

## 2. Daten

Da die Anforderungen an die Daten je nach Untersuchungsmethode variierten, wurden im Verlauf dieser Arbeit mehrere verschiedene Datensätze verwendet. Sie unterscheiden sich hinsichtlich der Länge der Zeitreihe (Anzahl der vorhandenen Jahre) und des Intervalls voneinander. Grundsätzlich kamen als Prediktanden die Temperatur sowie der Niederschlag und als Prediktoren Feldverteilungen des Bodendrucks sowie des 500-hPa-Geopotentials zur Anwendung. In einigen Fällen mußten vorhandene Lücken durch Interpolation geschlossen werden. Im folgenden sollen sowohl die Herkunft der Daten als auch die Vorgehensweise bei ihrer Bearbeitung dargelegt werden.

### 2.1 Temperatur

Sämtliche in dieser Arbeit verwendeten Temperaturdaten sind den Aufzeichnungen der Klimastation Berlin-Dahlem entnommen. Überwiegend wurden die monatlichen Klimamittelwerte nach *Riemer* (1976) verwendet. Hier findet man auch Informationen zu den verschiedenen Standorten der Station und zur Vorgehensweise in den Fällen, wenn keine Dahlemer Meßwerte vorlagen. Diese Reihe wurde seither weiter fortgeführt, wobei anzumerken ist, daß der Standort Podbielskiallee 62 im Juli 1997 nochmals verlegt werden mußte. Er befindet sich seitdem im nahegelegenen Botanischen Garten. Eine offizielle Anpassung der Meßreihe wurde bislang nicht vorgenommen, jedoch stellte *Gebert* (2000) eine befriedigende Übereinstimmung der Meßreihen beider Stationen fest. Zu einer weiteren, die Kontinuität der Daten möglicherweise beeinflussenden Änderung, kam es im Juni 1986. Die bis dahin üblichen Klimatermine 7, 14 und 21 Uhr wahrer Ortszeit wurden auf 7.30, 14.30 und 21.30 MEZ verschoben. Insgesamt umfaßt der Zeitraum, für den die monatlichen Klimamitteltemperaturen benutzt wurden, die Jahre 1909 bis 1999.

Für die Spektralanalysen wurden tägliche Daten benötigt. In diesem Fall kamen die wahren Tagesmittel der Station zur Anwendung, da sie entgegen den Klimatagesmitteltemperaturen nach *Riemer* bereits in digitaler Form vorlagen. Die Vorgehensweise bei ihrer Berechnung wurde im Laufe des Untersuchungszeitraumes von 1958 bis 1997 einmal leicht abgeändert: In den Jahren von 1958 bis März 1991 wurde das Mittel der stündlichen, in den darauffolgenden Jahren das Mittel der halbstündigen Meßwerte bestimmt. Die Kontinuität der Meßreihe dürfte dadurch jedoch nicht gefährdet sein, da der Unterschied beider Mittelungsverfahren zumeist äußerst gering ist.

### 2.2 Niederschlag

Beim Niederschlag basieren sämtliche verwendeten Zeitreihen auf den ebenfalls von *Riemer* (1973) veröffentlichten täglichen Niederschlagsmengen von Berlin-Dahlem seit 1909. Für die Fortführung dieser Reihe bis 1999 gilt das schon für die Temperatur Gesagte. Aus diesen Werten wurden sowohl die Monatssummen als auch die Anzahl der Tage mit meßbarem Niederschlag in jedem Monat konstruiert, d.h. in beiden Fällen jeweils zwölf Zeitreihen. Die täglichen Werte selbst wurden hingegen nicht unmittelbar verwendet. Bedingt durch die Kriegswirren konnten die Tageswerte vom 27. April 1945 bis zum 5. Mai 1945 weder gemessen noch durch Werte einer nahegelegenen Station

