

Lesser Black-Backed Gull Movement Patterns

COMP 551: Applied Machine Learning

Migration Modeling, Displacement Tolerance and Conservation

Vincent Antaki

vincent.antaki@mail.mcgill.ca
260745934

Chris Glasz

christopher.glasz@mail.mcgill.ca
260720944

Pulkit Khandelwal

pulkit.khandelwal@mail.mcgill.ca
260717316

Abstract—The phenomenon of bird migration is a popular field of study. The availability of light tracking devices, the wide assortment of species, and their diverse behaviors and habitats make the field a candidate of choice for time-series-based analysis. A data-driven analysis of bird migration can be used to extend our knowledge of the underlying causes of the path these birds travel and our potential influence on it. The MoveBank Data Repository¹ offers a wide selection of animal tracking data. Our work, using the data from Wikelski et. al [11], is focused on the migration of Lesser Black-Backed gulls, which annually travel over 7000km from their breeding grounds in Russia and Finland to Lake Victoria. Using machine learning algorithms and additional data annotations from the EnvDATA system [5], we perform various tasks to model and characterize their migration patterns and displacement tolerance.

I. INTRODUCTION

Bird migration has been studied in great detail. The process of birds finding their way back home from a position in a distant unfamiliar area to their local breeding ground (home site) is known as *Navigation* in a narrow sense. Birds are guided by cues or landmarks to help them in migration.

The behavior showed by these birds during migratory journeys is an important subject of study to help in their conservation, especially in the cases of birds which are on the verge of extinction. One important thing to note is that individual birds do not show reliable, homogeneous behavior in equivalent situations, which makes their study and analysis much more difficult, though still interesting.

Studies have been performed on birds which are displaced from their normal migratory routes during long distance journeys. Wallraff [10] performs experiments to see if olfactory cues are used by birds (adult Lesser Black-Backed Gulls) to correct such displacements. This study forms the basis of our project. We further investigate the problem by using additional cues (explained in Section II). We pose four questions and use machine learning techniques to predict and analyze the migratory patterns of these birds. We then scrutinize how these analyses will be helpful for the conservation of the given bird species, and generalize to migratory birds on the whole.

The rest of the paper is divided as follows: *Section II* details the various available cues, define some basic terms which will be helpful to understand the background required to understand the context of this project. *Section III* describes

the data used and its preprocessing. *Section IV* defines and concertize the problems dealt with. *Section V* describes the methodologies used in detail. *Section VI* gives the empirical results. *Section VII* discusses the results and concludes the study.

II. TERMINOLOGIES AND RELATED WORK

Gagliardo et. al [11] subjected adult Lesser Black-Backed Gulls (*Larus fuscus fuscus*) migrating from their Finnish/Russian breeding grounds to sensory manipulation, to determine the sensory systems required for navigation. They concluded the following from the experiments:

When translocated westwards and outside their migratory corridor birds with olfactory nerve section kept a clear directional preference (southerly) but were unable to compensate for the displacement, while intact birds and gulls with the ophthalmic branch of the trigeminal nerve sectioned oriented towards their population-specific migratory corridor. Thus, air-borne olfactory information seems to be important for migrating gulls to navigate successfully in some circumstances.

Current competing theories [10] to explain how the birds locate their position with respect to their destination propose the predominant use of either magnetic field intensity or olfactory cues. In fact, some lines of evidence suggest that the intensity of the Earth's magnetic field may play a role in the navigation of experienced migrants, in which case it may be used as a signal for latitudinal displacement, or as a signpost that a particular latitude has been reached.

The experiment suggests that intact olfactory nerves are necessary to correct for a longitudinal displacement when displaced outside the migratory corridor to an unfamiliar landscape that did not provide an alternative source of navigational information.

The authors of the paper suggest that "familiarity with the migratory corridors does not always in the ability to overcome the removal of navigational cues." They hint at possible exploration of different cues and their variations and this is what we have done in this project: explore other cues other than air-borne olfactory cues such as geomagnetic field, visual landscape, coriolis force, atmospheric chemical signals, infrasounds, electromagnetic sferics, and gravity.

¹<https://www.datarepository.movebank.org>

III. DATA DESCRIPTION

We used GPS data of Lesser Black-Backed Gulls, retrieved from the MoveBank database [8]. The data comes from a paper studying the birds' use of their olfactory nerves to navigate during the migratory season [1]. Each bird was fitted with a tracking device before being released, which tracked their global coordinates until their death. A handful of individuals were tracked for over five years, but the vast majority died early in the study, after being unable to find their accustomed nesting grounds. We focused much of our analysis on birds that survived multiple migration cycles, as they give more valuable information with regards to long-term behavior patterns and trends.

Each data point consists of a time stamp, the individual's identifier, and latitude/longitude coordinates. Additional information regarding individuals is also provided, which describes the group it belonged to in the study, date and location of its death, and when the tracker was deployed.

In addition to the location data from MoveBank, we obtained additional information using EnvDATA [5] (Surface temperature, population data, vegetation, etc.). These are described in the following sections.

We also calculated additional information using the given data, including estimated velocity and flight angle.

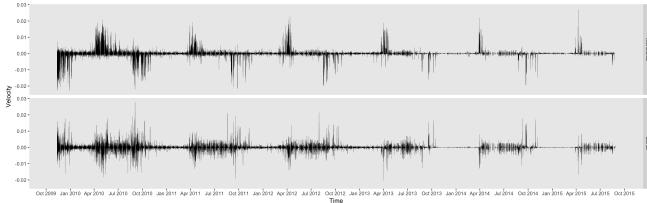


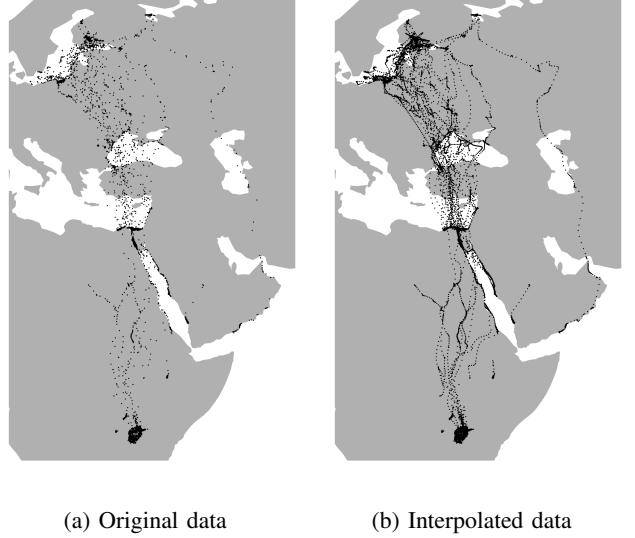
Fig. 1: Velocity over time, split into meridional and zonal components. It is clear from the data that Lesser Black-Backed Gulls nest in the winter, as their movement drops to near-zero. Mating takes place in the summer months, where they continue to move east and west in search of mates.

