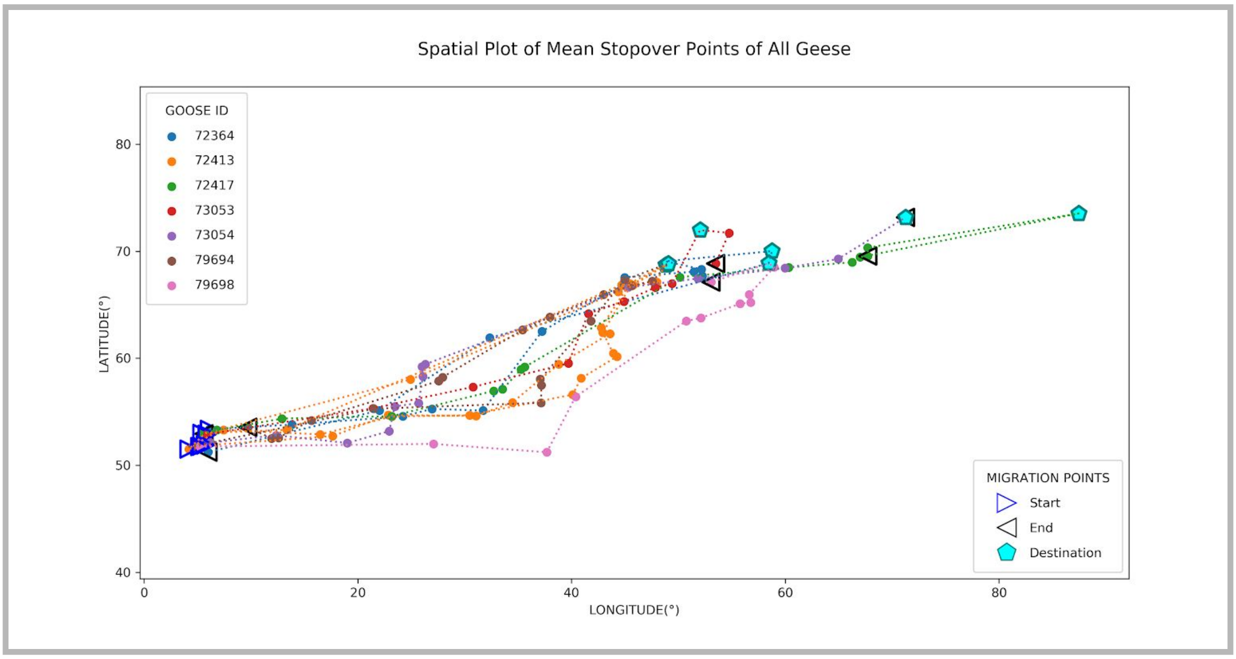

Figure 4 - Spatial plot for all seven geese

Description: Plot illustrates mean stopover points of all clusters for all seven geese (rendered in different colors) with legend displaying color associated with each goose id. Plot further highlights the starting and ending point of the recorded trajectory; and calls attention to the destination point in a blue pentagon.

Interpretation: Plot validates the idea previously discussed, that all geese follow a similar route over the same space in their to and fro movement of migration. However, destinations are different in general. Destination points on the plot are the locations of stopovers where the geese stayed for the longest duration of time. Even though the five of the geese settled in locations from 48° to 58° east, two geese (ID: 73504 and 72417) travelled further to 71° and 87° east. Literature review shows that after the flock of geese settle at the breeding site, non-breeder and failed breeder geese travel east towards the moulting sites and spend more than a month there before returning to wintering grounds (Kölzsch et al., 2016). For the breeding geese and the family, the moulting site is the same or near the breeding site (Hearn, 2004; Kölzsch et al., 2016). Table-1 entails summary statistics associated with spatial plot of all geese.

RELATIONSHIP - DEPARTURE TIME AND FLIGHT DISTANCE

In order to observe the relationship between departing time and flight distance, scatter plot is a suitable option as both variables are continuous. Departing time, which is the departing hour mark, is represented in the form of 24 ticks in order. Such a scatterplot is available in Figure-5.

