

Challenges in Understanding the Atmosphere

Vorbereitungsmaterial

Seminar: Einblicke in verschiedene aktuelle Forschungsgebiete der Statistik

Leitung: Prof. Dr. Thomas Augustin, Dr. Marco Cattaneo, Gero Walter, Andrea Wiencier

Wintersemester 2011/2012

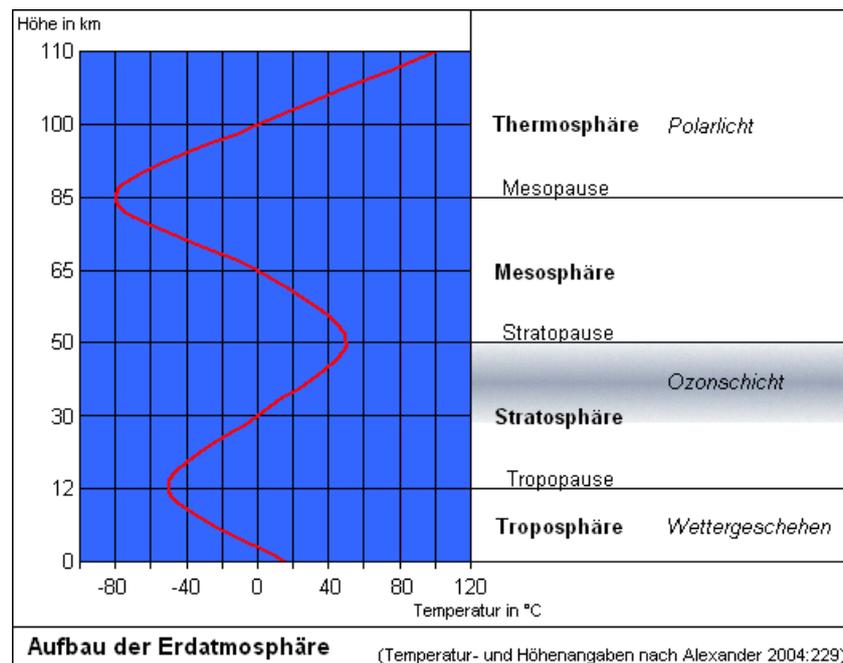
17.01.2012

Delic Edita

1. Allgemeine Beschreibung

1.1. Die Atmosphäre

Die Atmosphäre ist die Lufthülle der Erde, die von Gasen gebildet wird, welche durch die Erdanziehung auf ihrer Oberfläche gebunden werden. Die Atmosphäre (die Lufthülle der Erde) ist der Ort, in dem sich Klima und Wetter abspielen. Sie sorgt für den Temperatúrausgleich zwischen Äquator und Polen, d.h. ohne sie wäre es unerträglich heiß bzw. kalt. Sie schützt uns außerdem vor dem Einfall gesundheitsschädlicher Strahlung, z. B. vor der ultravioletten (UV) Strahlung.



Wie die Grafik zeigt, wird die Atmosphäre auf Grund des vertikalen Temperaturverlaufes in verschiedene Schichten unterteilt. Die äußersten "Hüllen" der Erde sind die Thermosphäre und die Mesosphäre. In der äußeren Thermosphäre herrschen Temperaturen von bis zu 1000°C. In etwa 80 km Höhe trennt die Mesopause die Thermosphäre von der Mesosphäre ab. Die Troposphäre ist für den Meteorologen die interessanteste Schicht, weil sich nur hier das gesamte Wettergeschehen abspielt. Diese unterste Schicht der Atmosphäre wird durch aufsteigende Warmluft erwärmt und kühlt sich an anderer Stelle wieder ab. Dadurch entstehen nicht nur Temperatur- sondern auch Luftdruckunterschiede, die wiederum Winde und Wolken bewirken. Diese Erscheinungen sind das Wetter.

Durch die Anziehungskraft der Erde werden auf ihrer Oberfläche Gase gebunden - die Atmosphäre. Zusammensetzung der untersten 100 km (= Homosphäre):

- 78,08% Stickstoff
- 20,95% Sauerstoff
- 0,93% Argon

Gase, die einen Anteil von weniger als 1% haben, werden als Spurengase bezeichnet. Dazu gehören u.a. 0,03% Kohlenstoffdioxid, Wasserdampf und Ozon.