The data retrieved from MoveBank was both sparse and irregular. For proper analysis, we decided to interpolate between known times and locations to create data with regular intervals of two hours. We performed this interpolation using a monotone Hermite spline. In addition to providing regularity, this expanded our data by over 300%.

IV. PROBLEM DEFINITION

A. Migratory Patterns

In order to better understand the migration patterns of Lesser Black-Backed Gulls (and migratory birds as a whole), we investigated the classification problem of categorizing data points into different migratory stages. With a model capable of reliably performing this classification, we can improve conservation efforts - we may be able to better predict migration behaviors of flocks, or more readily recognize when a population is exhibiting abnormal migratory behavior, signaling some problem with the local environmental conditions.



(a) Original data

(b) Interpolated data

We aimed to classify migration periods into four categories (*Nesting*, *Mating*, *Southerly Migration*, and *Notherly Migration*), using information derived from environmental and GPS data. We attempted to restrict our models to use data that is not location-dependent (opting to use features like velocity and temperature, as opposed to latitude and longitude). We have done no cross-species analysis, but hope that the drawn conclusions will generalize to migratory birds as a broad group, and not exclusively to Lesser Black-Backed Gulls.

B. Outlier Detection

As mentioned, our data includes three different kinds of birds: the control group, a group with segmented olfactory nerves, and a group with segmented trigeminal nerves. If we are to do binary classification of our displaced gull data, could we differentiate them using features derived from a sequence of locations? We'll consider two possible binary classifications of our data. The first will be the control group vs surgically treated gulls. The second will be olfactory-impaired group vs control and trigeminal-treated groups. The results of Wikelski et. al [11] suggest that the second classification task should be more successful, as trigeminally-impaired gulls and the control group have been less affected by the displacement.

Being able to classify correctly could point toward the existence of characteristic movement patterns for displaced and impaired birds. From a conservation point of view, the ability to detect disrupted and disoriented birds can drastically help efforts to identify environmental factors disruptive to bird migration.

C. Trajectory Prediction

Different techniques can be used to describe the gulls migration patterns. One way to characterize them is to look at the expected behavior in certain configurations. We consider the regression problem of learning a function which, given a sequence of locations and time indicators, predicts the angle (ignoring the altitude dimension) and average speed in that

direction (hereafter referenced as the *displacement vector*) between two data points. Although it is likely a demanding task in terms quantity of data required, learning to do so on the non-impaired, non-displaced subset of our data is implicitly learning about the gulls' migration corridor and their expected behavior around it.

D. Compensation for accidental displacements and hot-spot corridors

We notice that many birds remain at some particular locations for some considerable amount of time during their migratory journey. These locations can be described as hot-spots. We analyze these hot-spots to understand why the majority of birds spend considerable time at these locations. In the study from which our data was drawn, the birds are released from three different initial locations. Can we say that even though the birds are released from different locations, they still visit the common hot-spots during their journey? If so, why? What leads them to congregate there? What are the cues which help them reach these locations?

With time series and clustering models using local cues at the hot-spot regions we can predict what environmental factors help the birds navigate to the hot-spots. Once reached, can we improve those factors to help their conservation so that more and more birds are able to complete their migratory journey (unless dead because of old age).

These results can be extrapolated to see how birds which are accidentally displaced from their breeding grounds are able to get back to the normal migratory route without becoming lost.

V. METHODOLOGY

A. Migratory Patterns

At a glance, it would appear that the long-term movement of migratory birds is structured enough to easily classify behavior into clear categories based on the time of year or location. However, a more careful inspection of the data reveals that many individuals do not follow these patterns. Simple features such as these are insufficient for classification - different individuals nest at different latitudes (most travel all the way to Lake Victoria, but many choose to nest on the coast of Saudi Arabia, and some don't even fly south of Cairo), and some gulls simply stay in their nesting location year-round (Fig. 3). The number of individuals that stuck to "correct" migratory patterns is vanishingly small, and simple threshold-based classification is hopelessly insufficient to describe the data as a whole.

First, we attempted to classify the migratory periods using an unsupervised approach: K-Means clustering. After several rounds of experimentation, we found that the velocity of the birds, decomposed into its meridional (North-South) and zonal (East-West) components, was the most valuable feature for clustering. Meridional velocity is clearly helpful in determining whether (and in which direction) birds are migrating. The zonal velocity is also valuable for determining whether birds are nesting or mating - during mating season, the gulls travel east and west in search of partners, while

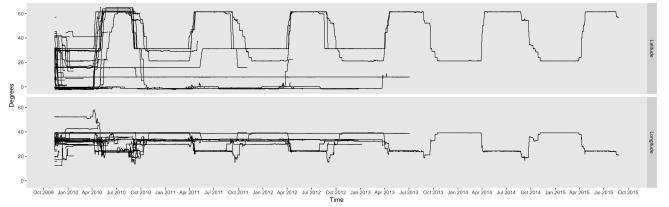


Fig. 3: Latitude and longitude location over time, in degrees. Here, we can see that many individuals do not return north in the summertime. It is not clear whether this is caused by the surgical alterations to the birds' olfactory centers, longitudinal relocation, or some other factor.

during nesting season, they barely move at all (this can be observed in Fig. 1). For K-Means clustering to properly distinguish between these periods, however, we did need to incorporate the birds' latitude. This causes the algorithm to misclassify when birds nest abnormally far north, but it does perform quite well in most cases.

We had far more success with supervised learning. We hand-labeled one well-behaved individual's movement (ID 91916A), and used it to train a Gaussian-kernel SVM. This particular bird was one of the longest-lived in the study, meaning it provided a fair amount of data on its own, enough to train an SVM with reasonable performance. We found that, using the SVM, we only needed to use the decomposed velocity described above, and that the classification was much more robust to divergent behavior.

B. Outlier Detection

For this binary classification task, we built as a feature a 2-dimensional histogram of speed and direction. The direction is expressed in radians, between $-\pi$ and π . The speed is expressed in meters per second. Each of the two axes are separated in p subsections which are set at a regular interval between the minimum and the maximum observed value for that axis in the training data. This results in p^2 bins to classify each vector of speed and direction. After having computed the histogram for a series of data points, we normalize each bin by the sum of all bins to obtain a histogram of density which is independent of the number of data points contained in our path (as far as we do not consider sampling bias). We considered two values for $p \in (10, 15)$.

