



Modellverfahren

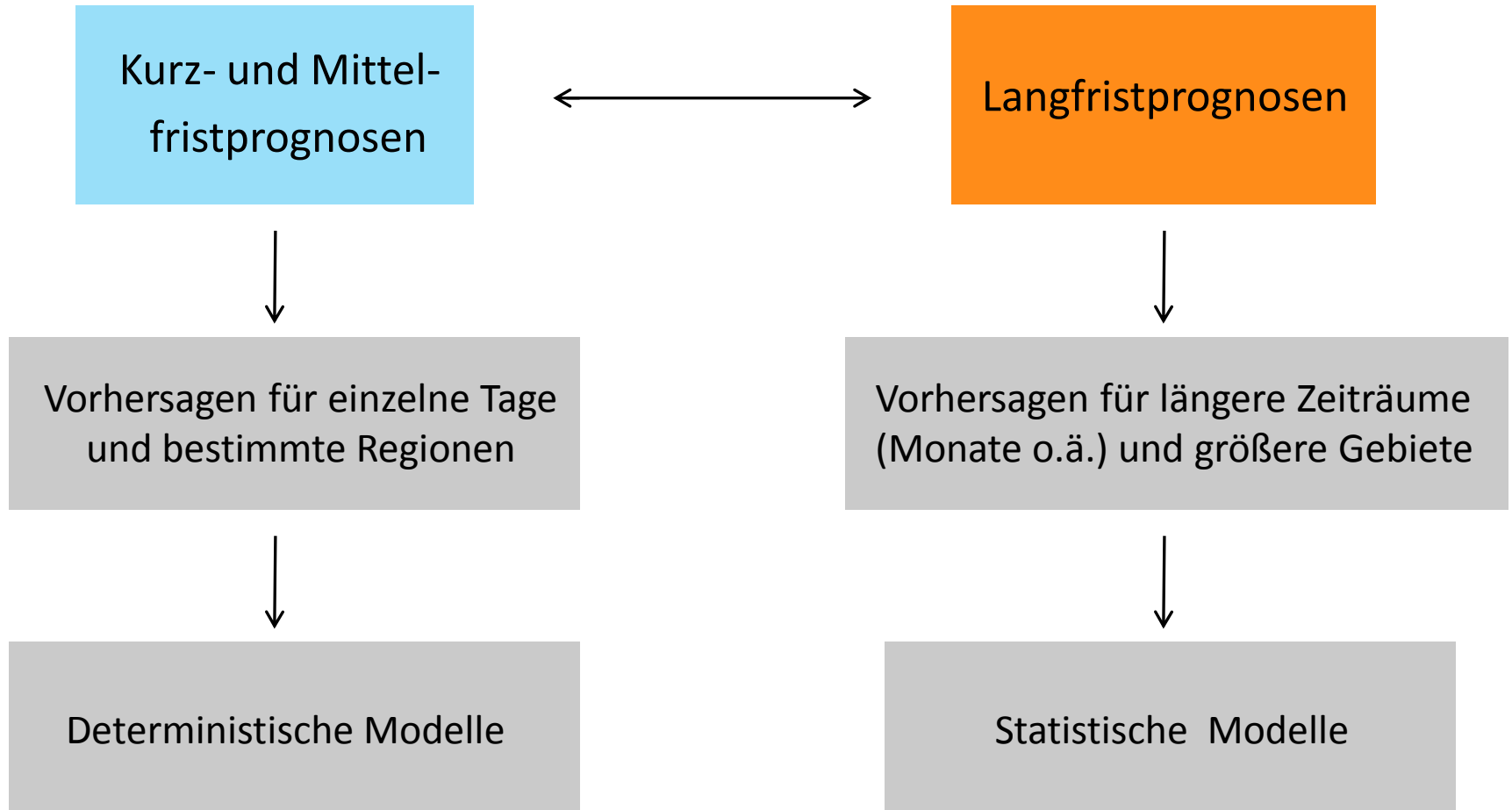
Long-Range-Forecast-System (LRFS)

Von Lars Thieme (M.Sc.)

