




Effects of uncertainty and variability on population declines and IUCN Red List classifications

Pamela Rueda-Cediel,^{1,2} Kurt E. Anderson,¹ Tracey J. Regan,^{3,4} and Helen M. Regan ^{1*}

¹Evolution, Ecology and Organismal Biology Department, University of California-Riverside, 900 University Avenue, Riverside, CA 92521, U.S.A.

²College of Biological Sciences, University of Minnesota, 315 Ecology Building, 1987 Upper Buford Circle, St. Paul, MN 55108, U.S.A.

³Arthur Rylah Institute for Environmental Research, The Department of Environment, Land, Water and Planning, Heidelberg, VIC 3084, Australia

⁴School of Biosciences, University of Melbourne, Melbourne, VIC 3010, Australia

Abstract: *The International Union for Conservation of Nature (IUCN) Red List Categories and Criteria is a quantitative framework for classifying species according to extinction risk. Population models may be used to estimate extinction risk or population declines. Uncertainty and variability arise in threat classifications through measurement and process error in empirical data and uncertainty in the models used to estimate extinction risk and population declines. Furthermore, species traits are known to affect extinction risk. We investigated the effects of measurement and process error, model type, population growth rate, and age at first reproduction on the reliability of risk classifications based on projected population declines on IUCN Red List classifications. We used an age-structured population model to simulate true population trajectories with different growth rates, reproductive ages and levels of variation, and subjected them to measurement error. We evaluated the ability of scalar and matrix models parameterized with these simulated time series to accurately capture the IUCN Red List classification generated with true population declines. Under all levels of measurement error tested and low process error, classifications were reasonably accurate; scalar and matrix models yielded roughly the same rate of misclassifications, but the distribution of errors differed; matrix models led to greater overestimation of extinction risk than underestimations; process error tended to contribute to misclassifications to a greater extent than measurement error; and more misclassifications occurred for fast, rather than slow, life histories. These results indicate that classifications of highly threatened taxa (i.e., taxa with low growth rates) under criterion A are more likely to be reliable than for less threatened taxa when assessed with population models. Greater scrutiny needs to be placed on data used to parameterize population models for species with high growth rates, particularly when available evidence indicates a potential transition to higher risk categories.*

Keywords: life-history traits, population dynamics, red lists, risk assessment, threatened species

Efectos de la Incertidumbre y la Variabilidad sobre las Declinaciones de Población y las Clasificaciones de la Lista Roja de la UICN

Resumen: *Los Criterios y las Categorías de la Unión Internacional para la Conservación de la Naturaleza (UICN) son un marco de trabajo cuantitativo que se usa para clasificar a las especies de acuerdo al riesgo de extinción. Los modelos poblacionales pueden usarse para estimar el riesgo de extinción o las declinaciones de población. La incertidumbre y la variabilidad surgen en las clasificaciones de amenaza a través de errores de medida y del proceso en los datos empíricos y la incertidumbre en los modelos que se usan para estimar el riesgo de extinción y las declinaciones de población. Además, se sabe que las características de las especies afectan el riesgo de extinción. Investigamos los efectos de los errores de medida y proceso, tipo de modelo, tasa de crecimiento poblacional, y edad durante la primera reproducción sobre las clasificaciones de riesgo basadas en las declinaciones de población proyectadas por las clasificaciones de la Lista Roja de*

*email belen.regan@ucr.edu

Article impact statement: Population models can be used for reliable assessments of IUCN Red Lists under uncertainty and variability. Paper submitted February 16, 2017; revised manuscript accepted January 12, 2018.

la UICN. Utilizamos un modelo poblacional estructurado por edades para simular trayectorias verdaderas con diferentes tasas de crecimiento, edades reproductivas y niveles de variación, y los sometimos a errores de medición. Evaluamos la habilidad de los modelos escalares y de matriz, parametrizados con estas series de tiempo simuladas, para capturar con exactitud la clasificación de la Lista Roja de la UICN generada con declinaciones poblacionales verdaderas. En todos los niveles de error de medida evaluados y en el nivel bajo de error de proceso, las clasificaciones fueron razonablemente exactas; los modelos escalares y de matriz produjeron aproximadamente la misma tasa de clasificaciones equivocadas, pero la distribución de los errores difirió; los modelos de matriz resultaron en más sobreestimaciones del riesgo de extinción que subestimaciones; los errores de proceso tendieron a contribuir en mayor proporción a clasificaciones equivocadas que los errores de medida; y ocurrieron más clasificaciones equivocadas para las historias de vida rápidas que para las lentas. Estos resultados indican que las clasificaciones de los taxones con mayor amenaza (es decir, los taxones con tasas bajas de crecimiento) bajo el criterio A tienen mayor probabilidad de ser confiables que para los taxones menos amenazados cuando se evalúan con modelos poblacionales. Se debe realizar un mayor escrutinio en los datos que se usan para parametrizar los modelos poblacionales para las especies con una mayor tasa de crecimiento, particularmente cuando la evidencia disponible indica una transición potencial hacia categorías de mayor riesgo.

