Birds with highest discrepancy in observation patterns across years

Methodology

Jensen Distance for smoothed curves is chosen as the distance metric to quantify "distance" between two different observation curves. 4 such distances were computed for each bird:

- iNat-19 vs iNat-22 (Column Name - iNat)
- eBird-19 vs eBird-22 (Column Name - eBird)
- iNat-19 vs eBird-19 (Column Name - Y19)
- iNat-22 vs eBird-22 (Column Name - Y22)

Since there is an implicit assumption of yearly seasonality in our assumption, birds where there is little distance between datasets (Y19 and Y22 colums as defined above) but where there is appreciable distance across years (iNat and eBird columns as defined above) are extremely concerning.

To quantitatively filter for such cases, two metrics were devised (using columns as defined earlier):

1. Highest Discrepancy Difference which is: $$\frac{iNat + eBird}{2} - \frac{Y19 + Y22}{2} $$
2. Highest Discrepancy Ratio which is: $$\frac{iNat + eBird}{Y19 + Y22}$$

The words 'difference' and 'ratio' below would be used to describe these above mentioned measures.

To select anamolous birds, the thresholds for both difference and ratio must be met. The following examples help us understand the reason:

### Example 1

Bird 68EA281A has a difference of 0.038106 and a ratio of 1.817987. While the ratio is large, that is because of a small denominator in computation as seen from the graph below:
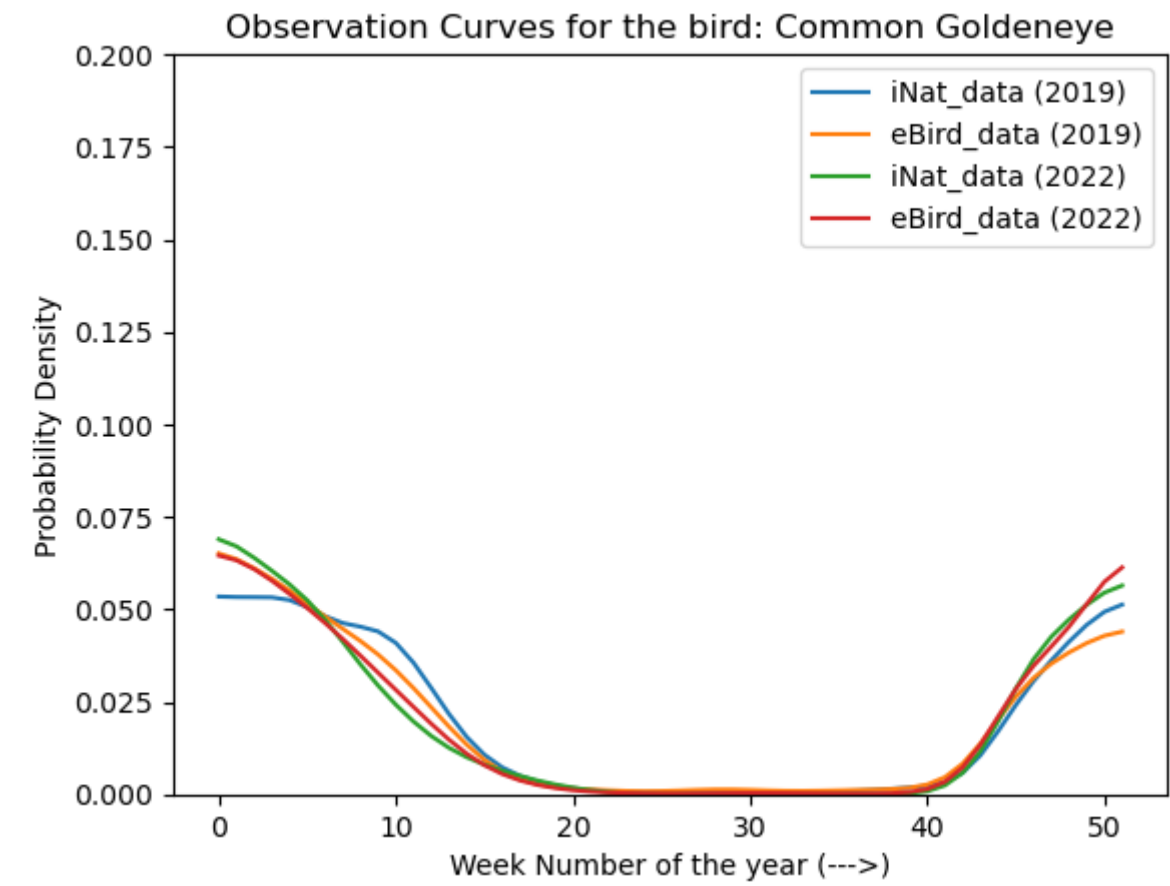
Example 2

Bird 135564AA has a difference of 0.113568 and a ratio of 1.374282. While the difference is large, that is because the curves are further apart from each other across both years and datasets.