Eine Vorstellung des verwendeten Vorhersageverfahrens für die veröffentlichten langfristigen Witterungsprognosen auf

www.langfristwetter.com

Einleitung



Einleitung



Statistische Modelle



Long-Range-Forecast-System



Input von relevanten
meteorologischen Größen

Betrachtung eines
Untersuchungszeitraums
(1964 – 2013)

Berechnung der Prognose
mittels statistischer
Verfahren

Modellverfahren



Eingangsparameter

| | |
|-----------------------------|--|
| Bodenluftdruck | Wichtigste steuernde Größe im bodennahen Bereich |
| 500-hPa-Zonalindex | Charakterisiert die Intensität des zonalen Windes im mittleren Troposphärenniveau |
| 200-hPa-Geopotential | Charakterisiert die Druck- und Temperaturverhältnisse vor allem im mittleren und <i>höheren</i> Troposphärenniveau |

Verlauf

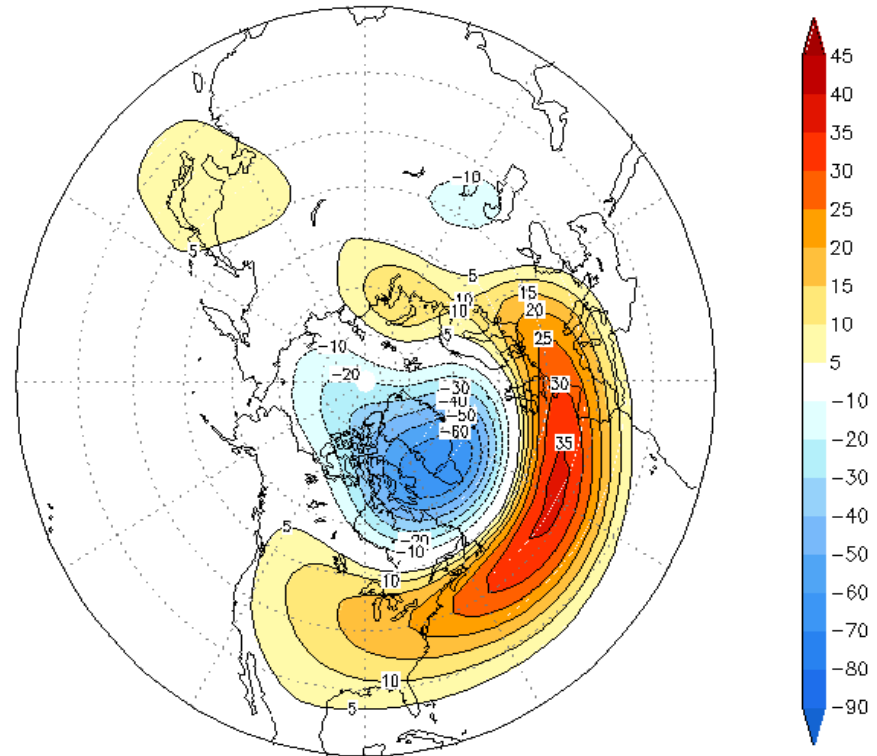


Hintergrund: Die Verteilung steuernder meteorologischer Parameter in den Vormonaten hat einen Einfluss auf die künftige Witterung.

Idee: Die atmosphärische Zirkulation vor kalten Monaten / Jahreszeiten unterscheidet sich von jener vor milden Zeitabschnitten.

Umsetzung: Erwähnte Parameter müssen von der gesamten Nordhemisphäre heran gezogen werden. Vor allem die Nordatlantik-Oszillation (NAO) hat bedeutenden Einfluss auf unsere Witterung. Sie ist v.a. gekennzeichnet durch die Luftdruckgegensätze zwischen den Regionen Island/Grönland und Azoren/Bermudas.

Loading Pattern for the North Atlantic Oscillation (NAO)
Geopotential Height at 500 hPa (m)



Verlauf



Messnetzdaten: Meteorologische Daten werden weltweit in verschiedenen Höhen erfasst und zwar an besonderen Punkten → Gitterpunkten

Gitterpunkte sind Stellen auf dem Erdball, wo sich bestimmte Breiten- und Längengrade schneiden

Im LRFS-Modell werden pro Parameter monatliche Daten an 592 Gitterpunkten verwendet

- entspricht bei 3 Parametern 1.776 Datensätzen
- bei Verwendung von 8 Vormonaten fließen somit **14.208 Datensätze** in eine Vorauswahl ein



1. Schritt

Selektion von Variablen mit Vorhersagepotential



Verlauf



Vorauswahl notwendig da:

- zu große Datenmenge für statistische Berechnung
- viele Informationen in den Variablen redundant vorhanden
- viele „unnötige“ Variablen → kein Vorhersagepotential
- Gefahr des „Overfittings“: Modell wird durch große Anzahl einfließender Variablen zu sehr an Untersuchungszeitraum adaptiert
 - lässt keinen Spielraum für Entwicklungen, die im Untersuchungszeitraum nicht erfasst wurden
 - hat in der Praxis das Versagen des Modells zur Folge

Vorauswahl...aber wie?

