

Möglichkeiten und Risiken der Probabilistik im Individuen- und Artenschutz

Autorin: Fränzi Korner-Nievergelt

Einführung

Der Ausbau der Windenergie-Infrastruktur fördert einerseits die Klimafreundlichkeit der Energieproduktion, hat jedoch gleichzeitig negative Effekte auf die Artenvielfalt durch die Erschließung vorher unerschlossener naturnaher Gebiete, durch Störung und durch Kollisionen von fliegenden Tieren mit den Rotorblättern. Damit besteht ein Dilemma zwischen Klima- und Artenschutz. Um dieses Dilemma bestmöglich zu lösen, ist es nötig, die Effekte von Windenergie auf die Artenvielfalt sowie die Nützlichkeit von Windenergie und Artenvielfalt auf den langfristigen Erhalt unserer Lebensgrundlagen zu quantifizieren und gegeneinander abzuwägen. Für diese komplexen Abwägungen können Modellierungen und Wahrscheinlichkeitsberechnungen (Probabilistik) zum Einsatz kommen. In diesem Aufsatz skizziere ich Möglichkeiten, Grenzen und Risiken der Probabilistik für die Beurteilung des Einflusses eines Windenergieprojekts auf Populationstrends von kollisionsgefährdeten Arten. Zuerst werde ich die Varianz von Kollisionsraten anhand eines Beispiels diskutieren (Beispiel 1 in Abbildung 1). Danach erläutere ich, welche Information über Kollisionsraten oder erwartete Populationswachstumsraten uns die Probabilistik zur Verfügung stellt (Beispiel 2). Ich argumentiere, dass die klassische statistische Signifikanz wenig hilfreich ist bei der Beurteilung, ob ein Schwellenwert überschritten wird oder nicht, und ich skizziere zwei Alternativen: 1) Relevanz- und Evidenzschwellenwerte (Abbildung 1) und 2) formelle Methoden der strukturierten Entscheidungsfindung (Abbildung 4).

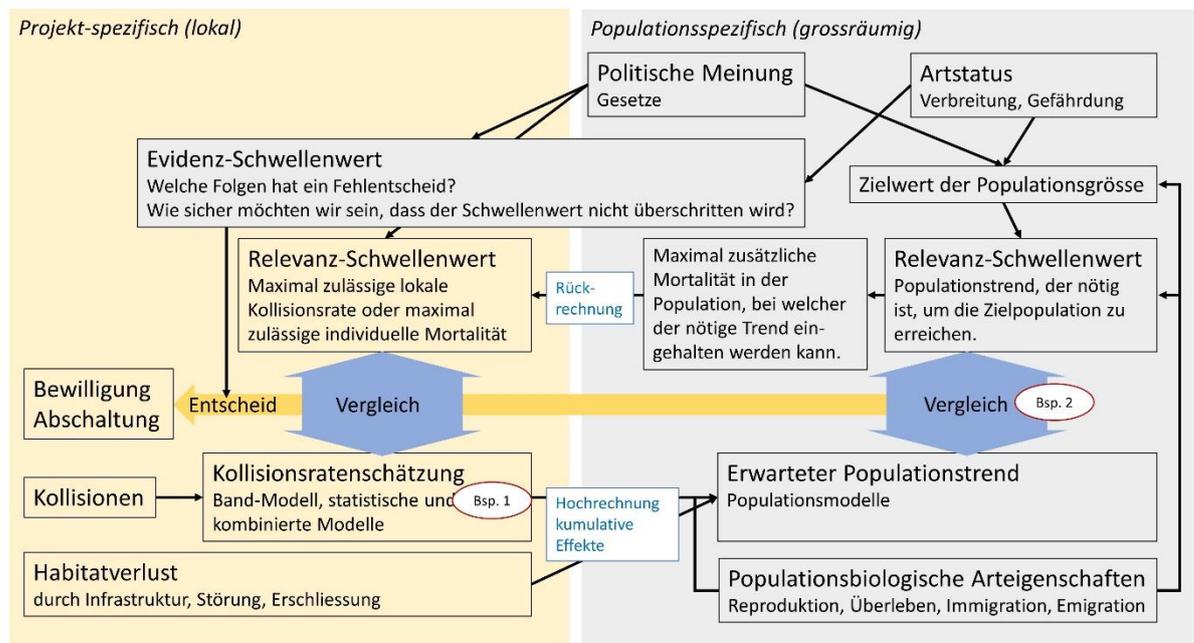


Abbildung 1. Schematischer möglicher Weg von der Kollisionsratenschätzung in einem lokalen (geplanten) Windenergieprojekt zum Entscheid über Bewilligung oder Abschaltung, unter Einbezug der politischen Meinung sowie des Gefährdungsstatus einer Art. Die runden Symbole zeigen, wo die ersten beiden in diesem Artikel verwendeten Beispiele anzusiedeln sind.

Unterschiedliche Rolle von Varianz und Unsicherheit in der Schätzung von Schlagrisiko und Schlagrate

Das (individuelle) Schlagrisiko ist definiert als Wahrscheinlichkeit, daß ein Individuum (z.B. ein Vogel oder eine Fledermaus) an einer Windenergieanlage während eines Referenzzeitraums verunglückt. Der Referenzzeitraum kann z.B. ein Durchflug durch einen Windpark, ein Nahrungssuchflug eines Rotmilans, ein Jahr oder die ganze Lebensspanne des Individuums sein. Auf einem Durchflug durch einen Windpark kann das Schlagrisiko sehr klein sein, jedoch kann für dasselbe Individuum die Wahrscheinlichkeit, an einer Windenergieanlage zu sterben, groß sein, wenn es viele Durchflüge tätigt.

Die Schlagrate ist definiert als die Anzahl Individuen, die an einer Windenergieanlage oder in einem Windpark pro Zeiteinheit (z.B. Jahr) verunglücken. Die Schlagrate wird bestimmt durch das Schlagrisiko der Individuen pro Zeiteinheit sowie die Individuendichte im Rotorbereich, welche wiederum bestimmt ist durch die Zahl der Individuen, die sich im Gebiet der Windenergieanlage bewegen, sowie das Flugverhalten der Individuen. Die Abschätzung der Einflüsse von Windparks auf Tierpopulationen beruhen auf artspezifischen Schlagraten, kumuliert über das Verbreitungsgebiet der Population oder der Art.

