



DS 2018

VAST CHALLENGE 1
REPORT

Assignment 1

Students :

Fangda ZHU
Yu XIANG
Taha HALAL
Henri HOURS

Professor :

Jean-Daniel FEKETE



November 4, 2018

Contents

0.1	Overview for the Mini-Challenge	1
0.2	Mini-Challenge 1	2
1	Map Preprocessing	3
2	Data Preprocessing	4
3	First Visualizations	6
3.1	All records locations	6
3.2	With Kasios records locations	7
3.3	Evolution of the locations of all Birds since 2007	8
3.4	Evolution of the locations of Blue Pipits since 2007	9

Introduction

Data visualization uses various tools and technologies to communicate information clearly and efficiently. With the goal of advancing the field of visual analytics, the Visual Analytics Science and Technology (VAST) Challenge is held annually through competition. In VAST Challenge 2018, the Mini-Challenge 1 (MC1) will ask the participants to apply visual analytics research and technologies to help a continuing troublesome state of affairs in a wildlife preserve. The background for the challenge and the task for MC1 are described in the following¹:

0.1 Overview for the Mini-Challenge

Mistford is a mid-size city located to the southwest of the Boonsong Lekagul Wildlife Preserve. The city has a small industrial area with four light-manufacturing endeavors. Mistford and the wildlife preserve are struggling with the possible endangerment of the Rose-Crested Blue Pipit, a locally loved bird. The bird's nesting pairs seem to have decreased alarmingly, prompting an investigation last year implicating Kasios Office Furniture, a Mistford manufacturing firm. The investigation by Mitch Vogel and his professors suggested that the Kasios Furniture manufacturing company may have been a primary contributor to the apparent reduction of the number of nesting pairs of the Rose-Crested Blue Pipit, a favorite bird of Mistford residents and Boonsong Lekagul Nature Preserve visitors. Kasios supposedly used the banned substance Methylosmolene in their manufacturing process. They surreptitiously dumped process waste in the northeast region of the Preserve and Methylosmolene was detected in their smokestack emissions.

Since the initial investigation, the situation has evolved: Kasios now claims that the analysis by Mitch Vogel and his professors was flawed and biased. To combat these

¹For more information on VAST Challenge 2018, see <http://www.vacommunity.org/VAST+Challenge+2018>

conclusions, Kasios has launched their own “investigation” into the Pipit situation, and they are now reporting that there are plenty of Rose-crested Blue Pipits happily living and nesting in the Preserve. To back up this claim, they have provided a set of Pipit bird calls, recently recorded across the Preserve, with locations of where they were recorded. Clearly, they claim, the Pipits are a thriving population.

0.2 Mini-Challenge 1

Kasios Office Furniture claims obviously contradicts to the work done by grad student Mitch from last year(2017). To further investigate the the birds’ situation, other data sources are needed since Mitch is not available to help validate Kasios’ claim and Mistford College does not have another Pipit expert they can call upon for help. Fortunately, Mistford College do have a collection of bird calls from the Preserve that has been vetted by various ornithology groups as having accurate identifications, and new techniques from machine learning and visual analytics can be applied to situations like this.

In the Mini-Challenge 1 (MC1), the tasks will be as follows:

1. Using the bird call collection and the included map of the Wildlife Preserve, characterize the patterns of all of the bird species in the Preserve over the time of the collection. It is assumed that we have a reasonable distribution of sensors and human collectors providing the recordings, so that the patterns are reasonably representative of the bird locations across the area.

Check whether any trends or anomalies in the patterns could be detected and present the result with a limit to 10 images and 1000 words.

2. Turn your attention to the set of bird calls supplied by Kasios. Does this set support the claim of Pipits being found across the Preserve? A machine learning approach using the bird call library may help the investigation. What is the role of visualization in the analysis of the Kasios bird calls?

Use machine learning techniques to analyze above questions. When presenting the result, limit the answer to 10 images and 1000 words.

3. Formulate a hypotheses concerning the state of the Rose Crested Blue Pipit. What are your primary pieces of evidence to support your assertion? What next steps should be taken in the investigation to either support or refute the Kasios claim that the Pipits are actually thriving across the Boonsong Lekagul Wildlife Preserve?

