

### 3.3 Statistische Methoden

#### 3.3.1 Mittlere Abundanz

Zur quantitativen Beschreibung des Vorkommens einer Art an einer Station wird häufig deren mittlere Abundanz, d.h. die Abundanz arithmetisch gemittelt über alle Proben einer Station, angegeben. Die auf die Fläche bezogene Heterogenität der Besiedlung ist durch die Standardabweichung gegeben. Die Standardabweichung der Besiedlung an einer Station ist ein fester, von Null verschiedener Wert und kein Maß für die Güte der Mittelwertbestimmung. Es kann also nicht das Ziel eines erhöhten Probenumfangs sein, die Standardabweichung einer Art in der Stichprobe zu reduzieren.

Die Schwankungsbreite der Mittelwertschätzung spiegelt sich im Standardfehler

$$SF = \frac{SA}{\sqrt{n}},$$

mit  $SA$  Standardabweichung und  $n$  Umfang der Stichprobe, wider. Bei einer annähernd normalverteilten Schätzung  $MW$ , Mittelwert der Stichprobe, liegt der wahre Mittelwert der Besiedlung mit etwa 70% Wahrscheinlichkeit zwischen den Werten  $MW \pm SF$ .

Ein Maß für die Güte (Präzision, Genauigkeit) der Mittelwertbestimmung ist der relative Standardfehler (index of precision, siehe ELLIOT 1977)

$$D = \frac{SF}{MW} = \frac{SA}{MW\sqrt{n}}.$$

Bei einer größeren Heterogenität der Besiedlung, also einer größeren Standardabweichung, sind daher mehr Proben nötig, um die gleiche Präzision bei der Mittelwertschätzung zu erhalten.

Wie nachzurechnen ist, ist der relative Standardfehler ein Wert zwischen 0 und 1, bzw. 0 und 100%. Fehlerwerte von über 100%, wie z.B. bei KOLBE (1995), beruhen möglicherweise auf einem anderen Ansatz. Nach ELLIOT (1977) ist ein relativer Standardfehler von 20% häufig eine vernünftige Grenze für einen noch zu tolerierenden Fehler bei der Mittelwertbestimmung. Bei einer mittleren Abundanz von beispielsweise 40 Individuen pro Quadratmeter ( $\text{Ind}/\text{m}^2$ ) bedeutet eine Abweichung von 20%, daß die mittlere Abundanz auch 32 oder 48  $\text{Ind}/\text{m}^2$  betragen kann. Allerdings werden von ELLIOT (1977) für diese Fehlerschranke keine Gründe oder Zitate ins Feld geführt.

Für manche ökologischen Aussagen reicht sicher bereits ein relativer Standardfehler von 50%. Damit wäre zumindest die Größenordnung einer mittleren Art-Abundanz zuverlässig bestimmt. Bei einer gemessenen mittleren Abundanz von beispielsweise 40  $\text{Ind}/\text{m}^2$  läge der wahre Wert dann mit hoher Wahrscheinlichkeit zwischen 20 und 60  $\text{Ind}/\text{m}^2$ .

Allerdings hängt die Aussagekraft des relativen Standardfehlers vom Stichprobenumfang und von der Verteilung der Mittelwertschätzung ab. So kann der relative Standardfehler u.U. erst bei einem Stichprobenumfang  $\geq 30$  vernünftig interpretiert werden (siehe auch ELLIOT 1977). Für einen kleinen Stichprobenumfang, gerade wenn die Arten, wie häufig zu verzeichnen ist, einer gehäuften oder

geklumpten Verteilung unterliegen, besitzt dann der Standardfehler nur eine geringere Aussagekraft.

