

## **Wunsch und Wirklichkeit – Wie lässt sich ein GVO-Monitoring kosteneffizient realisieren?**

CHRISTOPH BÜHLER

### **Zusammenfassung**

Ein Überwachungsprogramm, das auch nur annähernd alle postulierten, hypothetischen Wirkungen transgener Organismen auf die Umwelt erfasst, ist weder vernünftig noch finanzierbar. Selbst für das Monitoring unvorhergesehener Umwelteffekte von GVO (General Surveillance) ist es daher unvermeidlich, eine auf das spezifische Event zugeschnittene, gut begründete Auswahl der wichtigsten Zielgrößen zu treffen. Das Formulieren möglichst konkreter Anforderungen an die GVO-Überwachung betreffend Untersuchungsgebiet, Raumbezug der Messungen, Indikatorenwahl und an die Größe festzustellender Effekte ist deshalb der erste und vielleicht wichtigste Schritt zur Kosteneffizienz. Je präziser diese Anforderungen feststehen, desto besser lässt sich in einem zweiten Schritt ein passendes Monitoring-Programm darauf zuschneiden. Dort geht es in erster Linie um den optimierten Einsatz von Ressourcen, um die gewünschten Aussagen mit möglichst geringem Aufwand zu liefern. Hier spielen statistische Methoden, insbesondere die Power-Analyse, eine wichtige Rolle.

**Schlagerworte:** General Surveillance, Monitoring, Probenahme-Design, Kosteneffizienz, Power-Analyse

### **Summary**

A monitoring scheme that covers roughly all of the proposed hypothetical effects of GMO on the environment is neither reasonable nor can it ever be afforded. Even for general surveillance, which aims at unexpected effects it is unavoidable to focus on the most important elements for a specific event. Therefore, the criteria for determining the study area, the scale at which measurements are taken, the choice of indicators and the size of effects to be detected must be defined very clearly. This is the first and perhaps most important step towards cost-efficiency of a future monitoring program. The more precise these requirements can be set up, the easier it is to take the second step which is the design of a corresponding monitoring scheme itself. There, cost-efficiency is addressed directly by identifying the minimum effort to attain the required precision as stated. Here, statistical methods and power analysis in particular is an important tool.

## **1 Umweltveränderungen nachweisen – aber welche denn?**

Die wissenschaftliche Literatur enthält dutzende von denkbaren Wirkungswegen, wie transgene Organismen auf unsere Umwelt einwirken und sie verändern könnten (siehe z. B. SANVIDO et al. 2003; MARQUARD & DURKA 2005; CHAPMAN & BURKE 2006 und Referenzen in diesen Studien). Es ist deshalb unumgänglich, die Auswahl möglicher Schutzziele und Indikatoren für ein GVO-Monitoring fallweise auf die ökologisch und politisch bedeutendsten und aussagekräftigsten Inhalte zu reduzieren. Ein solcher Fokus steht eigentlich im Widerspruch zu den Anliegen des Allgemeinen Monitorings, das gerade die unerwarteten Auswirkungen an noch unbekanntem Ort aufdecken soll. Er trägt aber entscheidend zur Kosteneffizienz eines Monitorings bei. Dabei gibt es sicherlich nicht nur eine einzige *korrekte* Konstellation von Schutzgütern und Indikatoren für ein Monitoring-Programm. Zentraler Punkt ist eine begründete Auswahl einer Variante inklusive dem bewussten Verzicht auf mögliche Alternativen. Am Ende eines womöglich langwierigen Entscheidungsprozesses muss auf jeden Fall feststehen, welche Aussagen und Erkenntnisse von einem Überwachungsprogramm erwartet werden können und welche nicht (WEBER et al. 2004; BÜHLER 2006).

## **2 Wünsche: bitte klar formulieren**

### **2.1 Raumbezug**

Eine Umweltüberwachung besteht in ihrem Kern aus einer Stichprobenerhebung. Dabei muss schon bei der Planung klar sein, für welchen Stichprobenraum Aussagen zu machen sind, so dass der Erhebungsaufwand darauf konzentriert werden kann. Es sind mindestens drei wichtige Entscheide zu treffen. Sie betreffen erstens das Untersuchungsgebiet, zweitens die darin zu berücksichtigenden Lebensräume oder Nutzungstypen und drittens die räumliche Skala der Messungen.