When presenting the result, limit the answer to 10 images and 1000 words.

We aim to finish this MC1 while taking the course Visualization and Visual Analytics for Data Science 2018-2019 at Telecom ParisTech. MC1 is divided into several small assignments for the course. And this report is for the assignment 1 of the course, which is also the initial step for Mini-Challenge 1. In this report(i.e. assignment 1), we

1. Show the map in a nice way.
2. Show all the birds positions from AllBirdsv4.csv over the map
3. Improve the map to be more readable
4. Show the 15 positions of the birds from Kasios

1 Map Preprocessing

We used the skimage library to vectorize the image.

Because the map is only a support to situate the bird records locations, colors have to be neutral and lines very tight; so its better to vectorize contours of roads than vectorize roads (to visualize roads in this case, lines are to large).

We chose to display the lines of the grids (gray, dot lines), in order to locate the records in space.

The dumping site of Kasios is important as landmark (in red, because of the suspected negative impact on the environment).

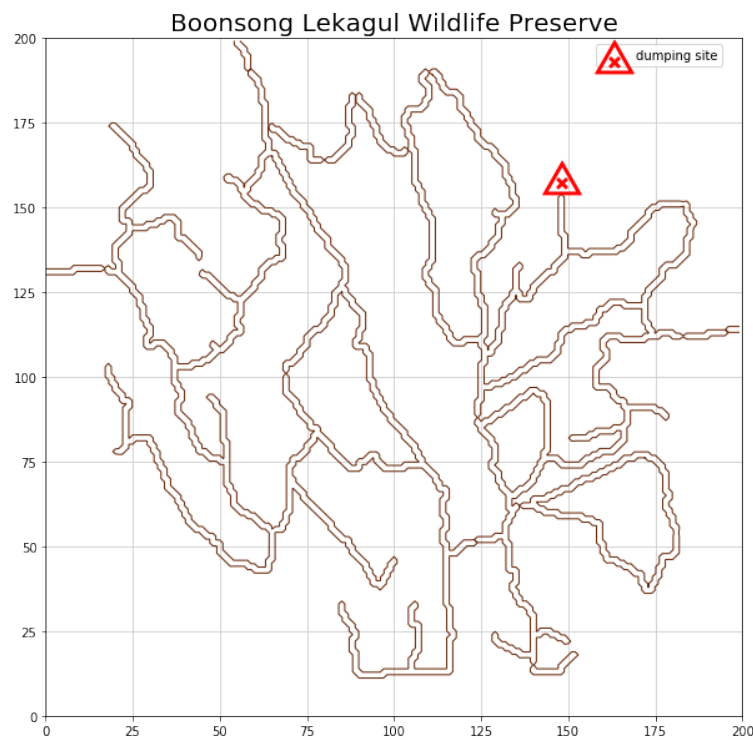


Figure 1: Map display

This map will serve as a background for all geographic informations.

2 Data Preprocessing

After loading the .csv, we checked and clean all data (missing values, lowercase or majuscule, good types,...)

We wanted to visualize columns with distinguishing Blue Pipits from other species. We chose plotting Blue Pipits in blue (with respect to their name) and other in green (we will keep this color distinction all along our study).