Die Atmosphäre ist aber kein isoliertes System, sondern steht mit der Hydrosphäre (Ozean und Wasserkreislauf auf Kontinenten und in der Atmosphäre), der Kryosphäre (Eis und Schnee), der Biosphäre (Tiere und Pflanzen), der Pedosphäre (Boden) und der Lithosphäre (Gestein) in Wechselwirkung. Diese Bestandteile des Klimasystems bewegen sich mit völlig unterschiedlicher Geschwindigkeit und sie haben drastisch unterschiedliche Wärmeleitfähigkeiten und Wärmekapazitäten. Die Dynamik des Klimasystems und die daraus folgende Statistik des Klimas werden daher durch die stark unterschiedlichen Zeitskalen der Komponenten geprägt. Die unterste Atmosphäre passt sich in Stunden den Bedingungen der Oberfläche an, die Tiefenzirkulation der Ozeane reagiert erst in Jahrhunderten voll auf die veränderte Zusammensetzung der Atmosphäre, und ein großes Inlandeisgebiet wie die Antarktis braucht dazu viele Jahrtausende.

1.2. Unterschiede zwischen Wetter und Klima

Die Weltorganisation für Meteorologie (WMO) definiert das Klima als die Statistik des Wetters über einen Zeitraum, der lang genug ist, um diese statistischen Eigenschaften auch bestimmen zu können. Während das Wetter den physikalischen Zustand der Atmosphäre zu einem bestimmten Zeitpunkt an einem bestimmten Ort beschreibt, ist Klima erst dann richtig gekennzeichnet, wenn die Wahrscheinlichkeit für Abweichungen vom Mittelwert angegeben werden kann, also auch Extremwerte Teil der Statistik sind. Zur Beschreibung des Klimas wird in der Regel eine Zeitspanne von 30 Jahren als Bezugszeitraum herangezogen. Die übliche Einteilung in Klimazonen folgt überwiegend dem Jahresgang der Temperatur und des Niederschlags.

Der Unterschied zwischen Wetter und Klima lässt sich vereinfacht ausdrücken als der Unterschied zwischen kurzfristigen Geschehnissen in der Atmosphäre an einem bestimmten Ort zu einer bestimmten Zeit (Wetter) und den Geschehnissen, die sich über einen längeren Zeitraum (mindestens 30 Jahre) vollziehen und die sich auf (über-)regionale, statistische Ausprägung bestimmter Variablen beziehen (Klima).

Wetter	Witterung	Klima
Momentaner Zustand der Atmosphäre (1 Stunde, 1 Tag)	Charakter des Wetters über einige Tage oder eine Jahreszeit	Mittlerer Zustand der Atmosphäre (z.B. 30-40 Jahre)

Klimaveränderung verstehen oder das Wetter für die nächste Woche voraussagen- jede dieser Tätigkeiten verbindet Kenntnisse der Atmosphäre mit der Datenanalyse und statistischen Wissenschaft. Der praktische Nutzen dieser Aktivitäten für unsere Gesellschaft und Wirtschaft sind ein fantastisch ausgeglichenes Zusammenspiel zwischen den Modellen, die diese geophysikalischen Prozesse und die Notwendigkeit der Statistik beschreiben um die Unsicherheiten zu quantifizieren und Daten die gesammelt werden zu assimilieren. Wir werden einen Einblick in die Bedeutung von Statistik, von der sekundären Rolle bei der Entwicklung und Bereitstellung von numerischen Klimamodellen geben bis zur Partnerschaft in dem Prozess der Wettervorhersage. Vor diesen Beispielen werden wir die grundlegenden physikalischen Prinzipien, die im Hintergrund der Wissenschaft über die Atmosphäre liegen, vorstellen.

2. Physikalische Gleichungen für die Atmosphäre

Die Bewegung und Evolution der Atmosphäre kann mit Hilfe von einem System von mehreren nichtlinearen partiellen Differentialgleichungen bestimmt werden. Diese Gleichungen werden gut als fundamentale Beschreibung des atmosphärischen Flusses akzeptiert und werden auch primitive Gleichungen genannt, wobei der Begriff „primitiv“ als „ursprünglich“ zu verstehen ist.

Die primitiven Gleichungen sind eine Reihe von nichtlinearen Differentialgleichungen, die ungefähre globalatmosphärische Strömungen verwenden und werden in den meisten atmosphärischen Modellen benutzt. Sie bestehen aus drei Gruppen von Gleichungen:

1. **Impulserhaltung:** Bestehend aus einer Form der Navier-Stokes-Gleichungen, die hydrodynamische Strömung beschreiben auf der Oberfläche einer Kugel
2. **Eine thermische Energie Gleichung:** Bezieht sich auf die allgemeine Temperatur des Systems und dient der Beschreibung des mittleren Temperaturfeldes
3. **Eine Kontinuitätsgleichung:** Stellvertretend für die Erhaltung der Masse

Im allgemeinem beziehen sich alle Formen der Grundgleichungen auf fünf Variablen u , v , ω , T , W .