Geese ID	Total Stops	Total Migration Distance (kms)	Maximum Flight Distance (kms)	Minimum Flight Distance (kms)	Average Flight Distance (kms)
72364	15	**10870**	1423	69	**496 ± 407**
72413	**33**	10670	**2244**	53	386 ± 319
72417	16	4339	1179	50	442 ± 341
73053	15	4090	1029	64	426 ± 284
73054	10	3721	1242	53	358 ± 285
79694	12	8623	1183	**44**	360 ± 328
79698	20	2204	1496	47	446 ± 445
Average	**17**	**6360**	**1400**	**54**	**416 ± 344**

Table 1 - Summary Statistics related to Spatial Plot

Result 03 - Scatter plot of all geese with color based on duration of stay

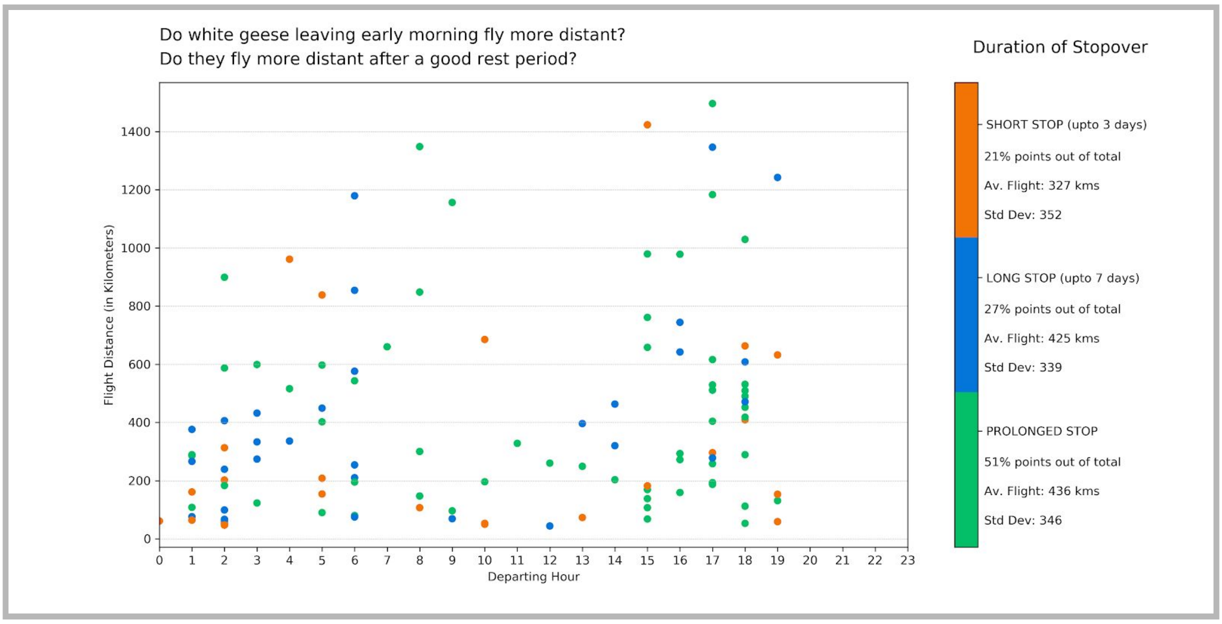

Figure 5 - Scatter plot (departing hour vs flight distance) of all geese with colors based on duration

Description: Departure hour is available on the principal axis starting from 00:00 hrs to 2300 hrs, whereas the axis of ordinates holds the scale of flight distance in kilometers. Colors are based on duration of stopover whose range is classified into three categories on 'manual' basis following guidelines suggested by Smetzer &

King, (2018) and Bawden et al., (2019). Three classes are Short stop (upto 3 days), Long stop (upto 7 days) and Prolonged stop (greater than 7 days).