We noticed that there were a few displacements that had very high average speed. This drove the maximum observed value quite high and resulted in a loss of resolution of our histogram feature. After further investigation into the origin of those values, it was discovered that some were results of the original displacement of the birds by the researchers from Finland to Germany or Russia. As a result, we ignored any displacement vector with norm above 12 m/s in the computation of our histograms. 185 vectors out of a total of 33284 were ignored because of this choice.

Using the Scikit Learn library [9], we tested 17 instances of classic machine learning models. The full list, including hyper-parameter settings, is available in table I in the ap-

pendix. We considered two baselines. The first predicts the most likely category while the second guesses randomly, with probabilities adjusted to the distribution of the training set. We performed a 10-fold cross validation. We've looked at the average accuracy and the F1 score (weighted by class distribution) of our models.

C. Trajectory Prediction

We used the data associated with the 54 non-displaced birds, for a total of 45911 data points. For this task, we attempt using feedforward neural nets to predict the angle and the speed of displacement. Implementation was done using the Lasagne library [4].

Our input is a fixed size sequence of time and location where we encoded the time as one-hot encoded week and hour of the day indexes. Doing so implies we trade some precision of the time (such as the day or the minute at which the data point occurs) for less dimension in our input. The week and the hour of the day one-hot encodings are of length 53 and 24 respectively, and the location used is the latitude and longitude as 2 floating point values.

Our algorithm contains an hyper-parameter for the "window length" which controls the number of previous data points taken in input by our discriminative function. Obviously, the windows length has a big impact on the dimensionality of the input (input dim = 79* window size). Therefore, we consider only window length $\in (1, 2)$ in our tests.

We trained on 32 example batches and performed a 5 fold validation, where the split is done in such a way as to not split an individual bird's data into multiple sets. We consider a certain class of parameterizable feed-forward nets whose architectures are characterized by the following parameters: the number of nodes per hidden layer (constant for all layers), the number of hidden layers (depth) and the proportion of the nodes by hidden layers which are sigmoid (the rest are rectified linear units (ReLUs)). Our output layer consist of two nodes, one of which has a sigmoid activation (for predicting the angle) and the other is a ReLU (for predicting speed). The output angle ($\in [-1, 1]$) is then multiplied by a π to effectively express the whole spectrum of possible angles in radians. The loss function considered is the average norm of the difference between the predicted displacement vector and the actual displacement vector.

Our grid search consists of the followings options : depth $\in \{1, 2, 3\}$, num_hiddens $\in \{150, 200\}$, proportion_of_sigmoid $\in \{0.0, 0.05, 0.1\}$, and two options for gradient descent. The first is to use Nesterov momentum with a learning rate of 0.01 and a momentum coefficient of 0.9, while the second is a typical gradient descent with a learning rate of $5 \cdot 10^{-4}$. Our baseline always predicts no movement. The baseline loss is consequently the average displacement vector norm. Its value is 0.557 m/s.

We considered the alternative to encode the angle in degrees, minutes, and seconds (adapting the loss function accordingly). This effectively doubles the dimensionality of our regression label. All attempts to use similar models on

this problem formulation resulted in the explosion of the gradients after few epochs.

D. Compensation for accidental displacements and hot-spot corridors

The hot-spot corridors can be seen in Fig. 17. These are the hot-spots which are obtained after running K-Means clustering over the GPS data. The cues used to understand these hot-spots are the human population density and vegetation cover. Wikelski et. al mention that regions of more vegetation cover is a natural stopping point for birds during migration. This also implies that population density is higher around these regions because such regions are habitable by humans. Fig. 18 shows the population density and Fig. 19 shows the vegetation cover in the area of the birds' migratory paths².

Time series models were built to predict the migratory journey of the birds. Given the previous timestamps and locations (both latitude and longitude), we tried to predict the next location for each birds. Additional features used were the human density population and vegetation cover data obtained from EnvDATA [5].

Three neural networks and deep learning models were used: a SimpleRNN, an LSTM (Long short-term memory) [6], and a GRU (Gated Recurrent Unit) [3]. The machine learning library Keras [2] was used to build all these models. We used the *adam* optimizer and *ReLU* activation units. These choices were obvious because of their popularity in the literature. For each of the models, we trained two versions: one had a single hidden layer, while the other had several hidden layers and thereby a deeper architecture. Squared error loss was used. Fig. 7 through Fig. 12 show all the different architectures used to train and test the models.

A train-validation split of 70:30 was used for the model evaluation. We report the training and validation errors along with the Root Mean Squared Error. Loss vs. Number of epochs plots helped us visualize how the different models performed during the training and validation process.

VI. RESULTS

A. Migratory Patterns

It is difficult to evaluate the performance of our migratory classification methods, as there is no labeled data. Though we hand-labeled one individual's movement, we are not experts in the migratory patterns of birds, and our intuition may be off-base. For proper validation of our results, we would require a data set of GPS locations labeled by ornithologists. That being said, we can offer a visual evaluation of our results.

In the unsupervised case, our K-Means classifier failed to fully capture the data - birds that differed from the standard migratory cycle were often improperly labeled (the difference between mating and nesting, in particular, was difficult for

²The human population density is expressed in persons per square kilometer and the vegetation cover is expressed according to NCEP NARR SFC Vegetation at Surface metric normalized between 0 and 1 as given in EnvDATA [5]

the algorithm to learn). Additionally, the expectation maximization algorithm frequently gets caught in local minima, causing it to group southerly migration together with mating (as shown in Fig. 4). This problem is alleviated with random restarts, but still occasionally presents itself. However, in most standard cases, the clustering approach performed reasonably well, and the simplicity of the algorithm may outweigh its poor performance in edge cases.

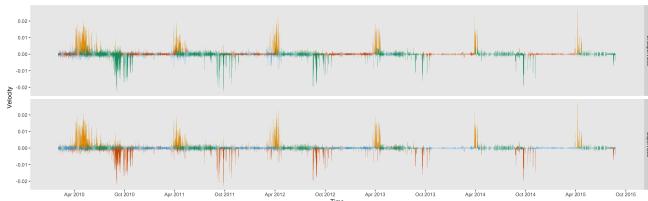


Fig. 4: Meridional velocity over time, colored by classified migratory behavior. Here, the clustering algorithm settled in a local minimum, and grouped southerly migration together with mating.

The SVM performed far better than clustering when classifying northern and southern migration, rarely making any mistakes at all. It had slightly more difficulty differentiating between mating and nesting seasons, as it was trained strictly on velocity data - if a bird does not move, it will usually mark it as nesting, even if it was surrounded by movement. This behavior is easily seen in Fig. 5. This problem may have been avoided with the use of an recurrent neural network, or possibly feeding a temporal window of observations to the SVM instead of individual points.