Beide Parameter, Schlagrisiko und -rate, können mit probabilistischen Methoden geschätzt werden. Das Band-Modell, ein mechanistisches Modell, und seine Weiterentwicklungen (z.B. Band et al., 2007; Masden & Cooc, 2016) schätzen das individuelle Schlagrisiko. Schlagraten können mittels Simulationen (z.B. Bastos et al., 2015) oder basierend auf Daten von systematischen Schlagopfernachsuchen und empirischer Modelle geschätzt werden (z.B. Huso et al., 2017; Niermann et al., 2011). Stehen neben den Schlagopferfunden auch Daten zur Anzahl Durchflüge zur Verfügung, kann auch das Schlagrisiko empirisch geschätzt werden (z.B. Aschwanden et al., 2018; Bastos et al., 2015; Grünkorn et al., 2016; Kerns et al., 2005). Empirische Daten zu Flugverhalten (z.B. GPS-Tracks), Schlagraten und -risiko bilden Grundlagen, um die mechanistischen Kollisionsmodelle oder Simulationen mit realistischen Parameterwerten zu versehen. So liefern z.B. Grünkorn et al. (2016), Aschwanden et al. (2018) und Schaub et al. (2020) konkrete Schätzwerte zum Ausweichverhalten. Das Ausweichverhalten ist maßgeblich mitbestimmend für das im Band-Modell geschätzte Schlagrisiko, deshalb sind artspezifische Daten zum Ausweichverhalten enorm wichtig (Chamberlain et al., 2006; Masden & Cooc, 2016).

Herausforderungen für die empirische Schätzung von Schlagrisiko und Schlagrate sind die hohe räumliche (Grünkorn et al., 2016; Kerns et al., 2005) und zeitliche (Aschwanden et al., 2018) Varianz in der effektiven Anzahl Schlagereignisse, die oft nicht unabhängig voneinander stattfinden, sowie die Seltenheit der Ereignisse und deren geringe Entdeckbarkeit (Huso et al., 2017). An einer Windenergieanlage finden an den meisten Tagen keine Kollisionen statt, während an ein paar wenigen Tagen gleich mehrere Kollisionen auftreten können, z.B. weil ein Schwarm den Rotorbereich durchfliegt, oder weil die Wetterbedingungen und schlechte Sicht Zugvögel zwingen, in Bodennähe zu fliegen (Aschwanden et al., 2018).

Durchschnittliche Kollisionsrisiken und -raten werden basierend auf Daten geschätzt, die an ausgewählten Windenergieanlagen während ausgewählter Zeiträume aufgenommen wurden. Basierend auf diesen Daten wird mittels statistischer Methoden räumlich und zeitlich extrapoliert, um Schlagraten für Windprojekte zu schätzen, wo keine Datenaufnahme stattgefunden hat, oder um die Schlagrate in der Zukunft zu

schätzen. Hinzu kommt, daß nur ein Teil der Schlagopfer gefunden wird und daraus auf die effektive Schlagopferzahl hochgerechnet werden muß. All diese Rechenschritte führen zu statistischer Unsicherheit in der Schätzung der durchschnittlichen Schlagrate bzw. des Schlagrisikos.

Im fiktiven Beispiel illustriere ich, wie die zeitliche und räumliche Varianz der Schlagopfer und die Größe der Stichprobe die statistische Unsicherheit beeinflussen. Wir schätzen für einen Windpark von 100 Windenergieanlagen die mittlere Anzahl Schlagopfer, einmal basierend auf einer Stichprobe von 10, einmal von 30 Windenergieanlagen. Im fiktiven Beispiel können wir die Schlagopferzahl an einer Windenergieanlage genau messen; in der Realität kommt bei diesem Schritt bereits eine Unsicherheit hinzu, weil wir nie sicher sein können, wie viele Schlagopfer nicht gefunden wurden. In Abbildung 2 sind für die 100 Windenergieanlagen für zwei Jahre die wahren, fiktiven Schlagopferzahlen in schwarz angegeben. Die Unterschiede in den Zahlen zeigen die Varianz in der Schlagopferzahl. Diese Varianz ist naturgegeben¹ und unabhängig von der Stichprobengröße, d.h. sie wird nicht kleiner, wenn wir intensiver beproben. Das bedeutet, daß wir auch mit einem sehr hohen Beprobungsaufwand und den besten Modellen nie genau vorhersagen können, wie die Schlagrate an einer Windenergieanlage in einem bestimmten Zeitraum sein wird. Wir können nur die *durchschnittlich* zu erwartende Schlagopferzahl schätzen.

In der kleinen Stichprobe (1) unseres Rechenbeispiels haben wir im ersten Jahr im Durchschnitt 0,2, in der großen Stichprobe 0,13 Schlagopfer gefunden. Im zweiten Jahr lagen die Durchschnitte der gleichen Stichproben bei 0,3 bzw. 0,8 Schlagopfer. Die Verteilung der Schlagopferzahlen für jede Stichprobe und jedes Jahr habe ich in Abbildung 2 als Histogramme dargestellt. Die Mittelwerte sowie die statistische Unsicherheit der Mittelwerte sind in der Abbildung als Kreise und horizontale Balken unterhalb der Histogramme dargestellt; die Balken sind 95% Unsicherheitsintervalle². In den meisten Fällen wird das Unsicherheitsintervall kleiner, wenn die Stichprobe grösser wird, wie im Beispiel im ersten Jahr. Wenn die Daten stark streuen, wie das bei Schlagopferzahlen oft der Fall ist, kann ein Unsicherheitsintervall zufälligerweise auch grösser werden, wenn mehr Daten hinzukommen, wenn im größeren Datensatz seltene Ereignisse miterfaßt wurden, die im kleineren Datensatz fehlten. Das ist im Beispiel im zweiten Jahr der Fall. Auch wenn im zweiten Jahr das Unsicherheitsintervall in der großen Stichprobe grösser ist als in der kleinen Stichprobe, ist jenes der großen Stichprobe dennoch zuverlässiger als jenes der kleinen Stichprobe. In kleinen Stichproben werden Varianzen und damit auch die statistische Unsicherheit im Durchschnitt unterschätzt (McNeish & Stapleton, 2016). Die Unterschiede in den Mittelwerten zwischen den Jahren werden durch die zeitlich variable Schlagrate verursacht. Wir können also, je nachdem wann und welche Anlagen wir beproben, unterschiedliche Schlüsse über die durchschnittliche Schlagrate ziehen.

¹ Es handelt sich hier um die „Restvarianz“, d.h. die Varianz, welche nicht durch Vorhersagevariablen wie Habitat, Zeit, Wetter u.a. erklärt werden kann. Diese ist bei Schlagopferzahlen naturgemäss meist gross.

² Angesichts der meist unterschätzten Unsicherheiten ist das Wort „Vertrauensintervall“ für das Intervall, das die statistische Unsicherheit beschreibt, irreführend: wir haben eigentlich kein Vertrauen in dieses Intervall, auch wenn es idealerweise die beste Schätzung der Unsicherheit ist, die wir basierend auf der aktuellen Datenlage haben. Deswegen haben Statistiker*innen vorgeschlagen, den Begriff „Vertrauensintervall“ durch „Unsicherheitsintervall“ oder „Kompatibilitätsintervall“ zu ersetzen (Gelman & Greenland, 2019; Greenland, 2019; Trafimow et al., 2018).