Eine gute Alternative zur Berechnung des Standardfehlers stellt die Bestimmung von Konfidenzintervallen für die Mittelwertschätzung dar (siehe ELLIOT 1977, BORTZ 1993 oder PFEIFER ET AL 1996b). Da dies allerdings auch nur bei größeren Stichproben sinnvoll ist, wird dieser Weg hier nicht weiter verfolgt. Es sei jedoch erwähnt, daß das oben erwähnte Intervall  $[MW - SF, MW + SF]$  als Konfidenzintervall aufgefaßt werden kann.

Der Median ist als Mittelwert zur Charakterisierung der Artabundanzen bei den schwach besiedelten Proben in der Elbe eher ungeeignet. Wird eine Art an einer Station in weniger als der Hälfte der Proben nachgewiesen, so ist der Median stets 0.

Die mittlere Gesamtabundanz pro Probe einer Station liefert eine Zahl, die einen groben Überblick über die Besiedlung an einer Station liefert. Die Überlegungen zu den statistischen Eigenschaften der Schätzung gelten hier analog, wie bei der mittleren Abundanz einer Art. Da die Gesamtabundanz einer Probe eine Summe von ähnlich verteilten Artenabundanzen ist, ist bei der mittleren Gesamtabundanz jedoch ein kleinerer relativer Standardfehler zu erwarten.

Wenn die mittlere Gesamtabundanz mit varianzanalytischen Methoden zum Vergleich von Stationen herangezogen wird, so ist ein tolerierbarer Standardfehler abhängig von dem Verhältnis der Varianzen an den Stationen zu den Differenzen der Stationsmittelwerte. Da jedoch die mittlere Gesamtabundanz nur zur Deduktion von größeren Veränderungen geeignet ist und eine Veränderung des Artenspektrums bei gleichbleibender Individuenzahl gar nicht indizieren kann, wird auf eine weitergehende Betrachtung dieser Größe verzichtet.

### 3.3.2 Arten-Areal-Kurve

Die Aussagekraft einer Besiedlungsanalyse drückt sich auch in ihrer Reproduzierbarkeit aus. Stützt sich eine Analyse auf ein gewisses Arteninventar, so sollte bei einer weiteren Beprobung dieses Arteninventar mit einer gewissen Sicherheit ebenfalls nachgewiesen werden können. Die Frage, welche Fläche im Mittel mindestens beprobt werden muß, um ein bestimmtes Artenspektrum nachzuweisen, wird in PFEIFER ET AL. (1996a) behandelt.

Dort wird ein modellgestützter Ansatz behandelt, der von der realitätsnahen Annahme ausgeht, dass die Individuen einer Art nicht gleichmäßig über die Fläche verteilt sind, sondern räumlich Poissonverteilt sind (siehe auch PFEIFER ET AL. 1992). In dieses Modell gehen als wesentliche Parameter die geschätzten mittleren Abundanzen der einzelnen Arten ein.

Modellhaft nachgestellt wird eine Beprobung, bei der die beprobte Fläche (Anzahl der Proben) sukzessive solange vergrößert wird, bis ein vorher definiertes Artenspektrum vollständig nachgewiesen werden kann. Diese *minimale Beprobungsfläche* (minimal area) ist selber eine Zufallsgröße. Der Erwartungswert dieser Zufallsgröße gibt an, wie groß im Mittel die minimal zu beprobende Fläche ist, um die vorgegebenen Arten zu finden.

Wird dieser Ansatz bei den vorliegenden Elbe-Daten auf eine einzelne Station bezogen, wird die in MINAREA dimensionslose Beprobungsfläche 1 hier mit 0.6 m<sup>2</sup> bzw. 6 Parallelproben identifiziert. Wird dagegen etwa eine Gruppe von beispielsweise 5 Stationen betrachtet, so ist die Bezugsfläche 5 x 0.6 m<sup>2</sup> = 3 m<sup>2</sup>, bzw. wieder 6 Parallelproben. Wegen dieses direkten Zusammenhangs wird die minimale Beprobungsfläche im Folgenden auch *notwendiger Probenumfang* genannt.