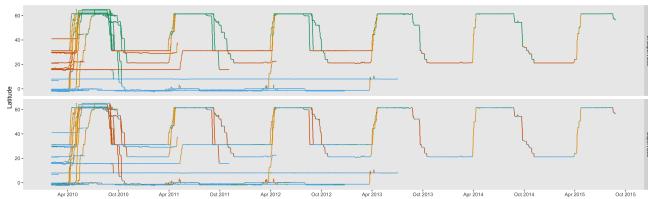


Fig. 5: Latitude over time, colored by classified migratory behavior. It is clear that the SVM had trouble differentiating between mating and nesting seasons, but handled migratory periods very well.

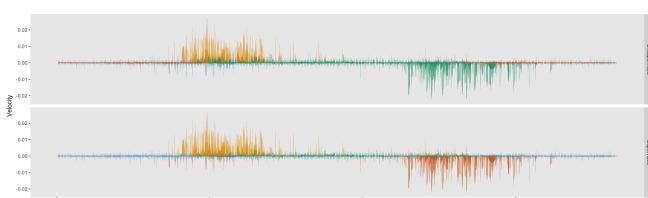


Fig. 6: Meridional velocity through the time of year, colored by classified migratory behavior.

B. Outlier Detection

For both tasks, the best model used $p = 10$ and not $p = 15$. For the "control group vs all" binary classification, the model that has both the best 10-fold cross-validated average accuracy and F1 score is the K-nearest neighbors algorithm with 10 neighbors. However, for both metrics, it beats the best baseline by close to 1%. To our opinion, this is not significant enough to consider that our algorithm has learned something meaningful for this task. For the "olfactory group vs all" binary classification, the decision tree model, our best performing model, had 0.662 average validation accuracy and 0.633 average F1 score compared to 0.569 and 0.560 for the first baseline and 0.585 and 0.439 for the second baseline. This indicates different navigation patterns for displaced lesser black-backed gulls in function of their olfactory capacity. It seems to corroborate the idea that the gulls need their olfactory senses to navigate properly and that the sectionning of the trigeminal nerve has induced less disorientation than the olfactory nerve sectionning.

See table II and table III in the appendix for the empirical results on both these tasks.

C. Trajectory Prediction

All our models perform spectacularly on the training set and quite poorly on the validation set. While the training loss is typically in the [1, 2] interval after initialization, it is systematically in the $1e - 6$ magnitude after only one epoch (for all models). The validation loss, although improving for the first epochs, decreases at a much slower rate. In that sense, the high variation observed for the validation loss is a result of the random initialization more than the variation in the quality of the training data. For all models that were considered, the average 5-fold cross-validated loss was in the [1.1, 1.5] interval. Considering our baseline is 0.557, we can consider these results as significantly inconclusive. The results have not been included in the present document because they were deemed unhelpful to and unworthy of the reader. Their probable causes will be discussed in Section VII.

D. Compensation for accidental displacements and hot-spot corridors

Fig. 15 shows the population density and Fig. 16 shows the vegetation cover in the area of the birds' migratory paths. The size of the marker in these two images show the proportional population density and vegetation cover. The bigger the size of the marker, the larger the population density and more vegetation cover. The corresponding plot of each of the locations (latitudes and longitudes) can be seen in Fig. 14. The hot-spots corridors can be seen in Fig. 17. There are 10 hot-spots which were the best obtained clusters.

Fig. 18 shows the population density and Fig. 19 shows the vegetation cover in the area of the birds' migratory paths across time. Tables IV and V gives a comparison of the two version of both the simple and deeper architectures. These plot were obtained by using the amazing ploy.lib python library [7]. SimpleRNN beats LSTM and GRU in

the simple version of the neural net architecture. The training and validation errors for SimpleRNN are 3.5745 and 3.3346 respectively. These are far better than that obtained by LSTM and GRU. But, can we do better? The answer is yes and for that we go much deeper.

The second version of the models implement a deeper model. The results are at the opposite end of the spectrum. LSTM performs the best followed by SimpleRNN and then GRU on the validation error. LSTM also performs the best on the RMSE error. It is least for LSTM (2.39) while SimpleRNN and GRU performs slightly worse than LSTM.

The interpretation of these results is discussed in detail in the following section.

VII. DISCUSSION

A. Migratory Patterns

Our migratory classification models performed well enough that we consider them to be successful, though there is room for improvement. We feel that the greatest strength of this work is the conclusion that velocity data alone is enough to classify migratory periods. This means that trained models may be transferred to work on similar species of birds indigenous to different regions without much loss in performance. This is crucial, as training a sophisticated model takes large amounts of data, and for many species of bird, this data simply does not exist.

Our results may have been improved with techniques that take into account surrounding data, such as recurrent neural networks or one-dimensional convolutional nets. We experimented with the former, using LSTM cells, but the model failed to capture long-term patterns, and took an obscene amount of time to train with our expanded data.

We expect that it may be possible to characterize migratory patterns with a generative model, but fear that our data would not be conducive to this particular task - the irregular time gaps in the raw data would pose a problem, and we expect that a generative model trained on the interpolated data would simply learn the particular interpolation function used.

B. Outlier Detection

Our results, although somewhat conclusive on the second classification task, are far from satisfying our hunger for a well-performing algorithm. There are multiples plausible causes for these (mixed) results. Considering the sparse number of examples and the irregularities in our data (e.g. pronounced uneven spacing of data points, anomalies), it is likely that our derived features are too noisy. Furthermore, looking at the accuracy on the training data, we can see that the some models are performing fairly well. This may suggest overfitting and but it most likely points to insufficient information to generate a well-defined regression function.

Furthermore, our naïve approach consisting of separating axes on a regular interval to create our histogram feature could potentially be poorly adapted to our data distribution; a strong majority of average speeds are low values. The exploration of some way to set our bin boundaries with finer resolution for small speed values and broader resolution for

higher speed values could allow compensation for the non-uniform distribution of average displacement speed in our dataset. Alternatively, usage of data interpolation as a way to smooth the time series, as done in Migratory Patterns section, could have reduced noise in our input features and possibly yielded better results.

C. Trajectory prediction

Our very overfitted results could a be direct consequence of the high dimensionality of our input encoding. It is clear that the two one-hot encodings (week and hour of the day) are responsible for most dimensions of our input and that they induce orthogonality between our examples. Two options could be considered to get around that problem. The first one would be to make the architecture deeper (the deepest considered had only 3 hidden layers) to enable proper decoding of input and the introduction of strong regularization to compensate our model's over-capacity. Alternatively, re-encoding our input could be a good alternative. The time, considered recurrent on a yearly basis, could be encoded as a single continuous variable between [-1,1] (ex. -1 would be the first second of January first and 1 the last second of the 31st of December. Though this ignores the fact that January and December are, in fact, quite close, it would get rid of our very cumbersome one-hot encodings.