Interpretation: Plot indicates no definite trend between the two variables under observation. Plot shows that there are few flights around noon hours and almost no flight before midnight. Plot also illustrates that geese tend to depart frequently early morning (0500-0600 hrs), also in the evening (1500-1800 hrs), however no trend is visible if they fly more distant or not. As we focus on the colors to explore the influence of the rest period on the following flight, again there is no explicit pattern. More than half of points (51%) are in the category of prolonged stopover, whereas long and short stopovers have 27% and 21%, respectively. Nonetheless summary statistics suggest that average flight distance of 436 ± 346 kms following a prolonged stopover is similar to long stopover's 425 ± 339 kms; both of which are considerably greater than short stopover's 327 ± 352 kms.

Figure-6 focuses on a similar relationship to identify if there is a particular goose which follows a different pattern and provides a hint of tendency as answer to the research question.

Result 04 - Scatter plot of all geese

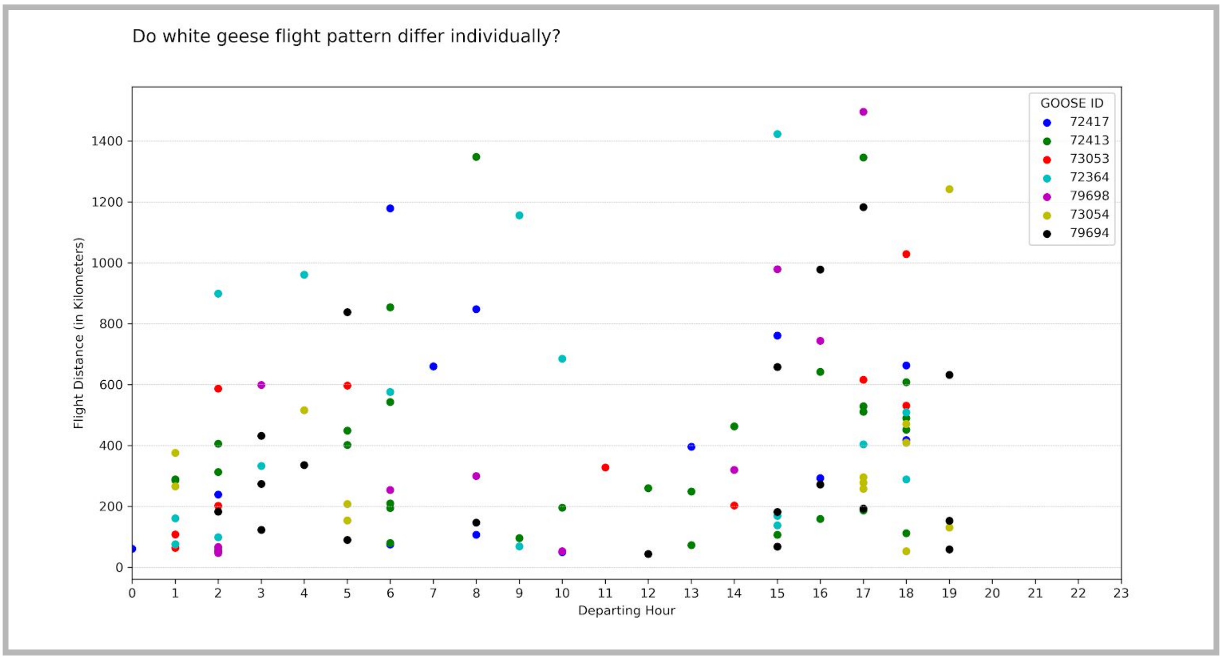

Figure 6 -Scatter plot (departing hour vs flight distance) of all geese

Description: Plot is exactly the same as figure-5, except the rendering, which is now based on geese wise distribution rather than stopover wise as in last figure.