Palabras Clave: características de historia de vida, dinámicas poblacionales, especies amenazadas, evaluación de riesgo, listas rojas

摘要: 世界自然保护联盟(IUCN)红色名录濒危等级和标准是一个根据物种灭绝风险进行分类的定量框架, 其中可能用到种群模型来估计灭绝风险或种群下降。经验数据测量和处理的误差以及用于估计灭绝风险和种群下降的理论模型的不确定性, 使濒危等级的划分也存在不确定性和可变性。此外, 物种自身的性状也会影响灭绝风险。我们根据IUCN红色名录濒危等级的预计种群下降, 探究了数据测量和处理的误差、模型类型、种群增长率和初次繁殖年龄对濒危等级划分的可靠性的影响。我们用年龄结构种群模型来模拟不同的增长率、繁殖年龄及变异程度下真实的种群变动轨迹, 同时考虑测量误差的影响。我们还评估了以这些模拟的时间序列为参数的标量模型和矩阵模型能否准确地估计出由真实的种群下降确定的IUCN红色名录濒危等级。在包含我们检验的所有水平的测量误差及较低水平的处理误差时, 濒危等级的划分都相当准确; 标量模型和矩阵模型划分濒危等级的错误率大致相同, 但错误的分布不同; 矩阵模型对灭绝风险更多的是高估而非低估; 相比于测量误差, 处理误差会导致更多错误; 相比于生活史慢的物种, 生活史快的物种更易被错误划分濒危等级。这些结果说明, 当用种群模型进行评估时, 依据标准A评估高度濒危的类群(即增长率低的类群)比评估较不濒危的类群更加可靠。最后, 我们应仔细检查增长率高的物种的种群模型参数数据, 特别是在有证据表明它们的濒危等级可能上升时。【翻译:胡怡思; 审校:魏辅文】

关键词: 种群动态, 红色名录, 风险评估, 受胁迫物种, 生活史性状

Introduction

Threat rankings for endangered species often rely on assessments of extinction risk and population declines (Andelman et al. 2004; Hunter et al. 2010). Although numerous classification protocols exist to assess threat status (Andelman et al. 2004), the International Union for Conservation of Nature (IUCN) Red List Categories and Criteria (IUCN 2012) has emerged as the global standard for quantitative classifications of extinction risk (Mace et al. 2008). The IUCN Red List Categories and Criteria is a framework for ranking species according to their extinction risk or proxies of extinction risk. The IUCN Red List has 5 sets of criteria: reduction in population size (criterion A), restricted distribution and decline (criterion B), small population size and decline (criterion C), very small or restricted population (criterion D), and probability of extinction (criterion E). Rules and thresholds are specified within each criterion that categorize taxa as either critically endangered (CR), endangered (EN), vulnerable (VU), near threatened (NT), least concern (LC),

or data deficient (DD). Two additional categories exist for evaluated taxa with adequate data: extinction (EX) and extinct in the wild (EW).

As of July 2017 there were 5,048 threatened species classified according to criterion A, for which population models were used in only 43 cases and mostly for species affected by commercial fisheries. However, this is likely an underestimate of the number of assessed taxa to which population models have been applied because it does not include assessments for which criterion A was applied but the highest threat category resulted from another criterion. The criteria rely on the best available evidence on risk of extinction or proxies for these measures such as the number and size of populations, estimates of population declines, and the spatial extent of the taxon. We focused on how uncertainty in population models and estimates of population declines can affect IUCN Red List classifications based on any one of the 4 subcriteria, A2–A4, under criterion A. Thresholds of decline in abundance over 10 years or 3 generations under subcriteria A2–A4 are >80% for CR, 50–80% for EN, and 30–50% for

VU. We did not consider criterion A1 because it involves different thresholds and requires the causes of population reduction to have ceased.

Estimates of population decline used in IUCN Red List assessments are inherently uncertain; hence, consideration of such assessments' reliability is fundamental to conservation prioritization (Taylor 1995; D'Eon-Eggertson et al. 2014). Erroneous risk assessments can result when uncertainty is not accounted for in the development of population models, potentially resulting in over- or underestimation of the actual population decline (Wilson et al. 2011; Connors et al. 2014). Misclassification errors can lead to consequences for endangered species conservation. For instance, if a critically endangered species is listed as least concern (an underestimation of extinction risk) intensive protective measures that might save the species would most likely not be executed, leading to further population declines from which the species may not be able to recover. Conversely, a vulnerable species listed as critically endangered (overestimation of extinction risk) may result in increased demand for the species due to erroneous perceptions of rarity, precipitating a decline. This pattern has been documented for butterflies in Papua New Guinea, where prices were higher for rarer species (Slone et al. 1997) and for species uplisted to a more restrictive appendix in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Rivalan et al. 2007). Errors in classifications can also compromise the realization of funding-allocation goals through conservation triage methods (Joseph et al. 2008). Such methods rely on accurate threat assessments to trade off the costs of managing highly imperiled species against possibly greater conservation benefits for less threatened species. Persistent misclassification errors in either direction can also erode confidence in threat classifications, which can lead to costly litigation when decisions based on such classifications impinge on human livelihood or when conservation actions for species are compromised (Akçakaya et al. 2017).

Uncertainty can arise in quantitative assessments of extinction risk in multiple ways: through measurement and systematic error in empirical data, through subjective judgment about quantities for which there is little data, and through uncertainty in the models used to estimate population declines and extinction risk (Regan et al. 2002). Although the adoption of population models has been recommended to minimize uncertainty associated with subjective judgment in estimates of extinction risk or population decline (Brook et al. 2002), uncertainty remains in the choice of model that best represents population dynamics and in the estimates of model parameters. Model choice depends on how much is known about the life history of the taxon and the amount of available data with which to estimate parameters (Regan et al. 2003). Two types of population models commonly used in conservation contexts are scalar and matrix models

(Morris & Doak 2002). Scalar population models assume all individuals are identical, and models are parameterized with a growth rate applied to the entire population. Matrix population models assume individuals differ in their contributions to population growth, and they are parameterized with survival and birth rates based on size, age, or life stage. Although matrix models are thought to convey greater realism than scalar models, they are potentially subject to greater uncertainty due to compounding errors across numerous parameters (Rueda-Cediel et al. 2015).

Natural variation, or process error, in data used to parameterize population models can exacerbate the uncertainty in population projections (Regan et al. 2002). When both natural variation and measurement error are coupled in population counts, greater uncertainty in modeled population projections can result (Rueda-Cediel et al. 2015). Populations exhibiting greater variability tend to have higher extinction risk (Tuljapurkar & Orzack 1980; Lande et al. 2003).

