Article title: Increasing Arctic tundra flooding threatens wildlife habitat and survival: impacts on the critically endangered Siberian crane

Journal name: Frontiers in Conservation Science

Author names: Paul J. Haverkamp, Inga Bysykatova-Harmey, Nikolai Germogenov, and Gabriela Schaepman-Strub

Corresponding author: Paul J. Haverkamp, pjhav17@gmail.com, University of Zurich

Supporting Information A. Landcover type classification methodology and results

Landsat 8 has coverage over the study area from 2013 through the present. Image selection was restricted to June for each year, as breeding occurs at this time. Images were then filtered for clouds and shadows using the quality assurance band, and a composite image was produced by taking the median value of each band over 2013-2019, for each pixel. To classify the imagery, a random forest decision tree classifier with 1000 trees was used in GEE. The random forest classifier is a nonparametric machine learning algorithm implementing many decision trees with multiple nodes. Each pixel is evaluated in all trees and classified, with final class decided by a majority of trees (Breiman 2001). Random forest classifiers have been used successfully in GEE in multiple studies, for example to classify wetlands (Amani et al. 2019) and land use (Midekisa et al. 2017). Training and testing data were derived from visual inspection of the image to train and validate the random forest classifier. We chose 50 points for training and 50 points for testing of each landcover type around the entire study area, for a total of 250 training and 250 testing pixels. After running the random forest classifier on the Landsat 8 visible and infrared bands, the accuracy of the resulting habitat map was validated by extracting the class of the test pixels to calculate a confusion matrix and the overall accuracy and kappa coefficient determined in GEE. The validation data found an overall accuracy of 0.88 and a kappa coefficient of 0.85.

Supporting Information Methodology B.

B.1. MaxEnt methodology

Before implementing the MaxEnt models, correlations between environmental variables were determined using the Correlations and Summary Stats tool in ArcGIS from the SDMToolbox (Brown 2014; Brown, Bennett, and French 2017), and no layers were found to be correlated. To prevent predicting areas in our model where Siberian cranes do not nest (Radosavljevic and Anderson 2014; R. P. Anderson and Raza 2010; Barve et al. 2011), we created a mask of the river floodplain areas using the maximum extent of rivers, and masked floodplains and areas along the sea coast for the modelling. Both areas can flood during the nesting season and only few nests have ever been identified here.

B.2. Snow depth methodology

We determined if there was a significant relationship between maximum snow depth and year from 2000-2018 using the ‘lm’ function in the stats package of R (R Core Team 2021) and plotted in ggplot2 (Wickham 2016). We used the ‘grubbs.test’ function in the outliers package (Komsta 2011) to check for outliers. To examine the relationship between yearly maximum snow depth and flooding, we calculated the Pearson correlation between the snow depth and percent of nest sites in flooded areas in the same year using the ‘cor.test’ function in the stats package of R (R Core Team 2021). To verify normality of variables, we examined the residuals of a linear model with percent of nest sites in flooded areas as the dependent variable and the percent of flooded nest sites as the independent variable, using the ‘qqnorm’ function in the stats package of R (R Core Team 2021), and plotted the relationship using ggplot with a 'lm' smoothing line from the ggplot2 package (Wickham 2016). Snow depth data downloaded from https://climexp.knmi.nl/selectdailyseries.cgi?id=someone@somewhere using the Chokurdah station, which is closest to the Indigirka breeding core.

Supporting Information C. Data tables and figures

Table C.1. MaxEnt model output for all models run with the habitat classification, distance to lakes, maximum slope within 150 m, and mean June land surface temperature (LST) environmental parameters. Models are sorted by lowest overall AICc per model group. Possible features are linear (L) and linear-quadratic (LQ), and regularization multiplier (RM) ranged from 0.75 - 1.25. Final selected model values are in bold.

