

**Table S1.** Descriptions and sources for variables included in models.

Source	Description	Pixel Size	Pixel count	Source
FGLAN DSAT	Temperate or sub-polar needleleaf forest	30m	2807647	LANDSAT 2017 satellite imagery classification (North American Environmental Atlas - Commission on Environmental Cooperation - <a href="http://www.cec.org/files/atlas/?z=8&amp;x=-120.2179&amp;y=49.8114&amp;lang=en&amp;layers=polbounds%2Clandcover2015ls%2Clandcover2010ls&amp;opacities=100%2C100%2C100&amp;labels=false">\"&gt;http://www.cec.org/files/atlas/?z=8&amp;x=-120.2179&amp;y=49.8114&amp;lang=en&amp;layers=polbounds%2Clandcover2015ls%2Clandcover2010ls&amp;opacities=100%2C100%2C100&amp;labels=false</a> )\
	Sub-polar taiga needleleaf forest		5438	Landsat 2017
	Temperate or sub-polar broadleaf deciduous forest		33753	Landsat 2017
	Mixed forest		66341	Landsat 2017
	Temperate or sub-polar shrubland		426084	Landsat 2017
	Temperate or sub-polar grassland (Combined with layer below)		2766154	Landsat 2017
	Sub-polar or polar shrubland-lichen-moss		463	Landsat 2017
	Sub-polar or polar grassland-lichen-moss		10997	Landsat 2017
	Wetland		1345	Landsat 2017
	Cropland		416361	Landsat 2017
	Barren lands		971984	Landsat 2017
	Urban		669632	Landsat 2017

	Water		614105	Landsat 2017
	Snow and Ice		38	Landsat 2017
	Grassland layer			Grasslands Conservation Council of British Columbia; <a href="https://a100.gov.bc.ca/pub/acat/public/viewReport.do?reportId=54174">https://a100.gov.bc.ca/pub/acat/public/viewReport.do?reportId=54174</a>
Agricultural land	Cropland (combined cultivated land C910, Fallow Land C920, Crop transition C930)			BC Department of Agriculture
	Tree fruits (C200)			BC Department of Agriculture
	Vines (C310) and berries (C320)			BC Department of Agriculture
	Recreation - Golf courses, ski hills, M320		4138	BC Department of Agriculture
Topography	Ruggedness	30m		SAGA GIS
	Elevation (m)	30m		Provincial DEM SRTM <i>Ministry of Forests, Lands, Natural Resources and Rural Development. Kamloops, BC 2019. Digital Elevation Model. Available: Geospatial Service British Columbia: <a href="https://www2.gov.bc.ca/gov/content/data/geographic-data-services/topographic-data/elevation/digital-elevation-model">https://www2.gov.bc.ca/gov/content/data/geographic-data-services/topographic-data/elevation/digital-elevation-model</a> .</i>
Climate	Mean Temperature (April to September)	1km		Chelsa Bioclim available at: <a href="http://chelsa-climate.org/bioclim/">http://chelsa-climate.org/bioclim/</a>
	Mean precipitation (April to September)	1km		Chelsa BioClim available at: <a href="http://chelsa-climate.org/bioclim/">http://chelsa-climate.org/bioclim/</a>
	Solar Insolation	30m		SAGA GIS
Road		30 m		GeoBC. Digital Road Atlas available at:

layer				<a href="https://www2.gov.bc.ca/gov/content/data/geographic-data-services/topographic-data/roads">https://www2.gov.bc.ca/gov/content/data/geographic-data-services/topographic-data/roads</a>

## **Supplement S1 Model selection**

These models operate on the principle that species' observations are not available for all sites and detection is imperfect, but environmental covariates and proxies are ubiquitous to inform predictions of occurrence. While incomplete detection could result in underestimates of reality, the predictive accuracy of these models can be used to assess model performance. Usually, such models have shown high performance specifically with machine-learning algorithms that make the best use of available data [1–3]. Different subsets of data can be pooled and used for calibration, or as training data to assess model performance [4,5]. According to comparative reviews, among the best performing SDMs are Random Forests [6,7], Maximum Entropy (MaxEnt - [8]) and Boosted Regression Trees (BRTs) or Generalized Boosting Models (GBMs; see [2,9–12]. We used 675 observations of rattlesnakes for modelling; of these, 75% were used for fitting and 25% for evaluation. Our data were presence-background data as defined by [13]. We also generated pseudo-absence data from a random sample of 5,000 background sites from the environmental data. We ran 10 replicates for the model in MaxEnt, which effectively randomly subsampled the rattlesnake observations 10 times.