Das Untersuchungsgebiet ist der geografische Geltungsbereich, für den das Monitoring Aussagen produziert. Dies können politische Einheiten oder auch biogeografisch charakterisierte Regionen sein. Innerhalb des Untersuchungsgebiets können sich Messungen wiederum auf einzelne Lebensräume oder Nutzungstypen beschränken, etwa das Ackerbaugebiet, das Landwirtschaftsgebiet oder das gesamte Offenland. Jegliche Gliederung ist möglich, so lange sie nachvollziehbar definiert ist. Die Skala der Messungen letztlich bestimmt die Raumeinheit, auf der Veränderungen tatsächlich im Feld gemessen werden. Die Wahl dieser Skala ist für die spätere Interpretation der Messresultate von großer Bedeutung. Es macht einen Unterschied, ob Veränderungen beschränkt auf voneinander isolierten Äckern (wie z. B. bei den Farm Scale Evaluations in Großbritannien, FIRBANK et al. 2003) oder aber auf der gesamten Ackerbaufläche z. B. eines Quadratkilometers gemessen wurden.

Bei der Wahl des Raumbezugs für ein GVO-Monitoring besteht oft ein Zwiespalt zwischen der Dauer, bis Effekte feststellbar werden, und der Repräsentativität der gewählten Skala der Messungen. Je stärker sich das Monitoring auf die effektiv mit GVO bebauten Flächen – Parzellen oder Schläge – konzentriert, desto schneller werden potentielle Effekte erkannt, weil sie wahrscheinlich dort zuerst stattfinden. Beispiele dafür wären die Veränderung der Bodenfruchtbarkeit oder der Abundanz von Nützlingen. Möglicherweise sind aber Effekte, die sich nur in einzelnen Parzellen des Ackerbaugebiets ereignen und vielleicht auch nur vorübergehend sind, von geringer ökologischer Relevanz. Erweitert man deshalb die Skala der Messungen und bezieht die Umgebung oder gar zusätzliche Lebensräume – z. B. Wiesen, Feldgehölze, Gewässer – ebenfalls mit ein, erhalten allfällige Veränderungen ein größeres ökologisches Gewicht, ihr Nachweis wird aber methodisch aufwändiger und dauert wahrscheinlich auch länger.

Antworten auf die Frage nach der passenden Messskala lassen sich finden, wenn man sich vorstellt, nach welchem räumlich-zeitlichem Muster ein Effekt von GVP auftreten könnte und dann prüft, welche dieser Wirkungsmuster im Monitoring unbedingt nachweisbar sein sollen. Effekte, die zwar rasch und deutlich, aber nur lokal und vorübergehend auftreten, sind für ein Allgemeines Monitoring womöglich von beschränktem Interesse. Die Tabelle 1 unternimmt einen Versuch, solche Wirkungsmuster zu definieren.

## 2.2 Schutzgüter und Schutzziele

Ein GVO-Monitoring soll mögliche Veränderungen der Umwelt erkennen. Umwelt ist allerdings ein sehr umfassender Begriff. Er schließt die belebte und unbelebte Umgebung mit ein: Boden Luft, Gewässer, Biotope, Tiere, Pflanzen und ihre Lebensräume. Für viele dieser Umweltgüter bestehen Rechtstexte, die einen Schutz verlangen. Oft beschreiben diese auch gleich die geltenden Schutzziele, d. h. welche Art von Beeinträchtigung zu vermeiden ist. Diese rechtlichen Vorgaben sind auch bei der Wahl der zu überwachenden Umweltgüter im GVO-Monitoring heranzuziehen. Maßgebend für die Wahl sind einerseits die Wahrscheinlichkeit, mit der es zu einer Beeinträchtigung des Schutzguts durch GVO kommen könnte, andererseits auch ganz pragmatisch die *technische* Eignung für Monitoring-Zwecke. Geeignete Schutzgüter lassen sich möglichst scharf definieren, weisen eindeutige Schutzziele auf und sind im Hinblick auf eine Erhebung gut quantifizierbar. Für die Schweiz hat AMMANN (2006) die rechtlich verankerten Schutzgüter und Schutzziele zusammengestellt und auf ihren Nutzen für ein GVO-Monitoring hin bewertet (Tab. 2).

Tab. 1: Charakterisierung vier verschiedener Typen der Wirkungsweise von GVP-Anbau auf die Umwelt aufgrund von drei Merkmalen. Die Liste erhebt keinen Anspruch auf Vollständigkeit. Die angeführten Beispiele sind fiktiv und dienen lediglich dem besseren Verständnis der Wirkungstypen.