Life-history traits also have an influence over population variability and, for a diverse range of species, life-history traits are good predictors of population dynamics (Rochet 2000; Morris et al. 2008; Van Allen et al. 2012). Age at first reproduction, proportion of repeat spawners, and fecundity have all been shown to correlate with population variability for fish, birds, mammals, reptiles, and amphibians (Rochet 2000; Stahl & Oli 2006; Van Allen et al. 2012). These traits have also been identified as proxies for population viability and growth rate (Rochet 2000; Stahl & Oli 2006). Age at first reproduction consistently influences population dynamics and viability (Lande et al. 2003; Morris et al. 2008; Saether et al. 2013), highlighting the need to understand how the interaction of age at first reproduction, environmental variability, and uncertainty affects the reliability of population viability assessments (PVA).

Life-history variation is commonly characterized along a fast to slow continuum. Slow life histories include traits such as low adult mortality, low growth rates, low fecundities, and late reproduction, whereas fast life histories include high adult mortality, high growth rates, high fecundity, and early age at first reproduction. Within this continuum, life histories at the slow end have been associated with greater risk of extinction (e.g., long gestation, small litters, and advanced age of first reproduction for mammals and birds [Purvis et al. 2000; Webb et al. 2002]).

Wilson et al. (2011), Regan et al. (2013), and Connors et al. (2014) evaluated the impact of different sources of uncertainty in PVA in the context of decision making. These researchers used simulated or real time-series data to construct state-space population models to examine the effect of process and measurement error on the outcome of extinction risk classifications. State-space models estimate measurement and process error to estimate true abundance from observed abundance. Results of these

Table 1. Parameter values used in model simulations of population trajectories.

| Parameter | Value |
|---|----------------------------|
| Input population growth rate (λ) | 0.900, 0.950, 1.000, 1.025 |
| Coefficient of variation for process error (PE) | 0.0, 0.1, 0.3, 0.5 |
| Coefficient of variation for measurement error (ME) | 0.0, 0.1, 0.3 |
| Age at first reproduction (years) (RA)* | 2, 5, 8 |
| Generation time (T_G) | 3.94, 6.73, 9.00 |
| Projection time (3 generations, rounded up)* | 12, 21, 27 |

*Ages at first reproduction are used to calculate generation times, which in turn are used to calculate projection times following criterion A of the IUCN Red List Categories and Criteria (IUCN 2012).

studies demonstrate that variability and uncertainty in data reduce the reliability of model projections and that these effects can translate to misclassification errors. All 3 studies are concerned solely with scalar models; the potential effects of compounding uncertainty over many parameters in matrix models on IUCN Red List classifications was not examined. State-space models, although powerful tools for modeling population trends under uncertainty and variability, are rarely used in applications of PVA, require long time-series data, and are difficult to apply to nonlinear problems (Bolker 2008). Hence, we followed more commonly used methods in PVA for parameterizing population models in which measurement error is subsumed into process error. We focused on both scalar and matrix models and identified the direction of the misclassification (Akçakaya et al. 1999; Morris & Doak 2002).

We used computer simulations to investigate the effects of model choice, process error, measurement error, population growth rate, and age at first reproduction on the reliability of IUCN Red List classifications made under criterion A when population models are used to project future population declines. We built on the methods and results of Rueda-Cediel et al. (2015), who compared the ability of age-structured and scalar models to project population percent decline for species when parameterized with the same variability and error-laden data sets. They found that scalar models exhibited the same or greater precision and accuracy in estimating population declines than matrix models under scenarios of high process and measurement error. Matrix models overestimated population declines when process and measurement error were high and underestimated declines for combinations of low to moderate process error and low measurement error. These biases tended to increase as the underlying population growth rate increased. However, Rueda-Cediel et al. (2015) did not examine the effect of age at first reproduction on population declines or evaluate the effects of species' traits, process and measurement error,

and model choice on IUCN Red List classifications. We applied the criteria A2–A4 thresholds from the IUCN Red List criteria to modeled population declines to determine the circumstances under which erroneous risk classifications occur and to determine the role of species' traits in the reliability of such risk classifications. In particular, we asked the following questions: How much process and measurement error cause misclassification errors and do these tend to over- or underestimate extinction risk? Which model, scalar or matrix, yields more reliable threat classifications under uncertainty? Which type of error, process or measurement, contributes the most to misclassification errors? And, which life histories incur the greatest threat misclassifications?

Methods

Below we provide a brief summary of our methods. Full details are presented in Supporting Information. Our method entailed 5 steps (Fig. 1). In step 1, the underlying true population dynamics are simulated with different levels of process error (Fig. 1). We simulated 1000 true population time series with stochastic age-structured matrix models with 10 age classes. Different true models were generated by varying age at first reproduction and intrinsic growth rate and incorporating different levels of process error (Table 1). Levels of process error and ages at first reproduction were chosen to correspond to ranges documented in natural populations of animals (Lande et al. 2003; Wilson et al. 2011; Pacifici et al. 2013).

We used 4 different input growth rates (λ): 0.9, 0.95, 1.0, and 1.025. A survival rate of 0.5 was assumed for all age classes. The age at first reproduction (RA) was set as 2, 5, or 8 years (Table 1). For all ages at first reproduction the same average fecundities were assigned to all breeding life stages. Fecundities were inversely calculated so that the dominant eigenvalue matched an a priori a specified growth rate (Table 1). Process error was incorporated into vital rates via β distributions characterized by the mean vital rate, as described above, and a variance specified by a predefined coefficient of variation (PE) (Table 1). Survival rates were sampled from a β distribution on the interval [0,1] and fecundities from a stretched β distribution bounded at 0 and 1,000. For simplicity and ease of interpretation, vital rates were assumed to be uncorrelated across age classes and through time. Density dependence was omitted because the main focus was on declining or stable populations that are less likely to be driven by density dependence (Gotelli 2008) and because IUCN's criterion A deals with declining populations. We constructed 48 models that reflected all combinations of growth rate, age of first reproduction, and process error (Table 1).

