

Auswertung großer Datenmengen

Tipps zum Umgang und zur Auswertung großer Datenmengen, wie sie im Dauermonitoring oder auch durch Störungen an WEA-Gondeln verursacht werden können.

Im Rahmen von Dauererfassungen werden teilweise sehr große Datenmengen mit 10000 bis 100000 Aufnahmen erhalten. Werden diese Datenmengen genauso wie einige wenige Aufnahmen bearbeitet, dann steigt der Aufwand enorm an. Dies lässt sich in der Regel nicht vor Auftraggebern verantworten und ist in der Regel auch nicht nötig. Ähnlich verhält es sich, wenn zwar bereits große Aufnahmemengen prognostiziert werden, dann aber zum Beispiel durch Störgeräusche deutlich größere Aufnahmezahlen für die Analyse anfallen.

Im Folgenden werden beispielhaft Vorgehensweisen gezeigt, um Ihnen Anregungen und Tipps zum Umgang mit großen Datenmengen zu geben. Dabei ist es immer das Ziel, die Aufnahmen möglichst gut zu prüfen, aber nicht zwingend auch alle einzelnen Aufnahmen und Bestimmungsergebnisse zu kontrollieren. Auf diese Weise wird es zwar immer einen Fehler im Hinblick auf die genaue Verteilung der Arten auf die Aufnahmen geben, aber dieser ist kontrolliert und dafür werden auch große Datenmengen in überschaubarer Zeit bearbeitet. Beachten Sie, dass direkte Vergleich von Arten untereinander zum Beispiel an Hand von Aufnahmezahlen nicht zielführend sind. Darstellung wie Kreisdiagramme und Angaben in Form von % der Arten an gesamter Aktivität sind auf Grund der unterschiedlichen Ortungsmodi und Jagdweisen eigentlich nicht korrekt.

Erste Schritte

Die ersten Schritte der Auswertung sind wie üblich das Einlesen der Daten in bcAdmin und eine automatische Rufsuche sowie Artbestimmung. Wir empfehlen gezielt mit den Session-Markierungen (bunte Fähnchen/Punkte) zu arbeiten und alle neuen Session entsprechend zu markieren. Bereits geprüfte Sessions können dann mit einem entsprechenden Status versehen werden. Dies erleichtert es später den Überblick über die Auswertungen zu behalten.



Hilfreich ist es, im Anschluss einige einzelne Tage genauer auszuwerten, um ein Gefühl für die vorkommenden Arten und die typischen Verwechslungen am Standort zu erhalten.

Vorgehen bei bodennahe Dauererfassung

Bei Erfassungen über längere Zeit zum Beispiel mit der Box-Erweiterung fallen in der Regel große Mengen an Aufnahmen mit Fledermauslauten an. Entsprechend des Standorts werden verschiedene Artengruppen häufiger oder seltener vertreten sein. Bei Offenlandstandorten sind meist Arten aus den Gattungen *Pipistrellus*, *Nyctalus* und *Eptesicus* häufig-

ger vertreten. Waldstandorte hingegen sind häufig geprägt durch die Gattungen *Pipistrellus* und *Myotis*. Nachdem das Artenspektrum grob bekannt ist, empfiehlt es sich zuerst seltene Nachweise und typische Fehler zu kontrollieren. Wir haben uns als Vorgehen angewöhnt, diesen Schritt auf zeitliche Einheiten wie z.B. 14 Tage oder 30 Tage anzuwenden. Zum einen ändern sich im Jahresverlauf die Artenzusammensetzungen, zum anderen bleibt die Menge an zu kontrollierenden Aufnahmen so auch etwas übersichtlicher.

Datengrundlage eines einfachen Auswertungstests

Das folgende Beispiel sind Aufnahmen, die mittels batcorder+Box-Erweiterung in einem Teil des Schlossgartens der Uni Münster aufgezeichnet wurden. Die Box war an einem Mast in ca. 5 m Höhe installiert, im Umfeld befanden sich Baumreihen und einzelne Obstbäume. Insgesamt wurden zwischen Ende Mai und Ende Oktober 35000 Aufnahmen verzeichnet. Für diese Anleitung wurden zwei Zeiträume exemplarisch ausgewertet. Dies waren der 01.08. bis 15.08. (3246 Aufnahmen) und der 15.09. bis 30.09. (3023 Aufnahmen).