Variablenmatrix

Jahresreihen →

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 6 | 6 | 3 | 8 | 1 | 4 | 3 | 2 | 4 | 9 | 5 | 9 | 4 | 4 | 0 | 9 | 6 | 3 | 7 | 2 |
| 3 | 2 | 8 | 1 | 9 | 9 | 3 | 4 | 5 | 3 | 1 | 1 | 5 | 7 | 4 | 9 | 3 | 7 | 9 | 3 | 4 |
| 0 | 5 | 5 | 9 | 9 | 1 | 8 | 3 | 8 | 9 | 2 | 4 | 8 | 5 | 8 | 5 | 2 | 7 | 3 | 2 | 1 |
| 1 | 7 | 7 | 0 | 2 | 1 | 3 | 6 | 4 | 7 | 1 | 9 | 6 | 7 | 4 | 2 | 0 | 2 | 1 | 8 | 7 |
| 0 | 5 | 5 | 2 | 2 | 2 | 4 | 6 | 7 | 8 | 7 | 3 | 1 | 3 | 8 | 6 | 1 | 2 | 3 | 7 | 1 |
| 4 | 5 | 3 | 2 | 4 | 1 | 8 | 8 | 1 | 8 | 3 | 6 | 7 | 9 | 5 | 8 | 5 | 4 | 3 | 0 | 2 |
| 0 | 3 | 8 | 7 | 7 | 7 | 4 | 0 | 9 | 7 | 8 | 5 | 4 | 6 | 2 | 7 | 1 | 9 | 3 | 3 | 4 |
| 1 | 4 | 6 | 7 | 5 | 7 | 2 | 2 | 0 | 7 | 1 | 9 | 7 | 5 | 3 | 6 | 6 | 5 | 7 | 7 | 2 |
| 6 | 8 | 8 | 4 | 7 | 4 | 7 | 4 | 4 | 2 | 2 | 9 | 1 | 7 | 2 | 7 | 7 | 3 | 4 | 8 | 9 |
| 2 | 3 | 5 | 8 | 8 | 8 | 7 | 8 | 7 | 4 | 9 | 7 | 1 | 6 | 9 | 9 | 1 | 7 | 6 | 5 | 9 |
| 7 | 0 | 9 | 7 | 3 | 5 | 3 | 5 | 4 | 0 | 1 | 4 | 2 | 5 | 5 | 3 | 8 | 0 | 3 | 3 | 3 |
| 6 | 7 | 2 | 5 | 0 | 8 | 1 | 9 | 7 | 7 | 6 | 1 | 9 | 8 | 9 | 6 | 5 | 6 | 8 | 5 | 6 |
| 4 | 6 | 3 | 6 | 6 | 3 | 5 | 9 | 2 | 6 | 5 | 8 | 9 | 4 | 0 | 1 | 7 | 0 | 1 | 9 | 5 |
| 6 | 6 | 8 | 0 | 3 | 7 | 5 | 2 | 3 | 1 | 5 | 7 | 8 | 7 | 8 | 6 | 1 | 8 | 9 | 7 | 9 |
| 8 | 1 | 9 | 2 | 3 | 6 | 5 | 2 | 1 | 7 | 8 | 6 | 1 | 0 | 8 | 2 | 7 | 6 | 5 | 6 | 9 |
| 3 | 1 | 2 | 3 | 9 | 8 | 1 | 5 | 3 | 2 | 4 | 1 | 1 | 7 | 2 | 2 | 1 | 9 | 6 | 6 | 6 |
| 5 | 6 | 7 | 5 | 9 | 3 | 5 | 9 | 2 | 7 | 5 | 9 | 7 | 1 | 5 | 1 | 7 | 6 | 4 | 3 | 2 |
| 3 | 5 | 3 | 6 | 6 | 6 | 8 | 3 | 4 | 9 | 3 | 2 | 3 | 2 | 4 | 2 | 4 | 9 | 7 | 0 | 6 |
| 8 | 1 | 0 | 7 | 3 | 1 | 4 | 8 | 3 | 7 | 9 | 1 | 4 | 6 | 5 | 6 | 1 | 1 | 1 | 7 | 6 |
| 3 | 2 | 6 | 8 | 7 | 3 | 7 | 7 | 7 | 4 | 4 | 2 | 6 | 4 | 1 | 8 | 2 | 5 | 8 | 7 | 7 |
| 6 | 5 | 5 | 9 | 4 | 8 | 5 | 8 | 4 | 2 | 8 | 6 | 8 | 2 | 6 | 5 | 1 | 3 | 4 | 1 | 1 |
| 8 | 5 | 3 | 9 | 8 | 3 | 0 | 8 | 7 | 7 | 6 | 2 | 6 | 1 | 9 | 0 | 9 | 7 | 5 | 1 | 1 |
| 0 | 7 | 6 | 8 | 6 | 2 | 4 | 5 | 4 | 4 | 7 | 5 | 5 | 1 | 8 | 4 | 8 | 3 | 3 | 7 | 8 |
| 1 | 7 | 0 | 5 | 2 | 6 | 5 | 4 | 1 | 6 | 3 | 4 | 2 | 7 | 4 | 1 | 4 | 8 | 2 | 8 | 5 |
| 7 | 7 | 8 | 6 | 4 | 6 | 6 | 3 | 1 | 1 | 0 | 0 | 6 | 3 | 7 | 6 | 4 | 9 | 9 | 1 | 2 |
| 4 | 1 | 3 | 1 | 2 | 6 | 1 | 4 | 9 | 1 | 9 | 4 | 2 | 6 | 6 | 8 | 7 | 2 | 8 | 1 | 5 |
| 8 | 2 | 7 | 3 | 5 | 7 | 6 | 6 | 2 | 8 | 5 | 0 | 1 | 1 | 3 | 9 | 1 | 2 | 1 | 0 | 8 |
| 3 | 5 | 7 | 1 | 3 | 5 | 2 | 3 | 8 | 2 | 3 | 5 | 5 | 6 | 8 | 7 | 8 | 5 | 3 | 1 | 0 |
| 4 | 4 | 4 | 8 | 3 | 7 | 6 | 8 | 5 | 0 | 5 | 6 | 0 | 7 | 4 | 9 | 2 | 8 | 0 | 7 | 2 |
| 9 | 4 | 1 | 1 | 9 | 5 | 4 | 0 | 5 | 2 | 2 | 1 | 7 | 8 | 7 | 3 | 6 | 5 | 3 | 8 | 5 |
| 5 | 3 | 8 | 3 | 9 | 1 | 4 | 8 | 6 | 2 | 3 | 7 | 4 | 4 | 8 | 3 | 8 | 7 | 3 | 2 | 1 |
| 6 | 9 | 1 | 3 | 1 | 2 | 5 | 5 | 6 | 4 | 8 | 7 | 1 | 3 | 6 | 8 | 5 | 8 | 8 | 3 | 7 |
| 8 | 9 | 9 | 5 | 6 | 9 | 2 | 7 | 4 | 4 | 1 | 1 | 6 | 1 | 2 | 7 | 8 | 0 | 2 | 1 | 4 |
| 6 | 5 | 7 | 4 | 7 | 7 | 8 | 4 | 2 | 2 | 7 | 7 | 0 | 9 | 9 | 7 | 8 | 5 | 0 | 1 | 2 |
| 7 | 9 | 5 | 2 | 2 | 0 | 7 | 1 | 8 | 2 | 9 | 6 | 9 | 5 | 4 | 6 | 7 | 0 | 8 | 1 | 1 |