Der Zusammenhang von Artenzahl und beprobter Fläche wird durch eine sogenannte Arten-Areal-Kurve beschrieben. Diese Kurven weisen zum Teil die gleiche Charakteristik auf, wie rein empirische Arten-Areal-Kurven (siehe PFEIFER ET AL. 1996a).

Die notwendigen Berechnungen werden mit dem Programm MINAREA (PFEIFER 1997) durchgeführt, eine von den Autoren PFEIFER und BÄUMER verbesserte Version, des in PFEIFER ET AL. (1996a) dokumentierten Algorithmus. Dieses Programm wurde der BfG von den genannten Autoren dankenswerterweise für diese Untersuchung zur Verfügung gestellt.

Eine Anwendung dieses Ansatzes im Zusammenhang mit Besiedlungsanalysen des Makrozoobenthos findet sich in PFEIFER ET AL. (1998).

Das Programm MINAREA erzeugt für die Abundanzen von 8 fiktiven Arten

| Art      | 1 | 2  | 3 | 4  | 5   | 6  | 7 | 8 |
|----------|---|----|---|----|-----|----|---|---|
| Abundanz | 2 | 34 | 1 | 14 | 156 | 26 | 6 | 4 |

eine wie in Abbildung 3.3-1 dargestellte Ausgabe.

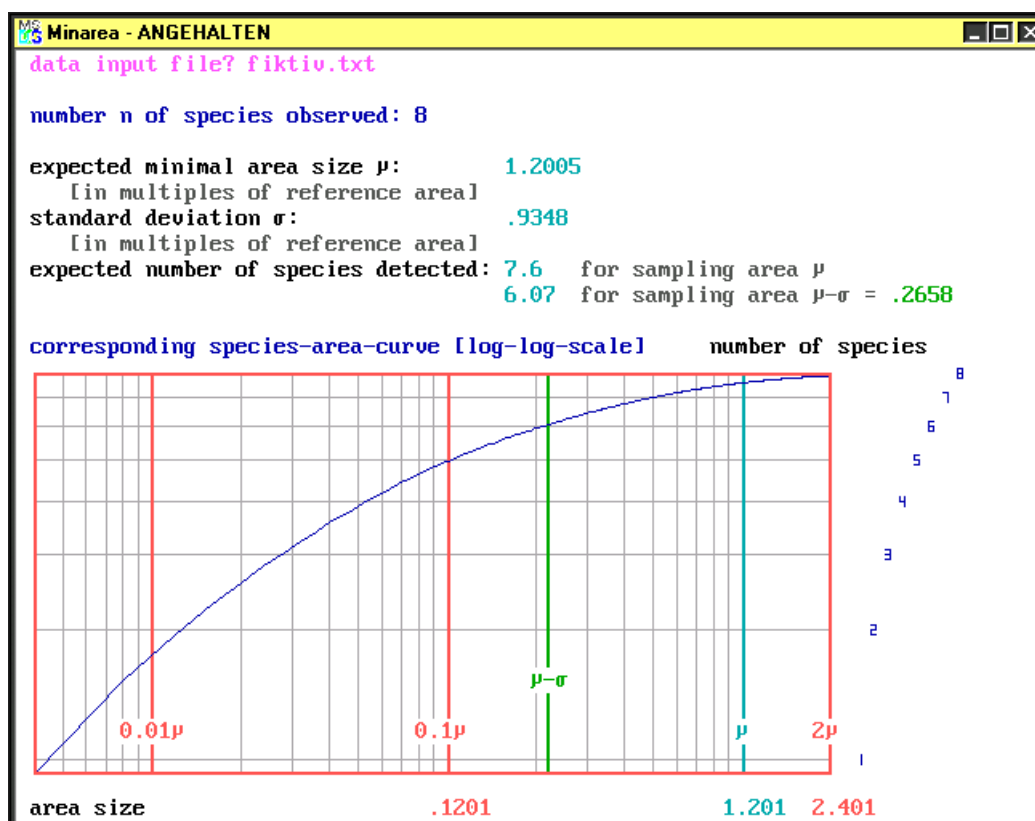


Abb. 3.3-1: Ausgabe des Programms MINAREA (PFEIFER 1997): reference area hier 6 Parallelproben pro Station, expected minimal area size hier  $1.2005 * 6 \approx 7.2$  Parallelproben pro Station.