Variables included	Features	RM	AICc	ΔAICc	Average AUC	Average AUCdiff	Number of model parameters
Habitat classification, Distance to lakes, Maximum slope within 150m	L	0.75	273.47	0.00	0.97	0.03	5
	LQ	0.75	294.61	21.14	0.96	0.03	5
	L	**1**	**274.30**	**0.83**	**0.97**	**0.03**	**5**
	LQ	1	296.88	23.42	0.96	0.04	5
	L	1.25	275.16	1.70	0.97	0.03	5
	LQ	1.25	298.61	25.14	0.96	0.04	5
Habitat classification, Distance to lakes	L	0.75	325.40	0.00	0.88	0.09	2
	LQ	0.75	345.73	20.33	0.85	0.01	2
	L	1	325.53	0.13	0.89	0.09	2
	LQ	1	348.77	23.37	0.84	0.01	2
	L	1.25	325.69	0.29	0.89	0.09	2
	LQ	1.25	351.00	25.60	0.83	0.01	2
Distance to lakes, Maximum slope within 150m	L	0.75	293.54	0.00	0.94	0.05	2
	LQ	0.75	338.56	45.02	0.94	0.06	2
	L	1	294.40	0.86	0.94	0.05	2
	LQ	1	342.75	49.22	0.94	0.06	2
	L	1.25	295.34	1.80	0.94	0.05	2
	LQ	1.25	345.91	52.37	0.94	0.06	2
Habitat classification, Maximum slope within 150m	L	0.75	335.53	0.00	0.86	0.07	2
	LQ	0.75	354.77	19.24	0.80	0.01	2
	L	1	336.11	0.57	0.86	0.07	2
	LQ	1	356.33	20.79	0.80	0.01	2
	L	1.25	336.72	1.18	0.86	0.07	2
	LQ	1.25	357.58	22.04	0.79	0.01	2

Table C.2. Yearly number and percent of known nesting sites located within June surface water based on the Global Surface Water (GSW) layer. Some nests were located in GSW areas with no observation, and thus percent of total nesting sites may change each year.

Year	Number of observed nests found within June surface water	Total number of observed nests with June GSW observation	% of observed nests within June surface water	Number of juveniles counted in Autumn at Momoge NR
2001	10	31	32.26	
2002	5	27	18.52	
2003	NA	NA	NA	
2004	4	29	13.79	
2005	11	37	29.73	
2006	4	25	16.00	
2007	7	30	23.33	245
2008	6	24	25.00	261
2009	7	28	25.00	301
2010	9	25	36.00	459
2011	13	37	35.14	359
2012	2	30	6.67	579
2013	7	37	18.92	
2014	12	26	46.15	
2015	12	37	32.43	
2016	17	37	45.95	
2017	17	31	54.84	29
2018	15	28	53.57	69

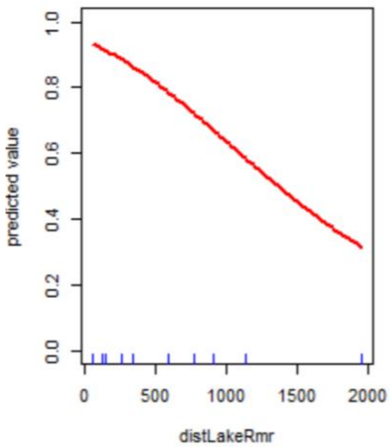

Figure C.1. Response curves for the three environmental variables within the final model, where ‘distLakeRmr’ is the distance to lakes, ‘LSclassptssmo_v2Rmr2’ is landcover type (where 1 = ice, 2 = clear water, 3 = turbid water, 4 = wetlands, 5 = other vegetation), and ‘maxslope5Rmr’ is the maximum slope within 150 m.

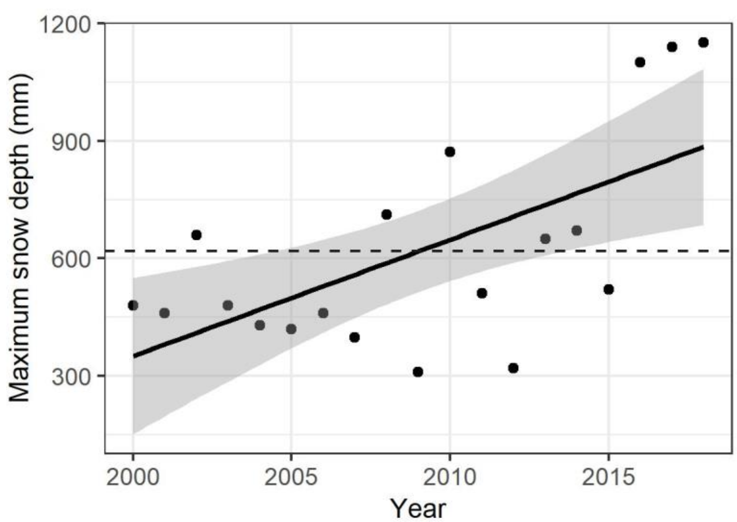

Figure C.2. Yearly maximum snow depth in the Indigirka breeding core area between September and May for the year before nesting season. Solid line represents linear fit of points ($R^2 = 0.39$), with gray areas representing 95% confidence intervals. Dashed line represents mean maximum snow depth over the years (618 mm ± 267). Maximum snow depth is significantly increasing over the study period (F-statistic = 10.9, p-value < 0.005).