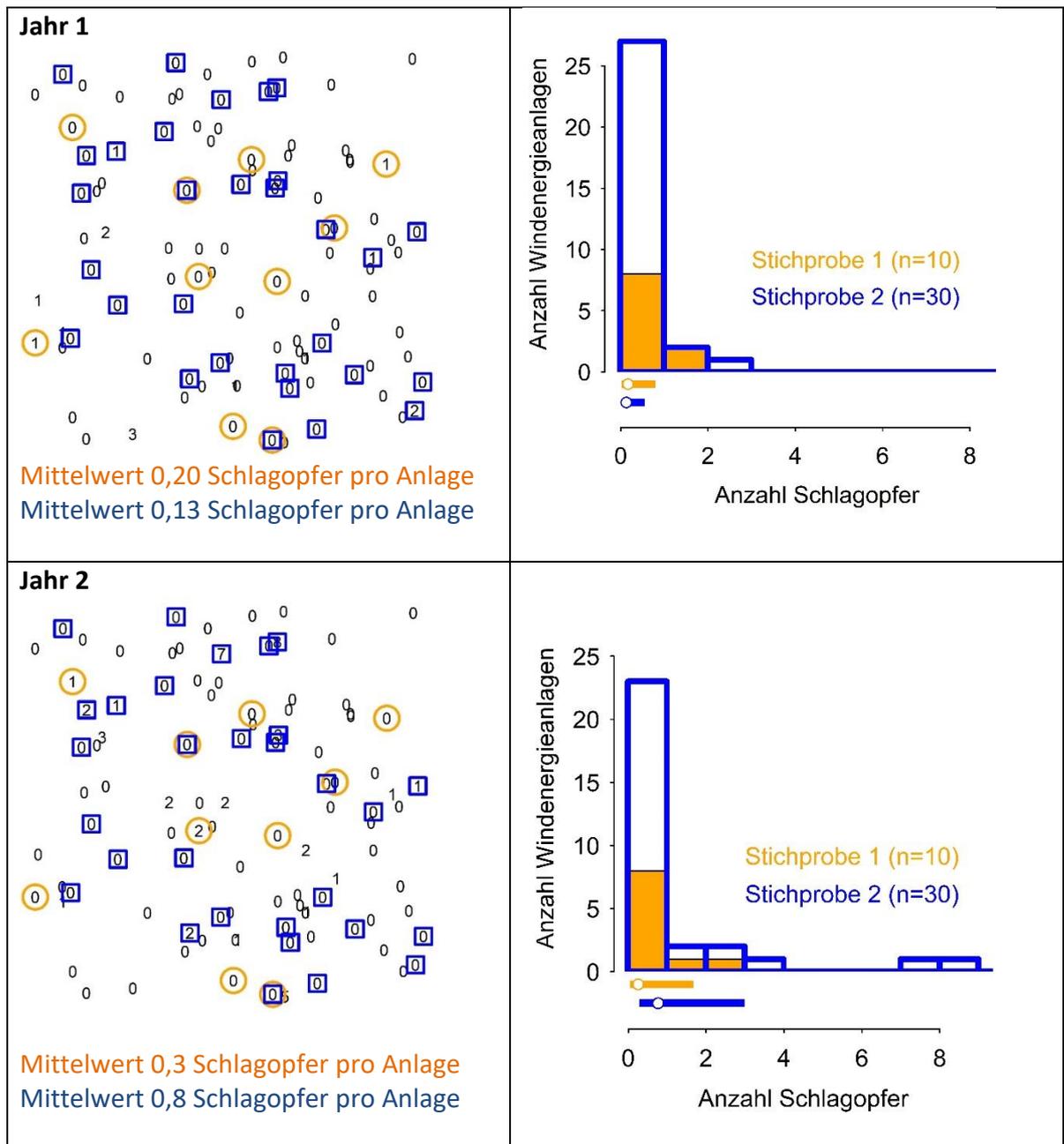


Abbildung 2. Illustration zur Varianz in der Schlagrate und zum Einfluß der Stichprobengröße auf die Unsicherheit der geschätzten durchschnittlichen Schlagrate. Die schwarzen Zahlen in der linken Abbildung sind hypothetische Schlagopferzahlen an im Raum verteilten Windenergieanlagen. Sie wurden als Zufallszahlen aus einer Negativbinomialverteilung gezogen, die pro Jahr eigene Mittelwerte hatten (zeitliche Varianz) jedoch keine räumliche Varianz (jede Anlage hatte denselben langjährigen Mittelwert). Die orangen Kreise zeigen eine Stichprobe aus 10, die blauen Vierecke eine Stichprobe aus 30 Windenergieanlagen, die in zwei Jahren beprobt wurden. Die Abbildungen rechts zeigen pro Jahr und Stichprobe die Verteilung der Daten (Histogramme) sowie darunter die Mittelwerte mit 95% Unsicherheitsintervallen.

Mit steigender Datenmenge wird die Unsicherheit in der geschätzten durchschnittlichen Schlagrate immer kleiner, d.h. die Präzision wird immer grösser. Hingegen bleibt die naturgegebene Varianz in den Schlagraten unabhängig von der Stichprobengröße und der Meßtechnik gleich groß. Eine große natürliche Varianz bedeutet, daß die effektive Schlagrate an einer Windenergieanlage in einem Jahr stark von der geschätzten durchschnittlichen Schlagrate abweichen kann. Im fiktiven Beispiel wurden an einer

Windenergieanlage im zweiten Jahr 8 Schlagopfer gefunden, obwohl aus allen Datensätzen eine durchschnittliche Schlagrate von unter 1 pro Jahr ermittelt wurde. Um vorherzusagen, wie stark die Schlagraten an einer Windenergieanlage in einem Jahr von der durchschnittlich erwarteten Schlagrate abweichen kann, muß die Varianz, d.h. die statistische Verteilung der Schlagopferzahlen um ihren Mittelwert herum bekannt sein. Daraus lassen sich sogenannte Vorhersageintervalle berechnen, die angeben, in welche Wertebereiche Schlagopferzahlen am ehesten zu liegen kommen können. Im Unterschied zu Unsicherheitsintervallen werden, falls die Modellstruktur nicht verfeinert wird, die Vorhersageintervalle nicht kleiner, wenn sich die Datengrundlage verbessert. Die Vorhersageintervalle beschreiben, wie die Schlagraten natürlicherweise streuen, während die Unsicherheitsintervalle angeben, wie präzise die durchschnittliche Schlagrate geschätzt werden kann.

Im obigen Beispiel habe ich als Zielvariable die Schlagrate verwendet. Die unterschiedlichen Rollen von Varianz einerseits und Unsicherheit in der Schätzung des Mittelwerts andererseits gibt es auch bei anderen Zielvariablen, wie dem Schlagrisiko oder der Populationswachstumsrate. Beides, Varianz und Unsicherheit, beeinflussen die Verknüpfung der verschiedenen in Abbildung 1 dargestellten Größen.

Möglichkeiten der Probabilistik zur Signifikanzbeurteilung des Einflusses auf das individuelle Sterberisiko und auf die Populationsentwicklung

Im Rahmen des Bewilligungsverfahrens für ein Windenergieprojekt wird eine Beurteilung verlangt, ob das individuelle Sterberisiko durch das Projekt „signifikant“ erhöht wird, und ob „signifikant“ negative Einflüsse auf die Populationsentwicklung zu erwarten sind. Diese Kriterien werfen mindestens drei Fragen auf: 1. Wie stark erhöht das Windenergieprojekt das individuelle Sterberisiko? 2. Welchen Einfluß hat dieses zusätzliche Sterberisiko auf die Entwicklung der Populationen? 3. Ab wann wird die Erhöhung des Sterberisikos bzw. der Einfluß auf das Populationswachstum als „signifikant“ erachtet?