D. Compensation for accidental displacements and hot-spot corridors

We obtained a good representation of the most important locations of birds' resting place during their migratory journey. These were clearly seen as 10 means of the clusters. On comparison with the plots of the population density and vegetation cover in Figures 18 and 19, one can see that there is high vegetation cover and high population density in these hot-spot corridors. These two cues are essential for the migration pattern of the birds as they look out for these markers during the course of their journey. Now, if we know where vegetation cover is high we can say that the birds are more likely to visit them in their normal migratory path. If some birds are displaced then they can search for such cues and reach these regions. From there they can get back on their normal migratory path. Also, a special attention should be given with those hot-spots regions so that they have their biodiversity are protected as much as possible. This will in turn help in the conservation of the birds by getting the birds back to the normal route.

Next, we also saw how machine learning techniques are used to model time series data of the different locations where the birds fly through along the path. Given the previous locations we have successfully predicted the future locations with a very high accuracy, as can be deduced from such low RMSE. This path prediction task will help us to study the detrimental environmental factors at those locations before hand. We can therefore help reduce man-made pollution and take efforts in improving the atmospheric, water and land conditions to better suit for the migration.

STATEMENT OF CONTRIBUTIONS

Vincent Antaki was responsible for outlier detection, the rather crude trajectory regression, an unmentioned unsuccessful attempt to model our data distribution with a deep belief network and writing the report abstract.

Christopher Glasz handled much of the data visualization, the supervised and unsupervised classification of migratory patterns, and wrote the description of the data.

Pulkit Khandelwal worked on the analysis of accidental displacements and hot-spot corridors, and wrote the introduction, terminology, and related work sections of the report.

We hereby state that all the work presented in this report is that of the authors unless otherwise referenced.

REFERENCES

- [1] Rami Al-Rfou et al. “Theano: A Python framework for fast computation of mathematical expressions”. In: *CoRR*, abs/1605.02688 (2016). URL: <http://arxiv.org/abs/1605.02688>.
- [2] François Chollet. “Keras”. In: (2015).
- [3] Junyoung Chung et al. “Gated feedback recurrent neural networks”. In: *CoRR*, abs/1502.02367 (2015).
- [4] Sander Dieleman et al. *Lasagne: First release*. Aug. 2015. DOI: 10.5281/zenodo.27878. URL: <http://dx.doi.org/10.5281/zenodo.27878>.
- [5] Somayeh Dodge et al. “The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data”. In: *Movement Ecology* 1.1 (2013), p. 3. ISSN: 2051-3933. DOI: 10.1186/2051-3933-1-3. URL: <http://dx.doi.org/10.1186/2051-3933-1-3>.
- [6] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.
- [7] Plotly Technologies Inc. “Collaborative data science”. In: (2015). URL: <https://plot.ly>.
- [8] Wikelski M et al. “Data from: True navigation in migrating gulls requires intact olfactory nerves”. In: (2015).
- [9] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [10] Hans G Wallraff. *Avian navigation: pigeon homing as a paradigm*. Springer Science & Business Media, 2005.
- [11] Gagliardo A Holand R Huttunen MJ Juvaste R Mueller I Tertitski G Thorup K Wild M Alanko M Bairlein F Cherenkov A Cameron A Flatz R Hannila J Hppop O Kangasniemi M Kranstauber B Penttinen M-L Safi K Semashko V Schmid H Wistbacka R Wikelski M Arriero E. “True navigation in migrating gulls requires intact olfactory nerves”. In: *Scientific Reports* 5:17061 (2015). DOI: 10.1038/srep17061.

APPENDIX

Model name	Hyper-parameters/comments
Decision Tree	max depth = 5
Nearest Neighbors 5	5nn, euclidian distance
Sigmoid SVM C=1	Sigmoid kernel, margin penalty=1
RBF SVM C=2	RBF kernel, margin penalty=2
balanced RBF SVM C=2	RBF kernel, margin penalty=2, examples weight by train set class distrib.
Nearest Neighbors 1	1nn, euclidian distance
Nearest Neighbors 10	10nn, euclidian distance
Nearest Neighbors 3	3nn, euclidian distance
Linear SVM C=0.025	Linear kernel, margin penalty=0.025
LDA	Linear Discriminant Analysis, singular value decomposition with tolerance $1e-4$
Linear SVM C=0.75	Linear kernel, margin penalty=0.75
Balanced RBF SVM C=1	Linear kernel, margin penalty=0.025, examples weight by train set class distrib.
Naive Bayes	Gaussian Kernel
RBF SVM C=1	RBF kernel, margin penalty=1
AdaBoost	n estimators=50, learning rate=1
Baseline clf2	Predicts most frequent label in training set
Sigmoid SVM C=2	Sigmoid kernel, margin penalty=2
Baseline clf1	Predicts randomly based on train set class distrib.
Random Forest	max depth = 5, n. estimators=10

TABLE I: Classic machine learning models considered for outlier detection and their respective hyper-parameters values.

Model	Train acc	Train f1	Valid acc	Valid f1
Decision Tree	0.948	0.947	0.454	0.426
Nearest Neighbors 5	0.727	0.691	0.562	0.482
Sigmoid SVM C=1	0.66	0.525	0.638	0.499
RBF SVM C=2	0.662	0.53	0.638	0.499
balanced RBF SVM C=2	0.606	0.596	0.485	0.457
Nearest Neighbors 1	1.0	1.0	0.531	0.508
Nearest Neighbors 10	0.692	0.614	0.646	0.555
Nearest Neighbors 3	0.783	0.772	0.546	0.502
Linear SVM C=0.025	0.66	0.525	0.638	0.499
LDA	0.954	0.954	0.546	0.515
Linear SVM C=0.75	0.66	0.525	0.638	0.499
Balanced RBF SVM C=1	0.569	0.533	0.408	0.345
Naive Bayes	1.0	1.0	0.377	0.367
RBF SVM C=1	0.66	0.525	0.638	0.499
AdaBoost	0.804	0.801	0.523	0.496
Baseline clf2	0.66	0.525	0.638	0.499
Sigmoid SVM C=2	0.66	0.525	0.638	0.499
Baseline clf1	0.537	0.535	0.562	0.546
Random Forest	0.827	0.808	0.592	0.497

TABLE II: 10-fold average training and validation accuracy and f1 score of different models for the outlier detection task control group vs. all using 100-bins displacement vector histogram

