1   **A Statistical Assessment of Population Trends for Data Deficient Mexican**

2   **Amphibians**

3

4   Esther Quintero, Subcoordinación de Especies Prioritarias, Dirección General de Análisis y

5   Prioridades, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico

6   D.F., Mexico

7   Anne E. Thessen, The Data Detektiv, Waltham, MA, USA

8   Paulina Arias-Caballero, Subcoordinación de Especies Prioritarias, Dirección General de

9   Análisis y Prioridades, Comisión Nacional para el Conocimiento y Uso de la

10  Biodiversidad, México D.F., México

11  Bárbara Ayala-Orozco, Subcoordinación de Especies Prioritarias, Dirección General de

12  Análisis y Prioridades, Comisión Nacional para el Conocimiento y Uso de la

13  Biodiversidad, México D.F., México

14  Corresponding author: Esther Quintero, Comisión Nacional para el Conocimiento y Uso de

15  la Biodiversidad, Liga Periférico - Insurgentes Sur, Núm. 4903, Col. Parques del Pedregal,

16  Tlalpan, 14010, Mexico, D.F. Tel. + 52 55 5004 4980 email

17  esther.quintero@conabio.gob.mx

1

# Abstract

**Background:** Mexico is the fourth richest country in amphibians and the second country with the highest quantity of threatened amphibian species, and this number could be higher as many species are too poorly known to be accurately assigned to a risk category. The absence of a risk status or an unknown population trend can slow or halt conservation action, so it is vital to develop tools that in the absence of specific demographic data can assess a species' risk of extinction, population trend, and to better understand which variables increase their vulnerability. Recent studies have demonstrated that the risk of species decline depends on extrinsic and intrinsic trait, thus including both of them for assessing extinction might render more accurate assessment of threat. **Methods:** In this study harvested data from the Encyclopedia of Life (EOL) and the published literature for Mexican amphibians and used these data to assess the population trend of some of the Mexican species that have been assigned to the Data Deficient category of the IUCN using Random Forests, a Machine Learning method that gives a prediction of complex processes and identifies the most important variables that account for the predictions. **Results:** Our results show that most data deficient Mexican amphibians have decreasing population trends. We found that Random Forests is a solid and accurate way to identify species with decreasing population trends when no demographic data is available. Moreover, we point the most important variables that make species more vulnerable for extinction. This exercise is a very valuable first step in assigning conservation priorities for poorly known species.

2

## Introduction

Amphibians are keystone to conservation and excellent bioindicators. Their extinction will trigger cascading effects on the ecosystem (Gardner 2001; Wake 1991; Wyman 1990). Among all terrestrial vertebrates, amphibians are the most threatened group with more "rapidly declining species" (Stuart et al. 2004). Mexico is the fourth richest country in amphibian species (Ochoa-Ochoa & Flores-Villela 2006) with around 375 documented species, although this number could be greatly underestimated (Flores-Villela & Canseco-Márquez 2004). At the same time, Mexico is the second country with the highest quantity of threatened amphibian species, 211 according to the IUCN (IUCN 2014), and this number could be even higher, as many species are too poorly known to be accurately assigned to a risk category. In Mexico, 38 (10%) amphibian species are currently listed as DD (Data Deficient) by the IUCN (2014) because specific data about a species are missing (i.e. geographic distribution, threats, population status, etc.).The absence of a risk status or the knowledge about the population trend can slow or halt conservation action, which for some species could have irreversible consequences. Therefore, it is of vital importance to develop tools that allow assessing species' risk in the absence of specific demographic data, as well as to better understand which variables increase vulnerability to extinction.

The first Global Amphibian Assessment (Stuart et al. 2004), found that amphibian declines are not random, but associated to ecological traits (i.e. stream associated species), geographic distribution (i.e. montane areas in the Neotropics, Australia and New Zealand), and specific taxonomic groups (i.e. Leptodactylidae, Bufonidae, Ambystomatidae, Hylidae, and Ranidae). Moreover, they divided the causes of decline in three groups: over-exploitation, defined as those declining due to heavy extraction (concentrated in species in

3

1 East and Southeastern Asia); reduced habitat, defined as those that were suffering from

2 extreme habitat loss (concentrated in Southeast Asia, West Africa, and the Caribbean); and

3 enigmatic declines, those that are declining even though suitable area remains (restricted

4 mostly to South America, Mesoamerica, Puerto Rico and Australia). Enigmatic declines

5 were found to be positively associated with streams at high elevation in the tropics, and

6 chytridiomycosis emerged as the most likely culprit.

7

8 Chytridiomycosis is a fungal disease caused by *Batrachochytrium dendrobatidis*, and has

9 been related to the decline of at least 43 species of amphibians in Latin America (Lips et al.