Interpretation: Again, there is no definite pattern to support the hypothesis that geese departing early morning fly more distant, nor any other indication to know what time of departure is associated with higher flight distance.

Scatter Plot highlights that there is no particular relation between Departing hour and Flight distance. Same for all geese as shown in figure-6. However, duration of stopover does have an influence over the following flight distances (based on figure-5). To explore further the average distances for each class of stopover, a group bar chart can help observe available trends. See figure-7 for such a bar chart.

Result 05 - Bar chart for average distances for each class of stopover

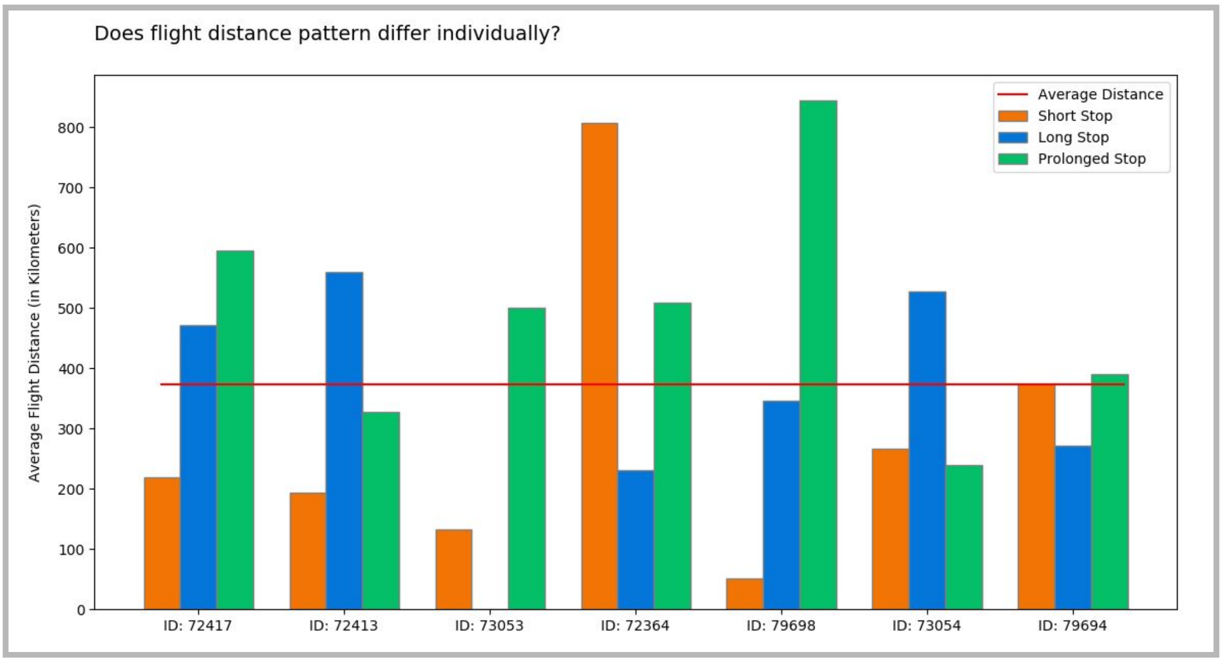

Figure 7 -Group bar chart for all geese illustrating their average distances for each stopover type

Description: Group bar chart represents each goose on the principal axis with their travelled average flight distances on the vertical axis. Groups illustrate each of the three stopover categories of short, long and prolonged in orange, blue and green, respectively. Average of all geese for all types of stopovers is also available (redline) as a reference to relate all average flight distances in general.

Interpretation: Plot indicates no definite pattern of whether a stopover duration is directly associated with more flight distance. Generally, prolonged duration's average distance is higher compared to others for most of the birds, however, it is not easy to claim that flight following a prolonged duration will be of greater average than others.