Model	Train acc	Train f1	Valid acc	Valid f1
Decision Tree	0.973	0.972	0.662	0.633
Nearest Neighbors 5	0.74	0.715	0.531	0.494
Sigmoid SVM C=1	0.654	0.518	0.585	0.439
RBF SVM C=2	0.656	0.522	0.585	0.439
balanced RBF SVM C=2	0.615	0.602	0.546	0.514
Nearest Neighbors 1	1.0	1.0	0.538	0.539
Nearest Neighbors 10	0.658	0.588	0.569	0.484
Nearest Neighbors 3	0.813	0.807	0.538	0.505
Linear SVM C=0.025	0.654	0.518	0.585	0.439
LDA	0.973	0.973	0.592	0.589
Linear SVM C=0.75	0.654	0.518	0.585	0.439
balanced RBF SVM C=1	0.587	0.536	0.508	0.415
Naive Bayes	1.0	1.0	0.638	0.61
RBF SVM C=1	0.654	0.518	0.585	0.439
AdaBoost	0.648	0.641	0.531	0.528
Baseline clf2	0.654	0.518	0.585	0.439
Sigmoid SVM C=2	0.654	0.518	0.585	0.439
Baseline clf1	0.565	0.568	0.569	0.56
Random Forest	0.85	0.836	0.6	0.517

TABLE III: 10-fold average training and validation accuracy and f1 score of different models for the outlier detection task olfactory group vs. all using 100-bins displacement vector histogram

Model	Train loss	Val loss	Train RMSE	Val RMSE
SimpleRNN	3.5745	3.3346	2.13	2.39
LSTM	308.7852	322.5411	24.90	25.33
GRU	2230.4742	2242.4645	24.90	25.33

TABLE IV: Training and Validation set error along with Root Mean Squared Error (RMSE) for different simple architectures

Model	Train loss	Val loss	Train RMSE	Val RMSE
SimpleRNN	2.6953	3.3793	2.29	2.41
LSTM	2.6285	3.3812	2.07	2.39
GRU	2.6957	3.3962	2.07	2.41

TABLE V: Training and Validation set error along with Root Mean Squared Error (RMSE) for different deep architectures