- u - ist die zonale Geschwindigkeit (Geschwindigkeit in der Ost / West Richtung Tangente der Kugel)
- v - ist die meridionale (Geschwindigkeit in Richtung Norden / Süden Richtung Tangente der Kugel)
- ω - ist die vertikale Geschwindigkeit in isobaren Koordinaten
- T - ist die Temperatur
- W – ist das fällbare Wasser (kondensierte, Art Niederschlag)

Das einfachste Beispiel einer nichtlinearen partiellen Differentialgleichung ist die Burgersgleichung. Die Gleichung ist eine nichtlineare partielle Differentialgleichung zweiter Ordnung für eine Funktion von zwei Variablen.

Viskose Burgersgleichung:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2}.$$

$$\frac{\partial u}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} (u^2) = \mu \frac{\partial^2 u}{\partial x^2}.$$

Der Parameter $\mu > 0$ kann hier als Viskositätsparameter interpretiert werden. Der Kehrwert der Viskosität ist die Fluidität, ein Maß für die Fließfähigkeit eines Fluids. Je größer die Viskosität (hochviskos), desto dickflüssiger (weniger fließfähig) ist das Fluid; je niedriger die Viskosität (niedrigviskos), desto dünnflüssiger (fließfähiger) ist es.

Wenn also eine ganze physische Beschreibung bekannt ist, warum ist dann Statistik erforderlich? Die kurze Antwort ist, dass diese Gleichungen einfach zu kompliziert sind, um auf eine genaue Weise jedes praktische Modell der Atmosphäre zu lösen. Eine präzisere Antwort ist mit der Größenordnung des Problems verbunden. Primitive Gleichungen geben die Beziehung zwischen den Bewegungen von großen und den Bewegungen von kleinen Maßstäben wieder. Selbst wenn man sich auf groß angelegte Bewegungen konzentrieren würde, sagen wir in einer Klimastudie, würde es notwendig sein, die Gleichungen auf einer viel detaillierteren Ebene zu lösen. Das ist, weil die Energie von kleineren Skalen zurück in die größere Skalenbewegungen fließt.

Ein anderer Grund ist, dass die primitiven Gleichungen empfindlich auf Anfangsbedingungen sind. Nehmen wir an, dass jemand die primitiven Gleichungen mit Hilfe von, ein wenig voneinander verschiedenen, anfängliche Bedingungen versucht zu lösen. Die nichtlinearen Bestandteile in den primitiven Gleichungen würden dazu neigen, kleine Unterschiede bei $t = 0$ in signifikante Abfahrten wie t zu verstärken. Schließlich würde es so aussehen als ob die zwei Lösungen voneinander unabhängig seien. Natürlich hat das auch eine enorme praktische Bedeutung für die Wettervorhersage. Man weiß nie den exakten Zustand der Atmosphäre, und obwohl diese Diskrepanz klein zu sein scheint, werden Unterschiede zwischen der Vorhersage und dem wahren Zustand der Atmosphäre exponentiell ansteigen je länger wir in einem Zeitraum vorhersagen. Diese Eigenschaft hat einen Einfluss auf eine Vielzahl von Klimaexperimenten, weil die erste Klimasituation auch von anfänglichen Bedingungen der Atmosphäre und des Ozeans abhängt.

Moderne Meteorologen frustriert das aber nicht, denn sie kennen mit der numerischen Berechnung ein mächtiges Hilfsmittel. Es erlaubt ihnen, sich durch den Zahlenschwung geschickt an die exakte Lösung der Gleichungen heranzupirschen. Solche mathematischen Näherungen sind zwar nie ganz genau, aber ein sehr wirksamer Kompromiss.