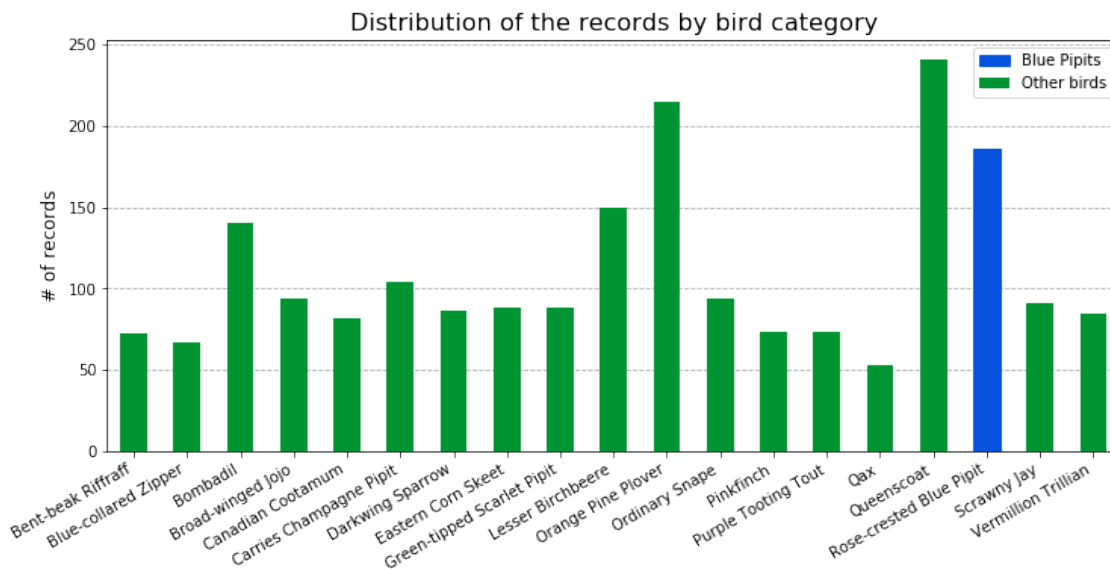


Figure 2: Distribution of the records by bird species

All categories are consistent, and the Blue pipits records constitute a sufficient percentage (8.94%) of the records to work.

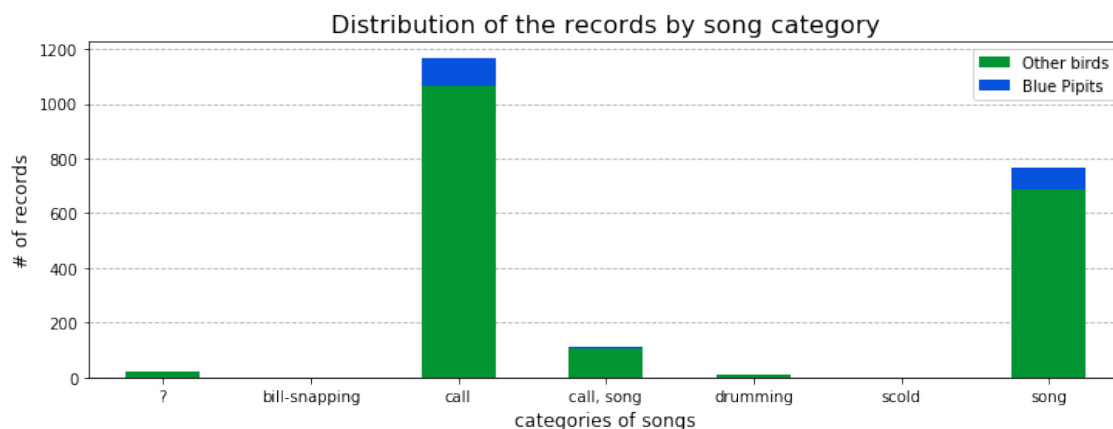


Figure 3: Distribution of the records by song category

Call & songs are the main categories. We can later remove others (to be checked), their audio wave will likely be outliers.