10 2006). In México, there is an association between higher elevations (from 939 to 3200 m)

11 and the prevalence of the infection. It has been found in pristine and disturbed areas alike,

12 but does not seem very common throughout tropical rain forests or lowland deserts (Frías-

13 Alvarez et al. 2008). The reason for this marked preference for high areas with temperate

14 climates may be that the optimal range of growth for this fungus is between 17-25 C

15 (Piotrowski et al. 2004; Longcore et al. 1999). A geographical survey for the presence of

16 chytridiomycosis in Mexico found the presence of the fungus in sites that have reported

17 "enigmatic declines" in amphibian populations (Frías-Alvarez et al. 2008). The finding by

18 these authors suggests that chytridiomycosis is a likely cause behind many of these

19 enigmatic declines.

20

21 Recent studies (Murray et al. 2011; Tingley et al. 2013) have demonstrated that the risk of

22 species decline depends on the specific threats they face, such as habitat loss, presence of

23 invasive species, and pathogens (extrinsic traits), and the species' own biological ability to

24 cope with these threats, such as clutch size and body size (intrinsic traits). Thus, including

4

1    intrinsic traits along with extrinsic threats for assessing extinction might render more

2    accurate assessment of threat (Tingley et al. 2013), and thus improve allocation of

3    resources (Cardillo & Meijaard 2012).

4

5    One of the most recognized efforts to assign risk categories to species is that of  the

6    International Union for Conservation of Nature (IUCN), which recognizes seven different

7    extinction risk categories for evaluated species: two of them are for species that are already

8    extinct (Extinct and Extinct in the wild), three are those considered as threatened categories

9    (Critically Endangered, Endangered, and Vulnerable), two are for those species that are not

10   yet threatened (Near Threatened and Least Concern), whereas the last one is for those with

11   not enough information to be evaluated (Data Deficient). The IUCN also lists species that

12   have not yet been evaluated (Not Evaluated).

13

14   IUCN's criteria for assigning a threat category to species are "quantitative in nature", but

15   the data quality and the uncertainty attached to any evaluation vary. Estimations,

16   inferences, projections, and suspected facts based on related data are acceptable, as long as

17   they can be supported and specified in the documentation. The Data Deficient category

18   (DD) is assigned to those species in which the available data is not enough to determine a

19   threat category, not even indirectly, for example through the status of their habitat or other

20   causal factors (IUCN 2012). Only approximately 75,000 out of the 2 million described

21   species are evaluated by the IUCN and one sixth of them are Data Deficient

22   (http://www.iucnredlist.org/about/summary-statistics), with 25% of all amphibians

23   classified as such (Stuart et al. 2004). By lacking a threat status, Data Deficient species are

24   not taken into account for conservation programs, potentially placing them at a higher risk

5

1   of extinction. Thus, it is clear that a more automated method of evaluating risk that can use

2   a wider variety of available data and still give accurate results is needed.

3

4   In this study we aim to harvest data from the Encyclopedia of Life (EOL www.eol.org) and

5   the published literature for Mexican amphibians and use these data to assess the population

6   trend of some of the Mexican species that have been assigned to the DD category of the

7   IUCN using Random Forests, a Machine Learning method algorithm that gives a prediction

8   of complex processes and identifies the most important variables that account for the

9   predictions (Breiman 2001; Cutler et al. 2007; Murray et al. 2011). A recent assessment of

10  DD mammals using and comparing multiple Machine Learning tools found that Random

11  forests perform very well for this type of predictions (Bland et al. 2014).  Focusing on such

12  a vulnerable and ecologically important group as the amphibians not only potentiates our

13  conservation efforts, but also has the potential to improve assessment of other ecologically

14  important groups for which we might lack demographic data.

15

16  **Methods**

17  ***Selecting traits for the analysis***

18  In order to assess the population traits of those species listed as Data Deficient, we selected

19  previously identified intrinsic traits that can predispose species to a greater degree of

20  vulnerability, as well as a series of extrinsic traits that have been associated to amphibian

21  decline (Stuart et al. 2004).

22

23  The extrinsic traits in our analysis were habitat use, habitat loss/degradation (one of the

24  biggest concerns for biodiversity (Millenium Ecossitem Assesment 2005; Brooks et al.

6

1    2002; Frías-Alvarez et al. 2008; Groombridge 1992; Parra-Olea et al. 1999; Wyman 1990),

2    presence of introduced species, presence of pollution, climatic fluctuations, harvest for pet

3    trade, desiccation of bodies of water, presence of chytridiomycosis, and presence of other

4    diseases that may decimate populations. The intrinsic traits selected for our analysis were

5    snout-vent length, ova size, clutch size, and development type as a way to understand their

6    life history and ecological preferences (Murray et al. 2011).

