

# Occupancy Modeling

Ethan Schoen



The background is a solid orange color. In the top-left corner, there are three vertical bars of varying heights, each composed of three overlapping circles. In the bottom-right corner, there are four vertical bars of varying heights, each composed of four overlapping circles.

**Goal: Determine reliability  
of MegaDetector for  
monthly presence/absence  
data of species at site**



# MegaDetector

- Computer vision model
  - Identifies animals vs. humans vs. empty
    - Does not classify species
- Some data loss due to low confidence
- How impactful is this data loss?



# Data Cleaning

- Grouping species
  - Handling misspellings and rare species
- Converting to actual .csv file
  - No commas within entry
- Selecting relevant columns
- Converting to proper format
- Calculating season

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	data	photo_group	check_date	photo_name	male	model	DateTimeOf	ExifImageWidth	ExifImageHeight	Timestamp	TimeOf	numAnimalsD	numAnimalsI	numAnimalsC	numAnimalsO	maxDetection	maxDetection	PhotoName	commonName	SpeciesData	Count
2	IB1	ARB1_2020_12_29	IMG_0001.JPG	BROWNING	BTG-4HX	2020-12-01	3840	2160	2020-12-01-04T			0	0	0	0	0	0	0	0	0	1
3	IB1	ARB1_2020_12_29	IMG_0004.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-04-04T	43786		1	Category: 1 Conf: 0.818 [0.809, 0.835]	0.3091	0.156	0.2545	0	0	NA	0	0.136
4	IB1	ARB1_2020_12_29	IMG_0005.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-04-04T	2420		1	Category: 1 Conf: 0.897 [0.860, 0.932]	0.0902	0.2123	0.2435	0	0	NA	0	0.987
5	IB1	ARB1_2020_12_29	IMG_0166.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-05-04T	53116		1	Category: 1 Conf: 0.884 [0.869, 0.900]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.986
6	IB1	ARB1_2020_12_29	IMG_0167.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-05-04T	4296		1	Category: 1 Conf: 0.986 [0.986, 0.986]	0.0208	0.0704	0.2376	0	0	NA	0	0.986
7	IB1	ARB1_2020_12_29	IMG_0168.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-05-04T	13075		1	Category: 1 Conf: 0.889 [0.860, 0.908]	0.0283	0.0638	0.2516	0	0	NA	0	0.989
8	IB1	ARB1_2020_12_29	IMG_0169.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-05-04T	197		1	Category: 1 Conf: 0.980 [0.952, 0.988]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.994
9	IB1	ARB1_2020_12_29	IMG_0170.JPG	BROWNING	BTG-4HX	2020-12-04	3840	2160	2020-12-05-04T	3214	0	NA	0	0	0	0	0	0	0	0	0
10	IB1	ARB1_2020_12_29	IMG_0331.JPG	BROWNING	BTG-4HX	2020-12-05	3840	2160	2020-12-06-04T	66636	1	Category: 1 Conf: 0.986 [0.980, 0.993]	0.0416	0.0702	0.2712	0	0	NA	0	0.986	0.986
11	IB1	ARB1_2020_12_29	IMG_0332.JPG	BROWNING	BTG-4HX	2020-12-05	3840	2160	2020-12-06-04T	6338	1	Category: 1 Conf: 0.868 [0.860, 0.881]	0.4448	0.0888	0.1719	0.0888	0.0888	0.0888	0.0888	0.0888	0.988
12	IB1	ARB1_2020_12_29	IMG_0333.JPG	BROWNING	BTG-4HX	2020-12-05	3840	2160	2020-12-06-04T	7143	1	Category: 1 Conf: 0.986 [0.986, 0.986]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.986
13	IB1	ARB1_2020_12_29	IMG_0334.JPG	BROWNING	BTG-4HX	2020-12-05	3840	2160	2020-12-06-04T	3688	2	Category: 1 Conf: 0.992 [0.992, 0.992]	0.0639	0.0843	0.1983	0.047	0.0554	0.0605	0.0605	0.0605	0.984
14	IB1	ARB1_2020_12_29	IMG_0335.JPG	BROWNING	BTG-4HX	2020-12-05	3840	2160	2020-12-06-04T	195	1	Category: 1 Conf: 0.991 [0.980, 0.994]	0.0294	0.0293	0.0435	0	0	NA	0	0.991	0.991
15	IB1	ARB1_2020_12_29	IMG_0336.JPG	BROWNING	BTG-4HX	2020-12-06	3840	2160	2020-12-06-04T	21280	3	Category: 1 Conf: 0.986 [0.986, 0.986]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.986
16	IB1	ARB1_2020_12_29	IMG_0369.JPG	BROWNING	BTG-4HX	2020-12-06	3840	2160	2020-12-06-04T	23250	2	Category: 1 Conf: 0.411 [0.411, 0.917]	0.2855	0.0214	0.1232	0.047	0.155	0.091	0.091	0.091	0.387