4 DISCUSSION

Identified results of no definite relationship between departing hour and flight distance, is also concluded by (Morbey & Hedenström, 2020) and (Moore, 1996) in case of birds like Piranga Rubra; and for white fronted geese by (Weegman, 2014) and (van Wijk et al., 2012). Greater duration of stopover tends to influence a longer flight duration, is also deduced by (Kölzsch et al., 2016) where they confirm a directly proportional relationship; however all durations in their study were above 5 days which is an almost 'prolonged stop' (greater than 7 days) in this study.

Clustering executed in the study is inspired from (van Wijk et al., 2012), where they worked with 13 geese and identified a total of 121 stopovers (an average of 10 per goose); with distances between stopovers ranging from 47 to 1741 kms (an average of 404 ± 331 kms). Our results provided in Table-1 nearly align with theirs, where the total average stops per goose is 17 with distances ranging from 44 to 2244 kms (an average of 416 ± 344 kms). Overall, similarity is encouraging, and a possible explanation for the difference of average stops per goose can be that they ignored (to cluster) the detours greater than 30kms if the goose arrives back within 8 hours, a condition not applied in this study, hence more number of average stops. Similarly, total migration distance in this study was identified as an average of 6360 kms compared to average identified as 5068 kms by (Xu et al., 2013). Even so, such a comparison is inexplicable as methodology, dataset and timeline are different.

Knowing that the goose (ID: 72413)'s recording is available for two migration cycles, a comparison in itself highlights that the total number of clusters in each cycle is relatively similar, 17 and 16, respectively in 2007 and 2008. Observing its spatial plot confirms that the goose follows an almost same route in both cycles, which validates the capital breeding nature of the goose as well as, in a sense validates the identified stopovers. However, validation of stopover locations in comparison with existing literature is unaccomplished, as no related study has been found during investigation.

In general, the validation process highlights the notion that research outcomes may result in dissimilar answers, as the methodology varies altogether. Even the distance threshold condition (30 kms), which is the core of this study, varies with 50 kms (F. Xu & Si, 2019) and 750 kms (density based) (Kölzsch et al., 2018) approach.

Discussed results were achieved with certain limitations. The main limitations are related to characteristics of the data such as the continuity because there was not a unique time gap between observations. Another limitation based on our methods is the difference of time zones which extends to multiple time zones and has not been catered for in the computation of departing hour. Limitation of the approach includes generation of many intermediate files which can be inconvenient to manage. A solution to this problem can be to replicate the attribute table in pandas dataframe. Generation of multiple dataframe can be easily managed and will not result in a risky approach of multiple files being overwritten or deleted in every run.

An anomaly in the developed program is that after application of limiting conditions for clustering, three points still remain in which one is with duration less than 48 hours and two with distance less than 30 kms. Duration one is acceptable as it is the first cluster of that goose (ID: 79694) flight path, which has to be considered by default. A possible answer to the other two with distance less than threshold is the mean representation of the cluster which can affect the mean to mean distance, compared to node to node distance on which the initial limiting conditions were applied.

Observing the early morning and late evening departure frequency from Figure 5 & 6, we can notice that geese departure is probably associated with the sun rise and sunset timings. An interesting way forward for the study is to relate observed variables with the accumulated photoperiod as discussed in (van Wijk et al., 2012). Knowing the capital breeder nature of geese, another important factor to relate is the land use. First, studying the favourable land use for geese sustenance (Polakowski & Kasprzykowski, 2016), and later integrating the distances between those land use areas to relate how geese adjust their departure time and flight distance to arrive at a preferable land use stopover site. An insightful comparison would be to relate spring with autumn migration in reference to the same research question to explore if a pattern is followed when geese are travelling in slower (spring) compared to a faster (autumn) migration (Kölzsch et al., 2016). Lasly, other attributes like climate, body mass and arrival time can be integrated to explore the relationship between departing time and flight distance (Schmaljohann, 2019).