Datensätze

Verlauf

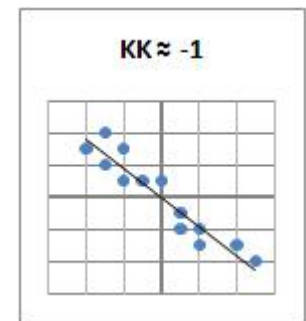
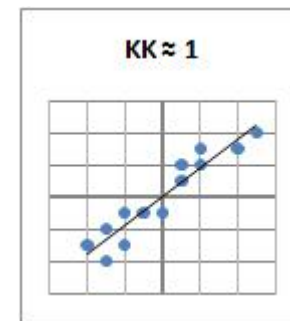
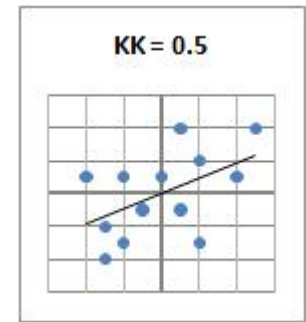
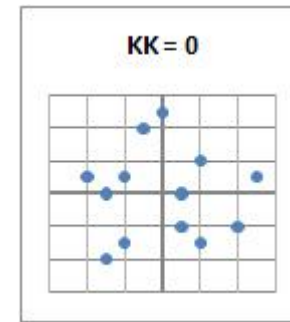


Korrelation

Eine Möglichkeit der Vorauswahl besteht in der Korrelation der Variablen mit der vorherzusagenden Größe (Prädikand). Sie kennzeichnet den linearen Zusammenhang zweier Größen. Die Stärke wird durch den **Korrelationskoeffizient** (KK) ausgedrückt.

