Occupancy Modeling

Ethan Schoen

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

	A	B	C	D	E	F	G	H	I	J
1	cAbbr	photo_group	check_date	photo_name	make	model	DateTimeOri	ExifImageWic	ExifImageHei	Timestamp
2	RB1	ARB1_2020_:	2020_12_29	IMG_0001.JP	BROWNING	BTC-6HDX	2020:12:03 1	3840	2160	2020-12-03T
3	RB1	ARB1_2020_:	2020_12_29	IMG_0004.JP	BROWNING	BTC-6HDX	2020:12:04 0	3840	2160	2020-12-04T
4	RB1	ARB1_2020_:	2020_12_29	IMG_0005.JP	BROWNING	BTC-6HDX	2020:12:04 0	3840	2160	2020-12-04T
5	RB1	ARB1_2020_:	2020_12_29	IMG_0166.JP	BROWNING	BTC-6HDX	2020:12:04 1	3840	2160	2020-12-05T
6	RB1	ARB1_2020_:	2020_12_29	IMG_0167.JP	BROWNING	BTC-6HDX	2020:12:04 1	3840	2160	2020-12-05T
7	RB1	ARB1_2020_:	2020_12_29	IMG_0168.JP	BROWNING	BTC-6HDX	2020:12:04 2	3840	2160	2020-12-05T
8	RB1	ARB1_2020_:	2020_12_29	IMG_0169.JP	BROWNING	BTC-6HDX	2020:12:04 2	3840	2160	2020-12-05T
9	RB1	ARB1_2020_:	2020_12_29	IMG_0170.JP	BROWNING	BTC-6HDX	2020:12:04 2	3840	2160	2020-12-05T
10	RB1	ARB1_2020_:	2020_12_29	IMG_0331.JP	BROWNING	BTC-6HDX	2020:12:05 1	3840	2160	2020-12-06T
11	RB1	ARB1_2020_:	2020_12_29	IMG_0332.JP	BROWNING	BTC-6HDX	2020:12:05 1	3840	2160	2020-12-06T
12	RB1	ARB1_2020_:	2020_12_29	IMG_0333.JP	BROWNING	BTC-6HDX	2020:12:05 2	3840	2160	2020-12-06T
13	RB1	ARB1_2020_:	2020_12_29	IMG_0334.JP	BROWNING	BTC-6HDX	2020:12:05 2	3840	2160	2020-12-06T
14	RB1	ARB1_2020_:	2020_12_29	IMG_0335.JP	BROWNING	BTC-6HDX	2020:12:05 2	3840	2160	2020-12-06T
15	RB1	ARB1_2020_:	2020_12_29	IMG_0336.JP	BROWNING	BTC-6HDX	2020:12:06 0	3840	2160	2020-12-06T
16	RB1	ARB1_2020_:	2020_12_29	IMG_0369.JP	BROWNING	BTC-6HDX	2020:12:06 1	3840	2160	2020-12-06T

	K	L	M	N	O	P	Q	R	S	T	U
1	TimeDif	numAnimalD	jsonAnimalD	numHumanD	jsonOtherDe	maxDetectio	maxDetectio	PhotoName	commonNam	SpeciesDeta	Count
