Estimating Wildlife Strike Costs at US Airports: A Machine Learning Approach

Levi Altringer[a,b]*
Jordan Navin[a]
Michael J. Begier[c]
Stephanie A. Shwiff[a]
Aaron Anderson[a]

[a]*US Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, CO, 80521, USA*

[b]*Department of Economics, Colorado State University, Fort Collins, CO, 80523, USA*

[c]*National Coordinator, Airport Wildlife Hazards Program, U.S. Department of Agriculture, Wildlife Services, WA, DC, 20250, USA*

*Corresponding author. Email: Levi.Altringer@usda.gov.

Introduction

- Costs reported to the National Wildlife Strike Database are largely missing.

- Previous imputation methods produce biased wildlife strike cost estimates.

- Machine learning (ML) techniques provide a more robust approach to cost imputation.

- ML estimates suggest that previous cost estimates may serve as a reasonable upper, as opposed to lower, bound.

Transportation Research Part D: Transport and Environment
Volume 97, August 2021, 102907

Estimating wildlife strike costs at US airports: A machine learning approach ☆

Levi Altringer [a,b] ✉, Jordan Navin [a], Michael J. Begier [c], Stephanie A. Shwiff [a], Aaron Anderson [a]

Show more ⌄

+ Add to Mendeley   ⭒ Share   🔖 Cite

https://doi.org/10.1016/j.trd.2021.102907   Get rights and content

https://doi.org/10.1016/j.trd.2021.102907

Motivation

▸ Q: How many aircraft have been destroyed due to wildlife strikes?

▸ Q: What part of the airplane is most frequently damaged?

▾ Q: What is the economic loss from wildlife strikes?

A: The reported costs for civil aircraft in USA totaled $900 million for the 30-year period, 1990 to 2019. When costs are adjusted reported strikes in which costs were not provided, and the estimated number of strikes that were not reported, losses could be as high as $500 million per year.

▸ Q: Who reports wildlife strikes?

▸ Q: How do I collect bird remains if I have a strike?

▸ Q: When was the first bird strike reported?

https://www.faa.gov/airports/airport_safety/wildlife/faq/

Motivation

Smithsonian Feather Lab identifies Cerulean Warbler struck by aircraft on April 28, 2020 as the 600th species of bird in the National Wildlife Strike Database

https://www.faa.gov/airports/airport_safety/wildlife/media/Wildlife-Strike-Report-1990-2020.pdf

Motivation

Table 23. Number of reported wildlife strikes indicating damage, a negative effect-on-flight (NEOF), aircraft downtime, repair costs, and other costs; and the mean losses per report in hours of downtime and inflation-adjusted U.S. dollars, civil aircraft, USA[1], 1990–2020.

Year	Number of reports indicating:					Mean losses per report[2]		
	Damage	NEOF	Aircraft down time	Repair costs	Other costs	Down-time (hours)	Repair costs ($)	Other costs ($)
1990	372	148	60	33	16	56.4	227,972	65,442
1991	400	184	61	49	25	79.8	77,311	41,674
1992	365	218	81	51	28	111.9	110,609	5,566
1993	399	240	67	57	19	277.9	93,957	9,917
1994	460	272	103	73	29	388.4	80,890	96,595
1995	497	307	95	62	33	96.3	532,204	232,338
1996	502	355	144	86	39	137.3	89,238	26,704
1997	578	379	182	126	47	230.7	80,441	42,107
1998	585	400	205	135	54	119.5	211,916	30,215
1999	706	446	282	179	79	148.8	116,628	22,093
2000	764	477	351	205	93	195.2	103,708	120,762
2001	645	434	293	157	65	142.6	299,883	41,133
2002	670	497	383	165	63	135.6	160,214	67,694
2003	632	437	355	172	81	111.8	170,293	45,095
2004	629	429	324	213	92	166.9	110,974	23,956
2005	610	454	327	227	125	88.0	281,949	81,224
2006	599	429	333	172	102	116.9	226,136	14,063
2007	568	453	364	178	135	165.2	182,011	35,040
2008	526	408	371	156	141	116.2	124,463	14,784
2009	605	520	563	195	193	80.8	394,863	15,467
2010	595	467	526	174	164	66.5	137,253	14,583
2011	542	498	526	179	208	70.8	247,649	15,955
2012	613	540	689	228	263	75.4	116,223	8,915
2013	609	521	802	238	304	75.7	69,006	13,165
2014	587	573	717	210	277	63.2	147,241	11,254
2015	622	545	704	207	289	48.0	158,595	20,382
2016	589	522	586	155	221	87.2	70,358	13,288
2017	667	568	633	192	262	50.3	185,925	13,154
2018	708	619	636	170	295	66.1	61,764	8,301
2019	739	618	702	167	243	67.8	105,486	21,502
2020	462	420	450	106	182	70.7	107,307	8,127
Total	17,845	13,378	11,915	4,717	4,167			
Mean	576	432	384	152	134	96.9	163,005	24,313