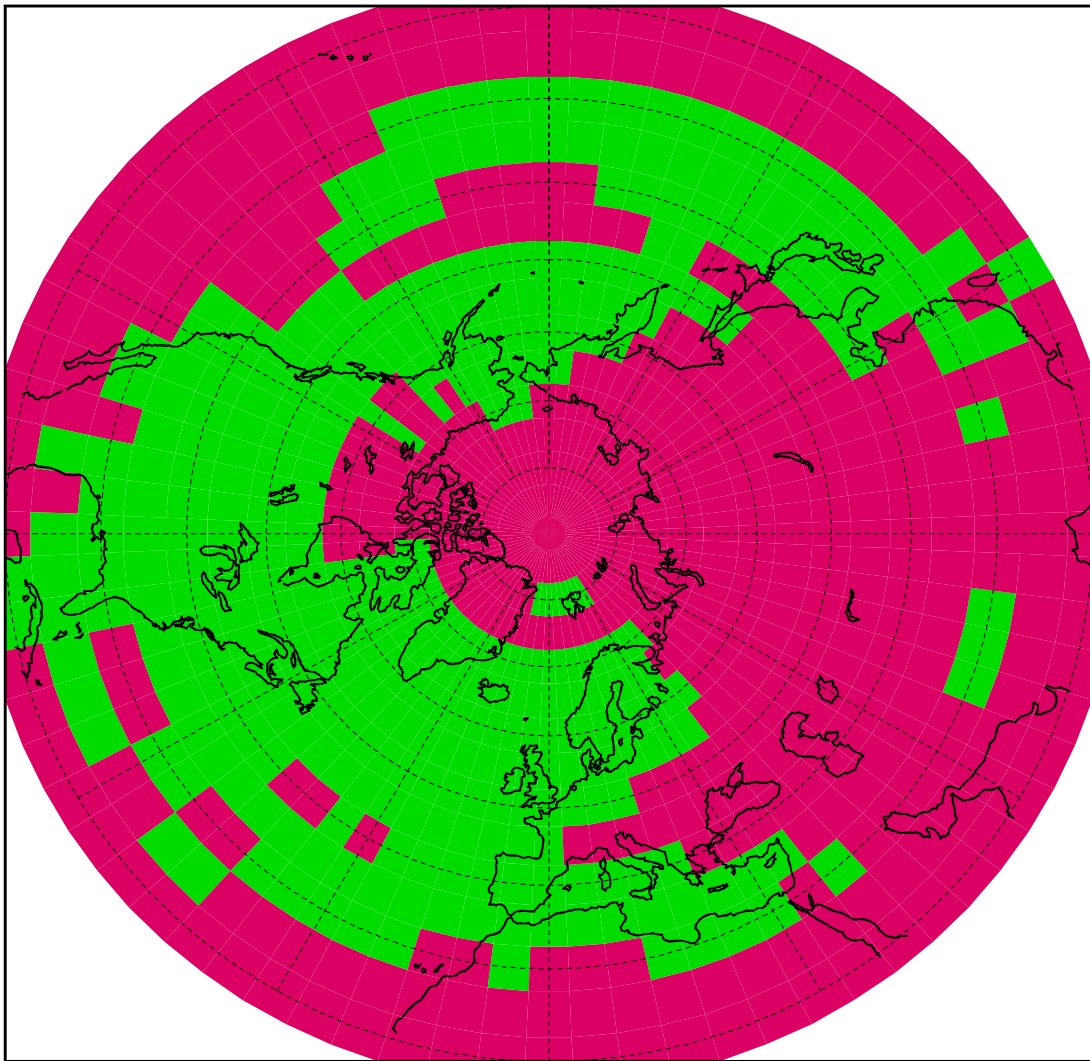
ersetzt werden. Im Gegensatz zur Temperatur weisen diese Daten daher eine Lücke auf. *Riemer* geht davon aus, daß der Wert für den 6. Mai die Gesamtmenge für den genannten Zeitraum angibt. Um trotzdem eine lückenlose Untersuchung durchführen zu können, wurde in dieser Arbeit demgegenüber davon ausgegangen, daß es während der neun betroffenen Tage vor dem 6. Mai trocken blieb. Der daraus möglicherweise resultierende Fehler ist jedoch vermutlich eher gering, da jeweils nur ein einzelnes Jahr in den Kalendermonaten April und Mai betroffen ist. Die in Betracht kommenden Untersuchungen wurden aber für einen 90-jährigen Zeitraum durchgeführt.