	A	B	C	D	E	F	G
1	locAbbr	Timestamp	commonName	Latitude	Longitude	maxDetection	Season
2	ARB1	2020-12-16T	Empty	34.2057	-118.161	0	10
3	ARB1	2020-12-16T	Empty	34.2057	-118.161	0.286	10
4	ARB1	2020-12-16T	Empty	34.2057	-118.161	0	10
5	ARB1	2020-12-16T	Virginia opos	34.2057	-118.161	0.991	10
6	ARB1	2020-12-16T	Empty	34.2057	-118.161	0.642	10
7	ARB1	2020-12-16T	Domestic ca	34.2057	-118.161	0	10
8	ARB1	2020-12-17T	Virginia opos	34.2057	-118.161	0.994	10
9	ARB1	2020-12-19T	Virginia opos	34.2057	-118.161	0.999	10
10	ARB1	2020-12-19T	Unknown	34.2057	-118.161	0.97	10
11	ARB1	2020-12-19T	Unknown	34.2057	-118.161	0.954	10
12	ARB1	2020-12-19T	Unknown	34.2057	-118.161	0.984	10
13	ARB1	2020-12-19T	Empty	34.2057	-118.161	0.577	10
14	ARB1	2020-12-19T	Unknown	34.2057	-118.161	0.956	10
15	ARB1	2020-12-19T	Virginia opos	34.2057	-118.161	0.997	10
16	ARB1	2020-12-19T	Empty	34.2057	-118.161	0	10

# Setting Up the Files

```
# For each entry in dataframe
for (i in 1:nrow(df_cur)) {
  # Get values of interest
  season <- df_cur$Season[i]
  relative_day <- df_cur$relative_day[i]
  location <- df_cur$locAbbr[i]

  # Get name of day column
  day_column_name <- paste0("Day_", relative_day)

  # Convert current entry's corresponding cell in cur_grid to 1
  cur_grid[cur_grid$Season == season & cur_grid$locAbbr == location, day_column_name] <- 1
}

# Add column specifying source
updated_grid <- cur_grid %>%
  mutate(Source = df_name)
# Add dataframe to list
results_list[[df_name]] <- updated_grid
}

# Put MD data below human data
complete_grid <- bind_rows(results_list)
```