Zeitlich Aufwand der Auswertung

Im Rahmen der ersten Korrektur wurden jeweils die Aufnahmen mit Arteinträgen Bbar, Malc, Pkuh, Stief, Pmid, Vmur und ein Teil der Pipistrelloiden gesichtet. Dazu wurden die **Aufnahmen nach diesen Arten gefiltert** und dann in der Rufvorschau von bcAdmin kontrolliert. **Fehler wurden über die Schnellasten behoben**. Dieser erste Prüfschritt dauerte für den August-Block 15 Minuten, für den September-Block 18 Minuten. Das bedeutet je Tag Erfassung ca. eine Minute für den ersten Schritt der Auswertung.

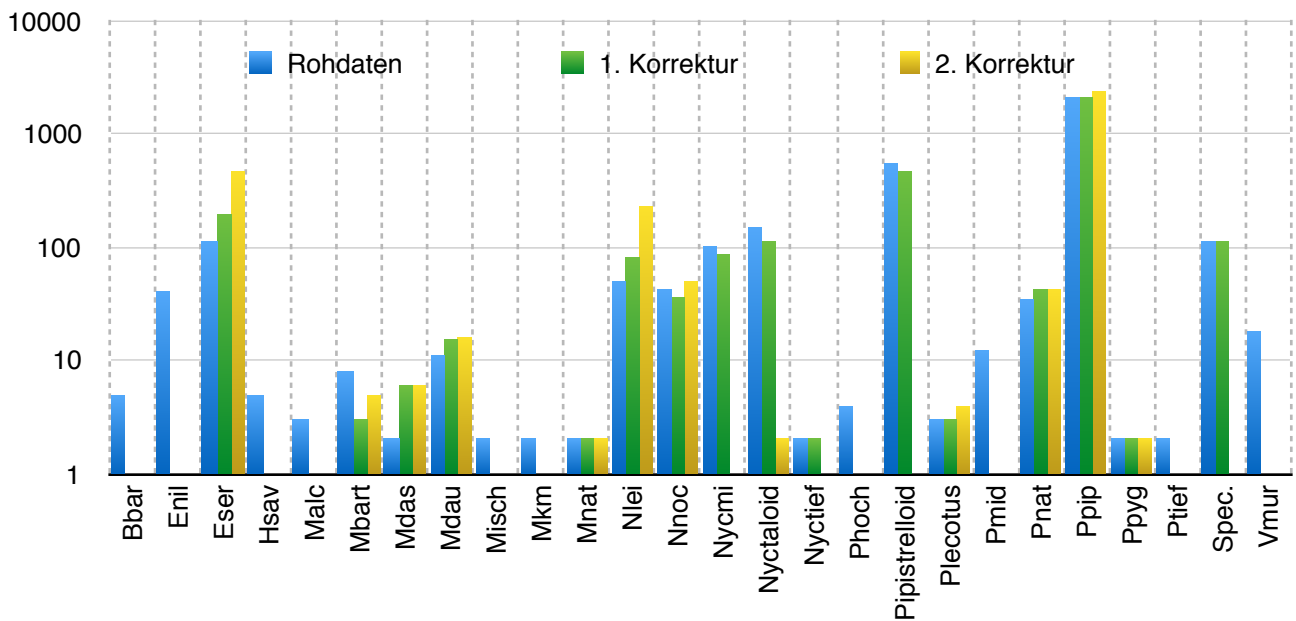
Im zweiten Korrekturschritt wurden dann die an diesem Standort häufig auftretenden Nyctaloide kontrolliert. Dazu wurden die **Aufnahmen mit zeitlichem Bezug gleichfarbig** in der Übersicht hinterlegt. So konnten Blöcke von Aufnahmen, die vermutlich von einem Tier stammen, schnell ausgewertet werden. Für diesen zweiten Schritt betrug der Zeitaufwand 35 respektive 17 Minuten. Der Zeitaufwand je Tag Erfassung lag demnach bei ein bis zwei Minuten.