7

8    *Automated Data Harvesting from Encyclopedia of Life*

9    Starting with a list of scientific names of Mexican amphibians (Table S1), relevant data and

10   text were harvested from EOL using TraitBank and the EOL API respectively (Parr et al.

11   2014). The code written for this project can be found at GitHub

12   (https://github.com/diatomsRcool/MexicanAmphibians). Data from EOL TraitBank was

13   retrieved by searching for taxon and measurement and downloaded as a .csv file. The API

14   was used to find the EOL identifier corresponding to each amphibian species. This

15   identifier, in combination with EOL chapters and keywords was used to filter and harvest

16   all relevant text data objects (Table 1). This process identified a subset of text data objects

17   for manual data extraction. Data from TraitBank and the text data objects were added to a

18   master spreadsheet for analysis (Table S1). Data gathered for this study that was not

19   already in TraitBank, was placed in a Darwin Core Archive and uploaded into EOL

20   TraitBank.

21

22   *Other sources of literature*

23   Data that was not available from EOL was obtained from the literature, and cited in Table

24   S2.  Data for threats (habitat loss/degradation, introduced species, pollution,

7

1    chytridiomycosis, climatic fluctuations, pet trade/harvest, desiccation of habitat, and other

2    diseases), as well as for population trend (decreasing, increasing, stable, unknown) were

3    obtained from the IUCN Red List (IUCN 2014).

4

5    *Data preparation*

6    A table was prepared with 302 rows and 16 columns (Table S1). Each row represented a

7    species and each column represented a trait of that species. Examination of this master table

8    revealed two traits (ova size and clutch size; Table 2) and four species (*Bolitoglossa*

9    *chinanteca*, *Dermophis oaxacae*, *Eleutherodactylus marnockii*, and *Eleutherodactylus*

10    *verruculatus*) to be particularly data deficient (defined as 10 or more missing traits). An

11    additional species was identified as being introduced (*Eleutherodactylus planirostris*).

12    These traits and species were removed from the data set.

13

14    All traits were coded into numeric categories (Table 3). Snout to Vent length classifications

15    followed (García & Ceballos 1994). In habitat use, we distinguished permanent water

16    associated from stream associated.  Threats were treated as present (1) or absent (0).

17    Chytridiomycosis was recorded as present in cases where it was reported as suspected.

18    Missing data were represented by a blank cell. From this table, we prepared a csv file for

19    missing data imputation in R. The scientific name, IUCN status, and population trend were

20    removed before imputation.

21

22    *Missing data imputation*

23    We used the mice package (van Buuren & Groothuis-Oudshoorn 2011) in R to impute

24    missing values (http://cran.r-project.org/web/packages/mice/mice.pdf). This was necessary

8

1   because the randomForest function did not tolerate missing values. All data were imported

2   into R as factors. The Snout-Vent Length, Habitat Use, and Development Type were

3   imputed as polytomous logistic regression (polyreg). The other traits were imputed using

4   logistic regression (logreg). Missing Population Trend data were not imputed. The data

5   before imputation can be found in Table S1. A summary of missing data can be found in

6   Table 2 and Table S1. Ten imputations were performed for each missing value. The final

7   imputed value was the mode of the 10 imputations. The data after imputation can be found

8   in Table S3.  The data set that includes the imputed data was used for predicting the

9   population trend for those species that were Data Deficient and Not Evaluated according to

10  the IUCN evaluation.

11

12  ***Predicting Population Trends***

13  We used the randomForest package in R (Liaw & Wiener 2002) to make predictions about

14  the population trends for each species of amphibian

15  (http://cran.rproject.org/web/packages/randomForest/randomForest.pdf). This was a two-

16  step process: the first step was using training data, i.e., traits for those species with known

17  population trends, to generate a random forest object. The second step was using the

18  random forest object to make predictions about population trends for those species listed as

19  Data Deficient and Not Evaluated by the IUCN (IUCN 2014).

20

21  The training data (including the imputed data) was read into R and given to the

22  randomForest (Liaw & Wiener 2002) function, which provided a random forest object as a

23  result. To test the efficacy of the random forest for prediction, we removed the Population

1     Trend data and made a prediction of population trend for the training data for comparison

2     to the observed population trend (Table 4).

3

4     The test data included all of the species with an unknown population trend and some

5     additional species, as needed, to balance the presence of categories for each trait, a

6     requirement for making predictions. The test data were read into R and given to the

7     randomForest and predict functions in the randomForest package (Liaw & Wiener 2002).

8     To ensure unbiased variable selection (Strobl et al. 2007), we used the cforest (Hothorn et

9     al. 2006a; Strobl et al. 2008; Strobl et al. 2007) and predict functions in the party package

10    (http://cran.r-project.org/web/packages/party/party.pdf). To better visualize the actual tree

11    structure, we used the ctree (Hothorn et al. 2006b) function in the party package to

12    visualize the interactions among the most important variables that determine the population

13    trend in the species included in our analysis.