Dauer bis zum Eintritt der Wirkung	Dauer der Wirkung	Ortsbezug	Fiktives Beispiel
kurz, unmittelbar	vorübergehend	lokal, begrenzt	Reduktion der Biomasse der Acker-Beikrautflora durch intensivierte Einsatz von Spritzmitteln
kurz, unmittelbar	bleibend (mehrere Jahre)	lokal, begrenzt oder auch ausstrahlend	Nachhaltige Schädigung einer seltenen Tagfalterart durch die Aufnahme von GVP-Pollen mit toxischer Wirkung
kurz, unmittelbar oder lang, verzögert	bleibend (mehrere Jahre)	lokal, begrenzt	Allmähliche Akkumulation toxischer Substanzen im Boden bis zum Erreichen eines Schwellenwerts
kurz, unmittelbar oder lang, verzögert	bleibend (mehrere Jahre)	lokal begrenzt oder auch ausstrahlend	Veränderungen in der Zusammensetzung der Beikrautflora, hervorgerufen durch Herbizidresistenz

Tab. 2: Aus Schweizer Rechtserlassen abgeleitete Schutzgüter und ihre Beurteilung bezüglich ihrem Nutzen für ein GVO-Monitoring gemäß der Analyse von AMMANN (2006). – Mit einem Stern versehene Schutzgüter sind für die Überwachung durch ein GVO-Monitoring gut geeignet. Als Beurteilungskriterien gelten die Qualität der Umschreibung im Recht, die Quantifizierbarkeit und der geographische Geltungsbereich.

Umwelt	Gewässer*
Tiere*	Biodiversität*
Nachhaltigkeit	Stoffhaushalt
Lebensraum*	Unerwünschte Verbreitung
Empfindliche Lebensräume*	Landwirtschaftliche Praxis
Boden	Biolandbau*

### 2.3 Effektgrößen

Ist die Wahl der zu überwachenden Schutzgüter vollzogen, bleibt zu bestimmen, mit welchen Indikatoren deren Entwicklung zu messen ist und welche Veränderungen dieser Indikatoren als ökologisch und damit auch politisch bedeutend einzustufen sind. Effekte dieser Größe muss das Monitoring erkennen können.

Umweltparameter und insbesondere die Bestände von Pflanzen oder Tieren variieren in Abhängigkeit vieler oft im Detail unbekannter Faktoren scheinbar zufällig von Jahr zu Jahr, von Ort zu Ort. Auch die Messung der Indikatoren selber produziert aus rein technischen Gründen schwankende Ergebnisse, weil sie in der Regel nicht perfekt wiederholbar ist. Aus diesen Gründen verlaufen auch langfristige Trends bei Indikatoren nicht stetig, sondern begleitet von scheinbar beliebigem *Rauschen*, einem Auf und Ab der Messwerte. Daraus ergeben sich drei Folgerungen: (1) Je geringer die effektive Veränderung, desto schwieriger – sprich aufwändiger – wird ihr Nachweis vor dem Hintergrund des vorhandenen Rauschens. (2) Veränderungen nachzuweisen, die kleiner sind, als allein aufgrund von Zufälligkeiten (Messfehler, Bearbeitereffekte) zu erwarten ist, ergibt ökologisch keinen Sinn. (3) Das Ausmaß der natürlichen Schwankungen eines Indikators bestimmt, wie zuverlässig beobachtete Veränderungen statistisch nachweisbar sind.

Damit ist erst die untere Grenze nachweisbarer Effekte umschrieben. Interessanter wäre eigentlich dasjenige Ausmaß, das eine längerfristige Schädigung des Schutzgutes anzeigt. Dass es hier kaum quantitative Angaben gibt, ist verständlich. Zum einen ist das umfangreiche Wissen für das Festlegen solcher Schwellenwerte kaum je vorhanden, zum anderen enthalten sie in jedem Fall eine subjektive Komponente. Was als schädlich oder lästig gilt, ist eine Frage der Bewertung. Trotzdem gibt es einige Ankerpunkte, die als Orientierungshilfe dienen:

- Aufzeichnungen über Veränderungen eines Indikators in früheren Zeiten, die heute als eindeutig negativ beurteilt werden (z. B. Arten- oder Bestandesverlust, Akkumulation von Schadstoffen im Boden).
- Bei vielen Biodiversitätsindikatoren zeigen die Erfahrungen, dass innerhalb des Zeitraums von ca. 10 Jahren Veränderungen von deutlich weniger als 10 % nur mit enormem Aufwand erkannt werden (PLATTNER et al. 2004).
- Bestehende Beispiele für Monitoring-Programme mit ähnlichen Zielen, nach denen man sich richten kann, finden sich zum Beispiel in SHANK et al. (2002), PERRY et al. (2003) und PLATTNER et al. (2004). Auch die Kriterien für die Einstufung von Arten in die Rote Liste geben einen Anhaltspunkt (IUCN 2001).