Um die beobachteten 8 Arten wieder zu finden, ist die Beprobungsfläche im Mittel mindestens auf 120,05% (expected minimal area size) der ursprünglichen Beprobungsfläche auszudehnen. Übertragen auf die Anzahl der Proben bedeutet das, dass etwa 20% mehr Parallelproben als zuvor genommen werden müssen, um genau diese 8 Arten wieder zu finden.

Die Arten-Areal-Kurve in doppelter logarithmischer Auftragung gibt an, bei welcher Probenfläche im Mittel wie viele Arten gefunden werden.

Bei diesem Ansatz bleibt unberücksichtigt, dass bei einer erneuten Beprobung evtl. noch weitere, andere Arten gefunden werden.

Bei der Anwendung dieses Ansatzes wird das Modell nicht mit den mittleren Abundanz in Ind /m<sup>2</sup> parametrisiert, sondern mit der Gesamtanzahl der gefundenen Individuen einer Art an einer Station oder einer Gruppe von Stationen.



### 3.3.3 Ordinationsverfahren

Die Besiedlungsdaten werden mit der Hauptkomponentenanalyse (*Principle Component Analysis*, PCA) und der Korrespondenzanalyse (*CO*rrespondence Analysis, COA) analysiert (GAUCH 1982, JONGMAN ET AL. 1995 oder BRAAK & VERDONSCHOT 1995). Bei diesen Methoden wird versucht, die wesentliche Information in den Daten in einigen wenigen Komponenten rechnerisch zusammenzufassen. Oft lassen sich die wichtigsten Charakteristika einer Besiedlungsstruktur bereits mit den beiden ersten, mit dem höchsten Informationsgehalt versehenen rechnerischen Komponenten (Hauptkomponenten bzw. Korrespondenzachsen) darstellen. Ein Ergebnis dieser sogenannten Ordinationsverfahren ist ein 2-dimensionaler Ordinationsplot (oder -diagramm), in dem die einzelnen Proben gemäß den wichtigsten Aspekten ihrer Besiedlung als Punkte bezüglich der Komponenten aufgetragen (ordiniert) sind.

Die Ordinationsergebnisse der Proben einer Station werden gemittelt und als Stationsmittelwert mit in das Ordinationsdiagramm aufgenommen. Über die Stationsmittelwerte können die einzelnen Stationen miteinander verglichen werden. Dabei gilt stets, dass die räumliche Nähe oder Distanz zweier Punkte (Proben oder Stationen), die Ähnlichkeit bzw. Unterschiedlichkeit der Besiedlung dieser Proben ausdrückt. Die Heterogenität der Besiedlung einer Station drückt sich in der Größe des „Sterns“ aus (siehe Abb. 3.3-2).

Die Abstände zwischen den Punkten im Ordinationsdiagramm sind genau wie die Unterschiede von Besiedlungsstrukturen nicht absolut, sondern immer relativ auf die Auswahl der Proben bezogen. Würde man beispielsweise die Proben der fünf limnischen Stationen aus Abbildung 3.3-2 mit Stationen aus dem marinen Bereich vergleichen, so sind die Unterschiede zwischen den fünf limnischen Stationen in Relation zu den ganz anderen marinen Stationen sehr viel kleiner. Die Sterne der limnischen Stationen sind dann wesentlich kleiner und die Abstände zwischen den Mittelwerten der limnischen Stationen geringer (vgl. auch Abb. 4.4-1 mit Abb. 4.4-2).