The initial population size was 1,000,000,000 (to avoid the effects of demographic stochasticity associated with

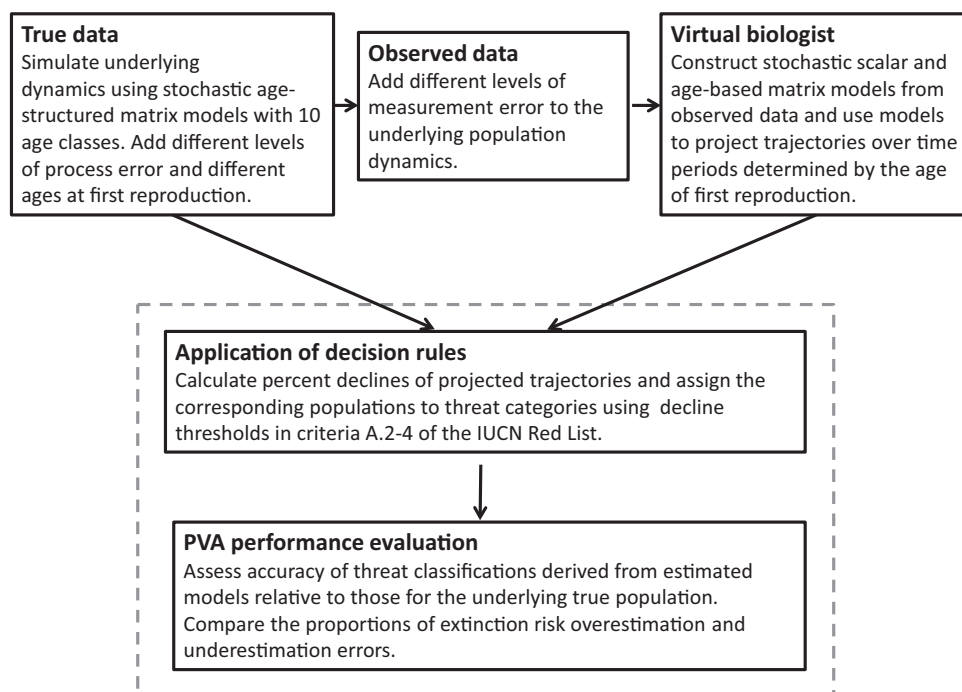


Figure 1. Summary of methods used to construct populations models. The dashed box indicates steps that apply the IUCN Red List criteria and the analysis of misclassification errors (PVA, population viability analysis).

small populations [Lande 1993]), and population dynamics were simulated for 50 years. We considered the final 20 years the true data that served as the baseline data to which measurement error was added and on which the estimated PVA models were based (step 2). A 20-year period was chosen because no effect of time-series length on population trajectories after 15–20 years has been shown (Rueda-Cediel et al. 2015). One thousand population trajectories were generated for each of the 48 models. These represented the true population dynamics in the absence of measurement error.

In step 2, observed time-series data, which represented the observations made by a virtual ecologist, were generated by adding measurement error to age-specific abundances in the underlying true time series (Fig. 1b). We assigned 3 coefficients of variation in measurement error (ME) (Table 1). Measurement error was incorporated into population counts for each of the 20 years in each population trajectory by randomly sampling from a normal distribution with the true abundance as the mean and the standard deviation calculated from this mean and the specified coefficient of variation (ME) (Table 1); distributions were bounded below by 0.

In step 3, we constructed scalar and age-based matrix models from each observed data set which then projected future population trajectories (Fig. 1). For the scalar model, 19 stochastic growth rates were calculated from 20-year population counts. The vital rates used in the estimated scalar population model were sampled, with replacement, from the 19 vital rates calculated for a given observed series.

Matrix models had 9 age classes, where the final stage in each life history was a composite class that included

all individuals older than 9 years (Akçakaya et al. 1999). Age-specific survival rates and fecundities were then calculated directly from the error-laden population data. We assumed fecundities were equal across the age classes. Under criterion A of the IUCN Red List Categories and Criteria, population trajectories are examined over 10 years or 3 generations (whichever is longer up to a maximum of 100 years). The parameterized estimated scalar and matrix models were used to project trajectories over 3 generations determined by age at first reproduction (Table 1). Likewise, population dynamics were projected a further 3 generations from the end of the 20-year sampling period with the true model 1,000 times. Hence, populations with an age of first reproduction of 2, 5, and 8 years were projected for 3 generations: 12, 21, and 27 years, respectively (Table 1). Refer to Supporting Information for calculation of generation time based on age of first reproduction. For each true simulation, 1000 trajectories were projected by each of the corresponding estimated scalar and matrix models.

In step 4, we applied IUCN criterion A to the population declines from the estimated and true projections (Fig. 1). We calculated the population percent decline for each true trajectory and the median percent decline from the corresponding estimated population projections and applied the thresholds under criterion A to assign a classification of CR, EN, or VU. The criteria for near threatened do not specify thresholds for population decline or require estimates of population decline, so we combined the near threatened and least concern categories in a catchall not threatened category because we were most interested in applying the decision rules for the threatened categories.

In step 5, a performance evaluation was conducted by comparing the classification derived from the estimated model with the true classification (CR, EN, VU, or not threatened) for each of the 144 combinations of parameter values from Table 1. Comparisons of threat classifications between the estimated and true population declines were divided into true positives and overestimation and under-estimation errors.