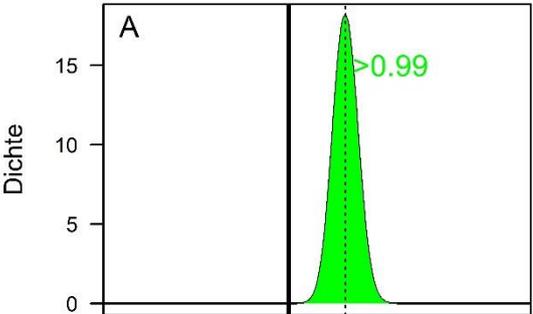
Probabilistik kann zu den ersten beiden Fragen Wahrscheinlichkeitsaussagen machen. Die Frage 3 nach der Signifikanz kann durch Probabilistik nicht beantwortet werden. Um zu beurteilen, ob negative Einflüsse „signifikant“ im Sinne von „bedeutsam“ sind, müssen die Konsequenzen des geschätzten Einflusses der Windenergieanlage mit einer Zielvorstellung verglichen werden. Die Zielvorstellung wird durch die Politik, z.B. in Gesetzen, oder von Interessensgruppen vorgegeben. Nachfolgend erläutere ich zuerst, welche Wahrscheinlichkeitsaussagen die Probabilistik zu Schlagrisiko und zur Schlagrate machen kann. Danach diskutiere ich Möglichkeiten, für die Signifikanzbeurteilung politische Meinungen mit einzubeziehen.

Empirisch ermittelte Schlagraten wie auch mechanistische Modelle können Aufschluß über die Erhöhung des individuellen Sterberisikos geben. Um die Konsequenzen der Erhöhung des individuellen Sterberisikos auf die Populationsentwicklung zu schätzen, können verschiedene Methoden der Populationsmodellierung verwendet werden (Bellebaum et al., 2013; Diffendorfer et al., 2017; Schaub, 2012).

Populationsmodellierungen resultieren in den meisten Fällen in der Schätzung der Populationswachstumsrate. Die Populationswachstumsrate entspricht dem (multiplikativen) Faktor, um den sich die Population jährlich verändert.

Wie beim oben beschriebenen fiktiven Beispiel für die Schätzung der Schlagrate, ergeben auch Populationsmodellierungen keine genauen Werte für die Wachstumsrate, sondern eine Schätzung mit statistischer Unsicherheit, die durch eine Wahrscheinlichkeitsverteilung abgebildet werden kann. Die für die Schätzung verwendeten probabilistischen Modelle sind eine Vereinfachung des in der Natur ablaufenden Prozesses. Die Modell-Annahmen reduzieren die große, komplizierte Welt auf eine kleine, einfache und berechenbare Welt. Die berechneten Wahrscheinlichkeiten gelten deshalb nur in dieser kleinen, vereinfachten Welt. Als Folge ist die in Studien ausgewiesene statistische Sicherheit überschätzt. Zudem beruhen die Schätzungen auf Daten in Stichproben, die aufgrund der natürlichen Varianz nicht um den wahren Parameterwert zentriert sein müssen. Um diese zusätzlichen und nicht meßbaren Unsicherheiten möglichst gering zu halten, brauchen wir eine große Datenmenge sowie realistische Modelle. Zentrale und zugängliche Datenbanken, welche Daten aus vielen Studien sammeln, könnten die Datengrundlage für probabilistische Berechnungen erhöhen und zur Validierung und Verbesserung der Modelle verwendet werden. Bei Entscheiden, die auf statistischen Resultaten beruhen, sollten wir uns immer bewußt sein, daß die statistische Unsicherheit unterschätzt ist, und entsprechend vorsichtig entscheiden.

In Abbildung 3 habe ich zu vier verschiedenen möglichen Resultaten für die Schätzung einer Populationswachstumsrate je drei Aussagen formuliert. Die ersten beiden Aussagen sind jene, die direkt aus der Probabilistik resultieren und die in der Praxis so verwendet werden, obwohl wir wissen, daß sie, wenn wir sie auf die komplexe reale Welt anwenden, möglicherweise nicht genau so zutreffen. Die dritte Aussage kommuniziert am besten, daß das Resultat nur in der kleinen, berechenbaren Modellwelt und für einen bestimmten Datensatz gilt.

<p>Resultat: A-posteriori Verteilung der geschätzten mittleren Populationswachstumsrate basierend auf den Daten und dem Modell</p>	<p>Mögliche Aussagen. Mit „Populationswachstumsrate“ ist hier immer die mittlere Populationswachstumsrate gemeint. Die A-posteriori Verteilungen zeigen die statistische Unsicherheit der geschätzten Populationswachstumsrate und <i>nicht</i> die Streuung der Populationswachstumsrate (z.B. von Jahr zu Jahr).</p>
	<p>„Mit großer Wahrscheinlichkeit zeigt das Populationswachstum erwünschte Werte.“ „Eine unerwünschte Populationsentwicklung können wir mit großer Wahrscheinlichkeit ausschließen.“ „Populationswachstumsraten, die mit den Daten und den Annahmen für die Berechnung am kompatibelsten sind, liegen im erwünschten Bereich.“</p>