[1] Includes strikes to U.S.-registered aircraft in foreign countries.

[2] See Table 18 for actual losses reported in total and by species of wildlife, 1990-2020.

Table 24. Projected annual losses in aircraft downtime (hours) and in repair and other costs (inflation-adjusted U.S. dollars) from wildlife strikes with civil aircraft, USA[1], 1990–2020. Losses are projected from mean reported losses per incident (Table 23). (Page 1 of 2).

Year	No. of adverse incidents[4]	Projected losses[2, 3]			
		Down-time (hours)	Repair costs (x $1 million)	Other costs (x $1 million)	Total costs (x $1 million)
1990	427	24,061	97	28	125
1991	484	38,601	37	20	58
1992	493	55,179	55	3	57
1993	509	141,456	48	5	53
1994	582	226,070	47	56	103
1995	656	63,149	349	152	502
1996	684	93,893	61	18	79
1997	783	180,606	63	33	96
1998	807	96,438	171	24	195
1999	981	145,946	114	22	136
2000	1,114	217,436	116	135	250
2001	977	139,314	293	40	333
2002	1,101	149,299	176	75	251
2003	997	111,490	170	45	215
2004	952	158,849	106	23	128
2005	979	86,161	276	80	356
2006	942	110,027	213	13	226
2007	978	161,606	178	34	212
2008	906	105,243	113	13	126
2009	1,186	95,857	468	18	487
2010	1,126	74,875	155	16	171
2011	1,146	81,107	284	18	302
2012	1,334	100,642	155	12	167
2013	1,447	109,550	100	19	119
2014	1,460	92,332	215	16	231
2015	1,455	69,877	231	30	260
2016	1,339	116,813	94	18	112
2017	1,446	72,800	269	19	288
2018	1,582	104,611	98	13	111
2019	1,650	111,901	174	35	210
2020	1,073	75,861	115	9	124
Total	31,596	3,411,050	5,040	1,043	6,084
Mean	1,019	110,034	163	33	196

The 1990-2018 NWSD report writes that "actual costs are likely 2 or more times higher" due to the issue of underreporting.

Motivation

2009 EXAMPLE:

### With the "Miracle on the Hudson" incident:

Cost type	Number of Reported Values	Mean for Reported Values (2021 $)	Number of Adverse Incidents	Estimate after Imputation (Millions, 2021 $)
Repair Costs	195	$398,467	1,186	$472.6
Other Costs	193	$15,452		$18.3

Without the "Miracle on the Hudson" incident:

Cost type	Number of Reported Values	Mean for Reported Values (2021 $)	Number of Adverse Incidents	Estimate after Imputation (Millions, 2021 $)
Repair Costs	194	$174,016	1,186	$206.4
Other Costs	193	$15,452		$18.3