Beiden Ordinationsverfahren ist gemein, dass sie aus den vieldimensionalen Besiedlungsdaten die Hauptkomponenten bzw. Korrespondenzachsen extrahieren, die am meisten zur Erklärung der Variation in den Daten beitragen. Die maximal mögliche Anzahl der Komponenten (Achsen) ist gerade das Minimum aus Arten- und Probenzahl. Die Achsen sind nach der Größe des Informationsgehalts sortiert. Hier werden stets nur die beiden ersten Hauptachsen betrachtet. Der Informationsgehalt dieser beiden Achsen läßt sich in Prozent der Gesamtvariation ausdrücken. In dem Beispiel in Abbildung 3.3-2 erklären die beiden Komponenten 59% der Variation (= relative Summe der ersten beiden Eigenwerte), der Rest ist auf die in diesem Beispiel vorliegenden 16 übrigen Hauptkomponenten verteilt.

Bei beiden Ordinationsverfahren werden die Abundanzwerte berücksichtigt. Allerdings werden die Rohdaten zur Durchführung von PCA und COA mit der Funktion  $\ln(x + 1)$  transformiert, um den Einfluß der überproportional großen Abundanzen zu dämpfen (siehe auch CLARKE ET AL. 1994). Bei der PCA wird die Kovarianzmatrix analysiert.

Ein wesentlicher Unterschied zwischen PCA und COA liegt in der Gewichtung seltener Arten. Seltene und damit i.d.R. auch individuen schwache Arten werden bei der PCA kaum berücksichtigt, das Hauptgewicht liegt bei den individuenreichen Arten. Bei der COA hingegen haben seltene Arten ein viel größeres Gewicht. Einzelfunde können so einen außerordentlich großen Einfluß auf das Ordinationsergebnis haben. Daher muß bei der COA das Artenspektrum sinnvoll eingeschränkt werden, um fundierte Ergebnisse zu erhalten.