### 2.3 Feldverteilungen des Bodendrucks und des 500-hPa-Geopotentials

Jene Untersuchungen, bei denen die Feldverteilungen beider Druckniveaus verwendet wurden, beziehen sich auf Grund der Datenverfügbarkeit auf die Zeit nach dem Ende des Zweiten Weltkrieges. Im Falle der PCR wurden die monatlichen Mittelwerte des Bodendrucks und des 500-hPa-Geopotentials für den Zeitraum 1946 bis 1999 verwendet. Sie liegen für den Bereich 20 bis 90 Grad Nord in Form von Gitterpunktwerten im Abstand von 5 zu 5 Grad vor. Dabei stammen die Daten der Jahre 1946 bis 1994 vom Max-Planck-Institut für Meteorologie in Hamburg. Sie wurden von *Dettmann* (2000) übernommen und durch die Werte des Datensatzes „ds085.1“ vom „National Center for Atmospheric Research (NCAR)“ bis hin zum Jahr 1999 ergänzt. Diese Daten stehen im Internet kostenlos zur Verfügung.

Beide Datensätze wurden zunächst auf erkennbar fehlerhafte Einzelwerte hin untersucht. Hierfür wurde für jeden Gitterpunkt getestet, ob der Wert außerhalb eines vorher festgelegten Intervalls liegt. Zusätzlich wurden die Druckdifferenzen zu den ihn umgebenden Gitterpunkten berechnet, um die Größe der Gradienten abzuschätzen. Wurde dabei ein ebenfalls vorher festgelegter Schwellenwert nicht überschritten, so wurde von einem korrekten Wert ausgegangen. In einem zweiten Schritt wurde die Kontinuität beider Zeitreihen geprüft. Hierfür wurden die Druckwerte des Jahres 1994, das in beiden Reihen vorlag, miteinander verglichen. Die Anwendung beider Prüfverfahren deckte insgesamt nur wenige Schwächen in den vorliegenden Datensätzen auf. Die Kontinuitätsprüfung zeigte jedoch teils größere Diskrepanzen von mehr als 3 hPa zwischen den Werten beider Bodendruckfelder an einzelnen Gitterpunkten. Diese traten vorwiegend in Regionen auf, in denen der Luftdruck auf Meereshöhe reduziert werden muß, d.h. in hochgelegenen Gebieten wie dem Himalaja. Da ausschließlich die Gradienten der von NCAR stammenden Daten in diesen Gebieten auffällig hoch sind, sind allein hier Probleme bei der Datenreduktion zu vermuten. Somit ist davon auszugehen, daß nur in den letzten fünf Jahren der Zeitreihe, also bei einem relativ kleinen Teil der Gesamtlänge, teils fehlerhafte Werte verwendet wurden. Dennoch kann es an den betroffenen Gitterpunkten zu einer Verfälschung der Ergebnisse kommen.

Die für die Spektralanalyse benötigten täglichen Daten des Bodendrucks und des 500-hPa-Geopotentials im Zeitraum 1958 bis 1997 stammen ebenfalls von NCAR. Es handelt sich um kostenpflichtige 00z-Gitterpunktwerte, welche persönlich übersendet wurden. Sie standen ursprünglich für die gesamte Erde in 2,5 Grad Abständen zur Verfügung, verwendet wurde jedoch nur der Bereich 20 bis 90 Grad Nord in Schritten von jeweils 5 Grad, d.h. die außertropische Region der Nordhalbkugel.



**Abbildung 2.1:** Darstellung der Vollständigkeit des Datensatzes „ds010.1“ von NCAR für das Gebiet 20 bis 90 Grad nördlicher Breite. Für Gitterpunkte im Zentrum der grünen Quadranten liegen monatliche Mittelwerte des Bodendruckes für den gesamten Zeitraum 1909-1998 vor.

Die Länge des vorhandenen 54-jährigen Datensatzes war für einige Untersuchungen bei weitem nicht ausreichend. Da für die Zeit vor 1946 keine 500-hPa-Geopotentialdaten zur Verfügung stehen, konnte nur auf den Bodendruck zurückgegriffen werden. Diese Werte sind dem ebenfalls von NCAR im Internet kostenlos zur Verfügung gestellten Datensatz „ds010.1“ entnommen, welcher bis ins 19. Jahrhundert zurückreicht. Somit war es entsprechend des Beginns der Dahlemer Klimareihe im April 1908 möglich, monatliche Mittelwerte des Bodendruckes für den 90-jährigen Untersuchungszeitraum von 1909 bis 1998 zu verwenden. Die Qualität der Daten bis zum Jahre 1977 wurde von *Trenberth und Paolino* (1980) überprüft, die viele Korrekturen durchführten. Leider weist der Datensatz „ds010.1“ bis in die Nachkriegszeit des Zweiten Weltkrieges hinein große Lücken auf, so daß es in diesem Fall nicht möglich war, eine vollständige Feldverteilung mit 5 Grad Auflösung zu verwenden. Abbildung 2.1 stellt jene 389 Gitterpunkte dar, welche für den gesamten Untersuchungszeitraum einen monatlichen Mittelwert des Bodendruckes zwischen 975 und 1055 hPa liefern. Die Datenlücken befinden sich hauptsächlich im Seegebiet westlich von Nordamerika sowie