Total cost estimate for 2009 is reduce by over half upon the removal of one extremely rare, high cost incident.
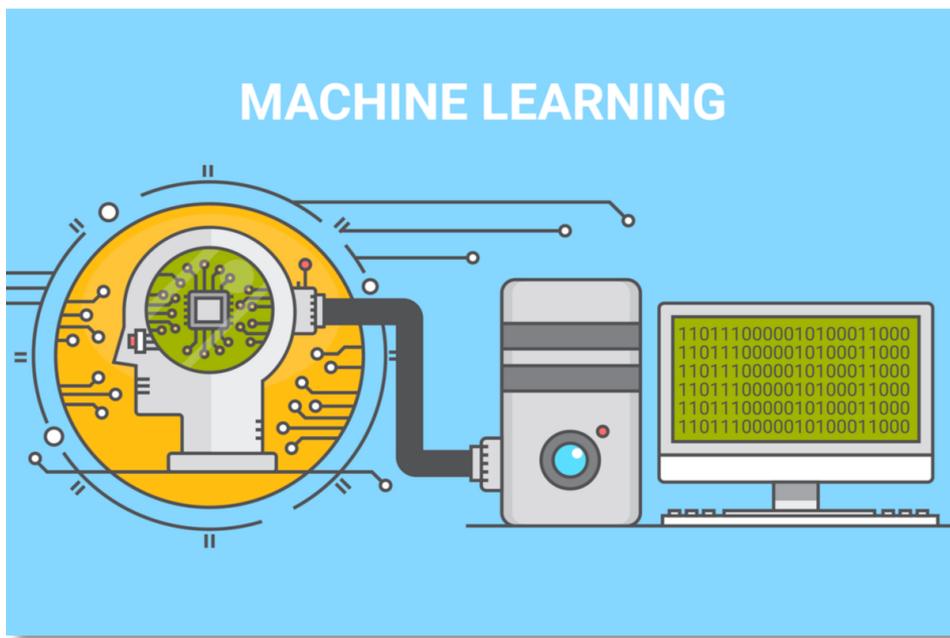
Motivation
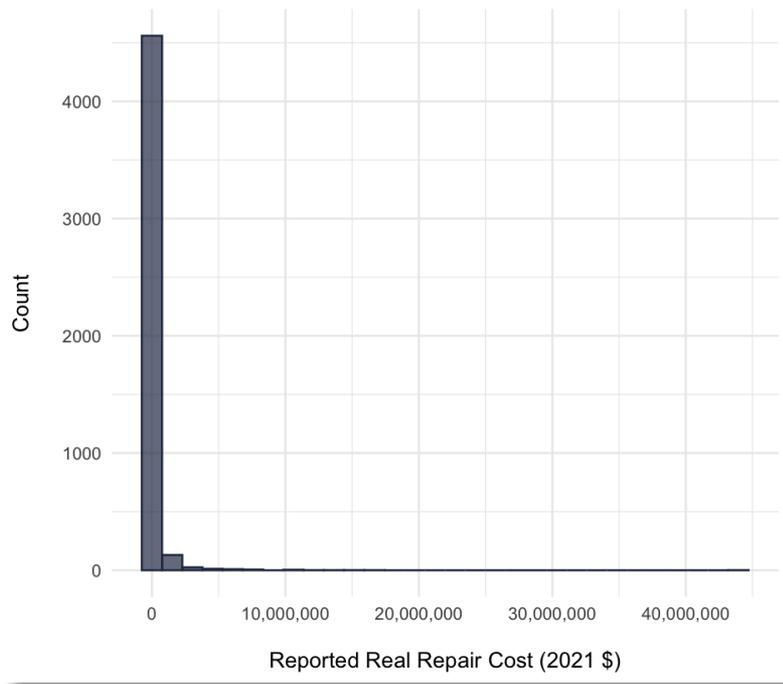