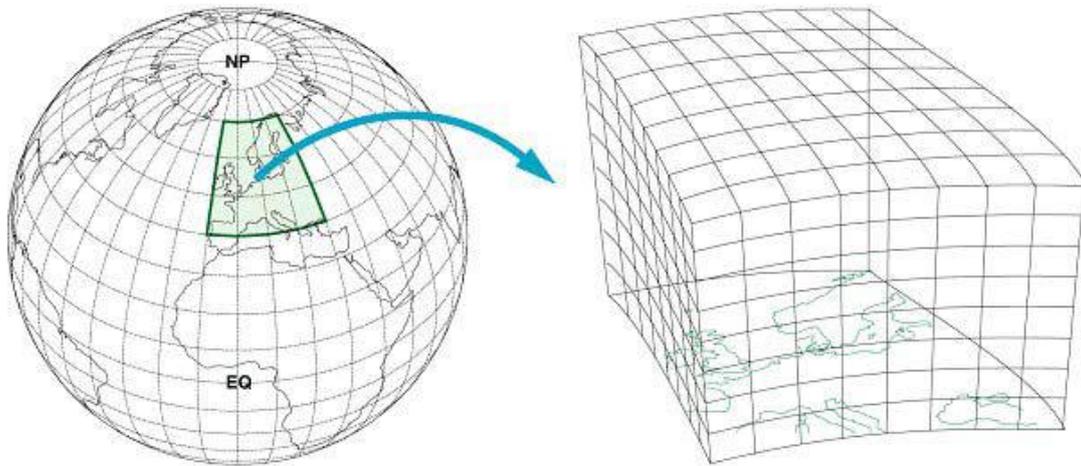
3. Wettervorhersage

3.1. Numerische Wettervorhersagen

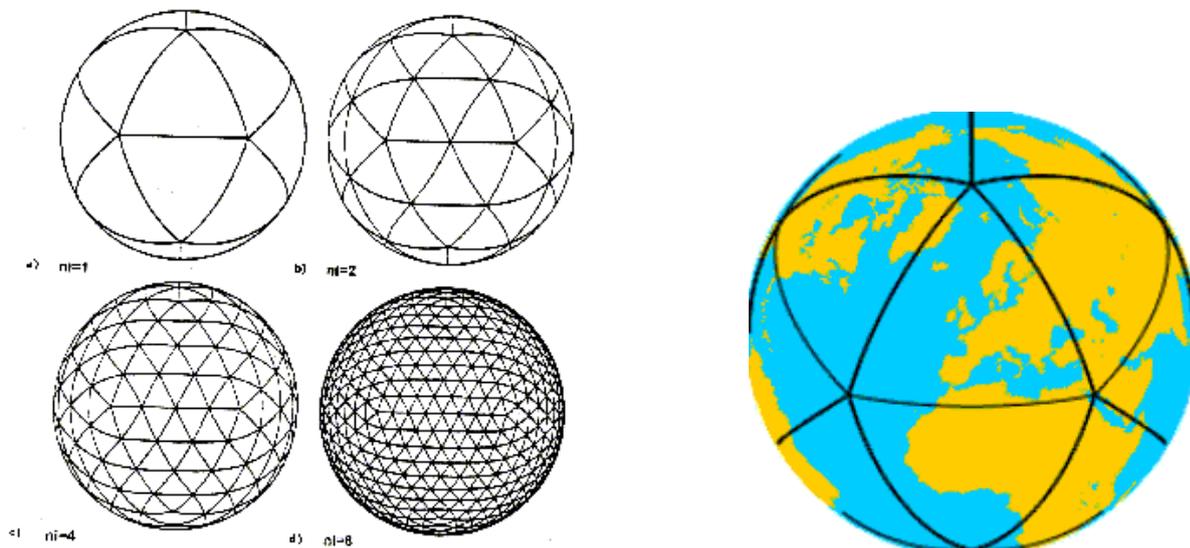
Numerische Wettervorhersagen sind rechnergestützte Wettervorhersagen. Aus dem Zustand der Atmosphäre zu einem gegebenen Anfangszeitpunkt wird durch numerische Lösung der relevanten Gleichungen der Zustand zu späteren Zeiten berechnet. Diese Berechnungen umfassen teilweise mehr als 14 Tage und sind die Basis aller heutigen Wettervorhersagen.

Wenn die Meteorologen ein numerisches Wettervorhersage-Modell entwickeln, dann überziehen sie zuerst das gewünschte Gebiet – oder gleich den kompletten Globus – vom Boden bis in etwa 35 Kilometer Höhe mit einem dreidimensionalen mathematischen Gitter. Beim DWD werden zwei Arten von Gittern verwendet: ein Rechteckgitter für das regionale Modell COSMO-Modell (früher LM genannt) und ein Dreiecksgitter für das globale Modell GME. Für jeden Kreuzungspunkt des Gitters stellen sie einen Satz von Näherungsformeln für die physikalischen Wettergrößen auf. In diese Gleichungen speisen sie alle in diesem Gebiet aktuell beobachteten Wetterdaten ein. Aus diesem Anfangszustand kann der Computer anschließend berechnen, wie sich die Atmosphäre Zeitschritt für Zeitschritt in die Zukunft hinein verändern wird. Als Ergebnis liefert er für jeden Gitterpunkt eine Vorhersage darüber, welchen Wert die Temperatur, der Druck, der Wind, die Feuchte

und andere physikalische Größen in einigen Stunden oder Tagen haben werden. Aus diesen Ergebnissen können die Meteorologen dann eine Wettervorhersage ableiten.