insbesondere im gesamten sibirischen und asiatischen Raum. Einige kleinere Lücken über dem Atlantik und dem Mittelmeerraum dürften auf Grund der hohen Interkorrelationen (hohe Korrelationen zwischen nahe beieinander liegenden Gitterpunkten auf Grund der Stetigkeit der Feldverteilung) vermutlich keinen nachhaltigen Informationsverlust bedingen.

Ein großes Problem stellte die Tatsache dar, daß der Monat Dezember 1944 im verwendeten Datensatz komplett fehlte. Es konnte nur durch Einfügung der Werte des Datensatzes „Global Climate Change Data“ für diesen Monat behoben werden. Diese Daten sind der „World Climate Disc“ von *Chadwyck Healey* („Climatic Research Unit at the University of East Anglia“) entnommen. Im Gegensatz zum NCAR-Datensatz liegen hier lediglich für jeden zehnten Längengrad Werte vor – die Breitenkreise hingegen im Abstand von 5 Grad. Um nicht noch eine große Zahl weiterer Gitterpunkte einzubüßen, wurden die Werte für sämtliche relevanten Gitterpunkte durch Interpolation bestimmt. Zumeist kam dabei ein bilinearer Ansatz zur Anwendung. Der Funktionswert am Gitterpunkt  $P(\lambda, \varphi)$  berechnet sich aus den Werten von vier, in der Regel nahegelegenen Gitterpunkten, die in Form eines Quadrats bzw. Rechtecks um den Zielpunkt angeordnet sein müssen, wie folgt (vergl. z.B. *Böhm*, 1999):

$$f(P) = A \cdot (\lambda - \lambda_1) + B \cdot (\varphi - \varphi_1) + C \cdot (\lambda - \lambda_1)(\varphi - \varphi_1) + D. \quad (2.1a)$$

Dabei ist  $\lambda$  die geographische Länge,  $\varphi$  die Breite. Die Koeffizienten ergeben sich zu:

$$A = \frac{f(\lambda_2, \varphi_1) - f(\lambda_1, \varphi_1)}{\Delta\lambda},$$

$$B = \frac{f(\lambda_1, \varphi_2) - f(\lambda_1, \varphi_1)}{\Delta\varphi},$$

$$C = \frac{[f(\lambda_2, \varphi_2) - f(\lambda_1, \varphi_2)] - [f(\lambda_2, \varphi_1) - f(\lambda_1, \varphi_1)]}{\Delta\lambda \cdot \Delta\varphi},$$
(2.1b)

$$D = f(\lambda_1, \varphi_1), \quad \text{mit } \lambda_1 \leq \lambda \leq \lambda_2 \quad \text{und} \quad \varphi_1 \leq \varphi \leq \varphi_2.$$

Standen die erforderlichen vier Gitterpunkte nicht zur Verfügung, so wurden lediglich zwei umliegende Werte zur linearen Interpolation herangezogen. Dabei wurden stets Gitterpunkte verwendet, die auf dem selben Längen- bzw. Breitenkreis liegen wie der, den es zu bestimmen galt. In einigen Fällen schien ein Eingreifen des Meteorologen trotz vollständig vorliegender Daten zweckmäßig zu sein. Lag die Achse eines Hoch- bzw. Tiefdruckgebietes direkt über einem Gitterpunkt, so wurde dieser ebenfalls nur linear interpoliert, um der Realität möglichst genau zu entsprechen.