The most important factor was determining the threshold at which rattlesnakes would be considered present (see [14,15]. Two types of error prediction can occur in SDMs; false negatives (omission errors) which underestimate species' distributions, and false positives (commission errors) which overestimate species' distributions. These errors are then measured using specificity (the proportion of correctly classified absences) and sensitivity (representing the proportion of correctly classified presences) in the model results. To evaluate the quality of a model in predicting a species distribution, data were divided into two groups: 1) a training dataset, used to create the initial model and 2) an independent test dataset which was used to test model quality [16] .

The first type of model platform we used was Random Forests [6,7,17]. Random Forests is part of the regression tree family but deploys bagging, rather than boosting and both the response variable and predictor variables can be either categorical or continuous. Computationally, Random Forests have many advantages including the fact that they are quite fast to train and predict, they can handle regression and classification, the generalization error is built into the model, they require only one or two tuning parameters, are used for high-dimensional solutions and are implemented in parallel. From a statistical perspective, other attributes include their visualization, missing value imputation, differential class weighting and evaluation of variable importance.

Rather than using a regressive relationship, the second platform, Maximum Entropy (MaxEnt) is a machine-learning model that seeks the solution that is the most uniform between the species and environmental variables [18,19]. It “estimates a target probability distribution by finding the probability distribution of maximum entropy (i.e., that is, most spread out, or closest to uniform), subject to a set of constraints that represent incomplete information about the target distribution” [18]. The algorithm assigns the highest probability possible to each pixel in the area of study (the sum of which must equal to one) based on the species’ occurrences and a number of variables. The advantages of MaxEnt include the fact that it relies on presence-background data, provides continuous output, input data can be either categorical or continuous, it makes no prior assumptions about the distribution of the response curves and is not as sensitive to sample size as are most other algorithms [9,18,20,21]. The resulting maps are in logistic format, with probabilities of suitability varying between 0 and 1, and can be interpreted as an estimate of the probability of presence [18]. Although MaxEnt tends to over-fit data, a regularization function can be used to prevent this (the default value is 1 but this can be changed, especially if different landscape scenarios or climate change scenarios are modeled). Moreover, predictions outside of the training data have high uncertainty —but this criticism applies to all predictive species distribution models [18,22]. Generalized Boosting Models (GBMs) or Boosted Regression Trees (BRTs) combine information from both statistical and machine learning modeling. Instead of producing a single ‘best model’ as in regression, they use a boosting technique originating from machine learning to combine hundreds or thousands of single regression trees (which use recursive binary splits to model the relationship between a response variable and predictors) and which optimize model performance [23]. It can fit a wide variety of models including logistic regression. To control for over-fitting we used a regularization factor of 1 (as in the default version of MaxEnt) which we considered was appropriate for our model. We stipulated a logistic output; thus, within the area of the training variable values are predicted probabilities whereas outside this range they represent relative suitability (see [19]).

All of the model platforms that we deployed have been evaluated as having high performance. However, choosing which individual SDM to use can be challenging because of variation in performance according to different species, regions and applications. Therefore, it has been recommended by many authors that an average of several different models is used (an ensemble - see [24,25], although not all researchers are in agreement. We

therefore chose to use an ensemble model by overlaying the lattices and creating average values using the BIOMOD2 Package in R.