Damit das numerische Wettermodell realistische Ergebnisse liefert, muss das Gitterraster sehr feinmaschig sein. Für jeden Rasterpunkt müssen möglichst viele Näherungsformeln berechnet werden, sonst wäre das Modell zu einfach für die Komplexität der Natur. Die Folge ist: Die Zahl der notwendigen Rechenschritte und der zu verarbeitenden Daten explodiert. Deshalb wächst die Leistungsfähigkeit der „Numerischen Wettervorhersage“ (NWV) stetig mit der Rechenleistung der Supercomputer. Das globale Modell des DWD überzieht den Globus mit einem Gitter aus fast gleich großen Dreiecken. Die durchschnittliche Länge einer Dreiecksseite beträgt derzeit etwa 30 Kilometer. Damit legt dieses Modell ein Netz aus 655 362 Gitterpunkten und doppelt so vielen Dreiecken um die Erde. In der dritten Dimension unterteilt es die Atmosphäre in 60 Stockwerke. Für jedes Stockwerk berechnet es die Lufttemperatur, den Wind, die Feuchte, das Wasser und das Eis in den Wolken, sowie Regen und Schnee. Im untersten Stockwerk kommt noch der Bodendruck hinzu. Das alles erzeugt eine gewaltige Datenflut. Für jeden Zeitschritt und für jeden Gitterpunkt muss der Computer etwa 5000 Rechenoperationen durchführen – und zwar für jedes der 60 Stockwerke in der Atmosphäre.



Die Entwickler der numerischen Wettermodelle arbeiteten in den letzten Jahren intensiv an der Optimierung der Vorhersage von Niederschlägen. Vor allem die Prognosemöglichkeit kleinräumiger aber starker Gewitter mit Schauern und Hagel hat sich verbessert. Trotz der enorm gestiegenen Leistungsfähigkeit hat die numerische Wettervorhersage auf diesem Gebiet in den letzten 15 Jahren aber keinen entscheidenden Durchbruch geschafft. Denn dafür bräuchten die Modelle eine Maschenweite von nur ein bis zwei Kilometern, was der Computertechnik eine gewaltige Leistungssteigerung abverlangt. Aber auch die mathematische Modellierung der physikalischen Prozesse in den Wolken muss noch viel genauer werden - dazu bedarf es weiterer Grundlagenforschung. Der Deutsche Wetterdienst hat 2007 mit seinem sehr hoch auflösenden lokalen Modell COSMO-DE für Kurzzeit-Vorhersage eine „Schallmauer“ durchbrochen. Seine Maschenweite liegt bei 2,8 Kilometern. Dieses lokale Kurzzeit-Modell verarbeitet auch Daten aus dem Wetterradar und befähigt die Meteorologen erstmals Unwetterwarnungen auf Basis der numerischen Wettervorhersage für die kommenden 18 Stunden herauszugeben. Das lokale Modell läuft ebenso wie das globale und regionale Modell auf dem Großrechner des Deutschen Meteorologischen Rechenzentrums (DMRZ) in der Zentrale des DWD in Offenbach. Es beschränkt sich dabei nur auf Deutschland und die nähere Umgebung, um die Rechner nicht zu überlasten.

3.2. Globale Modelle (GCM)

Ein globales Klimamodell beschreibt die wichtigsten klimarelevanten physikalischen Vorgänge in der Erdatmosphäre, den Ozeanen und auf der Erdoberfläche und deren Wechselwirkungen, wobei die Prozesse in sehr vereinfachter Form abgebildet werden. Vor allem die Prozesse in der Biosphäre werden im Augenblick noch als statische Größen und Parameter vorgegeben. Bestenfalls sind typisierte Jahresgänge enthalten. Diese Größen sind aber Systemgrößen und sollten sich während der Simulation dem globalen Wandel anpassen können, um realistische Projektionen auf die Zukunft abgeben zu können. Solche Rückkopplungsprozesse von gekoppelten Systemen sind im Augenblick die große Herausforderung in der Modellierung. Die Modelle sind so umfangreich, dass sie nur in sehr grober Auflösung (mehrere hundert Kilometer Gitterweite) betrieben werden können.