	August-Block	September-Block
Aufnahmen	3246	3023
Zeit für erste Prüfung	15 Minuten	18 Minuten
Zeit für zweite Prüfung	35 Minuten	17 Minuten
Zeit je Tag	ca. 3 Minuten	ca. 2 Minuten
Zeit je Aufnahme	1 Sekunde	< 1 Sekunde
Hochrechnung Aufwand (154 Tage, 35736 Aufnahmen)	8 Stunden	

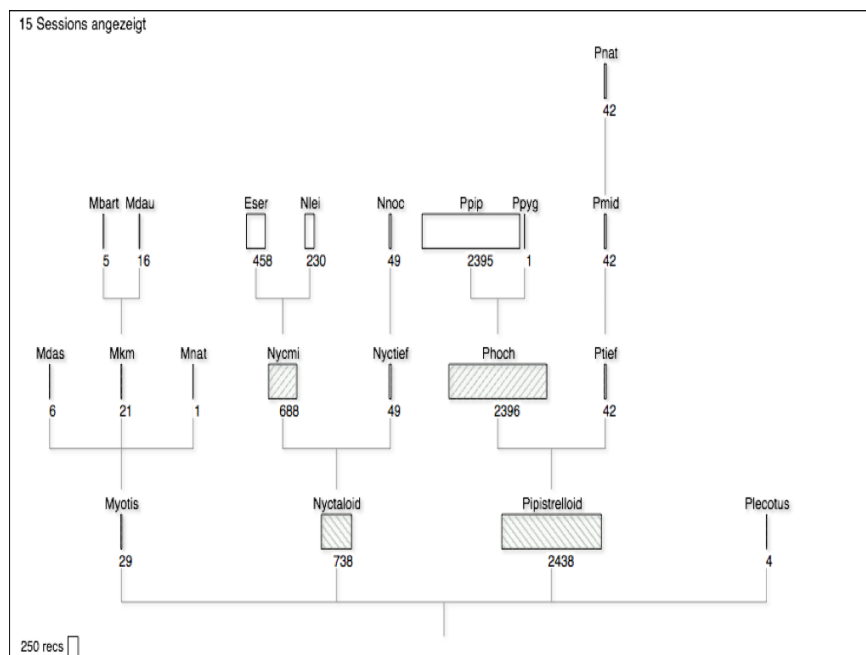
Qualität der Auswertung

Neben dem zeitlichen Aufwand ist die Qualität der Bestimmung ein wichtiges Maß. Im vorliegenden Fall wurde auf eine Bestimmung jeder einzelnen Aufnahme verzichtet, jedoch ist der Fortschritt der Bestimmung dokumentiert.