## References

1. Craig, E.; Huettmann, F. Using “Blackbox” Algorithms Such as TreeNET and Random Forests for Data-Mining and for Finding Meaningful Patterns, Relationships and Outliers in Complex Ecological Data. *Intell, Data Anal.* **2009**, 65–84, doi:10.4018/978-1-59904-982-3.ch004.
2. Hegel, T.M.; Cushman, S.A.; Evans, J.; Huettmann, F. Current State of the Art for Statistical Modelling of Species Distributions. In *Spatial Complexity, Informatics, and Wildlife Conservation*; Cushman, S. A. and Huettmann, F., Ed.; Springer, New York, NY, USA; 2010; pp. 273–311 ISBN 9784431877714.
3. Drew, C. A., Wiersma, Y. F., Huettmann, F. Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications; Springer-Verlag, New York, NY, USA; 2011.4. Breiman, L. Statistical Modelling: The Two Cultures. *Statist. Sci.* **2001**, 16, 199–215.
5. Huettmann, F., Gottschalk, T. Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. In *Predictive Modeling in Landscape Ecology*; Springer: New York, NY, USA, 2011; pp. 189–208.
6. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning: Methods and Applications*, Zhang, C., Ma. Y. Eds.; Springer Science+Business Media, LLC, Berlin, Germany, 2012, pp. 157-175. doi: 10.1007/978-1-4419-9326-7\_5.
7. Breiman, L. Random Forests. *Mach. Learn.* **2001**, 45, 5–32, doi:10.1023/A:1010933404324.
8. Phillips, S.J.; Anderson, R.P.; Dudík, M.; Schapire, R.E.; Blair, M.E. Opening the Black Box: An Open-Source Release of Maxent. *Ecography* **2017**, 40, 887–893, doi:10.1111/ecog.03049.
9. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudik, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; et al. Novel Methods Improve Prediction of Species’ Distributions from Occurrence Data. *Ecography* **2006**, 29, 129–151, doi:10.1111/j.2006.0906-7590.04596.x.
10. Stryzowska, K.M.; Johnson, G.; Mendoza, L.R.; Langen, T.A. Species Distribution Modeling of the Threatened Blanding’s Turtle’s ( *Emydoidea Blandingii* ) Range Edge as a Tool for Conservation Planning. *J.Herpetol.* **2016**, 50, 366–373, doi:10.1670/15-089.

11. Archis, J.N.; Akcali, C.; Stuart, B.L.; Kikuchi, D.; Chunco, A.J. Is the Future Already Here? The Impact of Climate Change on the Distribution of the Eastern Coral Snake (*Micrurus fulvius*). *PeerJ* **2018**, *6*, e4647, doi:10.7717/peerj.4647.
12. Ward, R.J.; Griffiths, R.A.; Wilkinson, J.W.; Cornish, N. Optimising Monitoring Efforts for Secretive Snakes: A Comparison of Occupancy and N-Mixture Models for Assessment of Population Status. *Sci. Rep.* **2017**, *7*, 18074 doi:10.1038/s41598-017-18343-5.
13. Guillera-Arroita, G.; Lahoz-Monfort, J.J.; Elith, J.; Gordon, A.; Kujala, H.; Lentini, P.E.; McCarthy, M.A.; Tingley, R.; Wintle, B.A. Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications. *Glob. Ecol. Biogeog.* **2015**, *24*, 276–292, doi:10.1111/geb.12268.
14. Liu, C.; White, M.; Newell, G.; Griffioen, P. Species Distribution Modelling for Conservation Planning in Victoria, Australia. *Ecol. Modell.* **2013**, *249*, 68–74, doi:10.1016/j.ecolmodel.2012.07.003.
15. Liu, C.; Newell, G.; White, M. On the Selection of Thresholds for Predicting Species Occurrence with Presence-Only Data. *Ecol. Evol.* **2016**, *6*, 337–348, doi:10.1002/ece3.1878.
16. Fielding, A.H.; Bell, J.F. A Review of Methods for the Assessment of Prediction Errors in Conservation Presence / Absence Models. *Environ. Conserv.* **1997**, *24*, 38–49, doi:10.1017/S0376892997000088.
17. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random Forests for Classification in Ecology. *Ecology* **2007**, *88*, 2783–2792, doi:10.1890/07-0539.1.
18. Phillips, S. J., Anderson, R. P., Schapire, R.E. Maximum Entropy Modeling of Species Geographic Distributions. *Ecol. Modell.* **2006**, *190*, 231–259, doi:10.1016/j.ecolmodel.2005.03.026.
19. Elith, J.; Phillips, S.J.; Hastie, T.; Dudík, M.; Chee, Y.E.; Yates, C.J. A Statistical Explanation of MaxEnt for Ecologists. *Divers. Distrib.* **2011**, *17*, 43–57, doi:10.1111/j.1472-4642.2010.00725.x.
20. Buermann, W.; Saatchi, S.; Smith, T.B.; Zutta, B.R.; Chaves, J.A.; Milá, B.; Graham, C.H. Predicting Species Distributions across the Amazonian and Andean Regions Using Remote Sensing Data. *J. Biogeog.* **2008**, *35*, 1160–1176, doi:10.1111/j.1365-2699.2007.01858.x.