Beispiele **globaler Klimamodelle** sind:

HadCM3 (Hadley coupled model, version 3)	Dieses Klimamodell wurde neben einigen anderen für den 3. und 4. Sachstandsbericht des IPCC verwendet.
HadGEM1 (Hadley global environment model 1)	Weiterentwicklung des HadCM3 Klimamodells. Verbessert wurde die Repräsentation des Einflusses von Wolken und von Seeis; ebenso die Abbildung der Parameter Wasserhaushalt, Atmosphärenchemie und die Effekte von Aerosolen. Die Repräsentation der Einflüsse von El Nino, des Monsuns sowie pazifischer Oberflächentemperaturen haben sich jedoch verschlechtert und sind Gegenstand laufender Forschungsarbeit, wobei bereits Fortschritte zu verzeichnen sind.
ECHAM4/OPYC3	Modell des Hamburger Max-Planck-Instituts für Meteorologie für den 3. Sachstandsbericht des IPCC
ECHAM5/OM1	Modell des Hamburger Max-Planck-Instituts für Meteorologie für den 4. Sachstandsbericht des IPCC

Ein in die Erdkugel platziertes Ikosaeder bildet den Kern der Gitterstruktur. Durch Verbindung der 12 Eckpunkte des Ikosaeders mit Großkreisabschnitten auf der Kugel entstehen 20 gleichseitige sphärische Dreiecke mit einer Seitenlänge von rund 7054 km. Durch sukzessive Unterteilung (zum Beispiel Halbierung) der sphärischen Dreiecke wird das Gitter der gewünschten Modellauflösung erzeugt. Als Maschenweite wird dabei die mittlere Seitenlänge der kleinsten sphärischen Dreiecke bezeichnet; sie beträgt beim GCM zu Zeit 30 km. Die Ecken der Dreiecke (Gitterpunkte) sind von sechs bzw. fünf Dreiecken umgeben. Die Gitterpunkte repräsentieren deshalb sechseckige (hexagonale) bzw. fünfeckige (pentagonale) Flächen. Das GCM-Gitter beschreibt also die Kugeloberfläche mit sehr vielen (bei 30 km Maschenweite: 655350) Sechsecken und genau 12 Fünfecken, die an den Eckpunkten des Ikosaeders liegen. Die mittlere Größe der Gitterelemente beträgt 778 km^2 (bei der Maschenweite von 30 km). Alle Modellvariablen wie Luftdruck, Temperatur, Wind, Wasserdampf, Wolkenwasser und Wolkeneis, sind als Mittelwerte über die Gitterelemente, d.h. über eine Fläche von 778 km^2 , anzusehen. Die wichtigsten Vorhersagevariablen des GME sind: Luftdruck am Erdboden, horizontale Windkomponenten, Temperatur, Wasserdampf, Wolkenwasser, Wolkeneis und Ozon auf 60 Modellschichten in der Atmosphäre vom Boden bis in etwa 34 km Höhe. Die Erdatmosphäre wird im GME also durch $655362 \times 60 = 39$ Millionen Gitterpunkten beschrieben.

3.3. Regionale Klimamodelle

Regionale Klimamodelle betrachten lediglich einen räumlichen Ausschnitt bzw. Sektor der globalen Atmosphäre, z. B. Europa. Numerische regionale Klimamodelle benötigen deshalb geeignete Randbedingungen an den Rändern des Simulationsgebietes. Diese Randbedingungen stammen aus Simulationen der globalen Klimamodelle. Man sagt deshalb, dass ein regionales Klimamodell durch ein globales Klimamodell angetrieben wird, was auch als "Nesting" oder "dynamic downscaling" bezeichnet wird und das Einbetten eines regionalen Modells mit einer hohen räumlichen Auflösung in ein globales Klimamodell mit einer geringen räumlichen Auflösung beinhaltet. Die Abstände der Gitternetzpunkte bei einem globalen Klimamodell sind in der Regel groß und liegen zwischen 150 und 500 km. Regionale Modelle verfügen über eine viel feinere Auflösung mit Gitternetzpunkten in einem Abstand von z. T. nur wenigen Kilometern (z. B. Modell REMO mit horizontaler Gitterauflösung von 10 mal 10 km und COSMO-CLM von 17 mal 17 km). Durch die Zunahme der Rechenkapazität moderner Supercomputer kann die räumliche Auflösung der Modelle ständig verbessert werden.