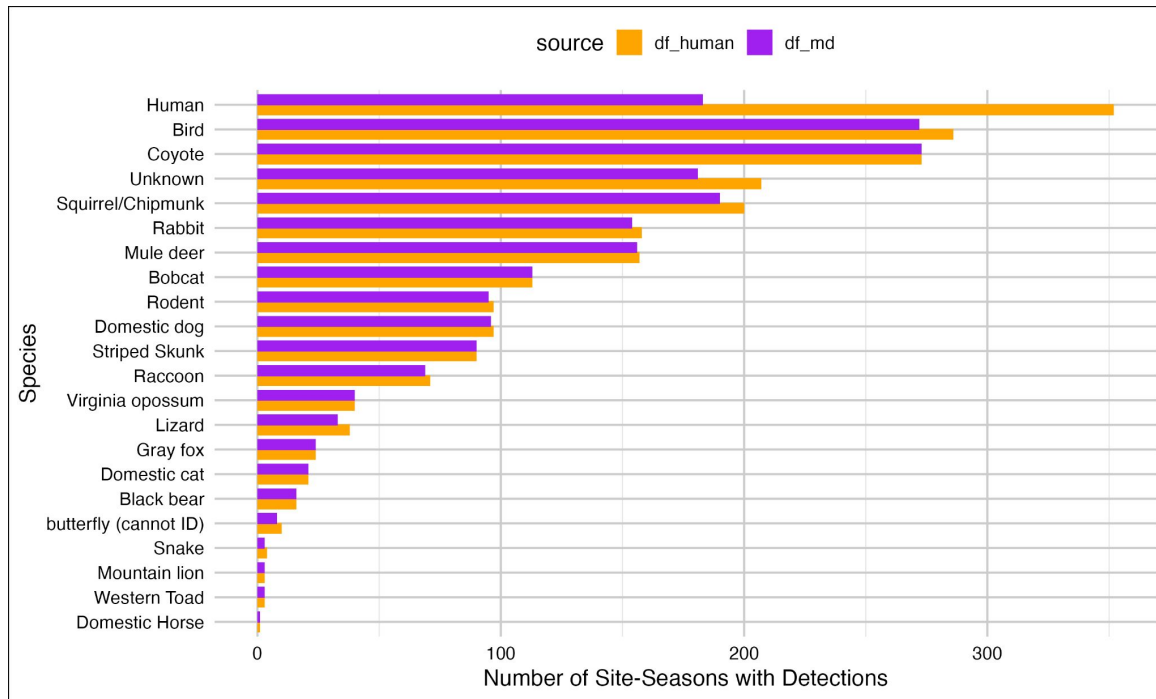
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Species	Season	Site	Latitude	Longitude	Day_1	Day_2	Day_3	Day_4	Day_5	Day_6	Day_7	Day_8	Day_9	Day_10	Treatment
400	Coyote	6	ORR1	34.1269	-118.209	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
401	Coyote	6	ORR2	34.1265	-118.21	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
402	Coyote	6	RHR1	34.218	-118.159	0	0	0	0	0	0	0	0	0	0	df_human
403	Coyote	6	RHR2	34.2174	-118.16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
404	Coyote	6	ROR1	34.1557	-118.168	NA	0	0	0	0	0	0	0	0	0	df_human
405	Coyote	6	ROR2	34.1545	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
406	Coyote	6	ROR3	34.1548	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
407	Coyote	6	ROSA	34.1541	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
408	Coyote	6	SAB1	34.1294	-118.164	NA	NA	NA	NA	0	0	0	0	0	0	df_human
409	Coyote	6	SPS1	34.1243	-118.167	1	0	0	0	0	0	0	0	0	1	df_human
410	Coyote	6	SPS2	34.1242	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
411	Coyote	6	SPS3	34.1244	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
412	Coyote	6	SPS4	34.1246	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
413	Coyote	6	SUN1	34.1771	-118.238	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
414	Coyote	6	SUN2	34.1761	-118.238	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
415	Coyote	6	TCR1	34.2523	-118.538	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
416	Coyote	6	TMC1	34.1767	-118.181	0	0	0	1	1	0	1	1	0	1	df_human
417	Coyote	6	TMC2	34.1756	-118.181	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
418	Coyote	6	WAW1	34.243	-118.311	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	df_human
419	Coyote	6	WAW2	34.2477	-118.308	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
420	Coyote	6	WAW3	34.2462	-118.304	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
421	Coyote	6	WAW4	34.2472	-118.305	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
422	Coyote	6	WHT1	34.2713	-119.351	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
423	Coyote	6	WPT1	34.2713	-118.518	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	df_human
424	Coyote	6	WTP1	34.2208	-118.518	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
425	Coyote	6	WWO1	34.2166	-118.286	0	0	0	0	0	0	0	0	0	0	df_human
426	Coyote	6	WWO2	34.2166	-118.286	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
427	Coyote	6	HOL1	34.0395	-118.216	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
428	Coyote	7	ABR1	34.2857	-118.161	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	df_human
429	Coyote	7	AST1	34.1382	-118.167	NA	0	1	0	0	0	0	1	0	0	df_human