Results

For matrix models, under- and overestimation biases in median percent decline estimates occurred across combinations of process and measurement error, growth rates, and age at first reproduction (Fig. 2 & Supporting Information). For low measurement (0.1) and process (≤ 0.1) error and high growth rates (≥ 1.0), population declines were increasingly underestimated as age at first reproduction increased. In contrast, for large measurement error (0.3), high growth rate (≥ 1.0), and all levels of process error, population declines were increasingly overestimated as age at first reproduction increased. For low growth rates (≤ 0.95) population declines were overestimated when measurement error and process error were both high, but these biases decreased as age at first reproduction increased. In contrast to the results from matrix models, scalar models did not exhibit these biases for any combination of parameters.

Both scalar and matrix models exhibited similar precision in percent population decline predictions (i.e., the spread in the differences between true and estimated population declines) as described by the interquartile range (IQR) of these differences. Precision generally decreased as process and measurement error increased. Age at first reproduction interacted with growth rate, process error, and, to a lesser extent, measurement error and resulted in changes to IQR. For instance, for high growth rates (≥ 1.0), the IQR increased with age at first reproduction when either process or measurement error were present (Fig. 2 & Supporting Information). On the one hand, for $\lambda = 1.025$, ME = 0, and PE = 0.1, the IQR for results generated with matrix models increased from 27.82% to 44.12% and 60.45%, when age at first reproduction increased from 2 to 5 and 8 years, respectively (Fig. 2g–i). However, when $\lambda \leq 0.95$ the IQR decreased as age at first reproduction increased for all combinations of measurement error and process error (Fig. 2a–c).

The underlying range of true IUCN Red List categories obtained under criterion A depended on the age at first reproduction and the underlying growth rate. In the absence of process and measurement error, taxa with ages of first reproduction of 5 and 8 years were classified as CR when $\lambda = 0.90$, EN when $\lambda = 0.95$, and not threatened when $\lambda = 1.0$ & 1.025 (Fig. 3 & Supporting Information). When the age of first reproduction was 2 years, taxa were

classified as EN when $\lambda = 0.90$, VU when $\lambda = 0.95$, and not threatened when $\lambda = 1.0$ and 1.025 (Fig. 3 & Supporting Information). As process error increased so did the span of the threat categories for the true populations; all threat categories were represented as process error increased from 0 to 0.5 across all growth rates and age of first reproduction, although in different proportions.

In most cases, accurate classifications were achieved under all levels of measurement error tested, for low levels of process error, and across all growth rates, ages at first reproduction, and model type. Process error and growth rate tended to have the biggest effects on classification errors (Fig. 3). However, the effect of growth rate and age at first reproduction on the accuracy of threat classifications is complex. For low growth rates (≤ 0.95), accuracy increased as age of first reproduction increased, whereas for high growth rates (≥ 1.0), accuracy decreased as age of first reproduction increased (Fig. 3 & Supporting Information). In other words, when the underlying true threat category was high, such as CR and EN, accuracy increased as age of first reproduction increased. And when the underlying true threat category was not threatened, accuracy increased as the age at first reproduction decreased.

As process error increased, over- and underestimation of extinction risk increased across all growth rates, ages at first reproduction, and measurement-error levels for both models. Although high levels of measurement error exacerbated over- and underestimation of extinction risk, the degree to which measurement error affected errors differed across the 2 types of models. Higher levels of overestimation of extinction risk were observed with matrix models than with scalar models for high measurement (0.3) and process (≥ 0.1) errors and high growth rates (≥ 0.95) across all ages of first reproduction evaluated (Fig. 3 & Supporting Information). Conversely, higher rates of underestimation errors were observed with scalar models than with matrix models for high measurement (0.3) and process (0.5) errors and over most growth rates (≥ 0.95) for all ages of first reproduction evaluated.

Discussion

Our assessment of IUCN Red List categories was robust to high levels of measurement error and low levels of process error and all growth rates, ages at first reproduction, and model types. This is despite the fact that large biases were observed in population declines estimated with matrix models. This is good news, and it is due to the interval nature of the categories under criterion A that accommodates ranges of percent population decline. For the most part, accuracy of threat classifications increased when age at first reproduction increased and population growth rate decreased. In general, process error rather than measurement error had