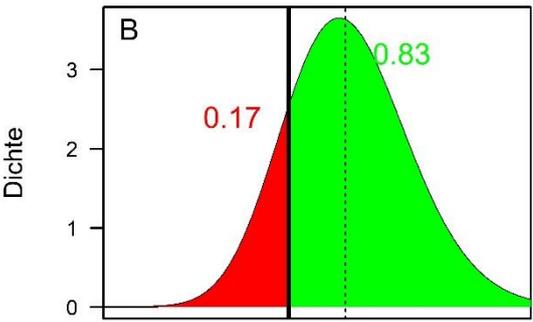
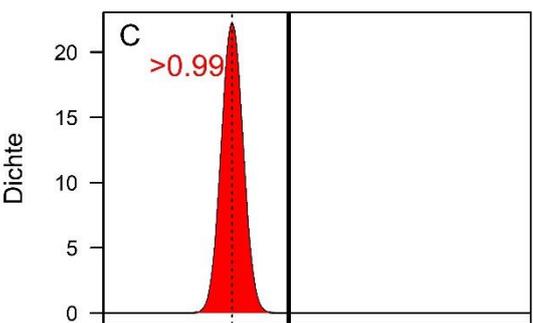
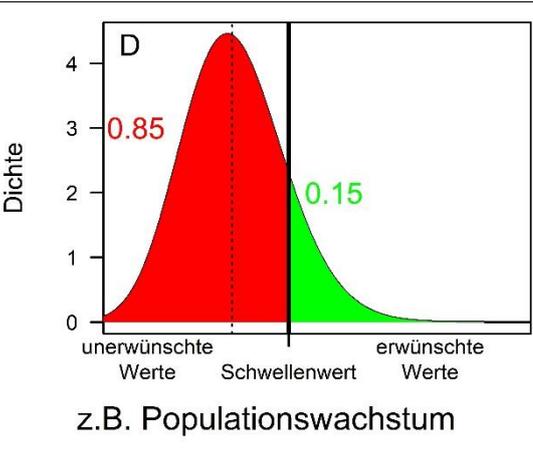
	<p>„Mit einer Wahrscheinlichkeit von 0.83 liegt das Populationswachstum im erwünschten Bereich.“</p> <p>„Mit einer Wahrscheinlichkeit von 0.17 zeigt das Populationswachstum unerwünschte Werte.“</p> <p>„17% der Populationswachstumsraten, die mit den Daten und den Modellannahmen am kompatibelsten sind, liegen im unerwünschten, 83% im erwünschten Bereich.“</p>
	<p>„Wir können ein erwünschtes Populationswachstum mit großer Wahrscheinlichkeit ausschließen.“</p> <p>„Das Populationswachstum zeigt mit großer Wahrscheinlichkeit unerwünschte Werte.“</p> <p>„Populationswachstumsraten, die mit den Daten und den Annahmen für die Berechnung am kompatibelsten sind, liegen im unerwünschten Bereich.“</p>
	<p>„Das Populationswachstum zeigt mit einer Wahrscheinlichkeit von 0.15 erwünschte Werte.“</p> <p>„Mit einer Wahrscheinlichkeit von 0.85 liegt das Populationswachstum im unerwünschten Bereich.“</p> <p>„85% der Populationswachstumsraten, die mit den Daten und den Modellannahmen am kompatibelsten sind, liegen im unerwünschten, 15% im erwünschten Bereich.“</p>

Abbildung 3. Vier mögliche Resultate (A-posteriori Verteilungen) einer statistischen Schätzung der Populationswachstumsrate einer betroffenen Tierart basierend auf probabilistischen Methoden. Dargestellt sind je eine präzise (A, C) und eine unpräzise (B, D) Schätzung der Populationswachstumsraten, die mehrheitlich erwünschte (A, B), resp. unerwünschte (C, D) Werte zeigen. Beim Schwellenwert handelt es sich um den Relevanz-Schwellenwert (siehe Text). In der rechten Spalte sind zu jedem Resultat drei verschiedene Formulierungsvarianten angegeben.

Wie entscheiden wir nun, ob der Einfluß der Windenergieanlage „signifikant“ ist? Der Begriff „Signifikanz“ führt in der Wissenschaft oft zu Verwirrung. Unter der im Rahmen von Bewilligungsverfahren geforderten Signifikanz des negativen Einflusses auf Individuen und Populationen ist kaum die statistische Signifikanz gemeint, sondern vielmehr, wie schwerwiegend der Einfluß beurteilt wird, also wie relevant der Einfluß ist. Die (klassische) *statistische* Signifikanz hat nichts mit Relevanz zu tun (Cohen, 1994; Field et al., 2004; Stahel, 2021). Für die Berechnung der *statistischen* Signifikanz werden die Konsequenzen (bzw. die Bedeutung) eines Effekts nicht berücksichtigt. Die *statistische* Signifikanz vermischt Effektgröße und statistische Unsicherheit, womit ihre

Interpretation, ohne zusätzliche Information wie Effektgröße, praktisch verunmöglicht wird. Unzählige Stimmen haben auf die Gefahr aufmerksam gemacht, Entscheide zu fällen, die teuer zu stehen kommen können, wenn sie alleine auf der herkömmlichen statistischen Signifikanz beruhen (z.B. Amrhein et al., 2017, 2019; Field et al., 2004; Ridley et al., 2007; Stahel, 2021; Trafimow et al., 2018).

Es gibt alternative Möglichkeiten, Entscheide basierend auf mit statistischer Unsicherheit behaftetem Wissen zu fällen. Ich skizziere hier zwei solche Möglichkeiten: 1) Relevanz- und Evidenz-Schwellenwerte und 2) Kostenfunktionen.

Der Entscheid, ob ein signifikanter (im Sinn von bedeutsamer) Einfluß vorhanden ist oder nicht, kann auf einer Kombination von Relevanz und Evidenz basieren (Stahel, 2021). Dabei müssen zwei Fragen geklärt werden. Erstens: Ab welchem Wert (d.h. ab welcher Effektgröße) beurteilen wir den Einfluß als *relevant*, also als bedeutend (Relevanz-Schwellenwert)? Allerdings, wie oben gezeigt, erhalten wir von der Statistik nicht nur einen einzelnen Wert, welcher mit dem Relevanz-Schwellenwert verglichen werden könnte, sondern eine Wahrscheinlichkeitsaussage (Abbildung 3). Daraus ergibt sich die zweite Frage: Ab welcher Evidenzstärke, d.h. ab welcher statistischen *Sicherheit*, gehen wir davon aus, daß der Relevanz-Schwellenwert tatsächlich überschritten wird (Evidenz-Schwellenwert)? Der Relevanz-Schwellenwert basiert auf Konsequenzen des Einflusses, z.B. Bedeutung des lokalen Verschwindens einer Population für die Erhaltung der Art, auf Zielvorstellungen über zukünftige Populationsgrößen, oder auf ethischen Grundsätzen, z.B. wie stark ein individuelles Sterberisiko erhöht werden darf (Abbildung 1). Um im zweiten Schritt den Evidenz-Schwellenwert festzulegen, müssen wir uns überlegen, wie schwerwiegend es ist, den falschen Entscheid zu fällen. Falsche Entscheide wären, wenn ein Projekt gebaut wird, obwohl damit der Relevanz-Schwellenwert überschritten wird. Ebenso wäre falsch, nicht zu bauen, wenn der Bau eigentlich keine Überschreitung des Relevanz-Schwellenwerts verursachen würde. Solche Abwägungen könnten je nach Gefährdungstatus einer Art unterschiedlich ausfallen. Für eine stark bedrohte Art möchten wir negative Effekte auf das Populationswachstum mit großer Wahrscheinlichkeit ausschließen können. Für eine solche Art würden wir bereits bei einem geringen Hinweis auf mögliche negative Effekte Schutzmaßnahmen ergreifen, d.h. wir würden nur bei Resultat A in Abbildung 3 ein Projekt bewilligen, nicht aber bei Resultat B (oder C oder D). Für eine weitverbreitete Art würden wir eine stärkere Evidenz (statistische Sicherheit) für negative Effekte verlangen (Resultat C und D), bevor ein Projekt gestoppt wird.

Ich sehe es als große Herausforderung an, die beiden Schwellenwerte so zu legen, daß sie den gesetzlichen und politischen Rahmenbedingungen entsprechen. Hilfreich wäre es, die Konsequenzen verschiedener Entscheide (z.B. ob ein Windprojekt gebaut wird oder nicht) für die Erhaltung der Arten, für das Klima und für die Energieversorgungssicherheit vorhersagen zu können. Für die vereinfachte Modell-Welt können wir vorhersagende Modelle erstellen, um die Konsequenzen der verschiedenen Handlungen auf unterschiedliche Aspekte unseres Lebens und unserer Umwelt abzuschätzen. Solche Vorhersagen können durch die politische Meinung bewertet werden, wozu sogenannte „Kosten“-Funktionen nützlich sein können (Abbildung 4).