5 CONCLUSION

Study identifies that there exist no particular relation between the departing time and following flight distance. Duration at stopover promotes greater flight distances, but with a weak correlation. Study also explored clustering of migration nodes, in the form of potential stopovers and examined how different tracks in the dataset differ individually. Answering the research question, study concludes that 'no', geese departing early morning do not fly more distance however we can say that 'yes', they do fly distant after a good rest period. Experience of this study gives rise to an opinion that departing time and flight distance are components influenced by many other factors like photoperiod, migration period, temperature, snow melt, wind pattern etc., and observing only them as it is - in relation with each other (as in this study) may result in no identified trend or misleading results.

6 REFERENCES

Bawden, J., Miller, R., Bawden, J., & Stack, R. (2019). Robert Stack. In *Conversations with Legendary Television Stars*. https://doi.org/10.2307/j.ctvhrd0f0.34

Drent, R., Both, C., Green, M., Madsen, J., & Piersma, T. (2003). Pay-offs and penalties of competing migratory schedules. *Oikos*. https://doi.org/10.1034/j.1600-0706.2003.12274.x

Kölzsch, A., Kleyheeg, E., Kruckenberg, H., Kaatz, M., & Blasius, B. (2018). A periodic markov model to formalize animal migration on a network. *Royal Society Open Science*. https://doi.org/10.1098/rsos.180438

Kölzsch, A., Müskens, G. J. D. M., Kruckenberg, H., Glazov, P., Weinzierl, R., Nolet, B. A., & Wikelski, M. (2016). Towards a new understanding of migration timing: slower spring than autumn migration in geese reflects different decision rules for stopover use and departure. *Oikos*. https://doi.org/10.1111/oik.03121

Kruckenberg H, Müskens GJDM, Ebbinge BS (2018) Data from: A periodic Markov model to formalise animal migration on a network [white-fronted goose data]. Movebank Data Repository. doi:10.5441/001/1.kk38017f

Moore, F. R. (1996). Time of Departure by Summer Tanagers (Piranga rubra) from a Stopover Site Following Spring Trans-Gulf Migration. *The Auk*. https://doi.org/10.2307/4088878

Morbey, Y. E., & Hedenström, A. (2020). Leave Earlier or Travel Faster? Optimal Mechanisms for Managing Arrival Time in Migratory Songbirds. *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2019.00492

Ojaste, I., Leito, A., Suorsa, P., Hedenström, A., Sepp, K., Leivits, M., Sellis, U., & Väli, Ü. (2019). *From northern Europe to Ethiopia : long-distance migration of Common Cranes (Grus grus) From northern Europe to Ethiopia : long-distance migration of Common Cranes (Grus grus). December*, 12–25. https://doi.org/10.13140/RG.2.2.35995.21282

Polakowski, M., & Kasprzykowski, Z. (2016). Differences in the use of foraging grounds by Greylag Goose Anser anser and White-fronted Goose Anser albifrons at a spring stopover site. *Avian Biology Research*. https://doi.org/10.3184/175815516X14739467542441

Schmaljohann, H. (2019). The start of migration correlates with arrival timing, and the total speed of migration increases with migration distance in migratory songbirds: A cross-continental analysis. *Movement Ecology*. https://doi.org/10.1186/s40462-019-0169-1

Smetzer, J. R., & King, D. I. (2018). Prolonged stopover and consequences of migratory strategy on local-scale movements within a regional songbird staging area. *The Auk*. https://doi.org/10.1642/auk-18-4.1

Spaans, B., Van Der Veer, W., & Ebbinge, B. S. (1999). Cost of incubation in a Greater White-fronted Goose. *Waterbirds*. https://doi.org/10.2307/1522007

Tang, M., Zhou, Y., Li, J., Wang, W., Cui, P., Hou, Y., Luo, Z., Li, J., Lei, F., & Yan, B. (2011). Exploring the wild birds' migration data for the disease spread study of H5N1: A clustering and association approach. *Knowledge and Information Systems*, *27*(2), 227–251. https://doi.org/10.1007/s10115-010-0308-x