### **3 Wirklichkeit: was liegt tatsächlich drin?**

Nur Wünschen, die eindeutig formuliert sind, kann gut entsprochen werden. Dies trifft auch für das GVO-Monitoring zu. Ressourcen für eine Datenerhebung, die nur im Glauben gestartet wird, spätere Bedürfnisse schon zu treffen, sind ziemlich sicher fehlinvestiert. In diesem Sinn trägt die Arbeit, die für das Formulieren der Anforderungen an ein GVO-Monitoring nötig ist, sehr wesentlich zu dessen Kosteneffizienz bei. Stehen die Anforderungen fest, beginnt die Suche nach Erhebungsmethoden, die mit möglichst wenig Aufwand die gesteckten Ziele erreichen. Die ökonomische Bedeutung dieses zweiten Schritts ist offensichtlich, da die Aufwandskomponente direkt angesprochen ist. Kosteneffizienz ist auch in diesem Schritt auf verschiedenen Ebenen zu erreichen.

#### **3.1 Auswahl des Erhebungsdesigns**

Mit Erhebungsdesign ist die Anordnung und Abfolge gemeint, mit der die Messungen erfolgen. Es legt fest, in welchem Rhythmus die Messungen vorgenommen werden, aber auch, wie die Messungen räumlich zueinander in Beziehung stehen. Für das Umweltmonitoring empfohlen sind allgemein Designvarianten, welche zumindest einen Teil der Messflächen bei jedem Erhebungsdurchgang wiederholen (URQUHARDT et al. 1998).

Welcher Erhebungsrhythmus aber letztlich am besten passt, hängt von den konkreten Bedürfnissen ab. Die Abbildung 1 skizziert als Beispiel zwei Varianten. Die eine berücksichtigt jährlich dieselbe Auswahl an Messflächen. Die andere Variante erhebt eine Stichprobe der fünffachen Größe, führt die Erhebung aber gestaffelt über fünf Teilstichproben innert fünf Jahren durch. Der Erhebungsaufwand ist identisch. Welche Variante eignet sich besser für den Nachweis von Veränderungen einer bestimmten Größe? Abbildung 2 zeigt die Resultate einer Power-Analyse aufgrund einer Computersimulation der Artendichte von Tagfaltern in 1 x 1 km Erhebungsflächen, wobei die Schätzung der Streuung der Messwerte auf bestehenden Freilanddaten beruht. Aus den Kurven geht hervor, dass ab einer Dauer von 10 bis 15 Jahren kaum mehr Unterschiede bezüglich Nachweisvermögen einer 10%-Veränderung bestehen. Kurzfristig, für einen möglichst raschen Nachweis ist Variante 1 besser, weil sie den statistischen Vorteil verbundener Stichproben von Beginn an nutzt. Allerdings bietet Variante 2 den großen Vorzug, dass ihre Resultate auf einer fünfmal größeren Stichprobe beruhen. Der Zustand des gemessenen Indikators – im Beispiel also der mittlere Artenreichtum von Tagfaltern pro Quadratkilometer – lässt sich deshalb wesentlich besser abschätzen. Zudem ist die Störung der Probeflächen durch die Erhebung geringer, da sie nur jedes fünfte Jahr aufgesucht werden.

Variante (1)

Aufnahmejahr	1	2	3	4	5	6	7	8	9	10	11
1. Teilstichprobe	S	S	S	S	S	S	S	S	S	S	S

Variante (2)

Aufnahmejahr	1	2	3	4	5	6	7	8	9	10	11
1. Teilstichprobe	S					S					S
2. Teilstichprobe		S					S				
3. Teilstichprobe			S					S			
4. Teilstichprobe				S					S		
5. Teilstichprobe					S					S	

Abb. 1: Schematische Darstellung von zwei Varianten der zeitlichen Abfolge einer Stichprobenerhebung. Das Jahr einer Erhebung ist durch ein „S“ symbolisiert. Variante 1 (oben) benutzt jedes Jahr genau dieselben Stichprobeflächen, während Variante 2 (unten) eine Gesamtstichprobe fünffacher Größe gestaffelt über 5 Jahre erhebt.