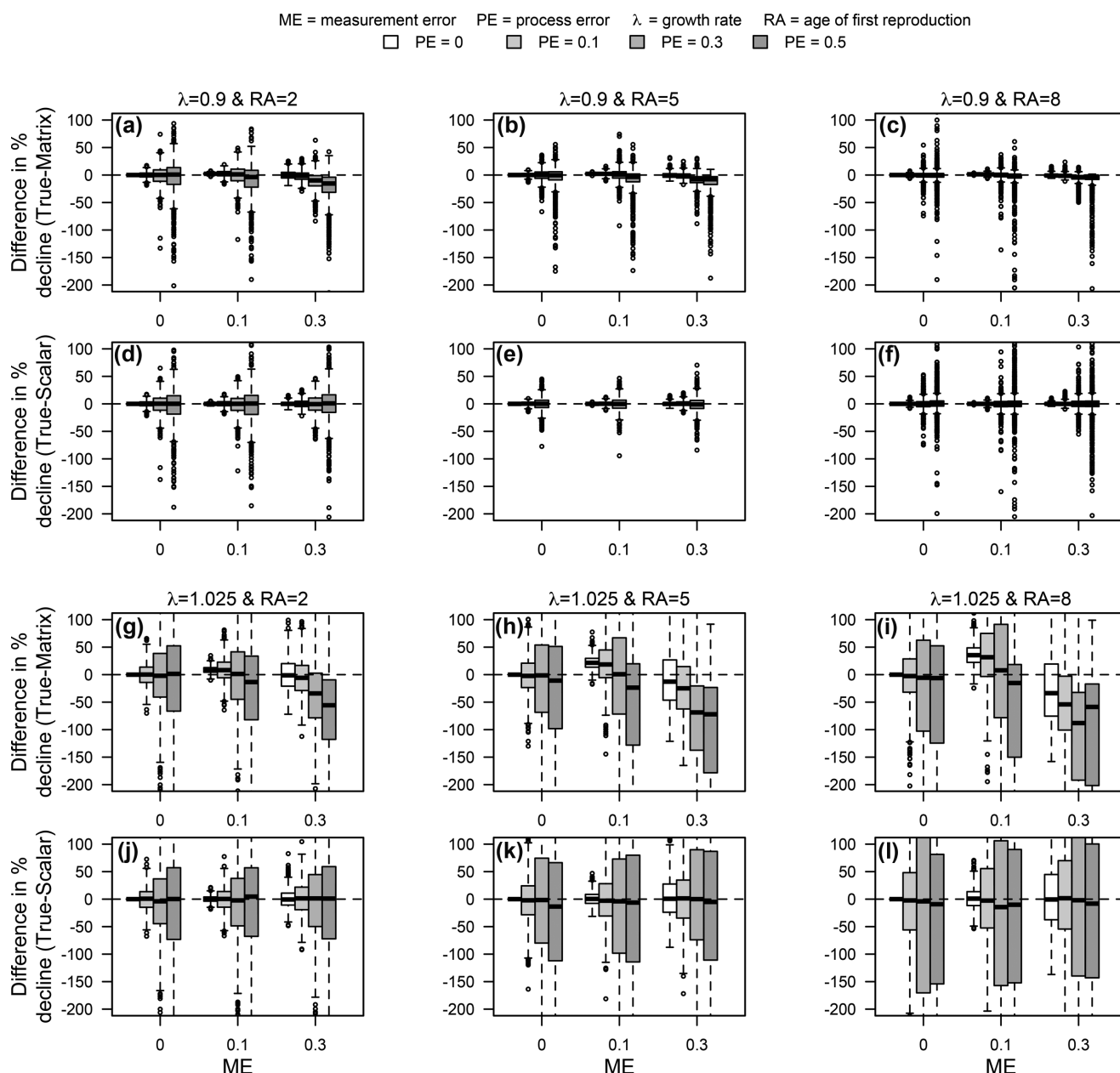


Figure 2. Difference in percent decline of population sizes between the true and median estimated model projections for matrix and scalar models when population growth rate (λ) is (a–f) 0.900 and (g–l) 1.025 across all ages at first reproduction (RA) (ME, coefficient of variation in measurement error; PE, coefficient of variation in process error; boxes, interquartile range (IQR); lower dashed vertical lines, 1st quartile – $1.5 \times$ IQR; upper dashed vertical lines, 3rd quartile – $1.5 \times$ IQR; thick black horizontal lines, median). Differences in percent decline are shown for PE = 0, 0.1, 0.3, and 0.5 for each ME value 0, 0.1, 0.3. Results for $\lambda = 0.95$ and 1.0 are in Supporting Information.

a stronger effect on the accuracy of threat classifications, and the strength of this effect depended on the age at first reproduction and the underlying growth rate.

Threat classifications based on projected population declines were more reliable for taxa at the slow end of the life-history continuum (later ages of first reproduction combined with low population growth rate) than they are for taxa at the fast end of the continuum (earlier

ages of first reproduction and high growth rate) (Fig. 3 & Supporting Information). This difficulty in accurately capturing the population dynamics and generating projections for species with early age at first reproduction and high variability in population estimates has been reported elsewhere (Conlisk et al. 2015).

In contrast to models with slow life-histories, matrix models with later ages of first reproduction combined