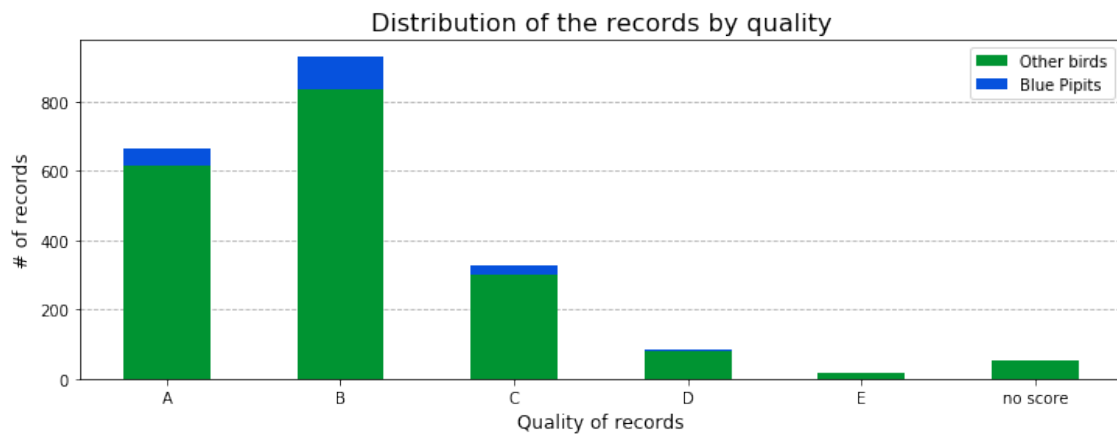


Figure 4: Distribution of the records by quality

We have sufficient Blue Pipit records with quality A, B and C. We will remove all D, E and ?-quality records, to save time and efficiency.

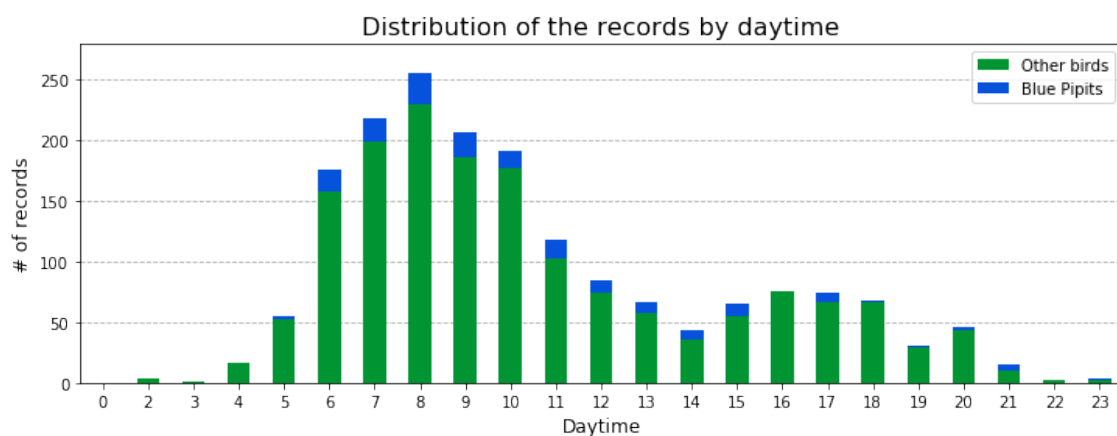


Figure 5: Distribution of the records by daytime

No particular conclusion.

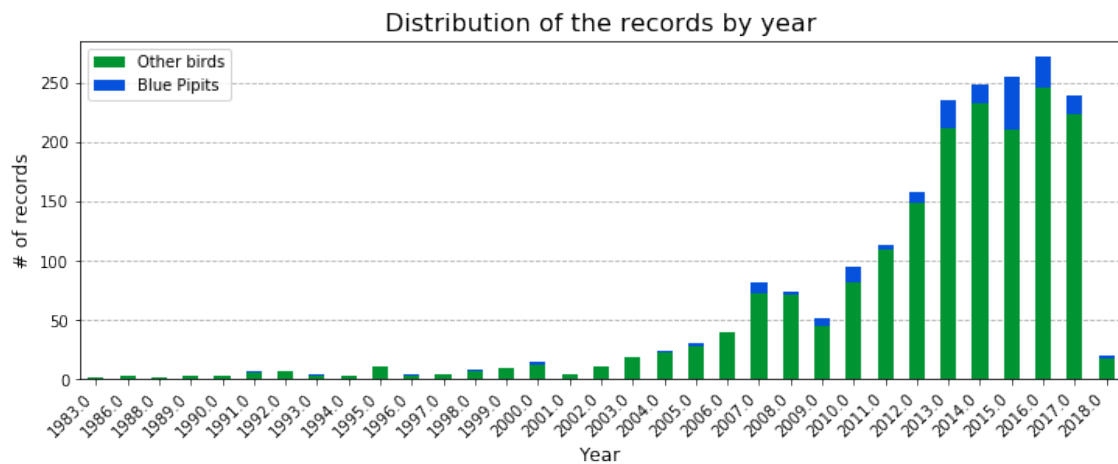


Figure 6: Distribution of the records by year

We don't have a lot of records for 2018. Furthermore, the interesting period to study Blue Pipits is after 2006.