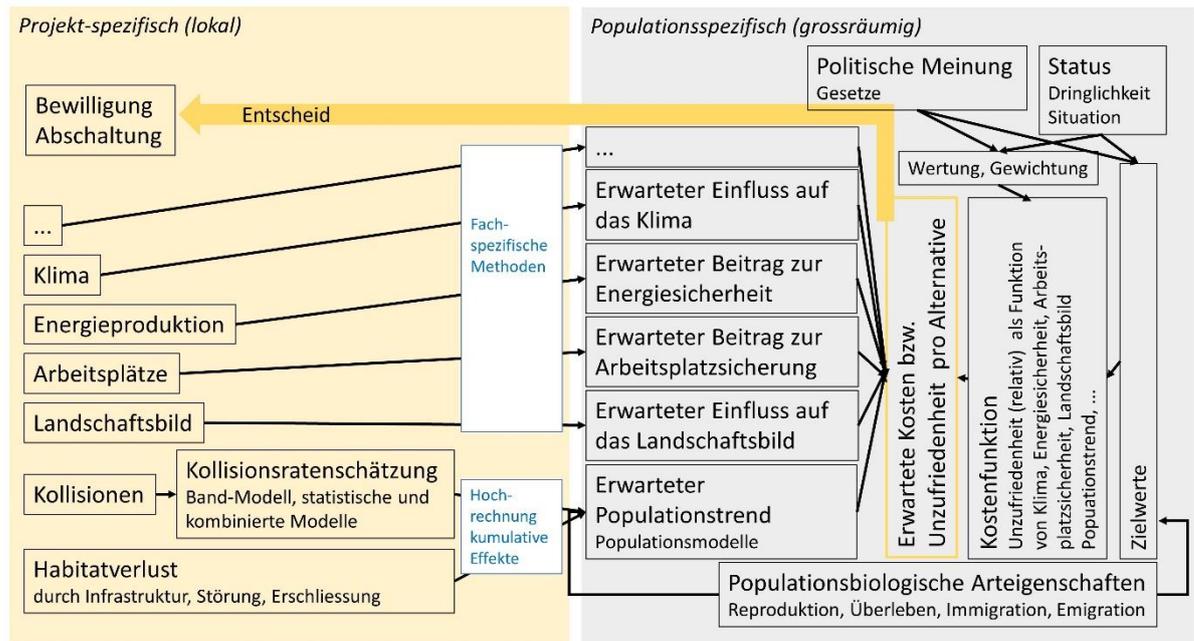


Abbildung 4. Möglicher Weg von der Kollisionsratenschätzung zum Management-Entscheid mittels Methoden der strukturierten Entscheidungsfindung über sogenannte Kostenfunktionen, welche die Unzufriedenheit mit den verschiedenen Szenarien formalisieren.

Kostenfunktionen quantifizieren ökonomische, ökologische und soziale „Kosten“. Die Kosten können in Euro angegeben werden oder auch relativ, also einheitslos. Wir könnten Kostenfunktionen auch „Unzufriedenheitsfunktionen“ nennen. Ist ihr Wert für eine Situation doppelt so hoch wie für die andere, sind wir mit erster Situation doppelt so unzufrieden wie mit letzterer. Unzufriedenheitsfunktionen werden an Workshops erarbeitet, an denen sich Vertreter der verschiedenen Interessensgruppen beteiligen (z.B. Bunnefeld et al. 2017). Im fiktiven Beispiel (Abbildung 5) stufen wir die Summe der ökonomischen, ökologischen und sozialen Kosten bzw. Unzufriedenheit ohne Windenergieprojekt doppelt so hoch ein wie mit Windenergieprojekt: Die Funktionswerte für das Szenario ohne Windenergieprojekt sind doppelt so hoch wie mit Windenergieprojekt. Im fiktiven Beispiel zeigen wir die Kostenfunktionen mit und ohne Windenergieprojekt in Abhängigkeit des Populationswachstums und wir deklarieren, dass die Kosten unterhalb des Relevanz-Schwellenwerts linear³ zunehmen. Aus den Kostenfunktionen können wir ablesen, welche Reduktion in der Populationswachstumsrate wir gleich werten wie die Summe der übrigen Kosten, die uns entstehen, wenn wir das Windenergieprojekt nicht bauen können. Um zu entscheiden, unter welchem Szenarium am wenigsten Kosten anfallen – oder wir am wenigsten unzufrieden sind – verrechnen wir die Kostenfunktion mit der geschätzten Populationswachstumsrate unter Berücksichtigung deren statistischer Unsicherheit⁴, je für das Szenarium mit und ohne Windenergieprojekt. Im fiktiven Beispiel erhalten wir erwartete Kosten von 5.1 mit und 3.8 ohne Windenergieprojekt (weisse Punkte in Abbildung 5 oben). Somit würden wir uns in diesem fiktiven Beispiel gegen den Bau des Windenergieprojekts entscheiden, wenn wir die zukünftig erwarteten Kosten klein halten möchten. Die Kostenfunktionen können sich zwischen Interessensgruppen

³ Eine exponentielle Zunahme der Kosten mit negativeren Populationswachstumsraten würde vermutlich der Realität besser entsprechen. Das Beispiel hier dient lediglich dazu, das Prinzip zu erklären.

⁴ Integration der Kostenfunktion über die Wahrscheinlichkeitsverteilung der Populationswachstumsrate

unterscheiden. In solchen Fällen kann die Darstellung der verschiedenen Kostenfunktionen Transparenz in den Entscheidungsprozeß bringen, indem sie visualisieren, für welchen Entscheid welche Interessen wie stark gewichtet werden.