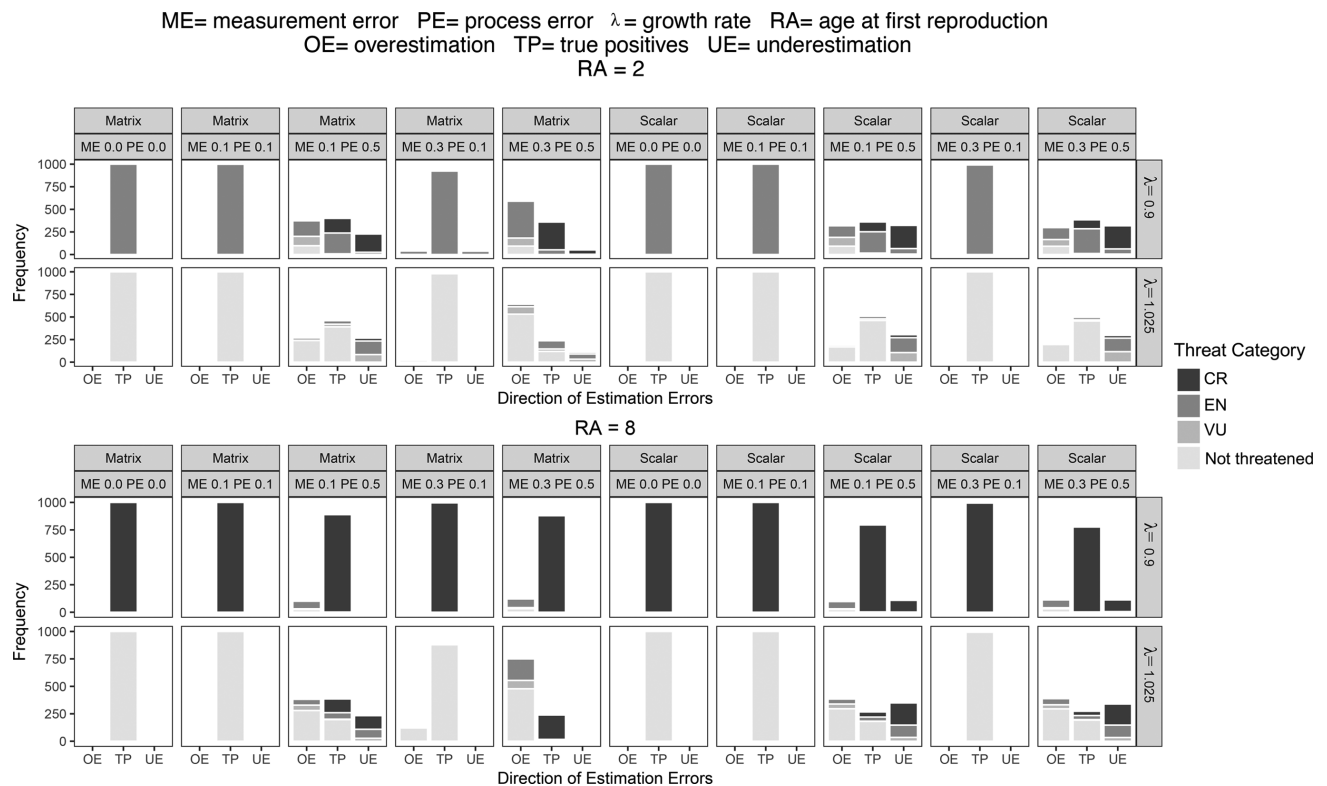


Figure 3. Frequency of threat classifications based on estimated population declines generated by scalar and matrix models relative to underlying true classifications for ages at first reproduction (RA) of 2 and 8 and all combinations of measurement and process error (performance categories: OE, overestimation; TP, true positives; UE, underestimation; threat classifications: CR, critically endangered; EN, endangered; VU, vulnerable; horizontal panels, specific population growth rates [λ] of 0.9 and 1.025; vertical panels, combinations of the coefficients of variation for measurement [ME] and process error [PE] for matrix and scalar models).

with high growth rates most dramatically overestimated population declines, leading to higher overestimation errors in classifications. This is due to increased variability in the underlying true dynamics as a result of time-lag-induced population cycles caused by increasing the age at first reproduction (Turchin 2003). Population dynamics under high growth rates exhibited higher variability than decreasing populations because process error was applied via a coefficient of variation.

Measurement error, when combined with higher levels of process error also drove matrix models to exhibit substantial decreases in accuracy, mostly in the direction of overestimation of population decline. The addition of measurement error to populations with higher growth rates could lead to unrealistically high estimated survival rates (i.e., > 1.0), resulting in a truncation of the distribution of survival rates and hence a bias in percent population declines and the corresponding IUCN Red List classifications that lead to more overestimation errors for matrix models than for scalar models. These results highlight the need for careful parameterization of matrix models because introduction and compounding of measurement error in matrix models can potentially lead to

overestimation of population declines. Employing more technically demanding parameterization schemes, such as state-space models (Wilson et al. 2011; Connors et al. 2014), that account for measurement and process error may be necessary to avoid these biases for population decline estimates, but they do not appear necessary to achieve greater accuracy in IUCN Red List assessments because the rate of true positives was roughly equal for declines generated with scalar and matrix models. Moreover, false negative rates generated in this study compared favorably with those generated by Connors et al. (2014) under high process error. Connors et al. (2014) similarly found that misclassification rates are the lowest for species with slow life histories and that misclassification rates for stable and declining populations are similar under process error for density-independent populations. This is encouraging because it shows a consistent pattern irrespective of the type of models and parameterization used.

Our results can help identify situations where IUCN Red List classifications based on population models to inform criterion A are most likely to be unreliable and whether the errors are likely to over- or underestimate

extinction risk for taxa. This should be useful to IUCN Red List assessors evaluating the quality of information used to assess species and to those constructing models to inform criterion A. These results indicate classifications of highly threatened taxa (i.e., taxa with low growth rates, under criterion A) are more likely to be reliable than for less threatened taxa when assessed with population models. Hence, greater scrutiny needs to be placed on data used to parameterize population models for species with high growth rates, particularly when available evidence indicates a potential transition to higher risk categories. The degree to which different errors are a concern to assessors depends on the magnitude of the error (e.g., truly CR and incorrectly assessed as not threatened versus truly CR and incorrectly assessed as EN) and the relative consequences of over- and underestimation of extinction risk. Our results can elucidate the likely magnitude of errors and the direction of such errors given a specific life history type and model employed. These results revealed that when population models are used under high process error, classifications are likely to be erroneous, particularly for species with fast life histories, and other information (e.g., for criteria B–D) should be used in the assessment. However, if process error is relatively low, then IUCN Red List assessments based on population declines are very likely to be accurate. This information can assist IUCN Red List assessors in deciding what information to put greatest reliance on and therefore which criteria to apply when making assessments under uncertainty.

Our results reveal 5 major insights that need to be considered in developing guidance to minimize misclassification errors under criterion A: under low to moderate levels of measurement error and low process error, classifications were reasonably accurate; scalar and matrix models yielded roughly the same rate of misclassifications, but the distribution of errors differed; matrix models led to greater overestimation of extinction risk than underestimation; process error tended to contribute to misclassifications to a greater extent than measurement error; and more misclassifications occurred for fast than slow life histories. Such an understanding of the potential under- and overestimation of extinction risk given by the criteria is critical to improving assessments. This is particularly pertinent for organizations around the world that have adopted the IUCN Red List criteria to guide environmental policy because there are real consequences of misclassification errors.

Acknowledgments

We thank J. Scott and C. Hilton-Taylor of the IUCN for assistance with extracting data from the IUCN Red List database.