Nachteil:

- Linearität wird bei Korrelation vorausgesetzt, die im atmosphärischem System nicht existent ist
- dennoch ist Technik geeignet, um Variablen mit prognoserelevanten Informationen herauszufiltern



Verlauf



Statistische Signifikanz

Ob ein Zusammenhang prognoserelevant sein könnte oder nicht, kann mit einem Signifikanztest des Korrelationskoeffizienten ermittelt werden.

Berechnet wird die Testgröße t für ein bestimmtes Signifikanzniveau.

- t-Werte sind in Tabellen ersichtlich
- t-Wert muss Grenzwert überschreiten

Bsp.: Für einen 50-jährigen Untersuchungszeitraum muss der KK zweier Variablen mindestens | **0.28** | betragen, damit ein 95 %-Signifikanzniveau erreicht wird.

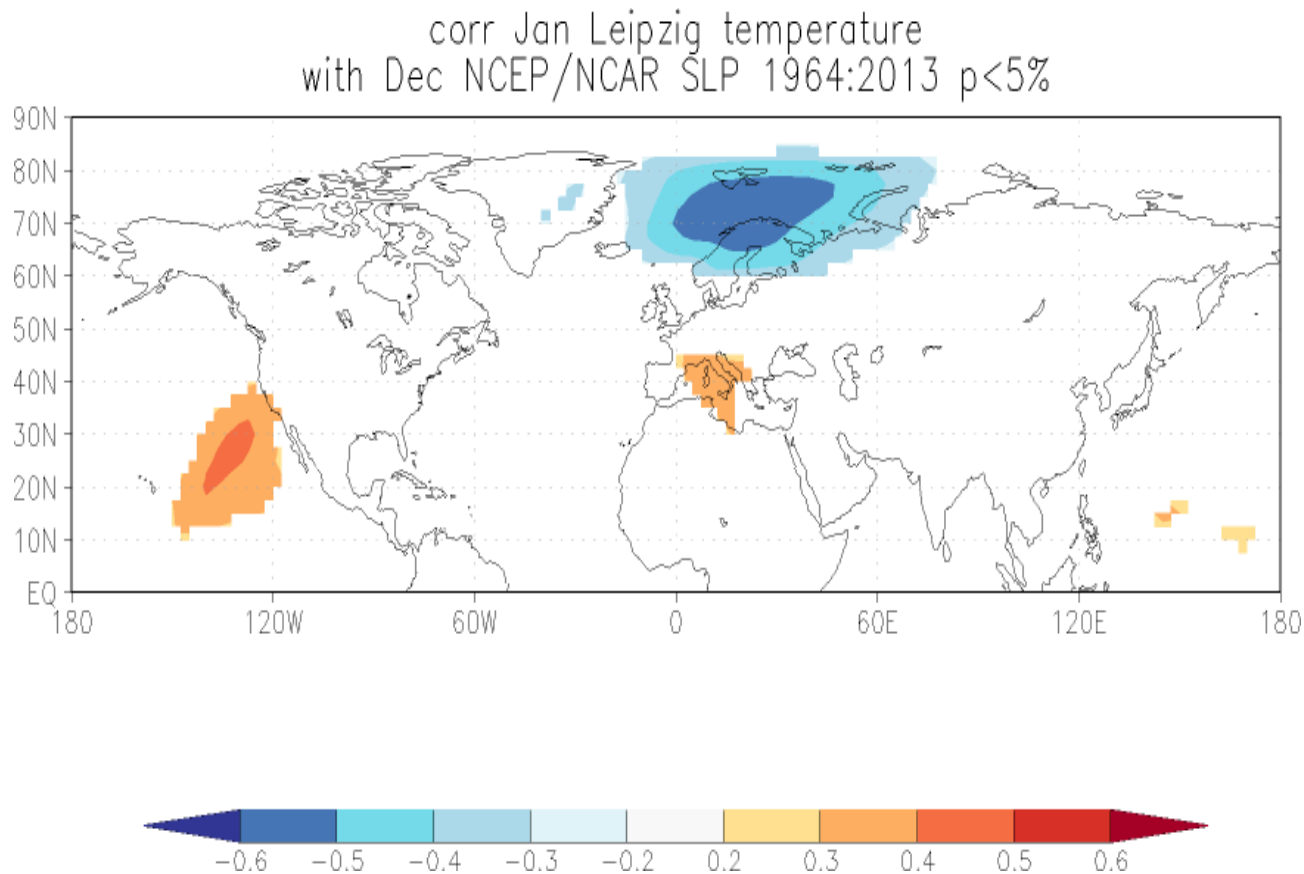
$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

r.....Korrelationskoeffizient
n.....Anzahl der Untersuchungsjahre
t.....Messgröße des t-Tests

Verlauf



Beispiel: Räumlicher Korrelationskoeffizient (KK) zwischen dem Bodenluftdruck im Dezember und der Temperatur Leipzigs im folgenden Januar. Farbige Flächen kennzeichnen Bereiche mit signifikantem KK und damit relevanten Prognosepotential.