Linear Scale
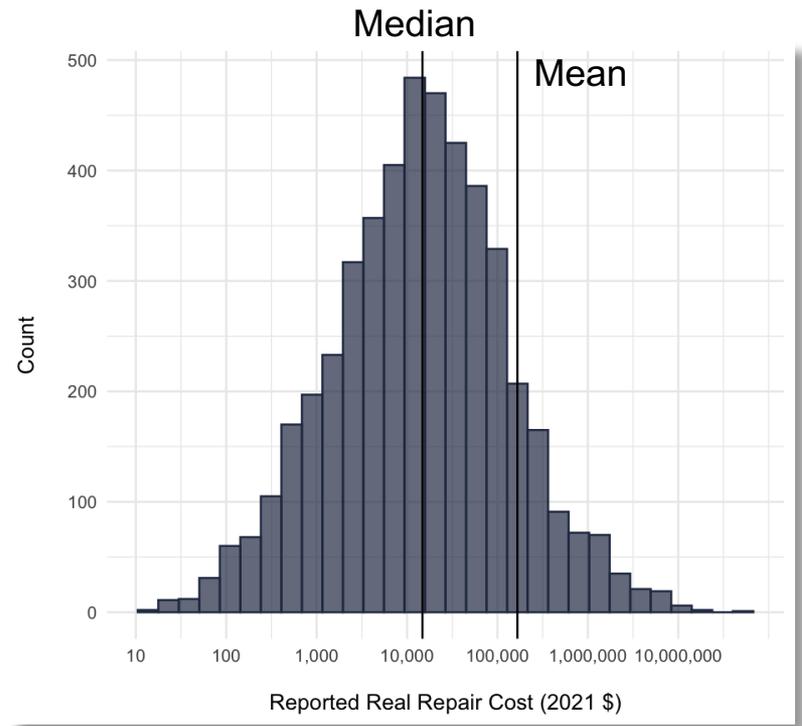
Log Scale

Research Question

Is there a better way to impute missing cost values in the NWSD and achieve a more robust lower bound cost estimate?
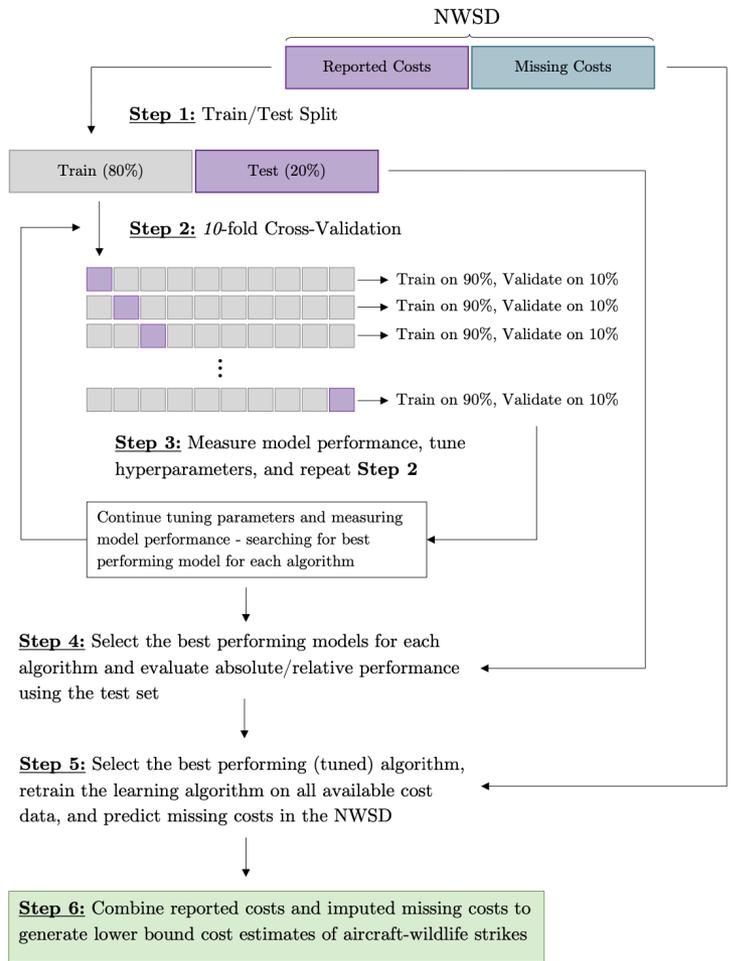
Machine Learning Algorithms

- Machine learning algorithms employed:
  - Random Forest (RF)
  - Artificial Neural Network (NN)

- Tested against the performance of simple Linear Regression (LR).