Die Anwendung COSMO-EU (COSMO Europa) überdeckt ganz Europa mit $665 \times 657 = 436905$ Gitterpunkten mit einer Maschenweite von 7 km. Vom Boden bis in etwa 24 km Höhe wird die Atmosphäre mit 40 Schichten aufgelöst, so dass COSMO-EU die Vorhersage an $436905 \times 40 = 17.5$ Millionen Gitterpunkten berechnet. Die horizontale Fläche der quadratischen Gitterelemente des COSMO-EU beträgt 49 km^2 , während sie im GME mit 778 km^2 etwa 16-mal größer ist. Deshalb kann COSMO-EU viele lokale Details der Landschaft, die einen prägenden Einfluss auf das Wetter haben, im Unterschied zum GME explizit erfassen. Die prognostischen atmosphärischen Variablen des COSMO-EU sind: Luftdruck, horizontale und vertikale Windkomponenten, Temperatur, Wasserdampf, Wolkenwasser, Wolkeneis, Regen, Schnee und turbulente kinetische Energie. Über Land werden zusätzlich die Temperatur und der Bodenwasser- und Eisgehalt für sieben Erdbodenschichten sowie Wassergehalt und Dichte des Schnees vorhergesagt. Über den Ozeanen wird die Meeresoberflächentemperatur einmal täglich aus Beobachtungen analysiert und während der COSMO-EU-Vorhersage festgehalten.

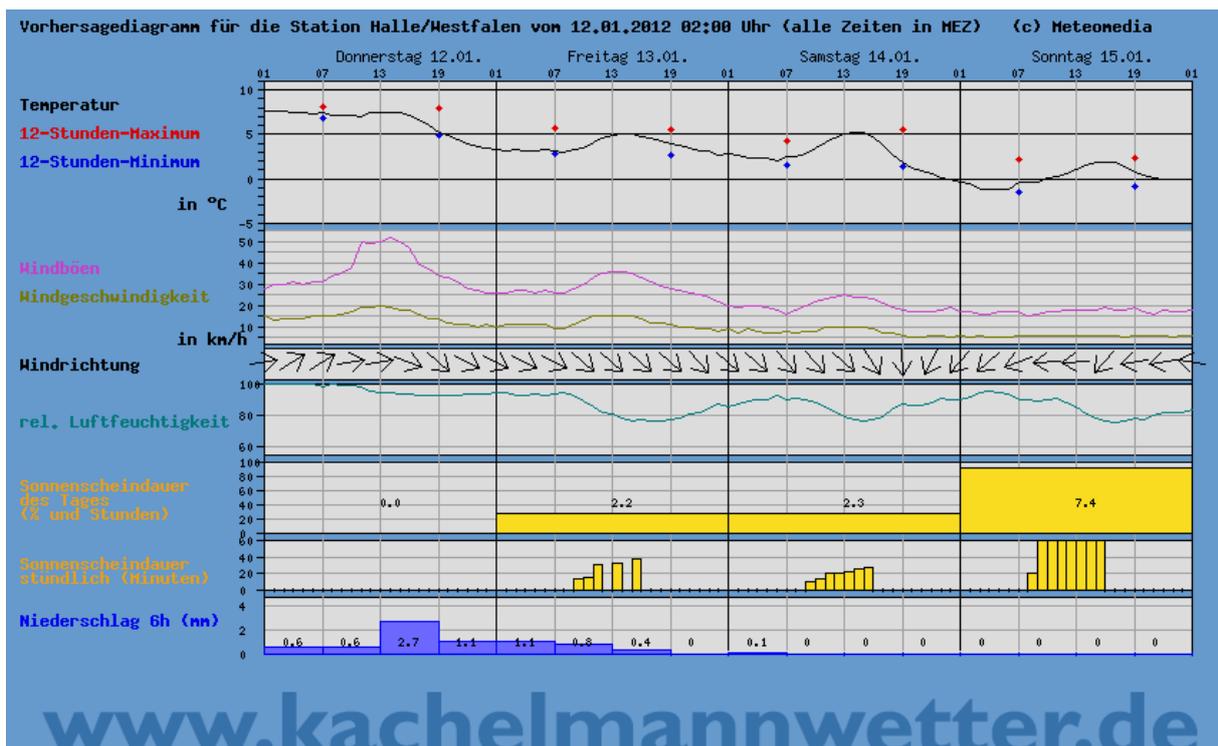
3.4. Model Output Statistics (MOS- Verfahren)

Model Output Statistics (MOS) ist ein statistisches Verfahren in der modernen Wettervorhersage, das in den 1960er/1970er Jahren in den USA entwickelt worden ist. Oftmals handelt es sich dabei um multilineare Regressionsgleichungen, die auf numerische Wettermodelle angewandt werden. Heutzutage finden MOS-Verfahren weltweit Verwendung und dienen als Hilfsmittel vor allem bei der lokalen Wetterprognose.