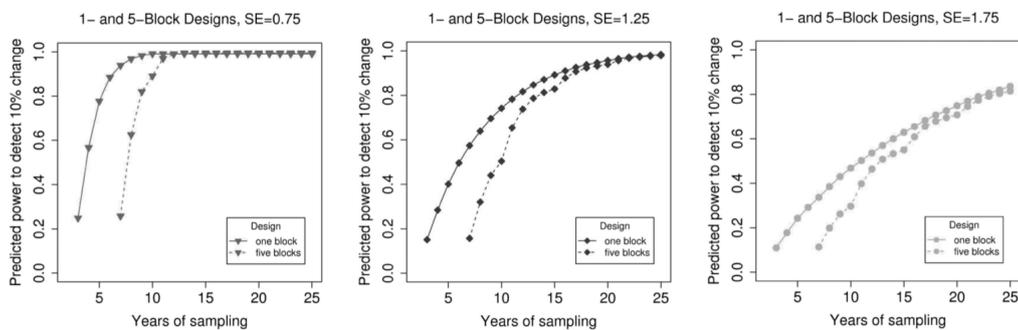


Abb. 2: Vergleich der Aussagekraft von zwei verschiedenen Erhebungsrhythmen (siehe Abb. 1) anhand von Computersimulationen der Artendichte von Tagfaltern in 1 x 1 km Quadranten bei unterschiedlicher Streuung der Messungen (SE). Das Ausmaß der Streuung wurde auf der Basis von Freilanddaten geschätzt (Biodiversitäts-Monitoring Schweiz, British Butterfly Monitoring Scheme). Aufgetragen ist die Dauer der Untersuchung in Jahren gegen die im Modell erreichte statistische Macht, eine Veränderung des mittleren Artenreichtums von 10 % nachzuweisen. Ausgezogene Linien: Variante 1 (jährlich dieselbe Stichprobe); Gestrichelte Linien: Variante 2 (gestaffelte Stichprobe). SE = Standardfehler des Mittelwerts der Werte eines Jahres. Je größer SE, desto länger dauert es, bis Variante 2 dieselbe Power erreicht wie Variante 1. Grafik übernommen aus PEARMAN (2006).

Nebst dem Erhebungsrhythmus gibt es auch für die räumliche Anordnung und gegenseitige Beziehung der Messflächen viele Varianten. Diese spielt vor allem dann eine Rolle, wenn die Erhebungen einen Vergleich zwischen verschiedenen Kategorien von Messflächen erlauben sollen, z. B. Messflächen mit und ohne GVP-Anbau. Hier braucht es weniger statistische Berechnungen als vielmehr qualitative Überlegungen zur Aussagekraft verschiedener Anordnungen. Bedeutende Kriterien für die Beurteilung verschiedener Design-Varianten sind:

- Kausalität: Rückschlüsse auf einen ursächlichen Zusammenhang zwischen dem Anbau von GVP und den beobachteten Unterschieden der Indikatorwerte sind möglich.
- Aussagemöglichkeiten: die Daten lassen sich vielseitig auswerten: verschiedene räumliche Skalen, Simultanvergleich, Zeitreihen etc.
- Reaktionszeit: Die Design-Variante erlaubt sehr rasch (innert Jahren) erste Aussagen zu Veränderungen.
- Flexibilität: das Monitoring bleibt auch dann effektiv, wenn sich Anbaudichten und -muster im Lauf der Zeit ändern.
- Synergien mit anderen Programmen: Die Erhebungsflächen oder -daten Dritter lassen sich nutzen.
- Umsetzbarkeit: Die methodischen Schwierigkeiten und/oder administrative Hindernisse sind gering.
- Aufwand: Der Arbeitsaufwand, und damit wahrscheinlich auch die Kosten, diese Design-Variante einzurichten und zu betreiben, sind verhältnismäßig gering.

### **3.2 Auswahl der Indikatoren und Messmethoden**

Komplexe Schutzgüter wie Biodiversität oder Bodenfruchtbarkeit lassen sich nicht direkt messen. Dazu dienen stellvertretende Indikatoren, die den Zustand des Schutzguts möglichst gut abbilden. Im Fall der Biodiversität kommen als Indikatoren einzelne Arten, eine Auswahl von Arten oder auch eine ganze Artengruppe (z. B. Tagfalter) in Frage. Die Auswahl und Definition von Indikatoren erfolgt häufig aufgrund von Risikoanalysen (siehe MEIER & HILBECK 2005), aber auch anhand einfacher Überlegungen zur ökologischen Bedeutung und zur Praktikabilität: Indikatoren, für die Standardverfahren bestehen, ein großes Angebot an Spezialisten verfügbar ist und die dazu noch eine hohe gesellschaftliche Bedeutung haben, sind einfacher zu realisieren. Letztere Argumente sind aus rein fachlicher Sicht zwar

nachrangig, für die Realisierbarkeit eines Monitoring-Programms aber dennoch bedeutend.

Um diese Indikatoren zu erheben, sind letztlich Messmethoden zu bestimmen, die dann im Feld zur Anwendung kommen. Auch hier steht der Anwender meist vor einer Auswahl. Zunächst stehen qualitative Eigenschaften im Vordergrund. Zentrale Voraussetzungen für die Eignung einer Methode sind insbesondere:

- Standardisierbarkeit: die Methode lässt sich so beschreiben, dass sie von zwei Bearbeitern genau gleich angewendet wird.
- Repräsentativität und Sensitivität / Auflösung: die Methode erfasst die entscheidende Eigenschaft des Indikators, z. B. ein möglichst vollständiges Artenspektrum.
- Robustheit: die Methode reagiert unempfindlich auf leicht abweichende Erhebungszeitpunkte, Witterungsbedingungen, Lage der Messfläche.