3 First Visualizations

3.1 All records locations

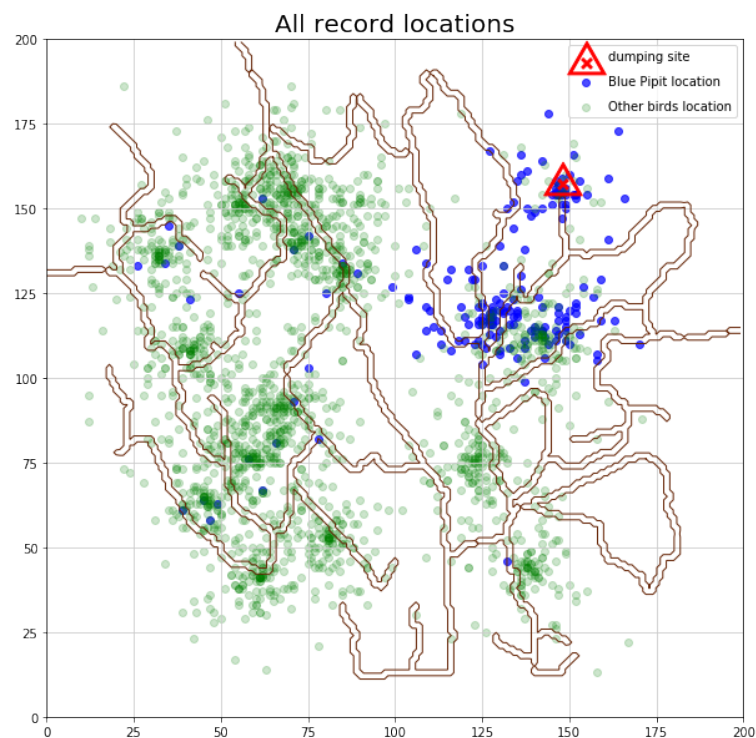


Figure 7: All bird record locations

We don't have a lot of records for 2018. Furthermore, the interesting period to study Blue Pipits is after 2006.

It seems that there are high density plots, birds have some habits and we will be able to study their behavior. It will have to be studies timely.

At first glance, pipit blues have nested or still nest in the area of dumping site.

3.2 With Kasios records locations

For Kasios records locations, we chose another color(yellow), neither green nor blue, because we have to confirm if their are Blue Pipits or not.

To be able to refer to a singular record, we add the ID of the record.