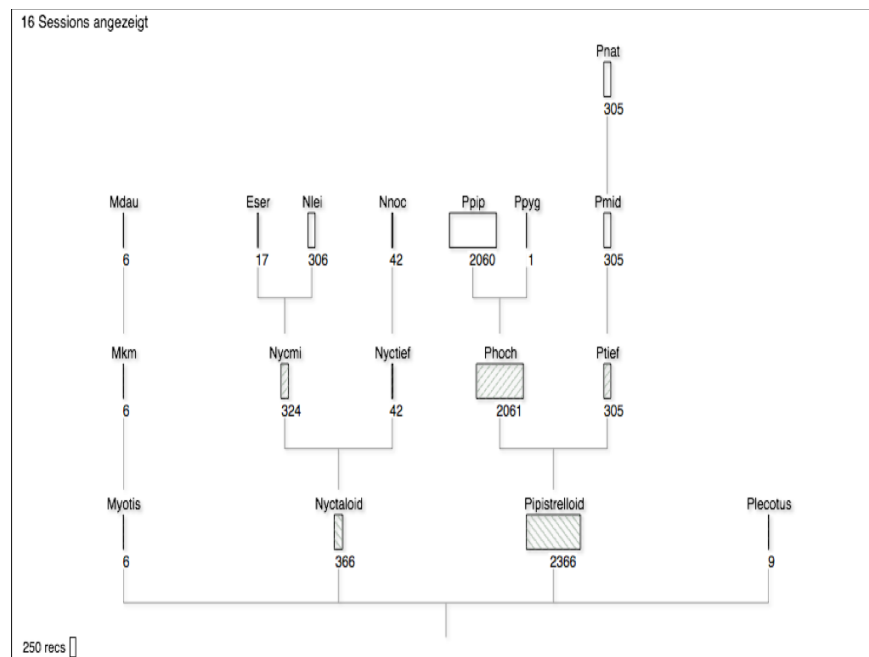
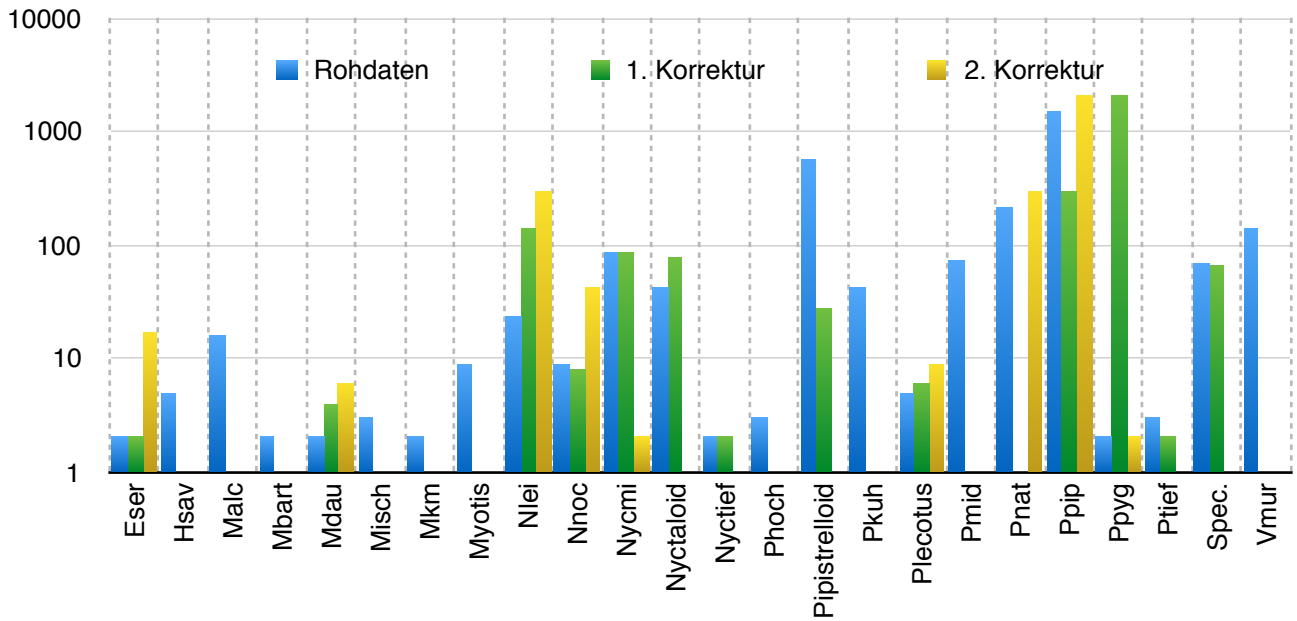
Für den August ergibt sich daraus dann folgendes Bild - die Balken geben die Anzahl Aufnahmen je Art/Gruppe an, die Farben die Korrekturschritte (in der Grafik werden Arten mit einer Aufnahme für eine bessere Darstellung mit 2 Aufnahmen gezeigt!)..



Der Artenbaum des Endergebnis sieht wie folgt aus:



Die selben Auswertungen für den September ergeben die folgenden Abbildungen:



Das endgültige Artenspektrum sieht für den Standort sehr vernünftig aus - sicherlich werden einzelne Verwechslungen innerhalb der Gattungen *Nyctalus* und *Eptesicus* möglich sein. Jedoch würde eine Korrektur dieser sehr aufwendig sein (einzelne Aufnahmen alle kontrollieren) und zum anderen wird sich daraus kaum ein deutlich besseres Ergebnis ableiten lassen.