Methoden, die eines dieser drei Kriterien schlecht erfüllen, kommen nicht in Frage. Weitere, vor allem kostenrelevante Kriterien wie Materialaufwand, Arbeitsaufwand, Schwierigkeitsgrad der Anwendung spielen zwar auch eine Rolle. Sie sind aber zunächst keine Ausschlusskriterien, bis Fragen zum Aufwand-Ertragsverhältnis geklärt sind.

### **3.3 Messintensität**

Jede Methode lässt sich bezüglich investiertem Aufwand pro Messung variieren. Dieser Punkt soll am Beispiel der Transektmethode für Tagfalter illustriert werden. Bei dieser Methode werden die Falter auf vorgegebenen Routen mit Keschern gefangen und bestimmt. Wie lang die Transekte sein müssen und wie viele Exkursionen pro Jahr notwendig sind, ist offen. Überlegungen zur Kosteneffizienz bringen hier Aufschluss. Die Abbildung 3 vergleicht zwei Varianten, die sich bezüglich der Anzahl Exkursionen unterscheiden. Zusätzlich ist eine dritte Variante mit stark verkürztem Transekt und einer höheren Zahl an Exkursionen gezeigt. Aus den Kurven geht hervor, dass die Variante mit langen Transekten und wenigen Exkursionen an mehr Standorten die größte Effizienz erzielt. Diese Bewertung betrachtet nur das Vermögen der Methode, Veränderungen des Artenreichtums nachzuweisen. Dass die drei Varianten ganz unterschiedliche Artenspektren liefern und verschiedene Auswertungsmöglichkeiten zulassen, müsste aber ebenfalls in eine Evaluation mit einfließen.

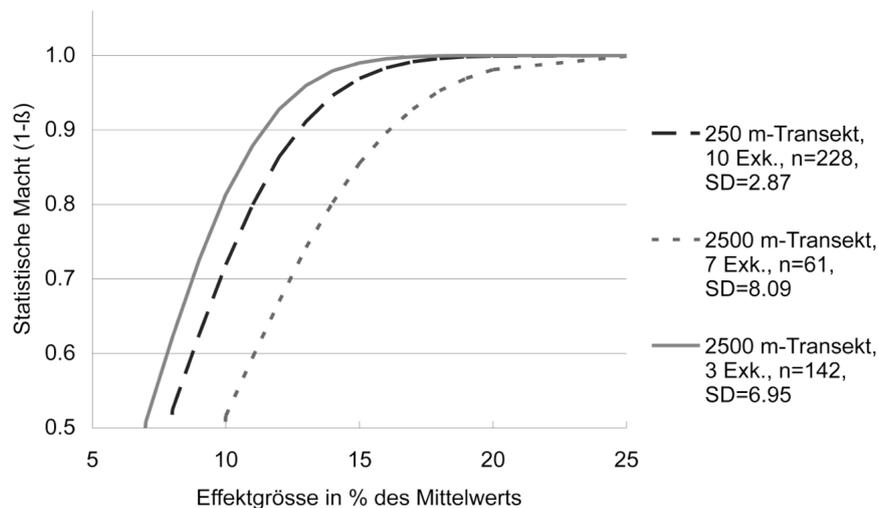


Abb. 3: Aussagekraft in Abhängigkeit der Verteilung des Aufwandes bei der Erhebung des Tagfalter-Artenreichtums. Aufgetragen ist die Größe des nachzuweisenden Effekts (in Prozent des mittleren Artenreichtums) gegen die geschätzte Statistische Macht ( $1-\beta$ ). Die Kurven bezeichnen drei Varianten bezüglich Länge des Transekts (m) und Zahl der Exkursionen (Exk.) pro Saison. Der Gesamtaufwand ist bei allen drei Varianten identisch und erlaubt eine stark unterschiedliche Anzahl an Transekten (n). Die Berechnungen der statistischen Macht erfolgten für jede Variante anhand von spezifischen Schätzungen zur Streuung der Messwerte auf Grundlage von bestehenden Daten aus dem Biodiversitäts-Monitoring Schweiz BDM und dem Programm LANAG des Kantons Aargau. Die Schätzungen beruhen auf der für die BDM-Daten noch unbestätigten Annahme, dass die räumliche Varianz gleich der zeitlichen Varianz sei. SD = Standardabweichung.