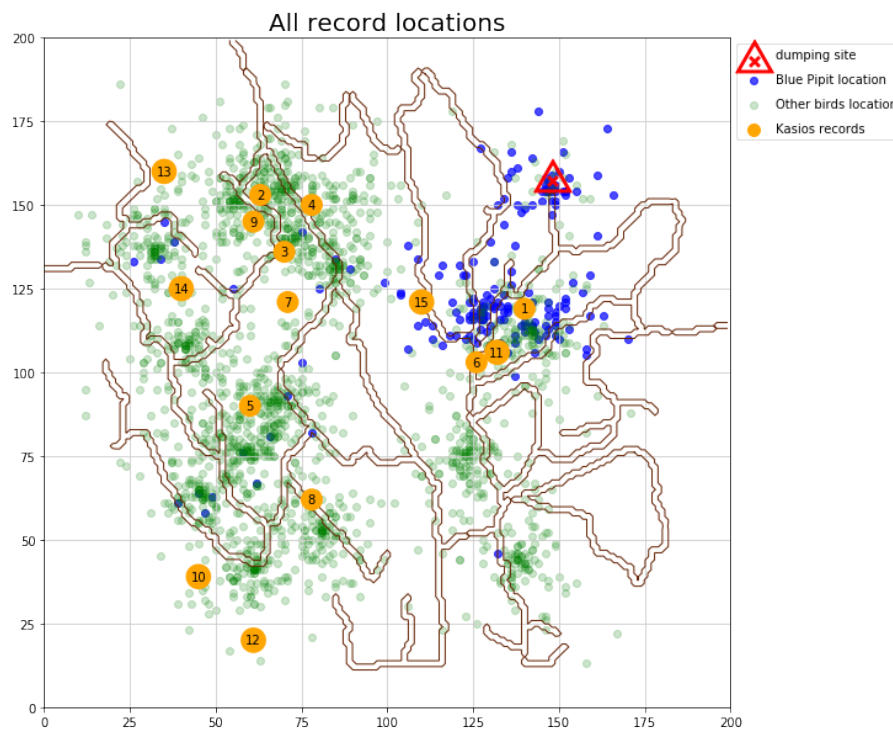
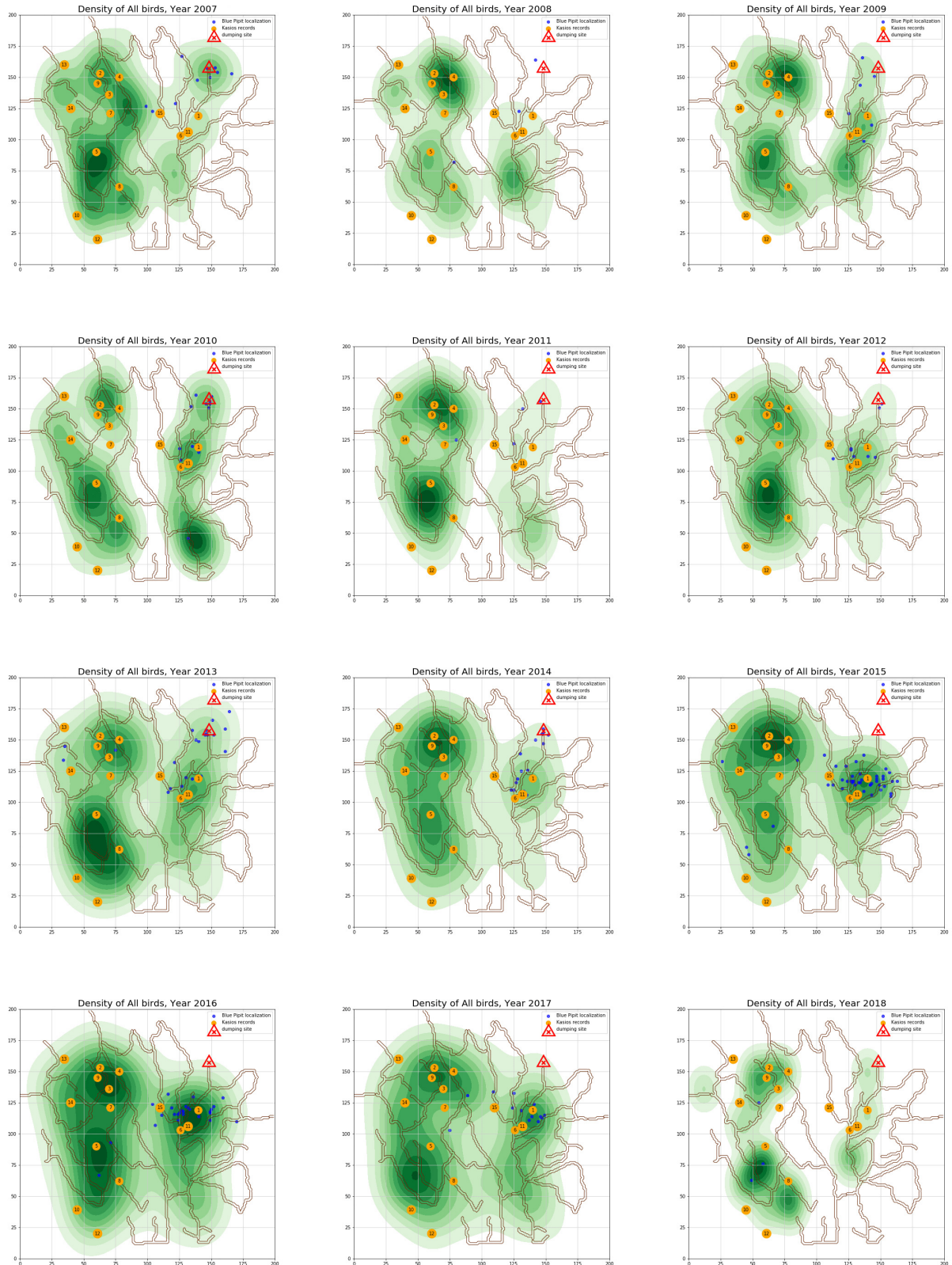