Verlauf



Trends in Datenreihen

Problem: Datenreihen – vor allem solche mit Temperatureinfluss – können hohe Korrelationen und damit Zusammenhänge vortäuschen, die ausschließlich durch einen gemeinsamen Trend zustande kommen

Trendbereinigung....ja oder nein?



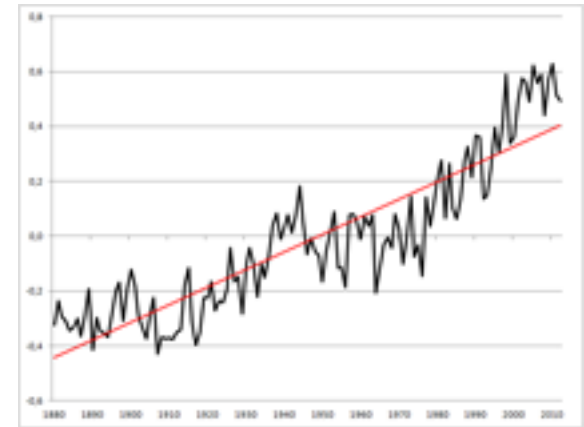
➤ nur wirkliche Zusammenhänge werden detektiert



- durch Bereinigung können Infos verloren gehen
- Modell würde Einfluss der Klimaerwärmung vernachlässigen → Prognosen wären „zu kalt“

Fazit: Auf Trendbereinigung wurde im Modell verzichtet

Daten mit Trend



Daten nach Trendbereinigung



Verlauf



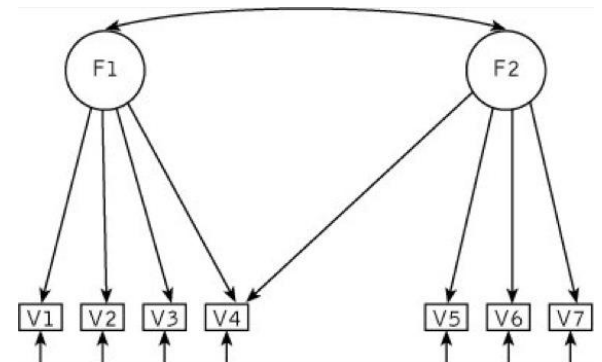
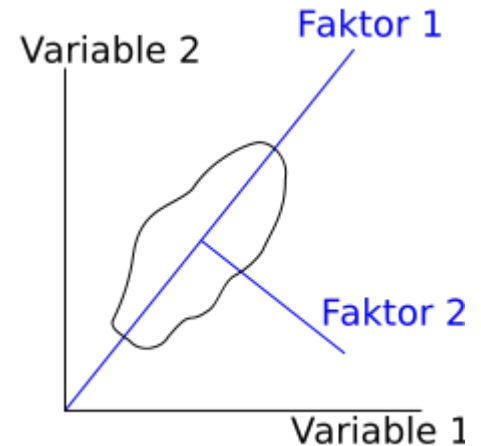
2. Schritt: Hauptkomponentenanalyse

Statistisches Verfahren zur Datenreduktion

- Vielzahl extrahierter Variablen wird durch HKA auf wenige aussagekräftige Hauptkomponenten reduziert
- Informationsgehalt der alten Variablen bleibt weitgehend erhalten
- für jeden Vormonat wird eine einzelne Hauptkomponente abgeleitet (gesamt 8) → korrelieren hoch mit Prädiktanden
- dienen als Prädiktoren für logistisches Regressionsmodell



Ordinale Regression



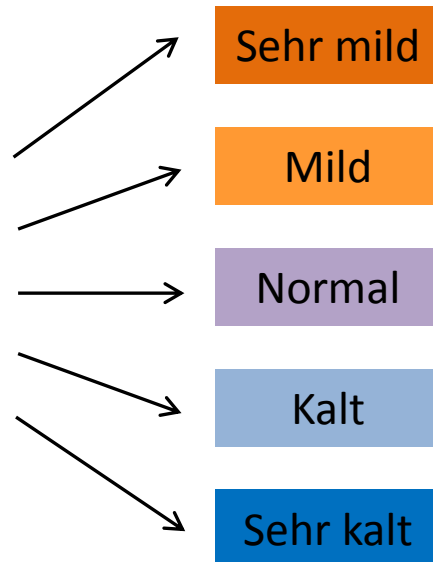
Verlauf



3. Schritt: Klassierung des Prädiktanden

➤ vor Durchführung einer ordinalen Regression muss der Prädiktand von einer metrischen in eine ordinale Skala mit 5 Klassen (Temperatur) bzw. 3 Klassen (Niederschlag) überführt werden