2	0	0	NA	0	NA	0	0	VID18494-00	Human	NA	1
3	43786	1	{'category': '1	'conf': 0.136	'bbox': [0.897	0.3091	0.1004	0.2045]}	0	NA	0.136
4	2420	1	{'category': '1	'conf': 0.997	'bbox': [0	0.09302	0.2123	0.2345]}	0	NA	0.997
5	53116	2	{'category': '1	'conf': 0.984	'bbox': [0.918	0.3439	0.07994	0.2117]}{'cat	'conf': 0.445	'bbox': [0.943	0.3811
6	4296	1	{'category': '1	'conf': 0.98	'bbox': [0.899	0.2828	0.09704	0.2376]}	0	NA	0.98
7	13375	1	{'category': '1	'conf': 0.989	'bbox': [0.906	0.2823	0.09038	0.2516]}	0	NA	0.989
8	157	1	{'category': '1	'conf': 0.994	'bbox': [0.923	0.3809	0.07439	0.1732]}	0	NA	0.994
9	3214	0	NA	0	NA	0	0	VID18494-00	Empty	NA	0
10	66636	1	{'category': '1	'conf': 0.98	'bbox': [0.930	0.4016	0.06732	0.1727]}	0	NA	0.98
11	6338	1	{'category': '1	'conf': 0.868	'bbox': [0.951	0.4428	0.04809	0.1173]}	0	NA	0.868
12	7143	1	{'category': '1	'conf': 0.995	'bbox': [0.916	0.3469	0.08215	0.2295]}	0	NA	0.995
13	3858	2	{'category': '1	'conf': 0.992	'bbox': [0.929	0.3639	0.06843	0.198]}{'cate	'conf': 0.554	'bbox': [0.660	0.3084
14	155	1	{'category': '1	'conf': 0.995	'bbox': [0.906	0.3284	0.09287	0.2445]}	0	NA	0.995
15	21280	3	{'category': '1	'conf': 0.999	'bbox': [0.010	0.2704	0.1795	0.2614]}{'cat	'conf': 0.932	'bbox': [0.665	0.3088
16	23250	2	{'category': '1	'conf': 0.411	'bbox': [0.917	0.2855	0.08214	0.2132]}{'cat	'conf': 0.15	'bbox': [0.919	0.3887

- 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
1	locAbbr	Timestamp	commonNam	Latitude	Longitude	maxDetectio	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

```r
# 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	SeasonNumb	Site	Latitude	Longitude	Day_1	Day_2	Day_3	Day_4	Day_5	Day_6	Day_7	Day_8	Day_9	Day_10
400	Coyote	6	OXY1	34.1269	-118.209	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