Figure 8: All bird record locations

The records #1, #6, #11, and #15 are located in a highly concentrated (current or passed) area of Blue pipits, in the vicinity of the dumping site.

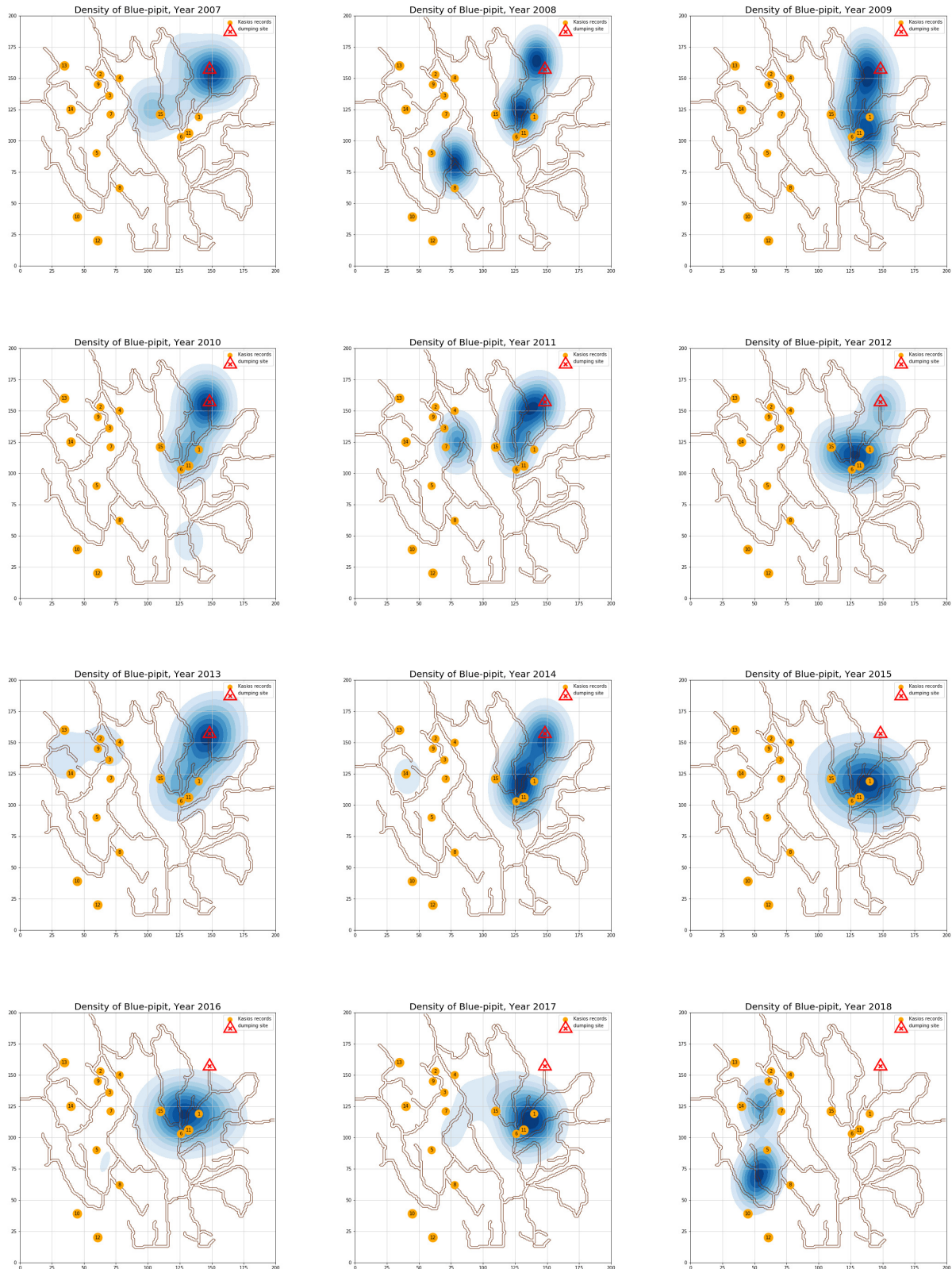
3.3 Evolution of the locations of all Birds since 2007