21. Elith, J.; Leathwick, J.R. Species Distribution Models: Ecological Explanation and Prediction across Space and Time. *Ann.l Rev. Ecol. Evol. System.* **2009**, *40*, 677–697, doi:10.1146/annurev.ecolsys.110308.120159.
22. Wisz, M.S.; Hijmans, R.J.; Li, J.; Peterson, A.T.; Graham, C.H.; Guisan, A.; Elith, J.; Dudík, M.; Ferrier, S.; Huettmann, F.; et al. Effects of Sample Size on the Performance of Species Distribution Models. *Divers. Distrib.* **2008**, *14*, 763–773, doi:10.1111/j.1472-4642.2008.00482.x.
23. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, *29*, 1189–1232.
24. Araújo, M.B.; New, M. Ensemble Forecasting of Species Distributions. *Trends in Ecol. Evol.* **2007**, *22*, 42–47, doi:10.1016/j.tree.2006.09.010.
25. Hao, T.; Elith, J.; Guillera-Arroita, G.; Lahoz-Monfort, J.J. A Review of Evidence about Use and Performance of Species Distribution Modelling Ensembles like BIOMOD. *Divers. Distrib.* **2019**, *25*, 839–852, doi:10.1111/ddi.12892.



**Table S2**

Supplementary Results

Pearson correlation coefficients for environmental covariates.

	Temperature	Precipitation	Annual insolation	Elevation	Slope	Ruggedness	Road density
Temperature	1						
Precipitation	-0.361	1					
Annual insolation	-0.049	0.219	1				
Elevation	-0.952	0.453	0.085	1			
Slope	-0.297	0.027	-0.232	0.273	1		
Ruggedness	-0.268	0.009	-0.244	0.243	0.888	1	
Road density	0.300	-0.064	0.030	-0.280	-0.285	-0.276	1

### Table S3: Thresholds

We used the predicted values for the sampled sighting locations to determine the threshold for suitable and unsuitable land cover for the ensemble from the three model platforms. For the overall ensemble model the mean prediction value at the sampled sighting locations was 0.446. Using this figure as a critical threshold (and counting all pixels with values greater than this) gave 1,766,551 pixels (1,589.9 km<sup>2</sup>) that were classified as habitat for rattlesnakes. While this area (1,589.9 km<sup>2</sup>) was close to the 'area of occupancy' used by COSEWIC (804-1,424 km<sup>2</sup>), the population estimate for this threshold (area of pixels multiplied by the average density from the three study areas) was unrealistically high (85,058 adults) and included many locations known not to be occupied by rattlesnakes. This estimate could represent the 'potential' population. Based on different thresholds for the ensemble model and using the average density from three study areas, the population was estimated at: 66,525 (0.5), 40,254 (0.6), 21,536 (0.7), 14,902 (0.75) and 9,722 (0.8). For all the remaining calculations and scenarios in this paper we used a threshold of 0.8 which resulted in an estimated 204,319 pixels classified as suitable or 183.9 km<sup>2</sup>. This gave a mean population of 9,722 ( $\pm$  3,009) adults from separate population estimates using the density values from the three intensive study areas (e.g., Osoyoos, 10,702; White Lake, 6,344; and Vernon, 12,118; see Table S3). This was very close to other provincial estimates, and was the most conservative.

Table S3. Adult population estimates for entire range of the western rattlesnake (*Crotalus oreganus*) in British Columbia, Canada using the Ensemble model from Random Forests, MaxEnt and Generalized Boosting Models and density estimates from three different study populations.

Threshold	# of suitable pixels	Osoyoos (58.2/km <sup>2</sup> )	White Lake (34.5/km <sup>2</sup> )	Vernon (65.9/km <sup>2</sup> )
<b>0.446</b>	1766551	92531.94	54851.41	104774.1
<b>0.5</b>	1398076	73231.22	43410.26	82919.89
<b>0.6</b>	845964	44311.59	26267.18	50174.12
<b>0.7</b>	452608	23707.61	14053.48	26844.18
<b>0.75</b>	313177	16404.21	9724.146	18574.53
<b>0.8</b>	204319	<b>10702.23</b>	<b>6344.105</b>	<b>12118.16</b>
<b>0.99</b>	12	0.62856	0.3726	0.71172

**Figure S1.** Response curves of rattlesnake occurrence in response to predictor variables used in individual models and ensemble model. Each graph shows how rattlesnakes' Relative Index of Occurrence changes based on different values of each variable.