14

15    **Results and Discussion**

16    Out of the 24 species classified as "Data Deficient" by the IUCN included in our analysis,

17    22 were predicted to be decreasing, and only two were classified as stable (Table 5). In

18    predicting Population Trend, the most important variables were Habitat Loss/Degradation,

19    Presence of Chytridiomycosis, Habitat Use, Development Type, Desiccation of Habitat,

20    and Presence of Introduced Species (Fig. 1). randomForest and cforest (results not shown)

21    show the same variables in the top 6 most important, which means that randomForest does

22    not have variable selection bias in this analysis. Therefore, we are confident in our results.

23

1    Our randomForest analysis accurately identifies species with decreasing population trend

2    (Table 4). In the training data there were 19 false positives and 7 false negatives (Precision

3    = 0.899. Recall = 0.960; F score of 0.929 where 1 is a perfect score) for the "decreasing"

4    category, which means that the method is more likely to flag a stable species as decreasing,

5    than it is to miss a decreasing species. Therefore, analyzing this kind of data with Random

6    Forest is a cautious (from the perspective of conservation groups) way to assess species

7    with unknown population trends, as it is unlikely that some of those whose population is

8    decreasing would be missed.

9

10    Habitat loss/degradation was the most critical variable (Fig. 2), which concurs with the vast

11    amount of information on the cause of species declines (Stuart et al. 2004). The results also

12    show that in the presence of habitat loss, the type of habitat use becomes an important

13    variable to determine the risk of a species. Species that depend on "Ephemeral Ponds", or

14    are "Stream Associated" seem to cope better when the surrounding habitat is lost, whereas

15    permanent water associated and terrestrial species are at higher risk (Fig. 2). However, if

16    chytridiomycosis is present, "Ephemeral Pond" and "Stream Associated" species seem to

17    have almost no possibilities of survival (Fig. 2).

18

19    Table 5 also shows the risk status according to the 2010 official Mexican National Red List

20    ((NOM-059-SEMARNAT-2010,

21    http://dof.gob.mx/nota_detalle_popup.php?codigo=5173091) for the Data Deficient species

22    included in the analysis. Of them, only one, *Bolitoglossa stuarti,* is categorized as

23    Endangered (A), while 5 of them (*Chiropterotriton mosaueri, Craugastor taylori,*

24    *Eleutherodactylus maurus, Eleutherodactylus pallidus,* and *Eleutherodactylus teretistes*)

1    are considered "Under Special Protection", which is the lowest risk status of the List. The

2    other 17 species have not been assessed at the national level and thus are not listed. Of all

3    these species only *Eleutherodactylus teretistes* results with a stable population trend, so it

4    would be advisable to assess the other 22 species for the new version of the official

5    Mexican National Red List, which is the only national policy instrument which foresees

6    law enforcement in order to protect Mexican threatened species. It thus becomes clear why

7    efforts like this are important in order to pinpoint priorities to fill the gaps needed to inform

8    public policy and advance in the conservation of the species.

9

10    In a similar study of Australian amphibians, Murray et al. (2011) found that Habitat Use

11    (ecological group) was the most important variable to determine population trend, followed

12    by the presence of chytridiomycosis and *Gambusia*, a predatory fish (defined by spatial

13    models of suitability). Contrary to what we did, these authors included extent of

14    occurrence, abundance, and testes mass, and the presence of *Gambusia*. Although our study

15    and that of Murray et al. (2011) found different variables as the most important to

16    determine population trend, both studies agree on the fact that habitat use and the presence

17    of chytridiomycosis are some of the most important variables. Moreover, our study concurs

18    with that of Murray et al. (2011) in that by integrating intrinsic and extrinsic factors that are

19    pertinent for the target region, one can get an accurate account of the population trend of a

20    given amphibian species, as well the risk factors that are most pressing for the different

21    ecological groups. In this study we have addressed the variables that are most pressing for

22    Mexican species, and our results show, just as in the Australian case, that this kind of

23    analysis can identify areas to focus limited conservation resources. Another important point

24    is the geographic importance of the analysis. Because Mexican and Australian amphibians

1  are subject to different extrinsic factors, an analysis appropriate for one does not necessarily

2  apply to the other.

3

4  **Conclusions**

5  The use of Random forests seems to be a very solid and accurate way to identify species

6  with decreasing population trends in the absence of demographic data. The kind of exercise

7  that we show here is an important first step when planning conservation priorities, as some

8  of the most endangered species might also be those for which most information is lacking,

9  thus falling through the cracks of conservation planning. Moreover, this method has the

10  advantage of not having to depend on aggregated museum locality data that may not have

11  been properly curated by experts, as is the case for some assessment efforts (Hjarding et al.

12  2014).