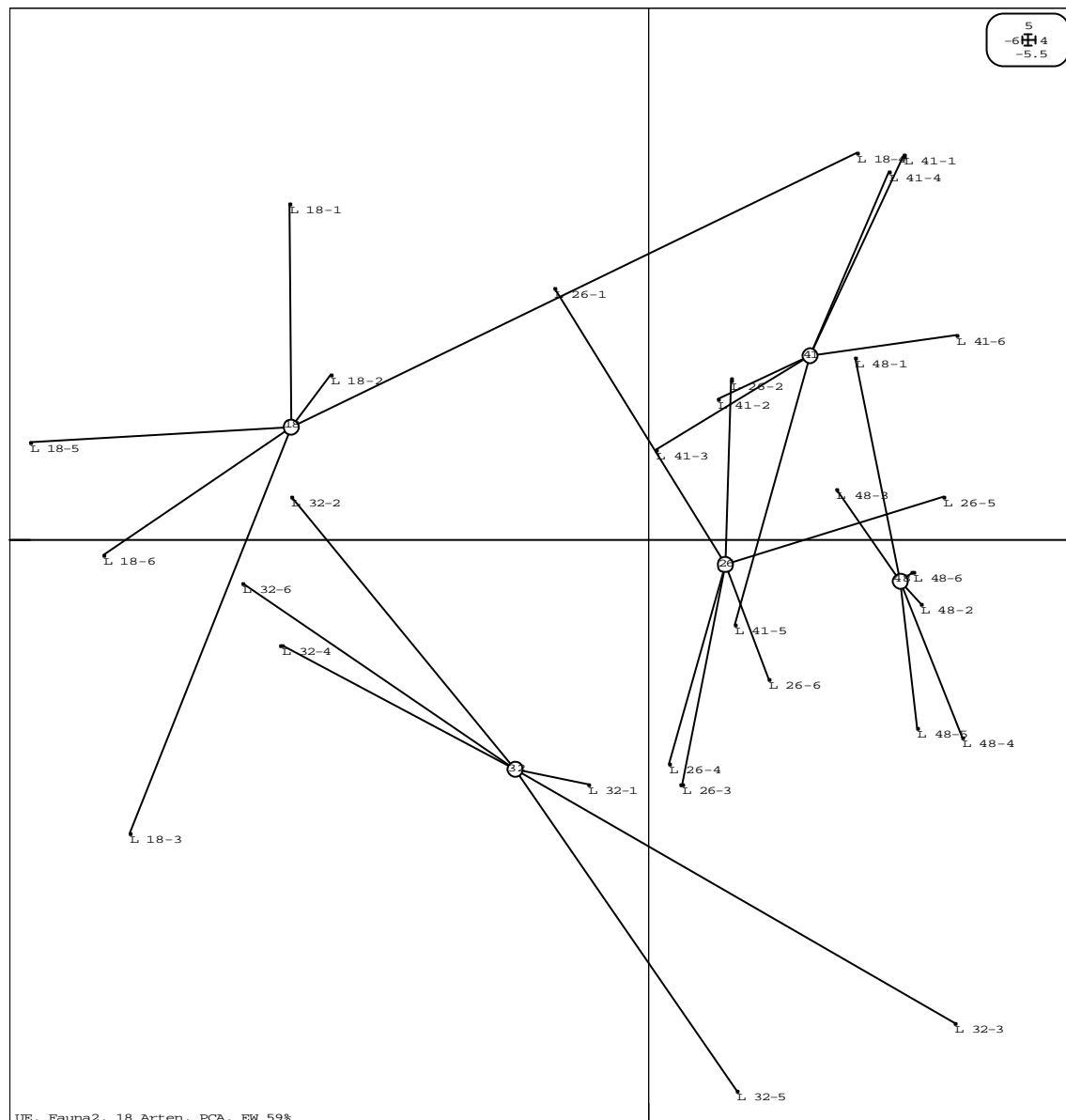


Abb. 3.3-2: Ordinationsplot (Sterndiagramm) der Fahrinnenstationen des limnischen Bereichs: die Ordinationsergebnisse der Proben (kleine Punkte mit der Bezeichnung der Probe) sind mit den Stationsmittelwerten (Kreise mit Stationsnummer) durch Strecken graphisch verbunden (Sterne), weitere Erläuterungen siehe Text.  
 (UE = Untere Elbe, Fauna2 = interne Bezeichnung der Datenquelle, 18 Arten = Anzahl der Arten, die in diese Ordination eingehen, EW 59% = relative Summe der ersten beiden Eigenwerte)

Die hier durchgeführten Berechnungen sowie Erfahrungen bei anderen Untersuchungen zeigen, dass

PCA und COA bei Makrozoobenthosuntersuchungen qualitativ oft sehr ähnliche Ergebnisse liefern. Das lässt den Schluss zu, dass sich Besiedlungsunterschiede genau so in den stetigen wie in den selteneren Arten zeigen. Weiterhin kann daraus geschlossen werden, dass sich Störungen oder Veränderungen der Biocoenose allein an den Abundanzen der „stetigeren“ (mehr als wenige Funde) Arten diagnostizieren lassen.

Die PCA und die COA werden mit dem Programmpaket ADE-4 (THIOULOUSE ET AL. 1999, THIOULOUSE ET AL. 1997) durchgeführt. Hiermit werden auch die wesentliche Bestandteile der Ordinationsplots generiert. Da PCA und COA oft sehr ähnliche Ergebnisse liefern, wurden zur besseren Vergleichbarkeit alle Besiedlungsanalysen mit der PCA durchgeführt. Damit entfallen zum Teil aufwendige Untersuchungen über ein für die COA geeignetes Artenspektrum, da seltene Arten bzw. Einzelfunde einen großen Einfluss auf die Ordination haben. Stichprobenartig wurden jedoch vergleichbare Berechnungen mit der COA durchgeführt, womit dann in etwa die gleichen Resultate wie mit der PCA erzielt werden konnten.