Supporting Information

Detailed methods (Appendix S1) and additional results (Appendix S2) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

- Akçakaya HR, Burgman MA, Ginzburg LR. 1999. Applied population ecology: Principles and computer exercises using Ramas Ecolab 2.0. Sinauer Associates, Sunderland, Massachusetts.
- Akçakaya HR, Keith DA, Burgman MA, Butchart SHM, Hoffmann M, Regan HM, Harrison I, Boakes E. 2017. Inferring extinctions III: A cost-benefit framework for listing extinct species. *Biological Conservation* **214**:336–342.
- Andelman SJ, Groves C, Regan HM. 2004. A review of protocols for selecting species at risk in the context of US Forest Service viability assessments. *Acta Oecologica* **26**:75–83.
- Bolker BM. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey.
- Brook BW, Burgman MA, Akçakaya HR, O'Grady JJ, Frankham R. 2002. Critiques of PVA ask the wrong questions: Throwing the heuristic baby out with the numerical bath water. *Conservation Biology* **16**:262–263.
- Conlisk E, Syphard AD, Franklin J, Regan HM. 2015. Predicting the impact of fire on a vulnerable multi-species community using dynamic vegetation model. *Ecological Modelling* **301**:27–39.
- Connors BM, Cooper AB, Peterman RM, Dulvy NK. 2014. The false classification of extinction risk in noisy environments. *Proceedings of the Royal Society B* **281**:20132935.
- D'Eon-Eggertson F, Dulvy NK, Peterman RM. 2014. Reliable identification of declining populations in an uncertain world. *Conservation Letters* **8**: 86–96.
- Gotelli NJ. 2008. A primer of ecology. 4th edition. Sinauer Associates, Sunderland, Massachusetts.
- Hunter CM, Caswell H, Runge MC, Regehr EV, Amstrup SC, Stirling I. 2010. Climate change threatens polar bear populations: a stochastic demographic analysis. *Ecology* **91**: 2883–2897.
- IUCN (International Union for Conservation of Nature). 2012. IUCN red list categories and criteria. Version 3.1. IUCN, Gland, Switzerland. Available from <http://www.iucnredlist.org/technical-documents/categories-and-criteria/2001-categories-criteria> (accessed July 2016).
- Joseph LN, Maloney RF, Possingham HP. 2008. Optimal allocation of resources among threatened species: a project prioritization protocol. *Conservation Biology* **23**:328–338.
- Lande R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist* **142**:911–927.
- Lande R, Steinar E, Saether B. 2003. Stochastic population dynamics in ecology and conservation. Oxford University Press, Oxford.
- Mace G, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, Milner-Gulland EJ, Stuart S. 2008. Quantification of extinction risk: IUCN's System for classifying threatened species. *Conservation Biology* **22**:1424–1442.
- Morris WF, Doak DF. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer Associates, Sunderland, Massachusetts.
- Morris WF, et al. 2008. Longevity can buffer plant and animal populations against changing climatic variability. *Ecology* **89**:19–25.
- Pacifici M, Santini L, Di Marco M, Baisero D, Francucci L, Grottolomarasini G, Visconti P, Rondinini C. 2013. Generation length for mammals. *Nature Conservation* **5**:87–94.

- Purvis A, Gittleman JL, Cowlishaw G, Mace GM. 2000. Predicting extinction risk in declining species. *Proceedings of the Royal Society of London B* **267**:1947–1952.
- Regan HM, Colyvan M, Burgman MA. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* **12**:618–628.
- Regan HM, Akçakaya HR, Ferson S, Root KV, Carroll S, Ginzburg LR. 2003. Treatments of uncertainty and variability in ecological risk assessment of single-species populations. *Human and Ecological Risk Assessment* **9**:889–906.
- Regan TJ, Taylor BL, Thompson GG, Cochrane J, Ralls K, Runge MC, Merrick R. 2013. Testing decision rules for categorizing species' extinction risk to help develop quantitative listing criteria for the U.S. Endangered Species Act. *Conservation Biology* **27**:821–831.
- Rivalan P, Delmas V, Angulo E, Bull LS, Hall RJ, Courchamp F, Rosser AM, Leader-Williams N. 2007. Can bans stimulate wildlife trade? *Nature* **447**: 529–530.
- Rochet MJ. 2000. May life history traits be used as indices of population viability? *Journal of Sea Research* **44**:145–157.
- Rueda-Cediel P, Anderson KE, Regan TJ, Franklin J, Regan HM. 2015. Combined influences of model choice, data, quality, and data quantity when estimating population trends. *PLoS ONE* **10** (e0132255) <https://doi.org/10.1371/journal.pone.0132255>.
- Saether BE, et al. 2013. How life history influences population dynamics in fluctuating environments. *The American Naturalist* **182**: 743–759.
- Slone TH, Orsak LJ, Malver O. 1997. A comparison of price, rarity and cost of butterfly specimens: Implications for the insect trade and for habitat conservation. *Ecological Economics* **21**:77–85.
- Stahl JT, Oli MK. 2006. Relative importance of avian life-history variables to population growth rate. *Ecological Modelling* **198**:23–39.
- Sterns SC. 1983. The influence of size and phylogeny on patterns of covariation among life-history traits in mammals. *Oikos* **41**:173–187.
- Taylor BL. 1995. The reliability of using population viability analysis for risk classification of species. *Conservation Biology* **9**:551–558.
- Tuljapourkar SD, Orzack SH. 1980. Population dynamics in variable environments. I. Long-run growth rates and extinction. *Theoretical Population Biology* **18**:314–342.
- Turchin P. 2003. *Complex population dynamics: A theoretical/empirical synthesis*. Princeton University Press, New Jersey.
- Van Allen BG, Dunham AE, Asquith CM, Rudolf VHW. 2012. Life history predicts risk of species decline in a stochastic world. *Proceedings of the Royal Society B* **279**:2691–2697.
- Webb JK, Brook BW, Shine R. 2002. What makes a species vulnerable to extinction? Comparative life-history traits of sympatric snakes. *Ecological Research* **17**:59–67.
- Wilson HB, Kendall BE, Possingham HP. 2011. Variability in population abundance and the classification of extinction risk. *Conservation Biology* **25**:747–757.