13

14  Adding intrinsic factors to this analysis, such as the ova and clutch size, as those variables

15  could also give us some information on how life history can affect the population trend of a

16  species when faced with certain extrinsic threats. However, as the amount of data we had

17  for those traits was so limited, we felt that including data that had mostly been statistically

18  generated could introduce an extra bias to our analysis. The fact that so little information on

19  the natural history of these endangered species is available is a major challenge that needs

20  to be addressed to successfully prevent their extinction. In addition, our aggregated data set

21  can be used to set data collection priorities to fill in gaps. Fortunately, as we show here, this

22  lack of information should not deter our efforts to assess risk status and assign priorities to

23  their conservation.

24

13

10   **Reference list**

11   Bland LM, Collen BEN, Orme CDL, and Bielby JON. 2014. Predicting the Conservation

12       Status of Data-Deficient Species. *Conservation Biology*: DOI: 10.1111/cobi.12372.

13   Breiman L. 2001. Random forests. *Machine Learning* 45:15–32.

14   Brooks TM, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Rylands AB, Konstant

15       WR, Flick P, Pilgrim J, Oldfield S, Magin G, and Hilton-Taylor C. 2002. Habitat

16       Loss and Extinction in the Hotspots of Biodiversity. *Conservation Biology* 16:909-

17       923.

18   Cardillo M, and Meijaard E. 2012. Are comparative studies of extinction risk useful for

19       conservation? *Trends in Ecology & Evolution* 27:167-171.

20   Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, and Lawler JJ. 2007.

21       Random Forests  for classification in ecology. *Ecology* 88:2783-2792.

22   Flores-Villela O, and Canseco-Márquez L. 2004. Nuevas especies y cambios taxonómicos

23       para la herpetofauna de México. *Acta Zoologica Mexicana* 20:115-144.

Frías-Alvarez P, Vredenburg V, Familiar-López M, Longcore J, González-Bernal E, Santos-Barrera G, Zambrano L, and Parra-Olea G. 2008. Chytridiomycosis Survey in Wild and Captive Mexican Amphibians. *EcoHealth* 5:18-26.

García A, and Ceballos G. 1994. *Guía de campo de los reptiles y anfibios de la costa de Jalisco*. México: Fundación Ecológica de Cuixmala, A.C. .

Gardner T. 2001. Declining amphibian populations: a global phenomenon in conservation biology. *Animal Biodiversity and Conservation* 24:25-44.

Groombridge B. 1992. *Global biodiversity: status of the Earth's living resources. World Conservation Monitoring Centre*. New York, NY, USA: Chapman & Hall.

Hjarding A, Tolley KA, and Burgess ND. 2014. Red List assessments of East African chameleons: a case study of why we need experts. *Oryx*.

Hothorn T, Buehlmann P, Dudoit S, Molinaro A, and Van Der Laan A. 2006a. Survival Ensembles. *Biostatistics* 7:355-373.

Hothorn T, Homik K, and Zeileis A. 2006b. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphic Statistics* 15:651-674.

IUCN. 2012. *IUCN Red List Categories and Criteria: Version 3.1*: Gland, Switzerland and Cambridge, UK: IUCN.

IUCN. 2014. The IUCN Red List of Threatened Species. Version 2014.1. *Available at http://www.iucnredlist.org* (accessed May 12 2014).

Piotrowski, J. S, Annis, S. L, and Longcore, J. E. 2004. Physiology of *Batrachochytrium dendronbatis*, a chytrid pathogen of amphibians. *Mycologia*:9-15.

Liaw A, and Wiener M. 2002. The rabdomforest package. *R News* 2:18-22.

Lips KR, Brem F, Brenes R, Reeve JD, Alford RA, Voyles J, Carey C, Livo L, Pessier AP, and Collins JP. 2006. Emerging infectious disease and the loss of biodiversity in a

15

Neotropical amphibian community. *Proceedings of the National Academy of Sciences of the United States of America* 103:3165-3170.

Longcore JE, Pessier AP, and Nichols DK. 1999. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia*:219-227.

Millenium Ecossitem Assesment. 2005. Ecosystems and Human Well-Being: Biodiversity Synthesis. World Resources Institute.

Murray KA, Rosauer D, McCallum H, and Skerratt LF. 2011. Integrating species traits with extrinsic threats: closing the gap between predicting and preventing species declines. *Proceedings of the Royal Society B: Biological Sciences* 278:1515-1523.

Ochoa-Ochoa LM, and Flores-Villela O. 2006. *Areas de Diversidad y Endemismo de la Herpetofauna Mexicana*. México, D.F. Universidad Nacional Autonoma de Mexico—Comision Nacional para el Conocimiento y Uso de la Biodiversidad.