Bei dem Stationsvergleich werden in einem Ordinationsplot die relativen Differenzen der Stationsmittelwerte mit der Streuung (Größe und Form der Sterne) an den einzelnen Stationen in Beziehung zu einander gesetzt. In der Abbildung 3.3-2 beispielsweise sind die Abstände von L26 zu L41 bzw. L32 in etwa gleich. Die Streubereiche (Strahlen der Sterne) zwischen L26 und L41 überschneiden sich jedoch, wogegen das bei L26 und L32 nicht der Fall ist. Daher wird man den Unterschied zwischen L32 und L26 als größer als den zwischen L26 und L41 einstufen.

### Signifikanztests

Die oben beschriebene Art des graphischen Mittelwertvergleichs wird rechnerisch durch die Varianzanalyse bewerkstelligt. Daher werden die Ordinationsergebnisse auf den beiden ersten Achsen mittels Varianzanalyse auf eine signifikante Abhängigkeit von der Station untersucht. Dabei werden zwei Stationen als *signifikant* unterschiedlich bezeichnet, wenn die Wahrscheinlichkeit für die Nullhypothese (p-Wert) bei mindestens einer Achse  $< 5\%$  beträgt. Das entspricht dem allgemein üblichen Signifikanzniveau (siehe KÖHLER ET AL. 1996). „Innovative Forschungen in einem relativ jungen Untersuchungsgebiet, bei den die Folgen einer fälschlichen Annahme ... vorerst zu vernachlässigen sind“, rechtfertigen durchaus ein höheres (= schwächeres) Signifikanzniveau (siehe BORTZ 1993, SACHS 1978). Daher wird hier in Ergänzung zu einem signifikanten Unterschied der Begriff eines *deutlichen* Unterschieds verwendet, sobald der p-Wert unterhalb der 20% Schranke liegt. Die p-Werte zu den multiplen Mittelwertvergleichen werden mit dem post-hoc-Test von Tuckey berechnet (siehe KÖHLER ET AL. 1996).

Für einen Test auf Ähnlichkeit von Stationen müsste der hier naturgemäß nicht bestimmbare Fehler zweiter Art berechnet werden. Statt dessen wird in diesem Bericht von großer bzw. deutlicher Ähnlichkeit gesprochen, sobald der p-Wert (die Wahrscheinlichkeit dafür, dass zwei Stationen gleich sind) für beide Achsen über 95% bzw. 80% liegt. Es sei jedoch darauf hingewiesen, dass es sich



hierbei nicht um statistische Fachtermini handelt, sondern nur um ein Darstellung von hohen p-Werten, die einen deutlichen Hinweis auf Ähnlichkeiten von Stationen liefern.

Weiterhin sei darauf hingewiesen, dass bei den durchgeführten Varianzanalysen die Voraussetzung der Normalverteilung (Shapiro-Wilks-W-Test, siehe ROYSTON 1982) nicht immer erfüllt ist. Dadurch wird die Genauigkeit einzelner p-Werte beeinträchtigt. Dies stellt aber im Zusammenspiel mit anderen, notwendigen Schritten, die die Komplexität der Besiedlungsdaten reduzieren, keine weitere Einschränkung der Güte der Aussagen dar. Für die alternative Anwendung des nicht parametrischen Kruskal-Wallis-Test in Ergänzung mit dem post-hoc-Test von Nemenyi (siehe KÖHLER ET AL 1996) bestand daher keine zwingende Notwendigkeit

Tab. 3.3-1: Übersicht über die verwendeten Signifikanzniveaus und Ähnlichkeitsangaben in diesem Bericht.