Teitelbaum, C. S., Fagan, W. F., Fleming, C. H., Dressler, G., Calabrese, J. M., Leimgruber, P., & Mueller, T. (2015). How far to go? Determinants of migration distance in land mammals. *Ecology Letters*. https://doi.org/10.1111/ele.12435

van Wijk, R. E., Kölzsch, A., Kruckenberg, H., Ebbinge, B. S., Müskens, G. J. D. M., & Nolet, B. A. (2012). Individually tracked geese follow peaks of temperature acceleration during spring migration. *Oikos*, *121*(5), 655–664. https://doi.org/10.1111/j.1600-0706.2011.20083.x

Weegman, M. D. (2014). The Demography of the Greenland White-fronted Goose. *University of Exeter*. https://ore.exeter.ac.uk/repository/handle/10871/16211

Xu, F., & Si, Y. (2019). The frost wave hypothesis: How the environment drives autumn departure of migratory waterfowl. *Ecological Indicators*, *101*(February), 1018–1025. https://doi.org/10.1016/j.ecolind.2019.02.024

Xu, Q., Luo, Z., Wei, Y., & Yan, B. (2013). Data mining approaches for habitats and stopovers discovery of migratory birds. *Data Science Journal*, *12*(April), 159–169. https://doi.org/10.2481/dsj.WDS-027

ANNEX - A DISTRIBUTION OF WORK

Category	Part	Task	Person
Python Development	Pre Processing	Indexing, Deleting Fields, Extracting Hours, Adding Fields	Alejandro
	Processing	Get Distance	Edwige
		Get Duration	Edwige
		Delineate Stops	Anu
		Clustering	Mateen
		Mean Coordinates	Anu
		Join Tables	Edwige
		Generated files management	Mateen
		Integrating all parts of Code	Anu + Mateen
	Post Processing	Spatial Plots	Anu
		Scatter Plots	Mateen
		Group Bar Chart	Alejandro
General		Inception and RQs	Group work by all
		Literature Review	
		Setup & Presentation	
		Code QC and Commenting	
		Report Writing & Formatting	
		Figures and Table	
		User Guide	

ANNEX - B QGIS SCRIPT - READ ME

Analysing 'Flight Distances' of White Fronted Goose

QGIS Script for the analysis of relation between Departing Time and Flight Distance, with the integration of 'duration at stopover' as the third parameter. Study is based on migration data of seven white fronted goose with spatial extent ranging from western europe to eastern russia, for the year 2007-2008.

Initiation

Script can be initiated for local system with following two steps:

1. Set variable 'data_dir' to the local directory of goose project
2. Set variable 'shape_path' to <filename> of goose dataset

Requirements

Script will work on any similar data as of move bank, with minimum attributes of

- 'timestamp'
- 'lat'
- 'long'
- 'tag_ident'

No additional QGIS related requirement is necessary.

Design

Script is designed in a function based approach where functions are developed for each task and are called upon as per requirement. Figure 1 illustrates the complete design workflow. Successful script run takes ~40 seconds to execute all processes.

Outputs

Successful script run will create following outputs:

- 7 x Spatial Plots (for each goose respectively)
- 1 x Spatial Plot (for all geese together)
- 1 x Scatter Plot (color based on duration)
- 1 x Scatter Plot (color based on goose tag_ident)
- 1 x Group Bar Chart (three classes of stopovers against average distance)

Intermediate Files

Following intermediate files are created and deleted on every run of the script, for each goose respectively

- <goose-id> (file with points only of this goose id)
- <goose-id_cluster> (file with mean coordinate information of each cluster)
- <goose-id_joined> (file after mean information is joined from other file)
- <goose-id_ignored> (file with all intermediate points removed)
- <goose-id_final> (final goose file with one departing point of each cluster)

Interface Files

All files are removed from the interface so the script can re-run without the need to delete files from QGIS interface, as they are to be overwritten in the next run.