Parr CS, Wilson N, Leary P, Schulz KS, Lans K, Walley L, Hammock JA, Goddard A, Rice J, Studer M, Holmes JTG, and Corrigan RJ Jr. 2014. TraitBank: Practical semantics for organism attribute data. http://www.semantic-web-journal.net/content/traitbank-practical-semantics-organism-attribute-data.

Parra-Olea G, García-París M, and Wake DB. 1999. Status of some populations of Mexican salamanders (Amphibia:Plethodontidae). *Revista de Biología Tropical*:217-223.

Strobl C, Boulesteix AL, Kneib T, Augustin T, and Zeileis A. 2008. Conditional Variable Importance for Random Forests. *BMC Bioinformatics* 9:307.

Strobl C, Boulesteix AL, Zeileis A, and Hothorn T. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 8.

16

1    Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, and Waller

2         RW. 2004. Status and Trends of Amphibian Declines and Extinctions Worldwide.

3         *Science* 306:1783-1786.

4    Tingley R, Hitchmough RA, and Chapple DG. 2013. Life-history traits and extrinsic threats

5         determine extinction risk in New Zealand lizards. *Biological Conservation* 165:62-

6         68.

7    van Buuren S, and Groothuis-Oudshoorn K. 2011. mice: Multivariate Imputation by

8         Chained Equations. *R Journal of Statistical Software* 45:1-67.

9    Wake DB. 1991. Declining amphibian populations. *Science*:860.

10   Wyman RL. 1990. What's Happening to the Amphibians? *Conservation Biology*:350-352.

11

12

13

Figure Titles and Captions

Figure 1
Title: Relative Importance of Variables for Predicting Population Trend
Caption: Bar graph showing the relative importance of all variables for predicting population trend. The individual variables are listed on the vertical axis. The horizontal axis shows the decrease in accuracy of the final result if the variable is removed. Important variables have a higher mean decrease in accuracy.