401	Coyote	6	OXY2	34.1265	-118.21	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
402	Coyote	6	RHR1	34.218	-118.159	0	0	0	0	0	0	0	0	0	0
403	Coyote	6	RHR2	34.2174	-118.16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
404	Coyote	6	ROS1	34.1557	-118.168	NA	0	0	0	0	0	0	0	0	0
405	Coyote	6	ROS2	34.1545	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
406	Coyote	6	ROS3	34.1548	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
407	Coyote	6	ROS4	34.1541	-118.168	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
408	Coyote	6	SAB1	34.1294	-118.164	NA	NA	NA	0	0	0	0	0	0	0
409	Coyote	6	SPS1	34.1243	-118.167	1	0	0	0	0	0	0	0	0	1
410	Coyote	6	SPS2	34.1242	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
411	Coyote	6	SPS3	34.1244	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
412	Coyote	6	SPS4	34.1246	-118.167	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
413	Coyote	6	SUN1	34.1771	-118.238	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
414	Coyote	6	SUN2	34.1761	-118.237	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
415	Coyote	6	TCR1	34.2523	-118.328	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
416	Coyote	6	TMC1	34.1757	-118.181	0	0	0	1	1	0	1	1	0	1
417	Coyote	6	TMC2	34.1756	-118.181	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
418	Coyote	6	WAV1	34.243	-118.311	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
419	Coyote	6	WAV2	34.2477	-118.306	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
420	Coyote	6	WAV3	34.2462	-118.294	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