To be able to compare with Blue Pipits locations, lets start studying all birds.



3.4 Evolution of the locations of Blue Pipits since 2007

To be able to understand the repartition of Blue Pipits, we need visualize its timely evolution (since 2007, according to the study of the yearly distribution).



It will be interesting to study especially records #1, #11, #6 and #15, and check if there are actually blue pipits records. Otherwise all blue pipits could have migrated to the area between (50, 150) and (50, 50).

Conclusion

With this first part, we have sufficient informations to orient the studies that will follow.

Firstly we are able to plot the records on a map.

Secondly, we have consistent samples with only records with quality A, B and C, and selecting only songs and calls

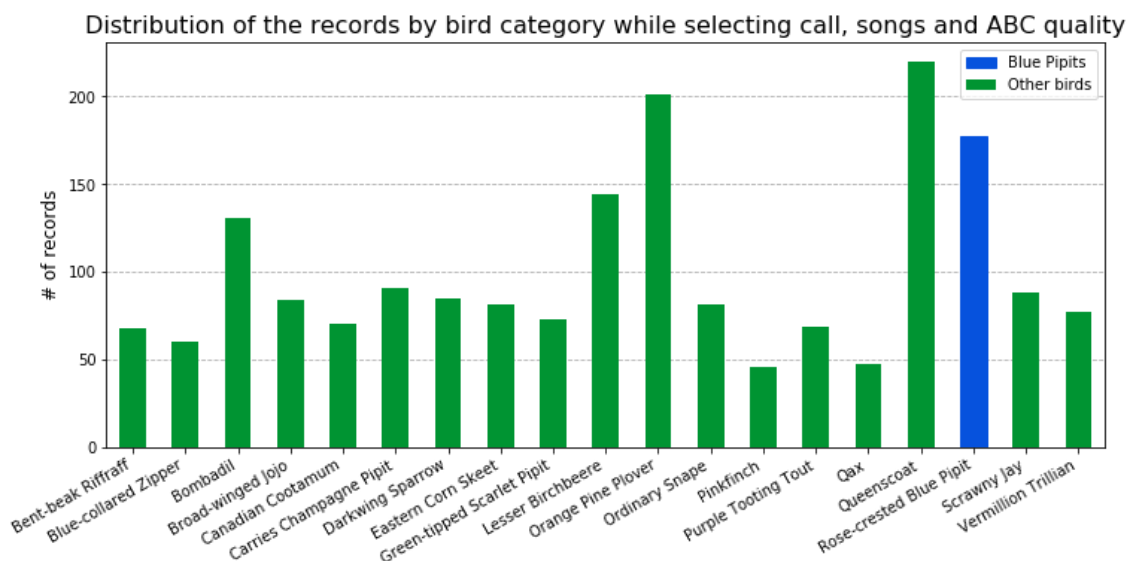


Figure 9: Distribution of the 1893 remaining records by species, while selecting only ABC quality and calls and songs

Endly, Blue Pipits nest or have nested IVO the dumping site, and we don't have any record of Blue Pipit in this area for 2018.

We now have to study the birds sounds, and checked if Kasios records are Blue Pipit records, especially records #1, #11, #6 and #15.