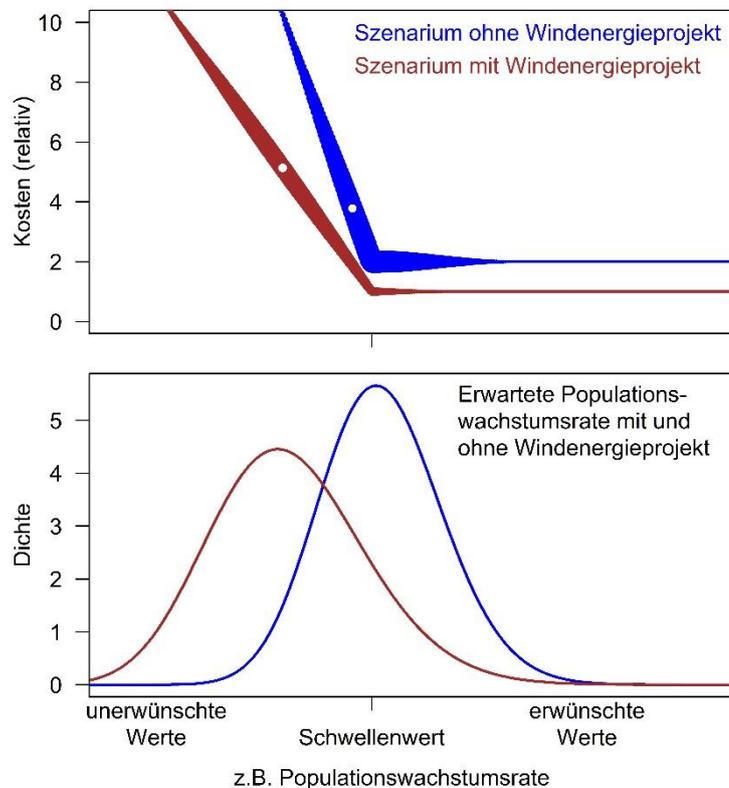


Abbildung 5. Fiktive Kostenfunktionen in Abhängigkeit der Populationswachstumsrate der untersuchten Tierart (oben) für die Szenarien mit und ohne Windenergieprojekt sowie die Wahrscheinlichkeitsverteilung der geschätzten Populationswachstumsraten (unten). Die untere x-Achse gilt auch für die obere Grafik. Die Liniendicke der Kostenfunktionen ist proportional zur Wahrscheinlichkeitsverteilung des Populationswachstums der Tierart (d.h. proportional zur Höhe der Linien in der unteren Grafik). Die weißen Punkte in den Kostenfunktionen zeigen für jede Entscheidung den Mittelwert der erwarteten Kosten. Der Abbildung liegen keine Daten zugrunde, sie dient der Illustration des Prinzips. Für die Erarbeitung von Kostenfunktionen sind aufwendige Workshops, Modellierungen und Befragungen von Experten und Interessengruppen nötig.

Kostenfunktionen können politische Interessen formalisieren, d.h. sichtbar machen, und für Transparenz bei Entscheidungen sorgen. Sie helfen, Entscheide so zu fällen, daß sie dem politischen Interesse am besten entsprechen, trotz unsicheren Entscheidungsgrundlagen (Runge et al., 2020).

Diskussion und Fazit

Bei solider Datengrundlage kann Probabilistik helfen, zu verstehen, welchen Einfluß Windenergieprojekte auf Arten haben. Zu wissen, in welchem Bereich Schlagrisiken und -raten in etwa liegen und welche Konsequenzen daraus für bestimmte Populationsparameter, wie Sterberate oder Wachstumsrate entstehen, kann Diskussionen versachlichen, indem Spekulationen vorgebeugt werden. Zudem liefert die Probabilistik Zutaten zu Vorhersage-Modellen von verschiedenen Aspekten unserer sozialen, ökologischen und ökonomischen Umwelt, die schlußendlich ermöglichen könnten, heute die bestmöglichen Entscheide zu fällen. Ich sehe jedoch auch verschiedene Risiken, die ich hier benenne und diskutiere.

Fehlinterpretation von Wahrscheinlichkeitsaussagen: Wahrscheinlichkeitsaussagen sind schwierig zu verstehen. Die Aussage, daß die mittlere Schlagrate mit einer Wahrscheinlichkeit von 30% über 5 liegt, bedeutet nicht, daß an 30% der Tage Schlagraten von über 5 auftreten. Der Unterschied zwischen statistischer Unsicherheit eines Parameters und der Varianz einer Variable ist fundamental, aber schwierig zu verstehen und zu kommunizieren, v.a. wenn beides in Wahrscheinlichkeiten (%) angegeben wird. So werden oft Unsicherheitsintervalle fälschlicherweise als Vorhersageintervalle interpretiert. Eine Verwechslung dieser beiden Größen führt zu Fehlentscheidungen. Um sicherzustellen, daß probabilistische Berechnungen richtig interpretiert werden, empfehle ich, professionelle Biostatistiker*innen während des ganzen Prozesses bis zur Entscheidung einzubeziehen.

Übertragbarkeit vom Mittelwert einer Studie auf andere Windenergieprojekte und über die Studienzeit hinaus. Ermittelte Parameterwerte sind nur dann auf einen anderen Standort oder in die Zukunft übertragbar, wenn das System, auf welches die Parameterwerte angewendet werden, sich nicht von jenem System unterscheidet, unter welchem die Daten gesammelt und das Modell entwickelt wurde. Wird in der Umgebung eine Hecke gefällt, kann sich das Flugverhalten von Greifvögeln oder Fledermäusen ändern, was wiederum die Schlagrate verändert. Probabilistische Methoden können deshalb lokale Gutachten wie auch ein Monitoring der Schlagrate und der Bestände sensibler Arten nicht ersetzen.

Das Modell macht (zu) stark vereinfachende Annahmen: Die berechneten Parameterwerte gelten in der vereinfachten Modell-Welt. Wie gut diese kleine Welt die komplexe reale Welt abbildet, können wir nur bedingt überprüfen. Neben den Standard-Methoden der Modelldiagnostik ist eine breite Diskussion der Modellannahmen wichtig, wie auch die Anwendung alternativer Modelle, die andere Annahmen machen. Diffendorfer et al. (2017) schlagen beispielsweise neben der Populationswachstumsrate weitere Indizes vor, die je auf unterschiedlichen ökologischen Aspekten, z.B. der Verbreitung oder dem Spezialisierungsgrad, basieren.

Fokussierung auf Kollisionen: Wenn ein großer Aufwand betrieben wird, um Schlagrisiko und -rate zu schätzen, besteht die Gefahr, daß andere Aspekte, die in einer Umweltverträglichkeitsprüfung berücksichtigt werden sollten, vernachlässigt oder sogar vergessen gehen.

Ausschaltung des Dialogs: Entscheidungen werden besser akzeptiert, wenn die betroffenen Personen ihr Interesse kundtun können und sehen, wie ihr Interesse für den Entscheid berücksichtigt worden ist. Wird der Entscheid einem komplizierten Algorithmus überlassen, wird er intransparent. Methoden der strukturierten Entscheidungsfindung, von denen ich hier ein Beispiel skizziert habe, finde ich deshalb vielversprechend, weil sich dabei die verschiedenen Interessengruppen an der Entscheidungsfindung aktiv beteiligen und Daten und Probabilistik Grundlagen für die Entscheidungsfindung liefern.

Probabilistik kann also verwendet werden, um aktuelles Wissen über relevante Parameter wie Sterbe- oder Populationswachstumsraten zu quantifizieren sowie um Konsequenzen der Veränderung dieser Parameter abzuschätzen. Solche Berechnungen

liefern wichtige Grundlagen für Entscheide im Arten- und Klimaschutz. Sie können jedoch beständiges Monitoring sowie Diskussionen unter den verschiedenen Interessensgruppen und den Einbezug von Artkenner*innen in der Entscheidungsfindung nicht ersetzen.

Dank: Valentin Amrhein, Pius Korner und Werner Stahel danke ich herzlich für wertvolle Anregungen zum Manuskript. Moritz Merker und Jannis Liedtke gaben Inputs zur Abbildung 1.