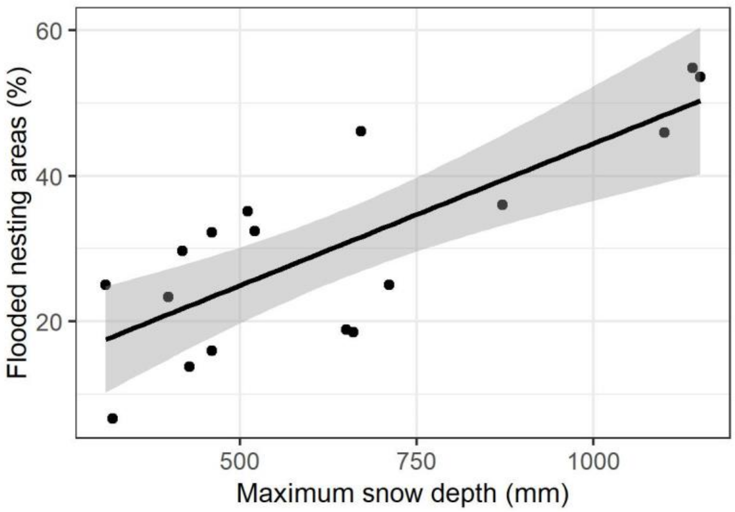

Figure C.3. Linear regression of maximum snow depth between the previous September and May and the percent of known nest sites in flooded areas in the Indigirka breeding core area in the same year. Solid line represents linear fit of points (R^2 = 0.61), with gray areas representing 95% confidence intervals. The percent of nests in flooded areas is significantly correlated with maximum snow depth (R = 0.78, t-value = 4.8, p-value < 0.0001), suggesting the amount of snowfall influences the surface water extent.

Supporting Information references

Amani, Meisam, Brian Brisco, Majid Afshar, S. Mohammad Mirmazloumi, Sahel Mahdavi, Sayyed Mohammad Javad Mirzadeh, Weimin Huang, and Jean Granger. 2019. "A Generalized Supervised Classification Scheme to Produce Provincial Wetland Inventory Maps: An Application of Google Earth Engine for Big Geo Data Processing." *Big Earth Data* 3 (4): 378–94. doi:10.1080/20964471.2019.1690404.

Anderson, Robert P., and Ali Raza. 2010. "The Effect of the Extent of the Study Region on GIS Models of Species Geographic Distributions and Estimates of Niche Evolution: Preliminary Tests with Montane Rodents (Genus Nephelomys) in Venezuela." *Journal of Biogeography* 37 (7). John Wiley & Sons, Ltd: 1378–93. doi:10.1111/J.1365-2699.2010.02290.X.

Barve, Narayani, Vijay Barve, Alberto Jiménez-Valverde, Andrés Lira-Noriega, Sean P. Maher, A. Townsend Peterson, Jorge Soberón, and Fabricio Villalobos. 2011. "The Crucial Role of the Accessible Area in Ecological Niche Modeling and Species Distribution Modeling." *Ecological Modelling* 222 (11). Elsevier: 1810–19. doi:10.1016/j.ecolmodel.2011.02.011.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1). Springer: 5–32. doi:10.1023/A:1010933404324.

Brown, Jason L. 2014. "SDMtoolbox: A Python-Based GIS Toolkit for Landscape Genetic, Biogeographic and Species Distribution Model Analyses." Edited by Barbara Anderson. *Methods in Ecology and Evolution* 5 (7). British Ecological Society: 694–700. doi:10.1111/2041-210X.12200.

Brown, Jason L., Joseph R. Bennett, and Connor M. French. 2017. "SDMtoolbox 2.0: The next Generation Python-Based GIS Toolkit for Landscape Genetic, Biogeographic and Species Distribution Model Analyses." *PeerJ* 2017 (12). PeerJ Inc.: e4095. doi:10.7717/peerj.4095.

Komsta, Lukasz. 2011. “Outliers: Tests for Outliers. R Package Version 0.14.” https://cran.r-project.org/package=outliers.

Midekisa, Alemayehu, Felix Holl, David J. Savory, Ricardo Andrade-Pacheco, Peter W. Gething, Adam Bennett, and Hugh J.W. W Sturrock. 2017. “Mapping Land Cover Change over Continental Africa Using Landsat and Google Earth Engine Cloud Computing.” *PLoS ONE* 12 (9). Public Library of Science. doi:10.1371/journal.pone.0184926.

R Core Team. 2021. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Radosavljevic, Aleksandar, and Robert P. Anderson. 2014. “Making Better MAXENT Models of Species Distributions: Complexity, Overfitting and Evaluation.” Edited by Miguel Araújo. *Journal of Biogeography* 41 (4). Blackwell Publishing Ltd: 629–43. doi:10.1111/jbi.12227.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.