- Could spend entire presentation explaining the RF and NN machine learning algorithms (please refer to manuscript).

- Will briefly describe the workflow and intuition behind machine learning methods.
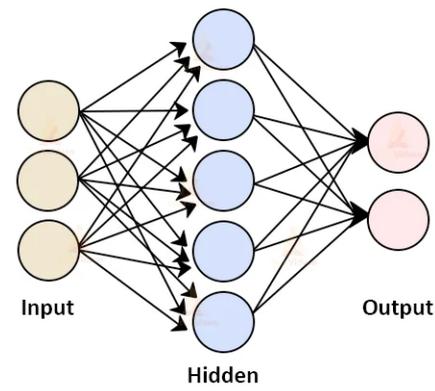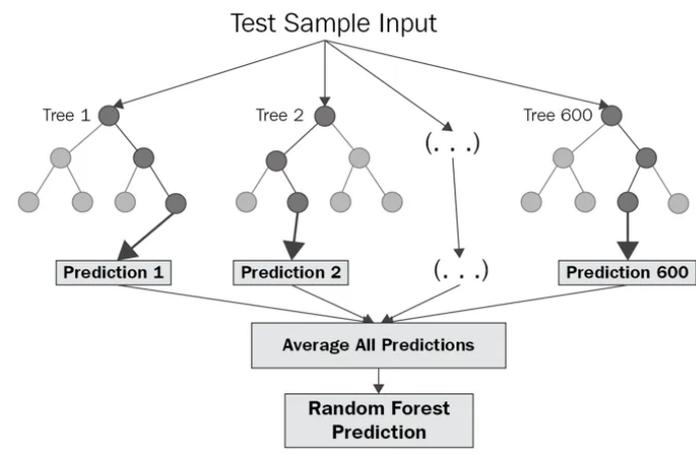
Machine Learning Methods (Very briefly…)

- Machine learning algorithms seek to predict a target variable from its observed features.

- Models are tuned and trained on a subset of available data—called the training set.

- Models are evaluated on the rest of the data that is withheld from the beginning—called the test set.

- Best performing model is selected and employed to predict new/missing observations.

- Machine learning methods have been shown to outperform more traditional statistical methods.

Data (Target Variable)

- National Wildlife Strike Database (paper based on last access on April 15th, 2020).

- Target variables are natural log of repair and other costs, separately.

- Data split into non-missing and missing cost information.

- Non-missing cost data is split into training (80%) and test (20%) sets.

	Mean	Std Dev	Min	Median	Max	N
			Repair Costs			
			(19,838 missing values)*			
Actual:						
All	$152,646	$926,856	$1.02	$13,670	$42,117,878	5,129
Training	$156,900	$968,301	$1.02	$13,557	$42,117,878	4,103
Test	$135,632	$738,383	$1.41	$13,936	$16,997,757	1,026
Log:						
All	9.40	2.34	0.02	9.52	17.56	5,129
Training	9.41	2.35	0.02	9.51	17.56	4,103
Test	9.38	2.31	0.35	9.54	16.65	1,026
			Other Costs			
			(231,445 missing values)			
Actual:						
All	$16,225	$149,036	$0.01**	$234	$6,419,450	5,860
Training	$15,770	$148,957	$0.01	$235	$6,419,450	4,688
Test	$18,049	$149,403	$0.01	$230	$4,370,334	1,172
Log:						
All	3.37	5.84	-4.61	5.46	15.67	5,860
Train	3.36	5.86	-4.61	5.46	15.67	4,688
Test	3.43	5.78	-4.61	5.44	15.29	1,172

NOTE: *The assumption is made that all non-damaging strikes reported to the NWSD ($N = 212,338$, 89.5%) have a repair cost of $0. Thus, the repair cost imputation exercise only pertains to damaging strikes ($N = 24,967$, 10.5%). This distinction is not made for the other cost imputation exercise, which is why the number of missing other costs values to predict is substantially larger. **Further, our log transformation of the target variable requires that we replace other cost values of $0 with $0.01.