Bei MOS-Verfahren handelt es sich um multilineare Regressionsgleichungen, die dazu dienen, die direkten Vorhersageergebnisse numerischer Wettermodelle (Direct Model Output, DMO) für einen bestimmten Ort anhand von Stationsmesswerten, die für den Ort vorhanden sein müssen, zu verifizieren bzw. zu verändern. So werden die Vorteile eines numerischen Modells mit den Messwerten vor Ort kombiniert, um eine möglichst genaue Prognose für den jeweiligen Standort zu erhalten.

Dabei werden mittels Regression diejenigen Parameter, Prädiktoren genannt, aus dem DMO herausgefiltert, welche einen signifikanten Einfluss auf andere meteorologische Größen haben, die z. B. das Bodenwetter umschreiben. Diese Parameter werden Prädiktanden genannt. Prinzipiell kann jedes numerische Wettermodell mit einem MOS-Verfahren gekoppelt werden; auch können mehrere, leicht unterschiedliche MOS-Verfahren auf dasselbe Modell angewandt werden. Voraussetzung ist allerdings, dass die Station, für welche eine Lokalprognose erstellt werden soll, über eine mindestens 18-monatige Messreihe verfügt.

Die Ergebnisse eines MOS-Laufes werden meist entweder in Diagrammform (MOS-Diagramm) dargestellt, was z. B. in den USA recht verbreitet ist, oder zu Generierung von Wettersymbolen verwendet, wie man es beispielsweise in Deutschland häufig findet. Dabei generiert ein Programm aus den MOS-Parametern die gewünschten und benötigten Symbole wie z. B. Wolke/Sonne/Niederschlag, Windpfeil, Regenrisiko etc.



3.5. Multiple lineare Regression

Bei der multiplen Regression wird der simultane Einfluss von mehreren unabhängigen Variablen (x_j) auf eine abhängige Variable (y) analysiert. Die multiple lineare Regression stellt eine Erweiterung der einfachen linearen Regression dar, die es erlaubt, mehrere UVs simultan in die Analyse einzubeziehen. Liegen zwei unabhängige Variablen vor, kann man sich die Regression vorstellen als das Hindurchlegen einer Fläche durch eine dreidimensionale Punktwolke.

$$\hat{Y} = b_0 + b_1 * X_1 + e_k \text{ (einfache Regressionsfunktion)}$$

$$\hat{Y} = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k + e_k \text{ (multiple Regressionsfunktion)}$$

\hat{Y} : abhängige Variable

X : unabhängige Variable

b_0 : konstantes Glied

b_k : Steigungsparametern

e_k : Residuengröße (Störgröße)

$$\hat{Y}_i = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k + e_i$$

Messniveau: **Abhängige Variable (y)** metrisch (z.B. Einkommen)

Unabhängige Variablen (xi) metrische und Dummy-Variablen

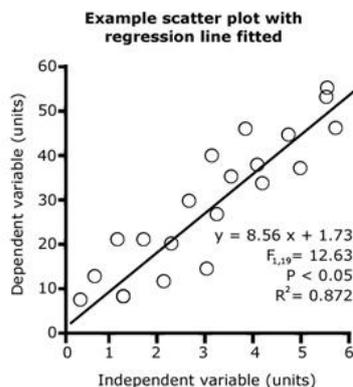
Regressionsgleichung: $\hat{Y} = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k + e_k$

Interesse: Lineare Zusammenhänge zwischen y und x_i

Schätzung: direkter linearer Einfluss der unabhängigen Variablen
(x_i) auf die abhängige Variable (y)

Optimierungsmethode: Kleinste-Quadrate-Kriterium $\sum(\hat{y}_i - y_i)^2 \rightarrow \min$
(Methode der besten Schätzung)

Graphisch:



Beispielhypothese:

Das Einkommen (y) einer Person steigt proportional zur
Dauer der besuchten Schuljahre (x_1)

Interpretation:

Nichtstandardisierter Regressionskoeffizient b_i

Standardisierter Regressionskoeffizient b_i : direkter Effekt des Prädiktors x_i auf y, relative Wichtigkeit von x_i

R^2 (Modellgüte, Bestimmtheitsmaß): Anteil erklärter Varianz

Bivariate Korrelation (Nullte Ordnung) (z.B. $[y x_1]$): Zusammenhang zwischen x_1 und y

Partielle Korrelation (z.B. $[y x_1]$): y und x_1 wurden um eine Drittfaktor bereinigt; Zusammenhang der Residuen

Semipartielle Korrelation (z.B. $[y x_1]$): nur x_1 wurde bereinigt; Zusammenhang von y mit den Residuen von x_1

F-Test:

Signifikanz des Gesamtmodells

t-Test:

Signifikanz der Prädiktoren