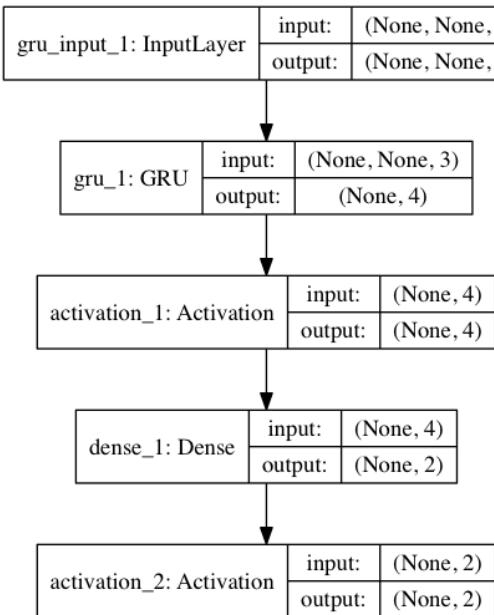


Fig. 7: Gated Recurrent Unit: Simple Architecture

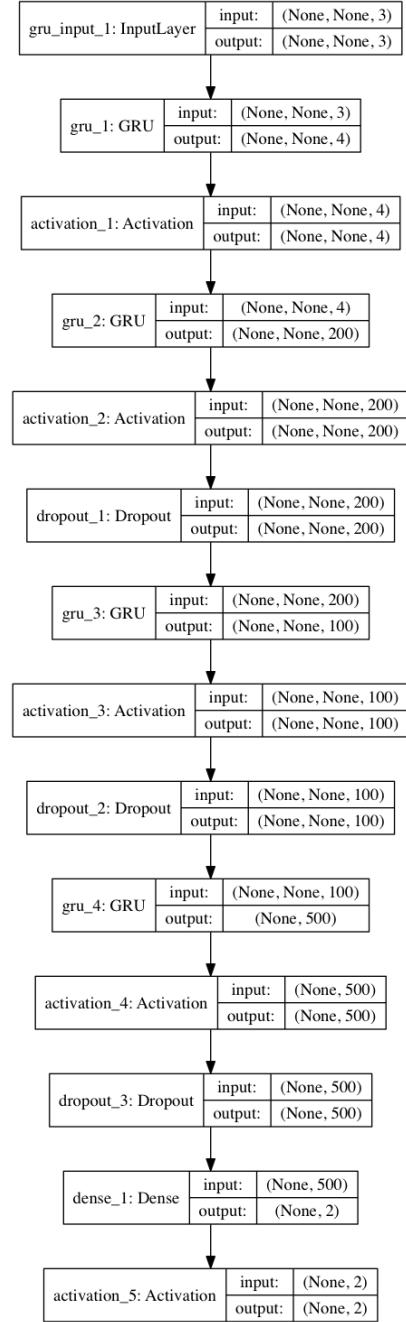


Fig. 8: Gated Recurrent Unit: Deeper Architecture

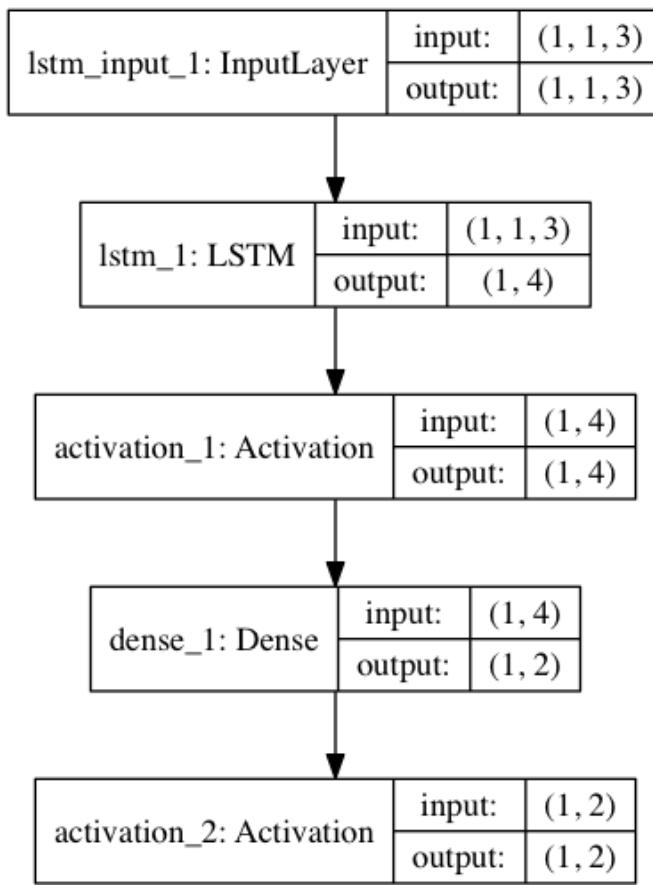


Fig. 9: Long short-term memory: Simple Architecture

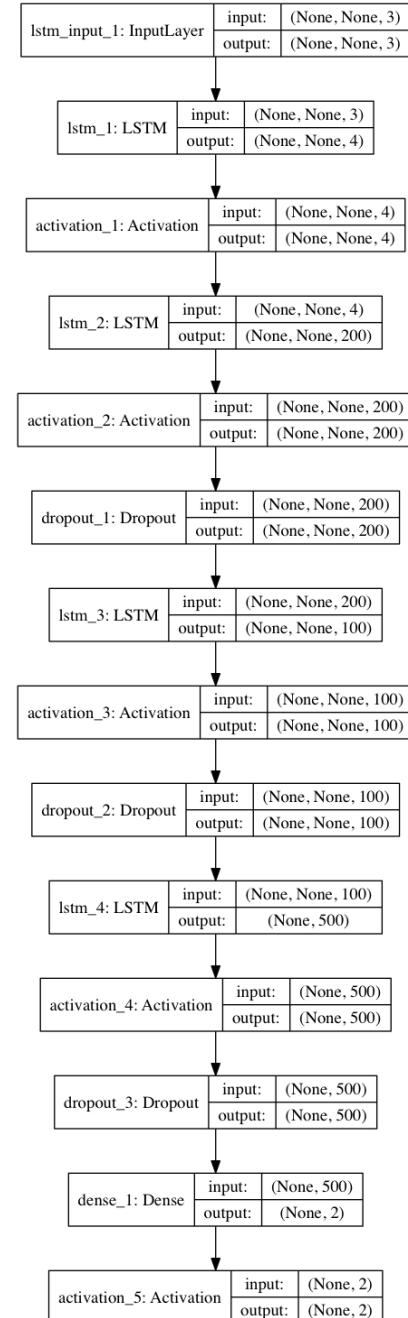


Fig. 10: Long short-term memory: Deeper Architecture

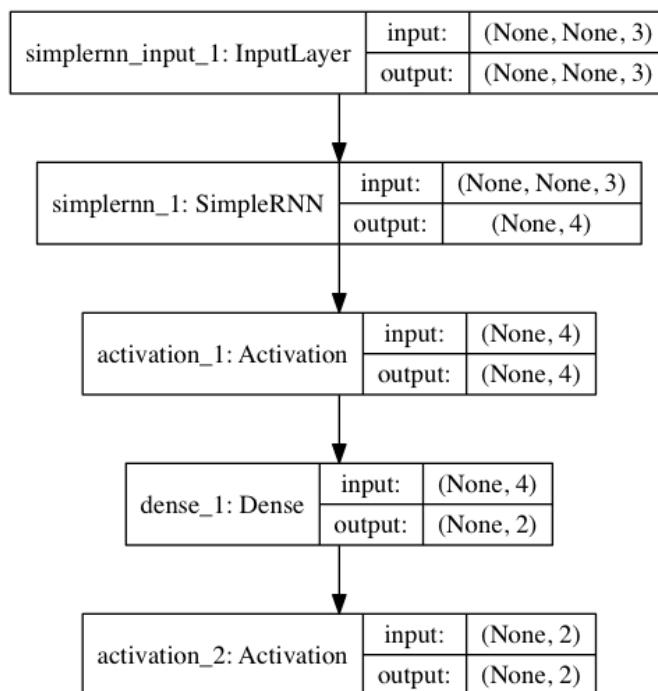


Fig. 11: SimpleRNN: Simple Architecture

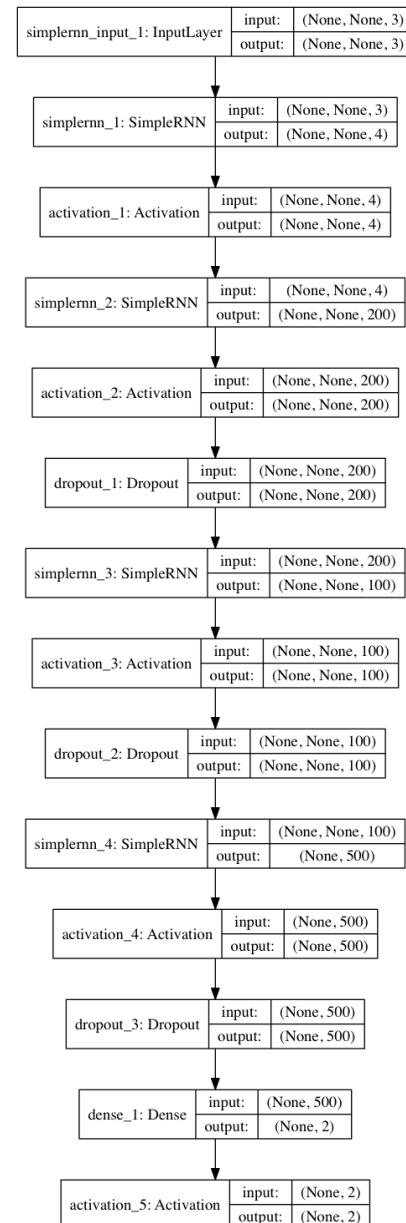


Fig. 12: SimpleRNN: Deeper Architecture

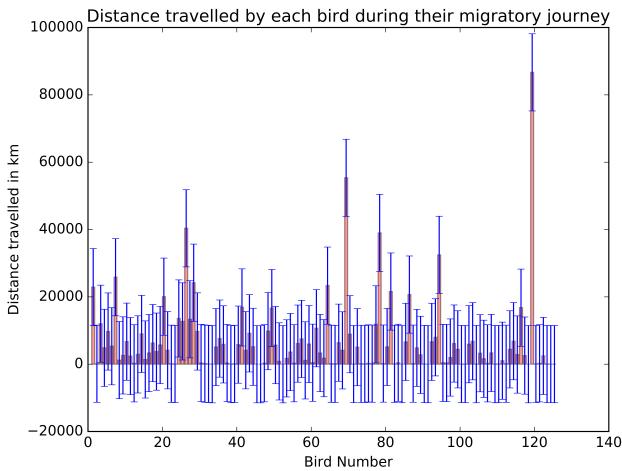


Fig. 13: Distance Traveled by each bird during its journey



Fig. 16: Vegetation Cover throughout the birds' migratory paths

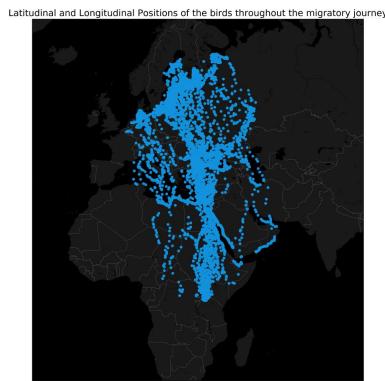


Fig. 14: Migratory path of the birds



Fig. 17: Clustered Hot-spots of the birds