### 3.4 Bestimmen der Stichprobengröße

Den Umfang der zu erhebenden Stichprobe festzulegen, ist gewöhnlich der letzte Schritt in der Versuchsplanung. Er erfolgt erst, wenn das Erhebungsdesign und die Messmethode feststehen und die Auswertungsroutine formuliert ist. Anhand der spezifizierten Mindestgröße für relevante Effekte ist das Berechnen der dazu erforderlichen Stichprobengröße eine weitgehend technische Angelegenheit ohne größeren Ermessensspielraum. Das Resultat der Berechnungen – der Stichprobenumfang – wird dabei erheblich von der Varianz der Messgröße beeinflusst. Diese gilt es zunächst zu schätzen. Idealerweise geschieht dies mittels Testerhebungen. Ist dies nicht möglich, müssen bestehende Daten aus methodisch möglichst ähnlichen Erhebungen dazu herangezogen werden. Nur wenn die geschätzte Varianz realistisch ist, d. h. wenn sie derjenigen der zukünftigen Erhebungen entspricht, ergibt sich auch die *korrekte* Stichprobengröße, da die Auswertungsroutine und die Schätzung des Stichprobenum-

fangs auf denselben statistischen Ansätzen beruhen. In diesem Arbeitsschritt wird sich zeigen, wie aufwändig die gewählten Indikatoren und Erhebungsmethoden tatsächlich sind.

Beim Schätzen des Stichprobenumfangs ebenfalls anzugeben ist die statistische Macht, englisch *Statistical Power*. Auch diese Größe wirkt sich auf den Stichprobenumfang aus. Die statistische Macht wird auch als  $(1 - \beta)$  ausgedrückt, wobei  $\beta$  den sog. Fehler 2. Art bezeichnet. Damit gemeint ist die Wahrscheinlichkeit, real existierende Effekte mittels statistischer Analyse auch tatsächlich nachzuweisen. Diese Toleranz ist vom Anwender zu bestimmen, ebenso wie der Fehler 1. Art, nämlich die zulässige Wahrscheinlichkeit, dass ein in der Analyse gefundener Effekt ein Irrtum ist. Umgekehrt lässt sich bei festgelegtem Stichprobenumfang die statistische Macht mittels einer sog. Power-Analyse berechnen. Eine geringe statistische Macht ist für ein Monitoring-Programm sehr ungünstig, weil Effekte selbst dann nicht ans Tageslicht kommen, wenn sie effektiv stattgefunden haben. Die Kenntnis der statistischen Macht einer Analyse ist für die Interpretation der Daten somit zentral (FAIRWEATHER 1991).

Vor diesem Hintergrund stellt sich die Frage nach der Kosteneffizienz beziehungsweise nach dem optimal eingesetzten Aufwand also doch noch einmal. Sowohl der Fehler 1. Art als auch derjenige 2. Art sind schlecht und ziehen unnötige Kosten nach sich. Entscheidend ist aber, dass sie sich gegenläufig zueinander verhalten. Je kleiner die Irrtumswahrscheinlichkeit für einen Fehler 1. Art, desto größer wird zwangsläufig ein Fehler 2. Art und umgekehrt. Oder anders ausgedrückt. Wer größte Sicherheit vor einem unerwünschten Ereignis will, muss einen gelegentlichen Fehlalarm in Kauf nehmen. Im Fall von Umweltwirkungen von GVO könnten die späteren Kosten und die übrigen Folgen auf die Umwelt, die ein zu Beginn verpasster Nachweis schädlicher Auswirkungen nach sich ziehen, weit größer sein als die Kosten, die ein Fehlalarm verursacht.

#### **4. Schlussfolgerungen**

Ein kosteneffizientes Monitoring ist nicht gleichzusetzen mit einem kostengünstigen Monitoring. Kosteneffizient meint den im Hinblick auf das Erreichen der Ziele optimalen Einsatz von Ressourcen. Ob ein Monitoring zu geringen Kosten möglich ist, ist von den Zielen abhängig.

Das Entwickeln eines GVO-Monitorings erfordert eine Reihe von Entscheidungen mit klar subjektiver Komponente. Gemeint sind damit aktuellen Werthaltungen, Risikoeinschätzungen und tolerierbaren Veränderungen in der Umwelt. Rechtliche Vorgaben geben zumindest die Schutzgüter und -ziele vor, aber oft nur qualitativ. Erforderlich

für ein kosteneffizientes Monitoring ist es, hier das weite Feld der Möglichkeiten einzuschränken und Prioritäten zu setzen.