Pseudo Code

Figure-1 illustrates the pseudo code which is as follows:

for given data:

- export birdwise files (to work on each individual file separately)
- execute pre-processing

  for each individual file (total 7):
    - compute distance between each pair (node-node)
    - temporary clustering based on distance (<30kms)
    - compute duration for each cluster
    - final clustering based on distance and time (>= 48 hours)
    - compute mean to represent each cluster
    - remove 'not stopovers' from file
    - extract departing point for each cluster
- combine all files into one for plotting
- plot

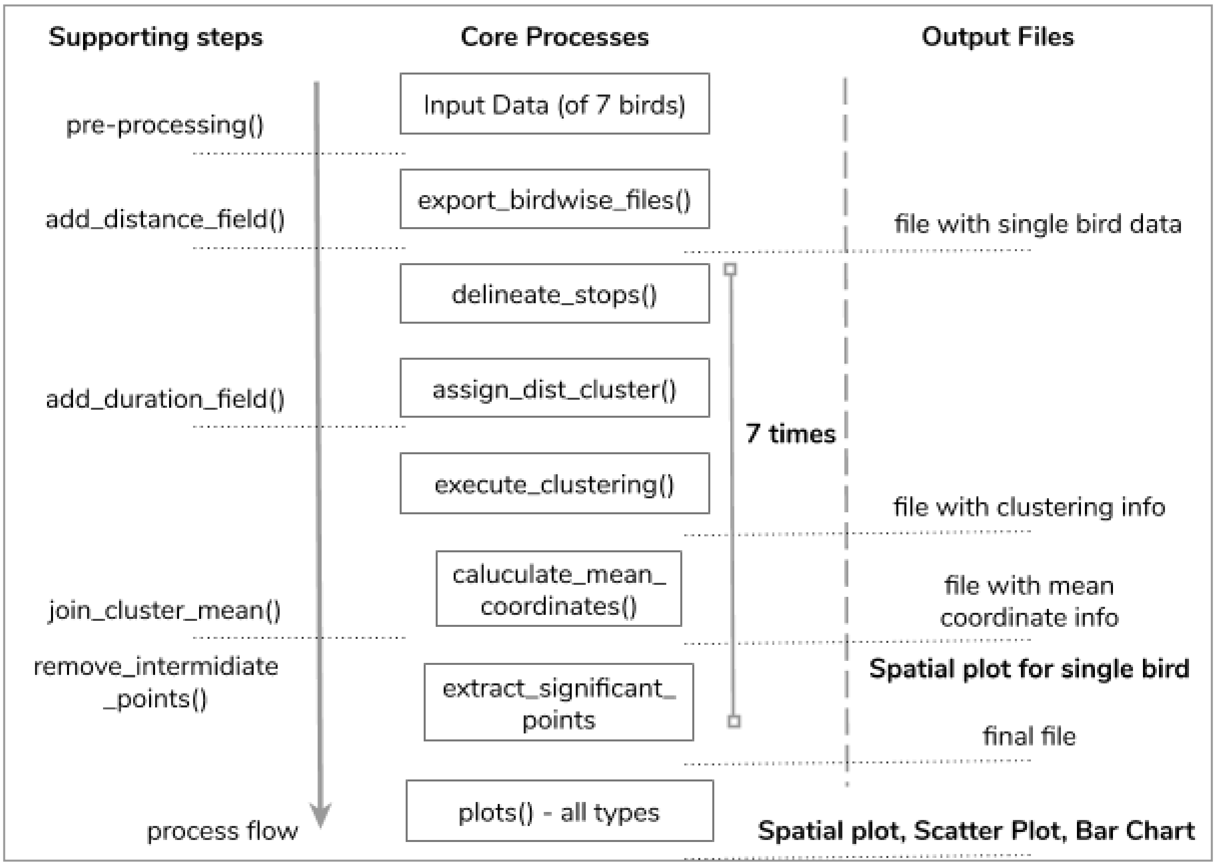

Figure-1 Figurative explanation of pseudo code