Figure 2
Title: Variable Interactions Visualized
Caption: Visualization showing the interaction tree for the four most important variables. The ovals represent the variable and contain the p value for the split at that variable. The numbers on the lines refer to the values for each variable listed in Table 3. The bar charts at the bottom show the likelihood of a stable population trend. The taller the darker portion of the bar, the more likely the group will have a stable population trend. The "n =" value in parentheses indicates the number of species falling in each category.
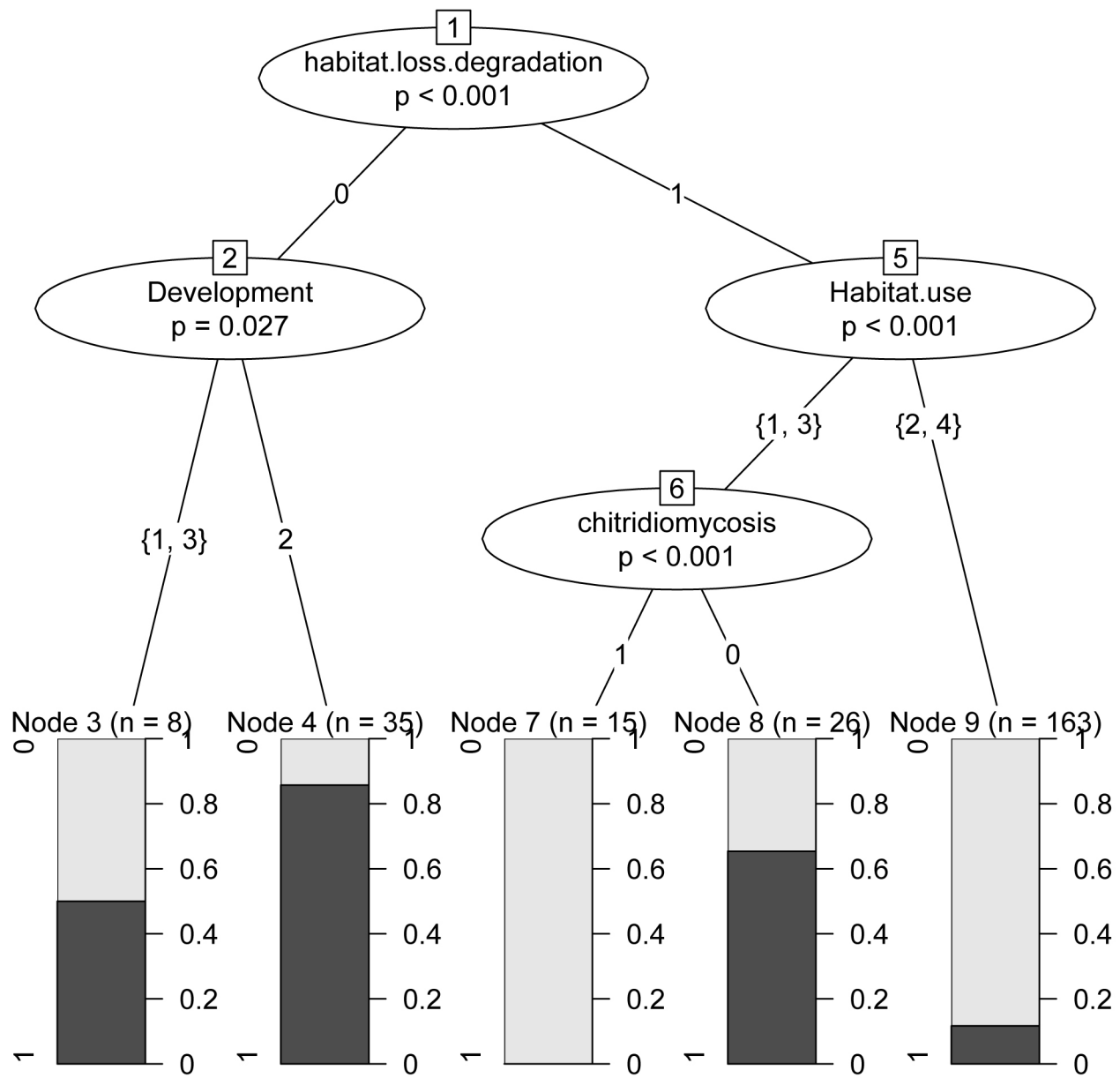
**Variable Importance**
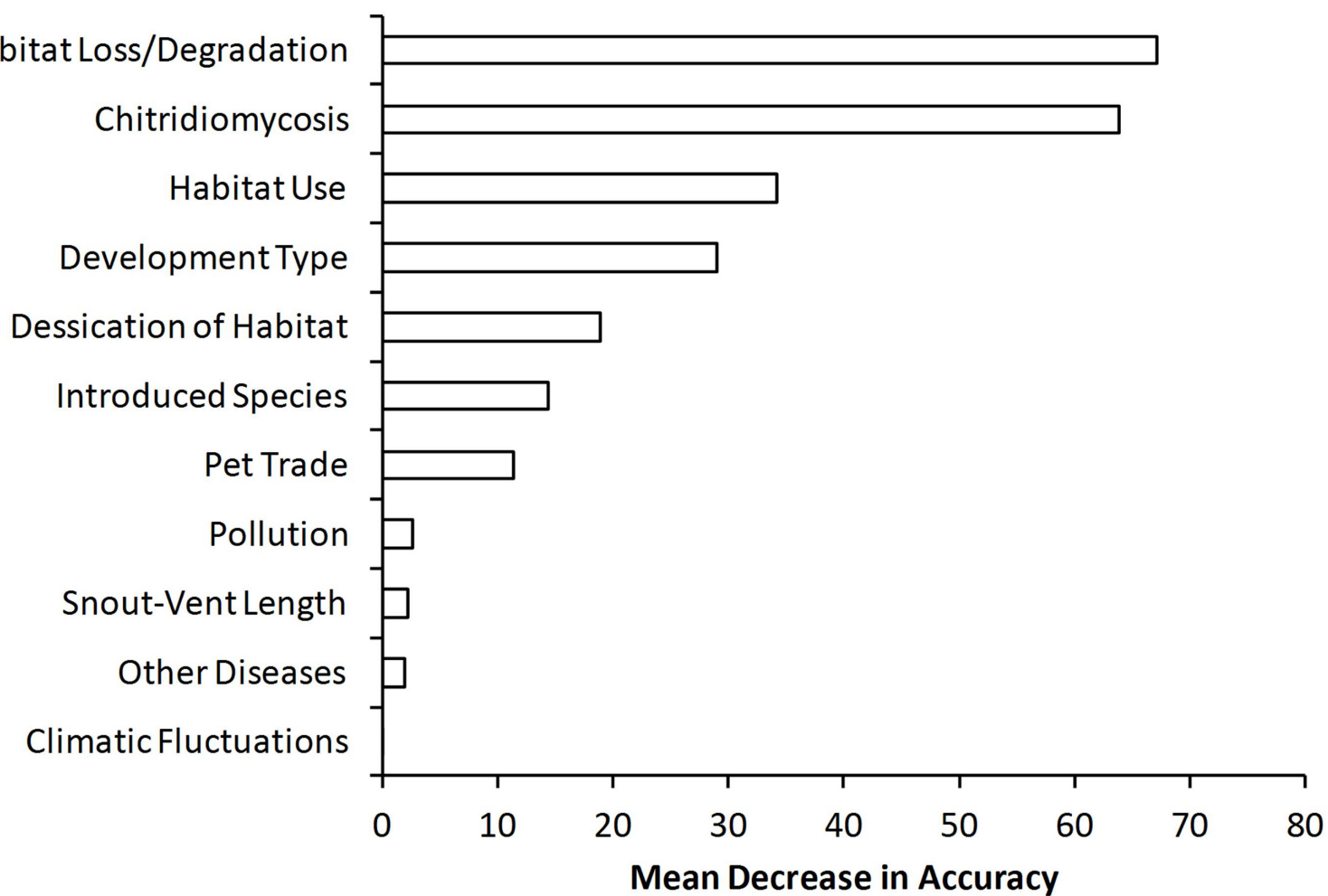
**Table 1**: EOL chapters and keywords used to filter and harvest relevant text data object for the study.