Observation Curves for the bird: Red Crossbill
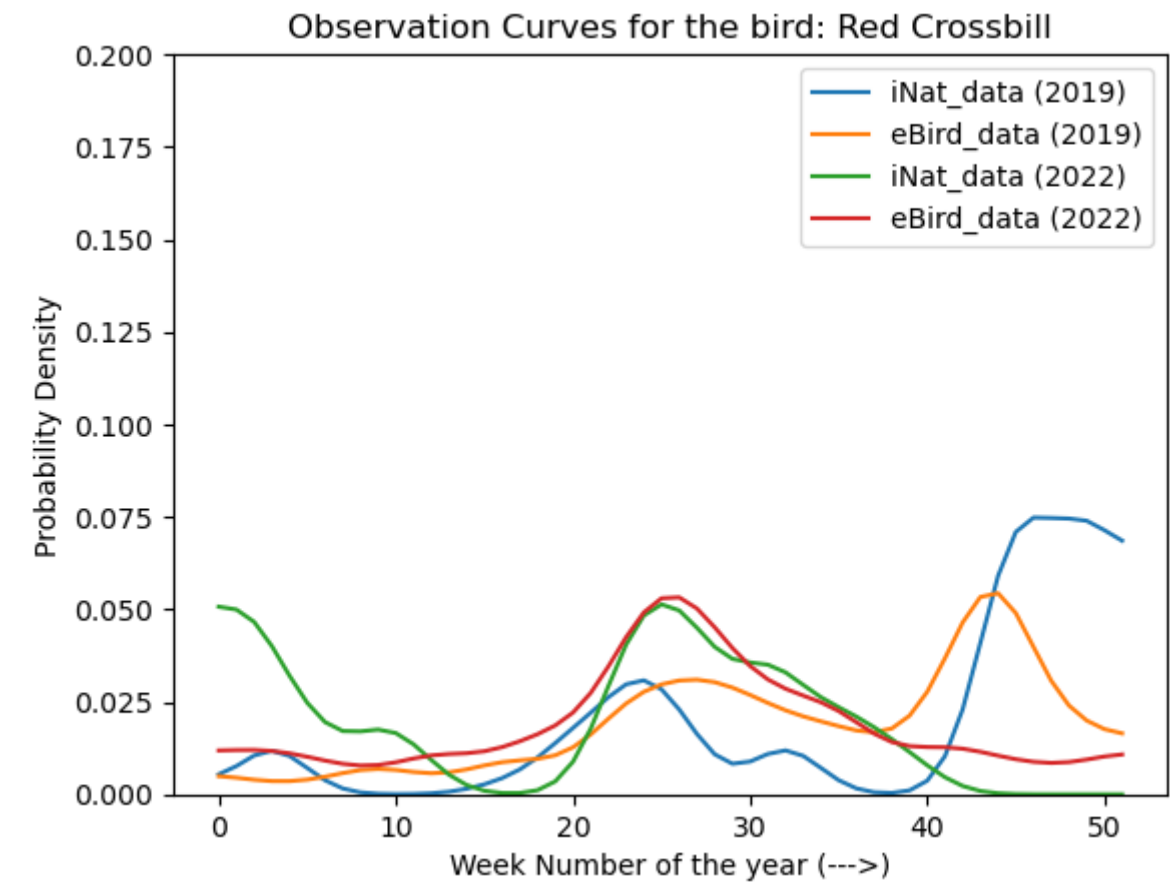
Hence, based on this, and after some tinkering with thresholds, birds with a ratio > 1.5 and difference > 0.08 were selected.

Anamolous Birds

Based on above the anamolous birds are:

	AvibaseID	iNat	eBird	Y19	Y22	Discrepancy diff	Discrepancy ratio
0	2DABF98F	0.459	0.447	0.124	0.154	0.314	3.267
1	AF0C7BDE	0.563	0.393	0.238	0.065	0.327	3.154
2	BCBD2EAE	0.551	0.494	0.201	0.167	0.339	2.838
3	4AAEF549	0.554	0.489	0.174	0.197	0.336	2.811
4	C5971E25	0.559	0.507	0.174	0.212	0.34	2.758
5	B48335B1	0.419	0.318	0.152	0.131	0.226	2.597
6	E1714A09	0.695	0.563	0.335	0.221	0.35	2.26
7	1456993D	0.42	0.337	0.223	0.135	0.199	2.109
8	4E192262	0.72	0.4	0.35	0.19	0.29	2.075
9	BB6B236F	0.216	0.231	0.134	0.086	0.114	2.038
10	0716A71C	0.329	0.221	0.171	0.104	0.137	1.999

	AvibaseID	iNat	eBird	Y19	Y22	Discrepancy diff	Discrepancy ratio
11	F6DF700B	0.512	0.386	0.256	0.202	0.22	1.96
12	28A4DA76	0.364	0.186	0.168	0.117	0.133	1.931
13	90B13ACF	0.438	0.17	0.237	0.126	0.123	1.678

The graphs for each of the birds is present below:

Observation Curves for the bird: 2DABF98F - Ruff

Observation Curves for the bird: AF0C7BDE - Long-tailed Duck

Observation Curves for the bird: BCBD2EAE - Harlequin Duck

Observation Curves for the bird: 4AAEF549 - Palm Warbler

Observation Curves for the bird: C5971E25 - Rock Sandpiper

Observation Curves for the bird: B48335B1 - Pine Siskin

Observation Curves for the bird: E1714A09 - Mandarin Duck

Observation Curves for the bird: 1456993D - Vermilion Flycatcher

Observation Curves for the bird: 4E192262 - Swamp Sparrow

Observation Curves for the bird: BB6B236F - Tropical Kingbird

Observation Curves for the bird: 0716A71C - Red-breasted Nuthatch

Observation Curves for the bird: F6DF700B - Sage Thrasher

Observation Curves for the bird: 28A4DA76 - Williamson's Sapsucker

Observation Curves for the bird: 90B13ACF - Rough-legged Hawk