- Expanding day columns
  - Mark inactive as NA and active as 0
- For each image of species, put 1 in cell for Season, Site, Day
  - Repeat for images with  $\geq 0.8$  confidence for MD
- Checking for accuracy
  - Took a surprising amount of time (and stress)
  - For photos with low confidence, often other photos that day that are above confidence threshold
- Run Occupancy Modeling File for each Species



# Results

Comparing number of Site-Seasons with detections from human-approved images with MD data. (Data loss per species)



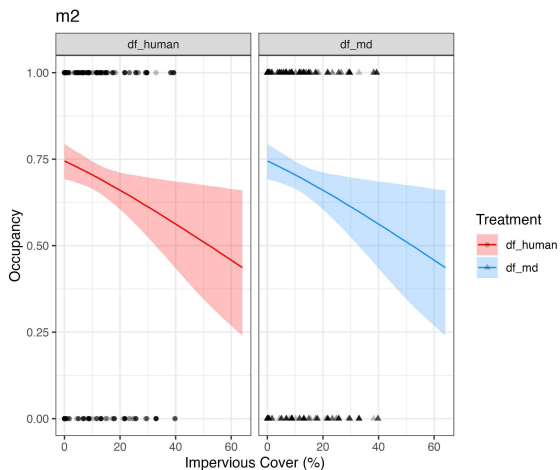


# Results

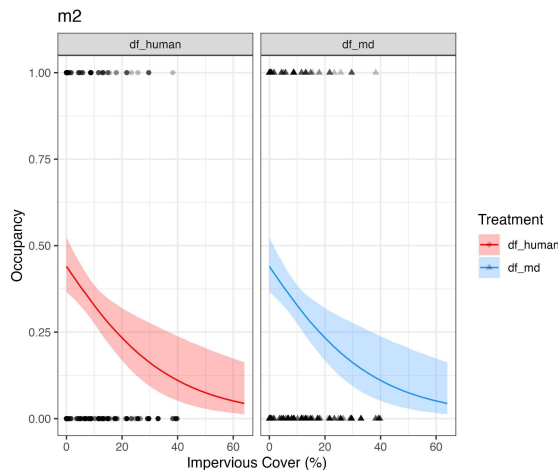
5 models:

1. m1: intercept-only (average)
2. m2: Impervious as predictor
3. m3: Treatment as predictor (human vs. MD)
4. m4: Combined Impervious and Treatment
5. m5: Interaction of Impervious and Treatment