| Trait Type | EOL Chapter | Keyword |
|---|---|---|
| Intrinsic | Size, Reproduction, Life Cycle | length, clutch, egg, breeding, development, reproduction, hibernation |
| Extrinsic | Distribution, Habitat | occur, range, inhabit, found, precipitation, wet, arid, dry, moist, temperature, temperate, tropic |

**Table 2:** Number of missing data points for each variable. The 30 "missing" data points for the IUCN status actually refer to the number of Data Deficient and Not Evaluated species.

| Trait | Missing Data |
|---|---|
| Snout-Vent Length | 11 |
| Habitat Use | 1 |
| Ova Size | 276 |
| Development | 5 |
| Clutch Size | 252 |
| Habitat Loss/Degradation | 4 |
| Introduced Species | 4 |
| Pollution | 4 |
| Chytridiomycosis | 4 |
| Climatic Fluctuations | 4 |
| Pet Harvest | 4 |
| Desiccation of Habitat | 4 |
| Other Diseases | 4 |
| IUCN status | 30 |
| Population Trend | 53 |

**Table 3:** Numeric categories codes for the traits used in the study.

| Trait | Category | Definition |
|---|---|---|
| Snout-Vent Length | 1 | up to 69 mm |
| | 2 | 70-120 mm |
| | 3 | 121-171 mm |
| | 4 | more than 172 mm |
| Habitat Use | 1 | ephemeral pond associated |
| | 2 | permanent water associated |
| | 3 | stream associated |
| | 4 | terrestrial |
| Development | 1 | direct development |
| | 2 | larval development |
| | 3 | paedomorphic |
| Habitat Loss/Degradation | 0 | absent |
| | 1 | present |
| Introduced Species | 0 | absent |
| | 1 | present |
| Pollution | 0 | absent |
| | 1 | present |
| Chytridiomycosis | 0 | absent |
| | 1 | present |
| Climatic Fluctuations | 0 | absent |
| | 1 | present |
| Pet Trade/harvest | 0 | absent |
| | 1 | present |
| Desiccation of Habitat | 0 | absent |
| | 1 | present |
| Other Diseases | 0 | absent |
| | 1 | present |
| Population Trend | 0 | Decreasing |
| | 1 | Stable |

**Table 4:** Confusion matrix obtained using the randomForest and predict functions in the randomForest package (Liaw & Wiener 2002) on the training data to predict population trend.

|  | PREDICTED | |
|---|---|---|
| **OBSERVED** | Decreasing | Stable |
| Decreasing | 170 | 7 |
| Stable | 19 | 51 |

**Table 5:** Predicted population trend for the 24 species classified as Data Deficient by the IUCN. The categories on the 2010 official Mexican National Red List (NOM-Semarnat-059-2010) are as follows: E, extinct; P, endangered; A, threatened; Pr, under special protection.

| Species | Population Trend predicted | Mexican Red List |
|---|---|---|
| *Bolitoglossa oaxacensis* | decreasing | - |
| *Bolitoglossa stuarti* | decreasing | A |
| *Bolitoglossa zapoteca* | decreasing | - |
| *Chiropterotriton mosaueri* | decreasing | Pr |
| *Craugastor amniscola* | decreasing | - |
| *Craugastor occidentalis* | decreasing | - |
| *Craugastor pelorus* | decreasing | - |
| *Craugastor taylori* | decreasing | Pr |
| *Eleutherodactylus maurus* | decreasing | Pr |
| *Eleutherodactylus pallidus* | decreasing | Pr |
| *Eleutherodactylus teretistes* | stable | Pr |
| *Exerodonta abdivita* | decreasing | - |
| *Exerodonta bivocata* | decreasing | - |
| *Lithobates lemosespinali* | decreasing | - |
| *Pseudoeurycea amuzga* | decreasing | - |
| *Pseudoeurycea maxima* | decreasing | - |
| *Pseudoeurycea mixcoatl* | decreasing | - |
| *Pseudoeurycea obesa* | decreasing | - |
| *Pseudoeurycea quetzalanensis* | decreasing | - |
| *Pseudoeurycea tlilicxitl* | decreasing | - |
| *Ptychohyla acrochorda* | decreasing | - |
| *Ptychohyla zophodes* | decreasing | - |
| *Thorius insperatus* | stable | - |