| Jahr | Temp Abw. Monat X |
|------|-------------------|
| 1965 | 1,7 |
| 1966 | -3,3 |
| 1967 | 0,9 |
| 1968 | -1,0 |
| 1969 | -0,2 |
| 1970 | -3,9 |
| 1971 | -1,4 |
| 1972 | -3,0 |
| 1973 | 0,2 |
| 1974 | 3,5 |
| 1975 | 5,1 |
| 1976 | 1,1 |
| 1977 | 0,5 |
| 1978 | 1,4 |
| 1979 | -4,1 |
| 1980 | -3,3 |



- als Definition der Klassengrenzen wird die Streuung zu Rate gezogen
- bei annähernd normalverteilten Daten befinden sich jeweils:
 - 20 % der untersten („Sehr kalt“) und obersten Daten („Sehr mild“) im Bereich $> \text{Streuung} * 0.84$ bzw. $< \text{Streuung} * (-0.84)$
 - 20 % um den Mittelwert („Normal“) im Bereich $< \text{Streuung} * 0.25$ bzw. $> \text{Streuung} * (-0.25)$
 - die restlichen Daten fallen in die Kategorien „Kalt“ bzw. „Mild“

Verlauf



4. Schritt: Ordinale Regression

Ziel: Ermittlung der Wahrscheinlichkeit des Eintretens einer bestimmten Ausprägung des Prädiktanden (bzw. Zugehörigkeit zu einer Kategorie) in Abhängigkeit von den Prediktoren

- Bestimmung der **wahrscheinlichsten** Klasse eines Monats mittels logit-Modells

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X$$

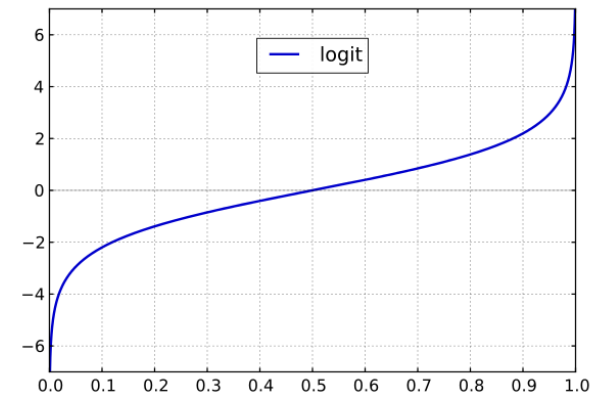
β_0Konstante

β_1Koeffizienten der Hauptkomponente

X.....Beobachtungsdaten der Hauptkomponente

p.....Wahrscheinlichkeit

Verlauf der logistischen Regression



Wahrscheinlichkeitsfunktion

$$p = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

Wird fortgesetzt....

Verifikation



2015

| Monat | 1. Mo | 2. Mo | 3. Mo | 4. Mo |
|-------|--------------------|--------------------|--------------------|--------------------|
| Jan | Eingetroffen | Eingetroffen | Nicht eingetroffen | Eingetroffen |
| Feb | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Eingetroffen |
| Mär | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Apr | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Mai | Keine Prognose | Keine Prognose | Eingetroffen | Eingetroffen |
| Jun | Keine Prognose | Keine Prognose | Keine Prognose | Keine Prognose |
| Jul | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Aug | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Sep | Keine Prognose | Nicht eingetroffen | Keine Prognose | Keine Prognose |
| Okt | Keine Prognose | Eingetroffen | Eingetroffen | Eingetroffen |
| Nov | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Dez | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |

2014

| Monat | 1. Mo | 2. Mo | 3. Mo | 4. Mo |
|-------|--------------------|--------------------|--------------------|--------------------|
| Jan | Eingetroffen | Eingetroffen | Eingetroffen | Keine Prognose |
| Feb | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Mär | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Apr | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Mai | Eingetroffen | Eingetroffen | Keine Prognose | Keine Prognose |
| Jun | Eingetroffen | Keine Prognose | Keine Prognose | Keine Prognose |
| Jul | Eingetroffen | Eingetroffen | Eingetroffen | Nicht eingetroffen |
| Aug | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |
| Sep | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |
| Okt | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Nov | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |
| Dez | Keine Prognose | Keine Prognose | Eingetroffen | Eingetroffen |