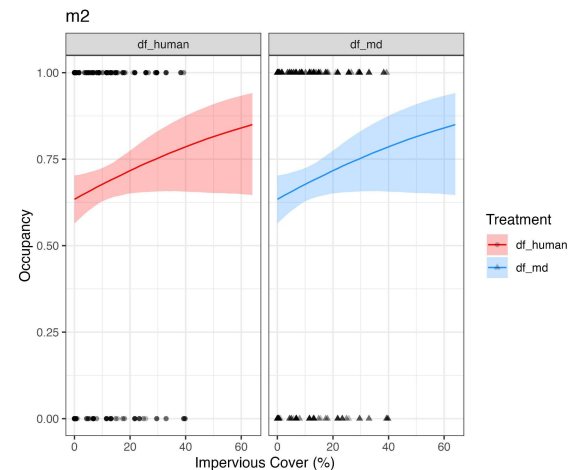
Bird



Bobcat



Coyote

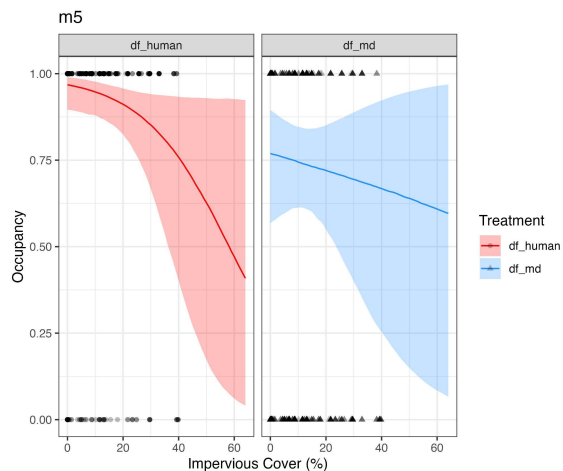


Interesting graphs of best models for some species

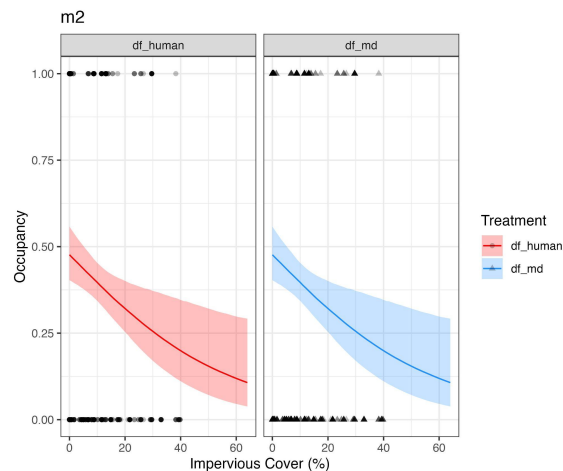


# Results

Human



Mule Deer



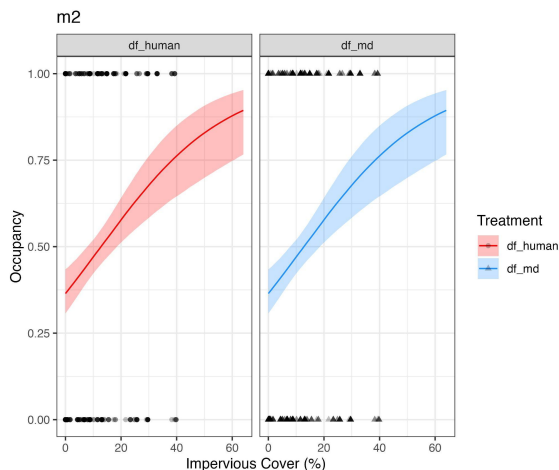




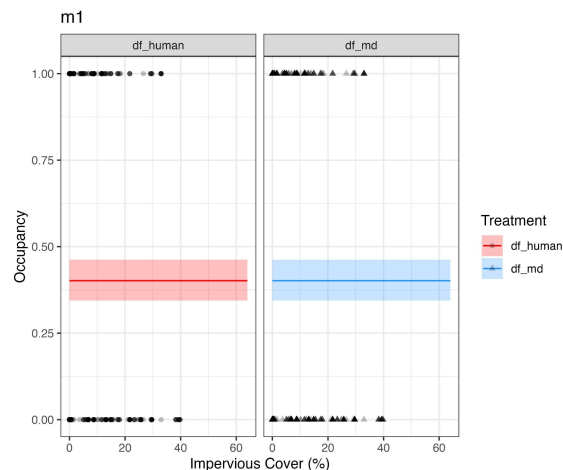
# Results

m2 typically the best model (Impervious alone is best predictor), which is good news! MD seems to not have a substantive impact on predicting occupancy rates.

Squirrel-Chipmunk



Rabbit





## Next Steps

- Explore more locations (not just Los Angeles)
- Test different confidence thresholds
- Check smaller species groups
- What else?