Data (Explanatory Features)

Variable	Variable Type	Description
Aircraft class	Categorical	Airplane; Helicopter; Glider; Ultralight; Missing
Engine type	Categorical	Reciprocating engine (piston); Turbofan; Turbojet; Turboprop; Turboshaft (helicopter); Missing
Aircraft mass	Categorical	2,250 kg or less; 2,251-5,700 kg; 5,701-27,000 kg; 27,001-272,000 kg; above 272,000 kg; Missing
Pilot warned	Categorical	No; Yes; Missing
Phase of flight	Categorical	Approach; Arrival; Climb; Departure; Descent; En route; Landing roll; Local; Parked; Taxi; Take-off run; Missing
Number seen	Categorical	1; 2-10; 11 or more; Missing
Number struck	Categorical	1; 2-10; 11 or more; Missing
Animal size	Categorical	Small; Medium; Large; Missing
Component struck	Dummies	A dummy variable for each of the following components of the aircraft: radome, windshield, nose, engine 1, engine 2, engine 3, engine 4, propeller, wing or rotor, fuselage, landing gear, tail, lights, other
Effect on flight	Categorical	Aborted takeoff; Engine Shutdown; Precautionary landing; Other; None; Missing
Damage*	Categorical	No; Yes
Damage type	Categorical	None; Minor; Substantial; Destroyed; Uncertain; Missing
Component damaged	Dummies	A dummy variable for each of the following components of the aircraft: radome, windshield, nose, engine 1, engine 2, engine 3, engine 4, propeller, wing or rotor, fuselage, landing gear, tail, lights, other
Ingested	Categorical	No; Yes; Missing
Cloud cover	Categorical	No clouds; Some clouds; Overcast; Missing
Time of day	Categorical	Dawn; Day; Dusk; Night; Missing

NOTE: *The damage variable is not used as an explanatory feature in models of repair cost for the reasons outlined in section 3.

- Explanatory features:
  1. Must be general enough so as not to uniquely identify individual observations.
  2. Must be detailed enough to provide explanatory power.

- Features were selected in an effort to encompass as many factors that may influence the cost of a wildlife strike as possible.

Results

- As an additional test, we trained and evaluated models without destroyed aircraft observations to determine the sensitivity of our models to extremely rare, high costs incidents.

- Each model (RF, NN, and LR) is evaluated on 100 random samples of the test set—65% without replacement—and average performance is obtained.

- We found that RF was, overall, the best performing model.

	Mean Square Error	Mean Absolute Error	R-Squared
	Mean (SD)	Mean (SD)	Mean (SD)
	Panel A: Repair Costs		
	Destroyed aircraft included		
LR	2.812 (0.098)	1.290 (0.023)	0.471 (0.015)
NN	2.732 (0.094)	1.270 (0.023)	0.486 (0.016)
RF	2.637 (0.088)	1.244 (0.022)	0.504 (0.015)
	Destroyed aircraft excluded		
LR	2.586 (0.090)	1.253 (0.021)	0.513 (0.013)
NN	2.483 (0.087)	1.227 (0.022)	0.532 (0.013)
RF	2.508 (0.090)	1.226 (0.023)	0.528 (0.013)
	Panel B: Other Costs		
	Destroyed aircraft included		
LR	2.104 (0.168)	0.929 (0.022)	0.937 (0.005)
NN	1.871 (0.110)	0.868 (0.021)	0.944 (0.004)
RF	1.822 (0.109)	0.838 (0.021)	0.945 (0.003)
	Destroyed aircraft excluded		
LR	1.724 (0.074)	0.884 (0.020)	0.949 (0.002)
NN	1.714 (0.070)	0.860 (0.019)	0.950 (0.002)
RF	1.626 (0.066)	0.806 (0.019)	0.952 (0.002)

NOTE: The means and standard deviations displayed in the table above are generated by testing each tuned algorithm—linear regression (LR), neural network (NN), and random forest (RF)—on 100 random samples of the test set. Figures 5 and 6 plot the complete distributions of these measures. Each of the 100 random samples was 65% of the respective test set—[1,026 ×0.65] = 667 randomly selected observations per sample for repair costs and [1,172 ×0.65] = 762 for other costs.