Referenzen

- Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature*, *567*, 305–307.
- Amrhein, V., Korner-Nievergelt, F., & Roth, T. (2017). The earth is flat ($p > 0.05$): Significance thresholds and the crisis of unreplicable research. *PeerJ*, *7*. <https://doi.org/10.7717/peerj.3544>
- Aschwanden, J., Stark, H., Peter, D., Steuri, T., Schmid, B., & Liechti, F. (2018). Bird collisions at wind turbines in a mountainous area related to bird movement intensities measured by radar. *Biological Conservation*, *220*, 228–236. <https://doi.org/10.1016/j.biocon.2018.01.005>
- Band, W., Madders, M., & Whitfield, D. P. (2007). Developing field and analytical methods to assess avian collision risk at wind farms. In M. de Lucas, Janss, G. F. E., & M. Ferrer (Eds.), *Birds and wind farms: risk assessment and mitigation* (pp. 259–275). Quercus.
- Bastos, R., Pihnanco, A., Santos, M., Fernandes, R. F., Vicente, J. R., Morinha, F., Honrado, J. P., Travassos, P., Barros, P., & Cabral, J. A. (2015). Evaluating the regional cumulative impact of wind farms on birds: how can spatially explicit dynamic modelling improve impact assessments and monitoring? *Journal of Applied Ecology*, online. <https://doi.org/10.1111/1365-2664.12451>
- Bellebaum, J., Korner-Nievergelt, F., Dürr, T., & Mammen, U. (2013). Wind turbine fatalities approach a level of concern in a raptor population. *Journal for Nature Conservation*, *21*(6). <https://doi.org/10.1016/j.jnc.2013.06.001>
- Bunnefeld, N., Nicholson, E., & Milner-Gulland, E. J. (2017). *Decision-Making in Conservation and Natural Resource Management: Models for Interdisciplinary Approaches* (N. Bunnefeld, E. Nicholson, & E. J. Milner-Gulland, Eds.). Cambridge University Press.
- Chamberlain, D. E., Rehfisch, M. R., Fox, A. D., Desholm, M., & Anthony, S. J. (2006). The effect of avoidance rates on bird mortality predictions made by wind turbine collision risk models. *Ibis*, *148*(SUPPL. 1), 198–202. <https://doi.org/10.1111/j.1474-919X.2006.00507.x>
- Cohen, J. (1994). The earth is round ($p > 0.05$). *American Psychologist*, *49*, 997–1003.
- Diffendorfer, J. E., Beston, J. A., Merrill, M. D., Stanton, J. C., Corum, M. D., Loss, S. R., Thogmartin, W. E., Johnson, D. H., Erickson, R. A., & Heist, K. W. (2017). A Method to Assess the Population-Level Consequences of Wind Energy Facilities on Bird and Bat Species. In J. Köppel (Ed.), *Wind Energy and Wildlife Interaction* (pp. 65–76).
- Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R., & Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical

- thresholds. *Ecology Letters*, 7(8), 669–675. <https://doi.org/10.1111/j.1461-0248.2004.00625.x>
- Gelman, A., & Greenland, S. (2019). Are confidence intervals better termed $\{\backslash\text{dq}\}$ uncertainty intervals $\{\backslash\text{dq}\}$? *BMJ (Clinical Research Ed.)*, 366, l5381. <https://doi.org/10.1136/bmj.l5381>
- Greenland, S. (2019). Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. *American Statistician*, 73(sup1), 106–114. <https://doi.org/10.1080/00031305.2018.1529625>
- Grünkorn, T., Blew, J., Coppack, T., Krüger, O., Nehls, G., Potiek, A., Reichenbach, M., von Rönn, J., Timmermann, H., & Weitekamp, S. (2016). *Ermittlung der Kollisionsraten von (Greif)Vögeln und Schaffung planungsbezogener Grundlagen für die Prognose und Bewertung des Kollisionsrisikos durch Windenergieanlagen (PROGRESS). Schlussbericht zum durch das Bundesministerium für Wirtschaft und Energie (Issue FKZ 0325300A-D).*
- Huso, M., Dalthorp, D., & Korner-Nievergelt, F. (2017). Statistical principles of post-construction fatality monitoring. In M. R. Perrow (Ed.), *Wildlife and Wind Farms, Conflicts and Solutions. Volume 2 Onshore: Monitoring and Mitigation* (pp. 84–102). Pelagic Publishing.
- Kerns, J., Erickson, W. P., Arnett, E. B., & [Nachname nicht vorhanden]. (2005). Bat and bird fatality at wind energy facilities in Pennsylvania and West Virginia. In E. B. Arnett (Ed.), *Relationships between bats and wind turbines in Pennsylvania and West Virginia: An assesment of fatality search protocols, patterns of fatality, and behavioural interactions with wind turbines. A final report prepared for the bats and wind energy cooper* (pp. 24–95). Bat Conservation International.
- Masden, E. A., & Cooc, A. S. C. P. (2016). Avian collision risk models for wind energy impact assessments. *Environmental Impact Assessment Review*, 56.
- McNeish, D. M., & Stapleton, L. M. (2016). The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. *Educational Psychology Review*, 28(2), 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- Niermann, I., Brinkmann, R., Korner-Nievergelt, F., & Behr, O. (2011). Systematische Schlagopfersuche - Methodische Rahmenbedingungen, statistische Analyseverfahren und Ergebnisse. In R. Brinkmann, O. Behr, I. Niermann, & M. Reich (Eds.), *Entwicklung von Methoden zur Untersuchung und Reduktion des Kollisionsrisikos von Fledermäusen an Onshore-Windenergieanlagen.* (pp. 40–114). Cuvillier Verlag.
- Ridley, J., Kolm, N., Freckelton, R. P., & Gage, M. J. G. (2007). An unexpected influence of widely used significance thresholds on the distribution of reported P-values. *Journal of Evolutionary Biology*, 20, 1082–1089.
- Runge, M. C., Converse, S. J., Lyons, J. E., & Smith, D. R. (2020). *Structured Decision Making*. Johns Hopkins University Press.
- Schaub, M. (2012). Spatial distribution of wind turbines is crucial for the survival of red kite populations. *Biological Conservation*, 155, 111–118.
- Schaub, T., Klaassen, R. H. G., Bouten, W., Schlaich, A. E., & Koks, B. J. (2020). Collision risk of Montagu's Harriers *Circus pygargus* with wind turbines derived from high-resolution GPS tracking. *Ibis*, 162(2), 520–534. <https://doi.org/10.1111/ibi.12788>
- Stahel, W. A. (2021). New relevance and significance measures to replace p-values. *PlosOne*, 16, e0252991. <https://doi.org/10.1371/journal.pone.0252991>

Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bilgiç, Y. K., Bono, R., Bradley, M. T., Briggs, W. M., Cepeda-Freyre, H. A., Chaigneau, S. E., Ciocca, D. R., Correa, J. C., Cousineau, D., de Boer, M. R., Dhar, S. S., Dolgov, I., Gómez-Benito, J., Grendar, M., ... Marmolejo-Ramos, F. (2018). Manipulating the alpha level cannot cure significance testing. *Frontiers in Psychology, 9*(MAY). <https://doi.org/10.3389/fpsyg.2018.00699>