2013

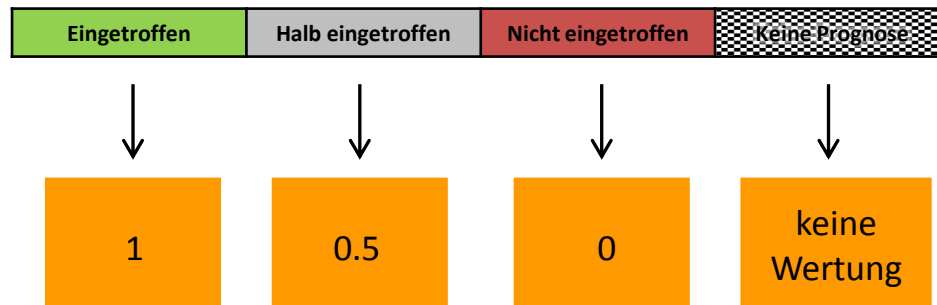
| Monat | 1. Mo | 2. Mo | 3. Mo | 4. Mo |
|-------|--------------------|--------------------|--------------------|--------------------|
| Jan | Eingetroffen | Keine Prognose | Keine Prognose | Keine Prognose |
| Feb | Nicht eingetroffen | Nicht eingetroffen | Keine Prognose | Keine Prognose |
| Mär | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Keine Prognose |
| Apr | Keine Prognose | Eingetroffen | Eingetroffen | Eingetroffen |
| Mai | Keine Prognose | Keine Prognose | Keine Prognose | Eingetroffen |
| Jun | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Jul | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Aug | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Sep | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Okt | Eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |
| Nov | Eingetroffen | Eingetroffen | Eingetroffen | Eingetroffen |
| Dez | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen | Nicht eingetroffen |





Verifikation

Codierung / Wertung



Eintreffwahrscheinlichkeit

Nach Jahren

| Jahr | Prognosen gesamt |
|------|------------------|
| 2015 | 71 % |
| 2014 | 65 % |
| 2013 | 67 % |

Nach Monatsvorlauf

| Jahre gesamt | | | |
|--------------|----------|----------|----------|
| 1. Monat | 2. Monat | 3. Monat | 4. Monat |
| 68 % | 64 % | 66 % | 71 % |

Verifikation



Verifikationsmaße

| | | |
|---|--|--|
| Mittlerer Fehler (ME bzw. BIAS) | $ME = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)$ | Beschreibt die durchschnittliche Abweichung von Vorhersagen gegenüber Beobachtungen. Bei negativem Wert ist Prognose im Mittel zu kalt. |
| Mittlerer absoluter Fehler (MAE) | $MAE = \frac{1}{N} \sum_{i=1}^N F_i - O_i $ | Beschreibt den mittleren Abstand der Vorhersagen gegenüber Beobachtungen. |
| Mittlerer quadratischer Fehler (MSE) | $MSE = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2$ | Beschreibt den mittleren quadrierten Abstand der Vorhersagen gegenüber Beobachtungen. Gewichtet größere Fehlprognosen stärker. |
| Scill Score (SS bzw. RV) | $SS = 1 - \frac{MSE}{MSE_r}$ | Verhältnis des MSE der Prognose zum MSE der Referenzprognose. Entscheidendes Maß für Beurteilung der Vorhersagegüte SS > 0 → Prognose besser als Referenz SS < 0 → Prognose schlechter als Referenz |

Verifikation



Verifikationsmaße

| | 1. Monat | 2. Monat | 3. Monat | 4. Monat |
|---------------------------------------|----------|----------|----------|----------|
| Mittlerer Fehler (BIAS) | -0.31 KE | -0.37 KE | -0.38 KE | -0.53 KE |
| Mittlerer absoluter Fehler | 0.86 KE | 1.00 KE | 0.91 KE | 0.94 KE |
| Mittlerer quadratischer Fehler | 1.47 KE | 1.97 KE | 1.85 KE | 1.65 KE |
| Scill Score | +0.31 | +0.10 | +0.18 | +0.23 |

Verifikation



Zusammenfassung

- Scill Score für alle vier Monatsprognosen positiv → Vorhersageleistung vorhanden
- schlechtesten Scill Score zeigt Zweimonatsprognose, danach wieder Anstieg → Scill Score ist unabhängig von Art der Monatsprognose
- BIAS bei Dreimonatsprognose am stärksten negativ → hier im Mittel die „kältesten“ Prognosen
- Eintreffwahrscheinlichkeiten der einzelnen Prognosen liegen im Mittel zwischen 64 % und 71 %