Sicherlich setzt die gezeigte, schnelle Auswertung eine gewisse Übung voraus. Aber auch bei weniger eingearbeiteten Anwendern wird die Auswertung in solchen Schritten sehr schnell sein. Selbst wenn der zeitliche Aufwand verdoppelt wird, sind solche Datensätze innerhalb von 2 bis 3 Arbeitstagen auswertbar. Hätte man alle Aufnahmen manuell kontrolliert (wenigstens 10 Sekunden je Aufnahme), wäre ein Aufwand von ca. 100 Stunden entstanden. Die Auswertung reduziert somit den Aufwand auf ca. 10%.

Gondelaufnahmen mit vielen Störungen

Ein anderes Beispiel für die Auswertung großer Datenmengen sind Gondelaufnahmen mit vielen Störungen (>> Fledermäuse). Wir haben hierzu als Beispiel einen Datensatz, bei dem je Nacht zwischen 400 und 3000 Aufnahmen aufgezeichnet wurden. Bei diesen handelte es sich um zahlreiche Störungen, hervorgerufen durch ein bei 20 kHz quietschendes Bauteil. Nachdem die automatische Artbestimmung viele der Störlaute auch einer Art der Gruppe Nyctaloid oder Pipistrelloid zugeordnet hatte, wurde nach einem Muster bei der Erkennung der Störungen bzw. die Störaufnahmen gesucht. Also zum Beispiel immer nur ein Ruf erkannt, oder immer nur eine kurze oder sehr lange Aufnahme. Solch ein einfaches Vorgehen erwies sich sehr schnell als nicht ausreichend.

Jedoch konnte basierend auf der Sichtung einer ersten Nacht mit Aufnahmen ein Verfahren entwickeln. Dieses erlaubt es mittels eines Aufwands von 5 bis 10 Minuten je erfasster Nacht an Aufnahmen ein sehr detailliertes Ergebnis der Aktivität zu erhalten. Es macht sich die Tatsache zu Nutze, dass die Störungen nur Bruchstückhaft vermessen werden und daher primär als Sozialrufe der Gattung *Pipistrellus* erkannt werden.

13:18:44	0,49	0	Spec. 0%
13:18:44	0,98	0	Pipistrelloid 95%
13:18:46	0,26	0	
13:18:46	0,39	0	Pipistrelloid 66%
13:18:46	0,36	0	Pipistrelloid 61%
13:18:48	0,26	0	
13:18:48	0,28	0	
13:18:50	0,70	0	Pipistrelloid 90%
13:18:50	0,29	0	Pipistrelloid 71%
13:18:52	1,00	0	Pipistrelloid 85%
13:18:52	1,06	0	Pipistrelloid 65%
13:18:54	0,76	0	Pipistrelloid 73%
13:18:56	0,39	0	Pipistrelloid 70%
13:18:56	0,42	0	Nnoc 82%
13:18:56	0,75	0	Pipistrelloid 90%, Nnoc 87%
13:18:56	0,37	0	Pipistrelloid 62%
13:18:58	0,27	0	
13:18:58	0,45	0	
13:18:58	0,26	0	

Bei dem im Folgenden beschriebenen Vorgehen werden unter Umständen einzelne Rufe, die stark von Störungen überdeckt sind, übersehen. Ebenso können vereinzelt Störaufnahmen ohne Rufe, die in direkter zeitlicher Folge zu echten Rufnahmen liegen, fälschlicherweise als Fledermauslaut mit in die Daten aufgenommen werden. Beide Fälle konnten in einzelnen genauer untersuchten Nächten nicht vorgefunden werden. Es ist davon auszugehen, dass ein Großteil der Aktivität richtig ermittelt werden kann, wie auch die Ergebnisse einiger beispielhaft ausgewerteter Nächte zeigen werden. Das als günstig erkannte Vorgehen ist wie folgt:

1. Automatische Rufsuche und Artbestimmung
2. In bcAdmin die farbige Markierungen von zeitlich zusammengehörigen Aufnahmen aktivieren (zB 20 Sekunden) und die Aufnahmetabelle nach Aufnahmezeit sortieren
3. Eine Nacht auswählen
4. In bcAdmin alle Aufnahme, die als Pipistrelloid bestimmt wurden, ausfiltern (ausblenden)
5. Dann sukzessive je zeitlichem Block eine Aufnahme auswählen und im Datei-Browser (Rufanzeige im bcAdmin) das Sonagramm prüfen
6. Wenn Rufe vorhanden, alle zeitlich zusammenhängenden Aufnahmen mit der Maus auswählen, die Art eingeben (über Schnellastenbelegung) und die Aufnahmen mit dem Attribut „Markiert“ versehen
7. Sind alle Aufnahmeblöcke der Nacht so untersucht, alle nicht „Markierten“ Aufnahmen einblenden (damit werden dann als die mit „Markiert“ versehenen ausgeblendet)
8. Alle Arteinträge löschen (Default Taste 0)
9. Alle Filter aufheben und so alle Aufnahmen anzeigen
10. Die Tabelle langsam von Oben nach unten Scrollen und Lücken in den durch Fettdruck hervorgehobenen Aufnahmeblöcke gegebenenfalls ergänzen (Art +Attribut „Markiert“)
11. Mit der nächsten Nacht bei 4. fortfahren

034739.raw	21:44:36	0,27	1	Nnoc 0%
034740.raw	21:44:36	0,27	1	Nnoc 0%
034741.raw	21:44:36	0,27	1	Nnoc 0%
034744.raw	21:44:38	0,27	1	Nnoc 0%
034743.raw	21:44:38	0,27	1	Nnoc 0%
034745.raw	21:45:08	0,26	0	Nnoc 0%
034746.raw	21:46:14	0,26	1	Nnoc 0%
034749.raw	21:46:22	0,27	1	Nnoc 0%
034747.raw	21:46:22	0,27	1	Nnoc 0%
034748.raw	21:46:22	0,26	0	Nnoc 0%
034750.raw	21:47:24	0,27	0	Nnoc 0%
034751.raw	21:49:40	0,26	1	Nnoc 0%
034753.raw	22:11:10	0,27	1	Nlei 0%
034752.raw	22:11:10	0,27	1	Nlei 0%
034754.raw	22:11:12	0,26	1	Nlei 0%
034755.raw	22:11:14	0,26	1	Nlei 0%
034756.raw	22:11:16	0,26	0	Nlei 0%
034757.raw	22:11:16	0,27	1	Nlei 0%
034758.raw	22:11:18	0,26	0	Nlei 0%
34759.raw	22:26:46	0,26	0	
34760.raw	22:26:48	0,61	4	
34761.raw	22:26:52	0,27	1	
34763.raw	22:26:54	0,49	2	
34762.raw	22:26:54	1,93	23	
34764.raw	22:27:00	0,26	0	
34765.raw	22:27:02	0,53	5	
034768.raw	22:32:42	0,27	1	Nnoc 0%
034766.raw	22:32:42	0,26	1	Nnoc 0%
034767.raw	22:32:42	0,27	1	Nnoc 0%
034769.raw	22:32:44	0,26	1	Nnoc 0%
034770.raw	22:32:44	0,27	1	Nnoc 0%
034771.raw	22:32:46	0,27	1	Nnoc 0%
034773.raw	22:35:40	0,26	1	Nnoc 0%
034775.raw	22:35:40	0,26	0	Nnoc 0%
034774.raw	22:35:40	0,27	1	Nnoc 0%
034772.raw	22:35:40	0,26	0	Nnoc 0%
034776.raw	22:35:40	0,26	1	Nnoc 0%
034777.raw	22:35:42	0,26	0	Nnoc 0%

Auf diese Art und Weise kann ein Aufnahmetag innerhalb von ca. 10 Minuten ausgewertet werden. Damit sind die Daten eines Monats innerhalb eines Arbeitstages durchgearbeitet.

Fazit

Bei geeignetem Vorgehen, angepasst an die vorliegenden Aufnahmen, lassen sich selbst sehr große Datenmengen in überschaubarer Zeit und damit kostengünstig auswerten. Dabei müssen unter Umständen Abstriche bei der Sicherheit der Ergebnisse gemacht werden. Wobei die Rate fehlerhaft bestimmter Arten im Rahmen bleiben wird. Wir gehen von einem Fehler von $\leq 10\%$ beim beschriebenen Vorgehen aus.