|   |                         |                     |
|---|-------------------------|---------------------|
| <b>Angaben für die Unterschiede:</b>                |                         |                     |
| Hoch (oder sehr) signifikant: $p < 0.01$            | signifikant: $p < 0.05$ | deutlich: $p < 0.2$ |
| jeweils in mindestens <i>einer</i> Ordinationsachse |                         |                     |
| <b>Angaben für die Ähnlichkeiten:</b>               |                         |                     |
| Sehr groß: $p > 0.99$                               | Groß: $p > 0.95$        | deutlich: $p > 0.8$ |
| jeweils in <i>beiden</i> Ordinationsachsen          |                         |                     |

Die Besiedlungsstrukturen wurden mit der PCA und anschließender Varianzanalyse nacheinander mit 3, 4, 5 und 6 Parallelproben untersucht. Dabei wurden in der Reihenfolge der Aufnahme stets die *ersten* 3, 4, bzw. 5 der 6 Parallelproben (siehe Anlage 8-2 und 8-3) berücksichtigt. Damit kann ein Eindruck davon gewonnen werden, wie viele Parallelproben benötigt werden, um Besiedlungsstrukturen signifikant unterscheiden zu können. Zur Durchführung einer Varianzanalyse sind aus mathematischen Gründen mindestens 3 Proben pro Station notwendig.

Wenn zwei Stationen unterschiedlich sind, können bei einer Vergrößerung des Probenumfangs kleinere p-Werte *erwartet* werden, was allerdings nicht zwingend der Fall sein muß. Allerdings erfüllt die hier vorgenommene Variation der Proben genau diese Erwartung. D.h., ein Unterschied, der mit 3 Proben als signifikant nachgewiesen wurde, ist auch bei 4,5 oder 6 Proben signifikant. Die Ergebnisse liefern daher einen deutlichen Hinweis auf einen notwendigen Probenumfang

Für die Ähnlichkeit von Stationen kann eine solche Erwartung nicht formuliert werden, da es sich hierbei nicht um einen Signifikanztest handelt. Daher kann auch keine Mindestprobenzahl zum

Nachweis von Ähnlichkeiten bestimmt werden.

Es muß auch darauf hingewiesen werden, dass die Ergebnisse für 3, 4 oder 5 Parallelproben von der speziellen Auswahl der Proben abhängen. Eine andere Auswahl könnte z.B. besagen, dass ein bestimmter Unterschied zwischen zwei Stationen nicht erst mit 5 sondern bereits mit 4 Proben signifikant ist. Ein Programm, das die notwendige Stichprobenanzahl zumindest für Teilaspekte unabhängig von dieser Reihenfolge berechnen kann (z.B. NCSS 1999), stand für die vorliegende Untersuchung noch nicht zur Verfügung.

Alle Berechnungen zur Varianzanalyse wurden mit dem Programm STATISTICA (STATSOFT 1997) durchgeführt.

Eine Alternative zu den oben beschriebenen Testverfahren ist ein Monte Carlo Permutationstest nach (MANLY 1991). Dabei wird ähnlich wie bei der Varianzanalyse untersucht, ob eine signifikante Abhängigkeit der Besiedlung von der Stationszugehörigkeit nachweisbar ist.

Dazu wird die Gesamtvariation zwischen den Stationen berechnet. Dieser Wert wird verglichen mit entsprechenden Gesamtvariationen, die sich ergeben, wenn die Proben zufällig (daher *Monte Carlo Methode*) auf die Stationen verteilt werden. Ergibt die tatsächliche Gesamtvariation eine signifikante Abweichung von den „zufälligen“ Gesamtvariationen, so kann von einer signifikanten Abhängigkeit der Besiedlung von den Stationen gesprochen werden.

Dieses Testverfahren ist inhaltlich nicht an eine Ordination gebunden. Allerdings ist dessen rechen-technischen Durchführung mit dem Programmpaket ADE-4 (THIOULOUSE ET AL 1999, THIOULOUSE ET AL 1997) mit der Berechnung einer PCA oder COA gekoppelt. Der Permutationstest wird hier aufgeführt, um die auf den Ordinationen beruhenden Signifikanztests abzusichern.