421	Coyote	6	WAV4	34.2472	-118.305	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
422	Coyote	6	WHT1	34.2713	-118.351	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
423	Coyote	6	WTP1	34.2213	-118.319	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
424	Coyote	6	WTP2	34.2208	-118.316	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
425	Coyote	6	WWD1	34.2166	-118.258	0	0	0	0	0	0	0	0	0	0
426	Coyote	6	WWD2	34.2166	-118.258	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
427	Coyote	6	HOL1	34.0395	-118.216	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
428	Coyote	7	ARB1	34.2057	-118.161	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
429	Coyote	7	AST1	34.1382	-118.167	NA	NA	0	1	0	0	0	1	0	0

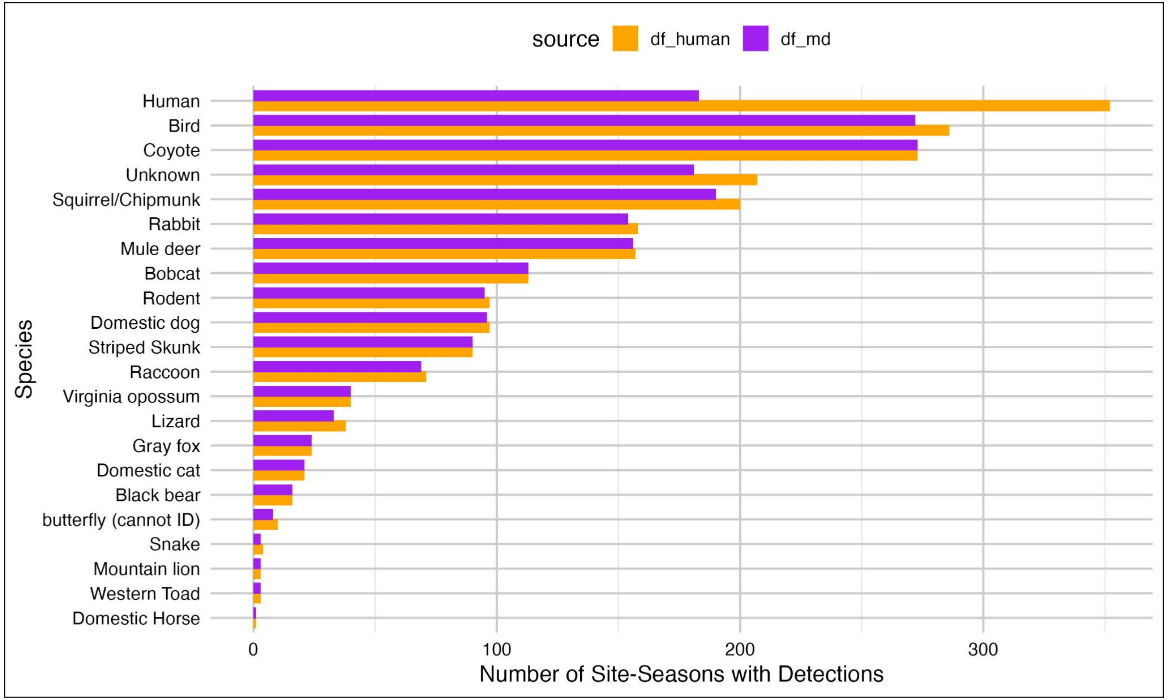
…

BJ	BK	BL	BM	BN
Day_57	Day_58	Day_59	Day_60	Treatment
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
0	0	0	0	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
1	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
0	0	0	0	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
NA	NA	NA	NA	df_human
0	0	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 ≥ 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)

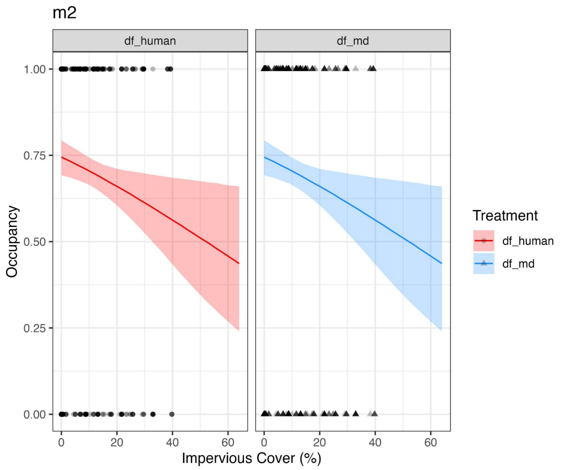


Results

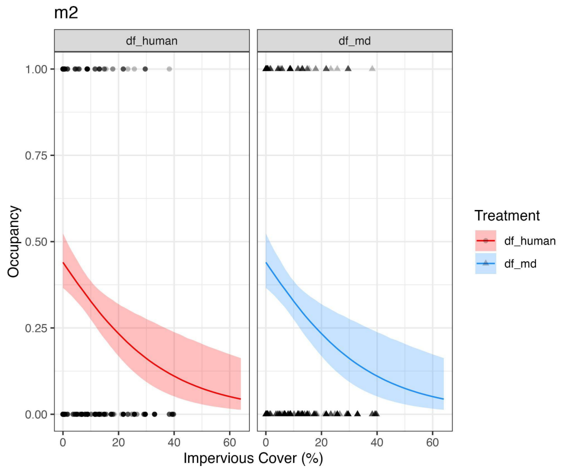
5 models:

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

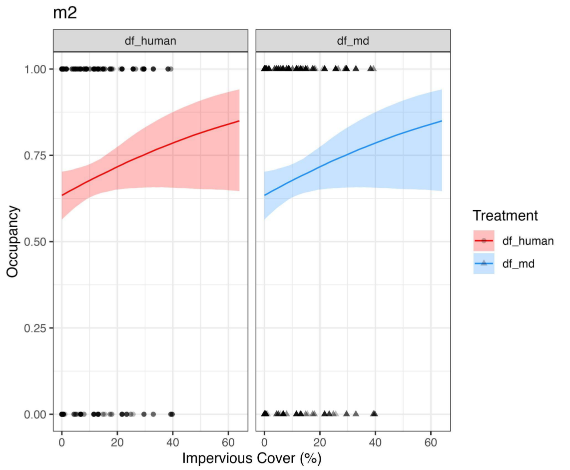
Bird



Bobcat



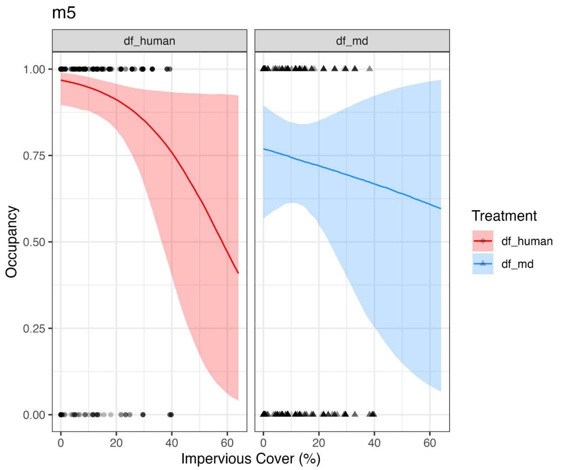
Coyote



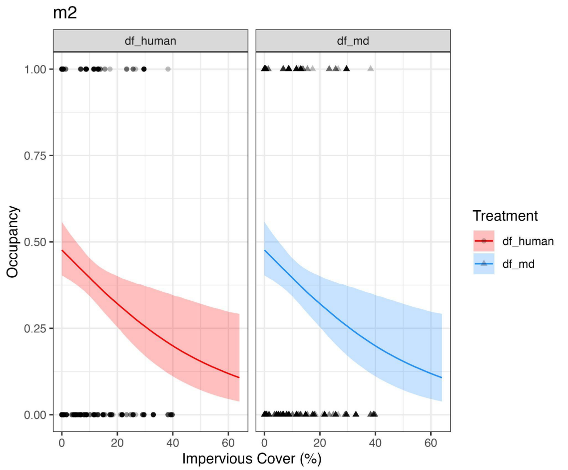
Interesting graphs of best models for some species

Results

Human



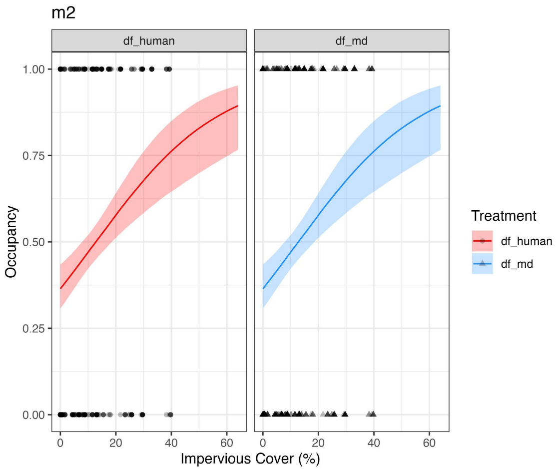
Mule Deer



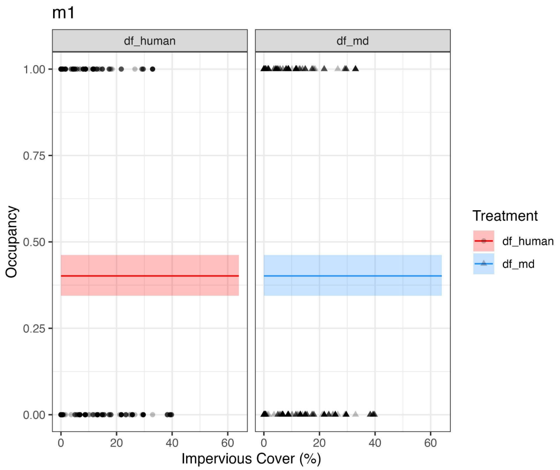
Results

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?