Fig. 15: Population Density throughout the birds' migratory paths

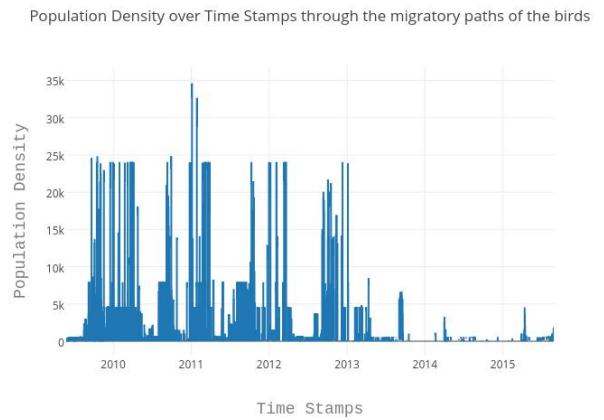


Fig. 18: Population Density across time during the migratory journey

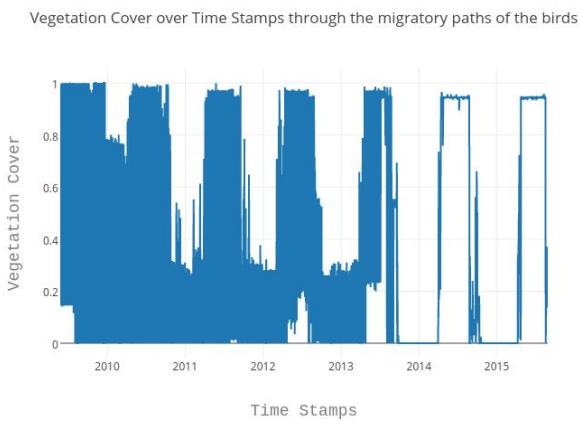


Fig. 19: Vegetation Cover across time during the migratory journey

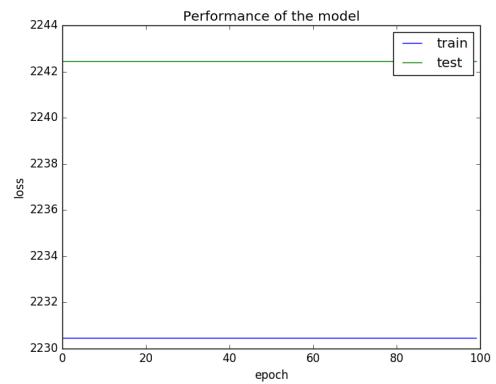


Fig. 22: Gated Recurrent Unit Simple Architecture: Loss vs. Epochs

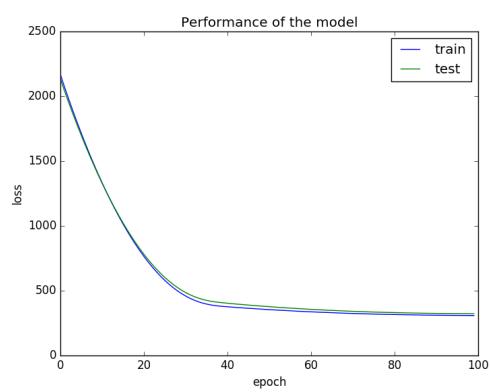


Fig. 20: LSTM Simple Architecture: Loss vs. Epochs

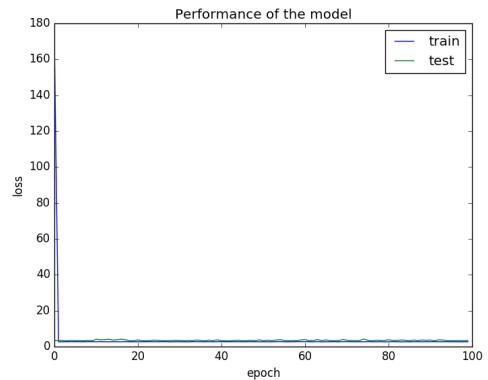


Fig. 23: Gated Recurrent Unit Deeper Architecture: Loss vs. Epochs

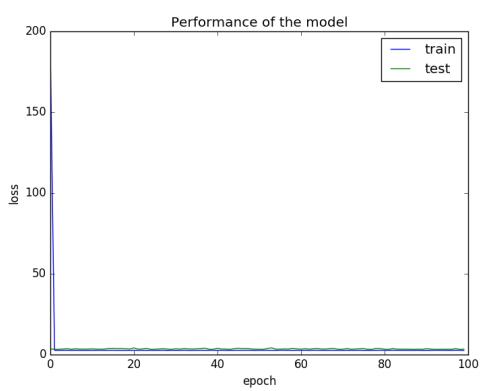


Fig. 21: LSTM Deeper Architecture: Loss vs. Epochs

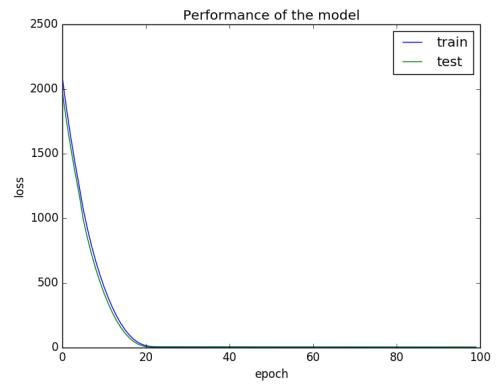


Fig. 24: SimpleRNN Simple Architecture: Loss vs. Epochs

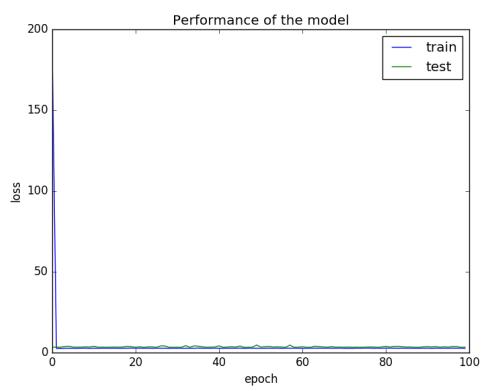


Fig. 25: SimpleRNN Deeper Architecture: Loss vs. Epochs