Während die Verwendung des lückenhaften Datensatzes bei mehreren Verfahren methodisch problemlos durchführbar war, konnte dieser im Falle einiger der getesteten Analogverfahren nicht unverändert benutzt werden. Für ihre Zwecke sollten möglichst großräumige Feldverteilungen des Gradienten betrachtet werden, zu deren Berechnung die vollständige Feldverteilung der Zustandsvariablen benötigt wird. Wie Abbildung 2.1 zeigt, fehlen im nordatlantisch-europäischen Raum nur wenige Daten. Daher wurde letztlich das Gebiet von 35 bis 70 Grad Nord und 80 Grad West bis 25 Grad Ost für diese speziellen Analogverfahren ausgewählt. Die Fehlwerte an den insgesamt elf betroffenen Gitterpunkten wurden ebenfalls durch bilineare Interpolation (Gl. 2.1) ersetzt. Sie verteilen sich auf sechs Kalendermonate des Zeitraumes November 1920 bis April 1947. Es sei noch einmal explizit auf die fehlenden Daten im Dezember 1944 hingewiesen. Für diesen Monat mußten natürlich wesentlich mehr Werte durch Interpolation abgeschätzt werden.

Eine Qualitätsprüfung für dieses nordatlantische Fenster, derart wie sie schon für den Datensatz „ds085.1“ durchgeführt wurde, deckte ungewöhnlich hohe Gradienten im grönländischen Raum auf. Wie ein Vergleich mit den Karten der *Berliner Wetterkarte* (1961-1998) zeigte, weichen die Werte direkt über dem Binneneis seit 1994 häufig stärker voneinander ab. Da davon auszugehen ist, daß es sich um Reduktionsfehler von NCAR handelt (wie schon beim Datensatz „ds085.1“; s.o.), wurde eine Korrektur dieser Daten für den Zeitraum 1994 bis 1998 nach der *Berliner Wetterkarte* durchgeführt. Diese Änderungen wurden auf den kompletten, 389 Gitterpunkte umfassenden nordhemisphärischen Datensatz übertragen.

Tabelle 2.1 auf der folgenden Seite ermöglicht einen vollständigen Überblick darüber, welche Datensätze und Untersuchungsgebiete (bei den Prediktoren) für die jeweilige Analysemethode verwendet wurden. Zusätzlich ist ihr zu entnehmen, in welchen Fällen der Temperatur der Niederschlag als Prediktand hinzugefügt wurde.

Untersuchungsmethode	Untersuchungszeitraum	Herkunft der Prediktoren & Untersuchungsgebiet	Prediktand Temperatur	Prediktand Niederschlag
Autokorrelations-Spektralanalyse	1958 – 1997		wahre Tagesmittel Berlin-Dahlem	
Kreuzspektrumanalyse	1958 – 1997	tägliche Werte des Bodendrucks & 500-hPa-Geop. Gitterpunkte in 5 zu 5° für 20 bis 90°N von NCAR	wahre Tagesmittel Berlin-Dahlem	
Analogverfahren: AFREG-, Ensemble- und „Mehrfach-Analog-Modell“	1909 – 1998	Dataset 010.1 von NCAR monatliche Mittel des Bodendrucks in 5 zu 5° für das Gebiet 35 bis 70°N & 80°W bis 25°E	monatliche Klimamittel Berlin-Dahlem	
Analogverfahren: „Vielfach-Analog-Modell“	1909 – 1998	Dataset 010.1 von NCAR monatliche Mittel des Bodendrucks in 5 zu 5° für 20 bis 90°N	monatliche Klimamittel Berlin-Dahlem	
PCR	1946 – 1999	bis 1994 MPI & Dataset 085.1 von NCAR monatliche Mittel des Bodendrucks & 500-hPa-Geop. in 5 zu 5° für 20 bis 90°N	monatliche Klimamittel Berlin-Dahlem	Anzahl der Tage pro Monat mit meßbarem Niederschlag Berlin-Dahlem
Multiple lineare Regression	1909 – 1998	Dataset 010.1 von NCAR monatliche Mittel des Bodendrucks in 5 zu 5° für 20 bis 90°N	monatliche Klimamittel Berlin-Dahlem	Anzahl der Tage pro Monat mit meßbarem Niederschlag Berlin-Dahlem
Clusteranalyse	1909 – 1998	Dataset 010.1 von NCAR monatliche Mittel des Bodendrucks in 5 zu 5° für 20 bis 90°N	monatliche Klimamittel Berlin-Dahlem	

**Tabelle 2.1:** Untersuchungsmethoden und verwendete Datensätze. Details zu den verschiedenen Datensätzen sind dem Text zu entnehmen.