Results
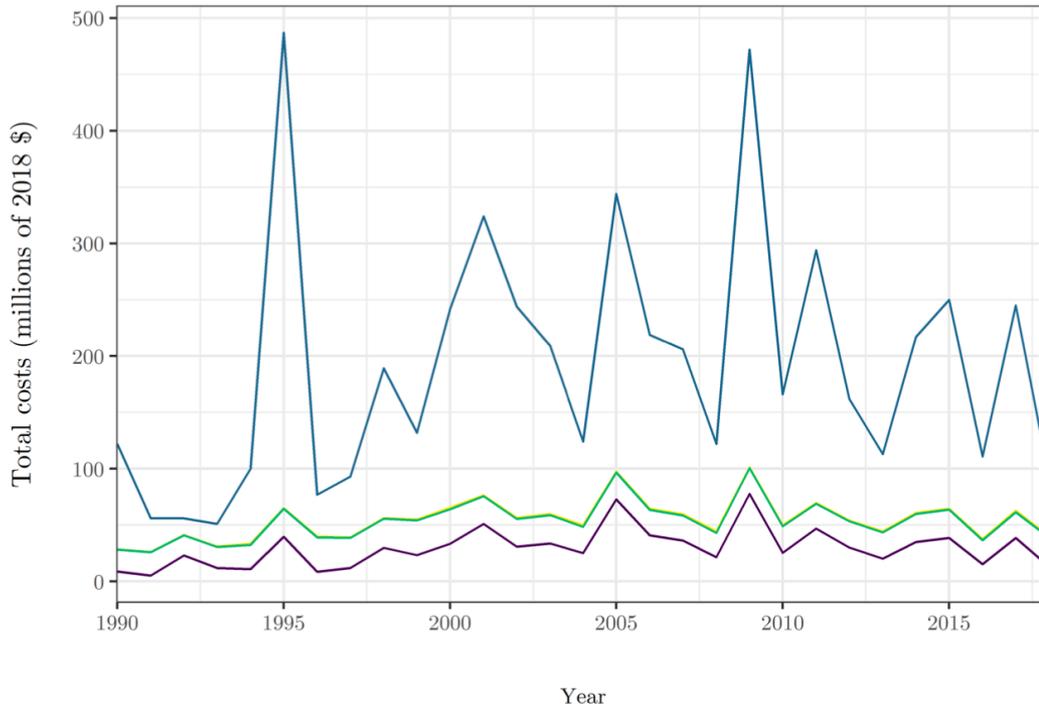
- Employ full trained RF model to impute missing cost values, but retain reported values.

- ML imputation increases average yearly reported costs by 80.1%—from $29.7 million to $54.3 million, specifically—for the 1990-2018 period.

- This stands in stark contrast to the 529.6% increase—from $29.7 million to $187 million—produced by previous imputation methods.

- Our ML estimates are effectively invariant to the inclusion/exclusion of extremely costly and rare destroyed aircraft incidents.

- Even if one assumes that wildlife strikes are underreported by a factor of 3, ML estimates still fall below previous lower bound estimates.

Discussion and Conclusion

- Main findings:

    1. Previous imputation methods and lower bound cost estimates are biased due to the skewed nature of reported cost data.

    2. ML imputation methods provide a dramatic improvement.

    3. ML estimates suggest that previous lower bound cost estimates may be upper, not lower, bound of economic burden of wildlife strikes to US civil aviation industry.

    4. ML methods could potentially be incorporated in future reports to more robustly estimate the costs associated with wildlife strikes (civil).

- Limitations:

    1. We are unable to completely measure all "indirect" or "external" costs associated with wildlife strike events. (Currently working on a couple of papers that address this issue.)

    2. Reported costs, and imputed costs, do not account for costs associated with the loss of human health and safety (mortality, morbidity, risk, liability, etc.).

    3. Wildlife management at airports is an important tool in reducing wildlife strike risk.

Thank You!