Auch die mehr technischen Aspekte und Entscheidungen betreffend Erhebungsdesign, Indikatoren, Messmethoden und Stichprobenumfang orientieren sich an den Zielen für das Monitoring. Hier sind zunächst qualitative Betrachtungen nötig, um festzustellen, welche Varianten die gesteckten Ziele überhaupt erfüllen können. Erst ganz am Schluss kommen statistische Verfahren zum Einsatz, welche die Aussagekraft verschiedener Varianten bei konstantem Erhebungsaufwand vergleichen.

### **Literatur:**

- AMMANN, D. (2006): Teilbericht Schutzziele. – In: Methodische Grundlagen der Langzeitüberwachung von GVO: Schutzziele, Indikatoren, Erhebungsmethoden. – Bundesamt für Umwelt: in Vorbereitung.
- BÜHLER, C. (2006): Biodiversity Monitoring in Switzerland: What can we learn for general surveillance of GM crops? – Journal für Verbraucherschutz und Lebensmittelsicherheit 1, Supplement 1: 37-41.
- CHAPMAN, M. A. & BURKE, J. M. (2006): Letting the gene out of the bottle: the population genetics of genetically modified crops. – New Phytologist 170: 429-443.
- FAIRWEATHER, P. G. (1991): Statistical Power and Design Requirements for Environmental Monitoring. – Australian Journal of Marine and Freshwater Research 42: 555-567.
- FIRBANK, L. G., HEARD, M. S., WOIWOD, I. P. & 16 weitere Autoren (2003): An introduction to the Farm-Scale Evaluations of genetically modified herbicide-tolerant crops. – Journal of Applied Ecology 40: 2-16.
- IUCN (2001): IUCN Red List Categories and Criteria, v3.1. [www.redlist.org](http://www.redlist.org). (Zugriff: 13. November 2006).
- MARQUARD, E. & DURKA, W. (2005): Auswirkungen des Anbaus gentechnisch veränderter Pflanzen auf Umwelt und Gesundheit: Potentielle Schäden und Monitoring. – Bericht erstellt im Auftrag des sächsischen Staatsministeriums für Umwelt und Landwirtschaft (SMUL): 207 S.
- MEIER, M. S. & HILBECK, A. (2005): Faunistische Indikatoren für das Monitoring der Umweltwirkungen gentechnisch veränderter Organismen (GVO). Verfahren zur Beurteilung und Auswahl. – Bonn (Bundesamt für Naturschutz) – Naturschutz und Biologische Vielfalt 29.
- PEARMAN, P. B. (2006): Power Analysis of Alternate Designs for Nation-Wide Monitoring Programs: A Central European Example. – Bericht zu Händen der Koordinationsstelle Biodiversitätsmonitoring Schweiz: 9 S.

- PERRY J. N., ROTHERY P., CLARK S. J., HEARD M. S. & HAWES C. (2003): Design, analysis and statistical power of the Farm-Scale Evaluations of genetically modified herbicide tolerant crops. – *Journal of Applied Ecology* 40: 17-31.
- PLATTNER, M., BIRRER, S. & WEBER, D. (2004): Data quality in monitoring plant species richness in Switzerland. – *Community Ecology* 5: 135-143.
- SANVIDO, O., BIGLER, F., WIDMER, F. & WINZELER, M. (2003): Ökologische Risiken von gentechnisch veränderten Pflanzen: Schwerpunkte künftiger Forschung. Projektbericht. – Reckenholz, Zürich (Eidgenössische Forschungsanstalt für Agrarökologie und Landbau FAL): 63 S.
- SHANK, C., SCHIECK, J. & FARR, D. (2002): The Alberta Forest Biodiversity Monitoring Program. Chapter 18. [www.abmp.arc.ab.ca/index.htm](http://www.abmp.arc.ab.ca/index.htm). (Zugriff: 13. November 2006. 87 S.
- URQUHART, S. N., PAULSEN, S. G. & LARSEN, D. P. (1998): Monitoring for Policy-Relevant Regional Trends over Time. – *Ecological Applications* 8: 246-257.
- WEBER, D., HINTERMANN, U. & ZANGGER, A. (2004): Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. – *Global Ecology and Biogeography* 13: 97-104.

### Der Autor

**Christoph Bühler** arbeitet als Biologe und Projektleiter in einem privaten Beratungsbüro. Haupttätigkeitsgebiet ist das Erstellen und Realisieren von Überwachungsprogrammen sowie die Auswertung der Daten, speziell für den Bereich Biodiversität.



Christoph Bühler  
 Hintermann & Weber AG  
 Ökologische Planung, Beratung & Forschung  
 Hauptstrasse 52  
 CH-4153 Reinach BL  
[buehler@hintermannweber.ch](mailto:buehler@hintermannweber.ch)

