



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

VERIFIKATION VON PUNKTPROGNOSEN FÜR DEN
GROSSRAUM WIEN

Verfasser

Manuel Kelemen

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:

A 415

Studienrichtung lt. Studienblatt:

Meteorologie

Betreuer:

o. Univ.-Prof. Dr. Reinhold Steinacker

„Fehler vermeidet man, indem man Erfahrung sammelt.
Erfahrung sammelt man, indem man Fehler macht.“

Laurence Johnston Peter
(1919-1990, kanadischer Universitätsprofessor & Pädagoge)

Zusammenfassung

Die Überprüfung und Quantifizierung einer erbrachten Leistung ist oftmals essentiell, sei es aus persönlichen oder aus wirtschaftlichen Gründen. Wobei speziell zweiteres von hoher Bedeutung ist, lässt sich doch damit entweder Geld einsparen oder aber auch Geld verdienen. Diese Überprüfung wird als Verifikation bezeichnet und wird selbstverständlich auch in der Meteorologie angewendet. Da eine solche Verifikation nur allzu gern subjektiv beeinflusst wird, ist eine exakte Vorgehensweise erforderlich, um brauchbare Ergebnisse zu bekommen. Diese Ergebnisse können dann für eine mögliche Fehlersuche und eine damit einhergehende Leistungssteigerung bzw. Optimierung herangezogen werden. Gerade in Zeiten automatisierter Prognosen und steigenden Anforderungen ist diese Überprüfung von immer größer werdender Relevanz.

In dieser Arbeit werden die Vorhersagen der Studenten des Instituts für Meteorologie und Geodynamik an der Universität Wien verifiziert und interpretiert. Neben gängigen Parametern wie dem allgemeinen Fehler werden diese Prognosen auch geeigneten Referenzprognosen gegenüber gestellt, wobei speziell den MOS-Vorhersagen, sogenannten automatisierten Punkt-Prognosen (in dieser Arbeit für Wien) besonderes Augenmerk geschenkt wird, um abschließend die alles entscheidende Frage „Mensch oder Maschine?“ - zumindest in dieser Arbeit - beantworten zu können.

Abstract

Sometimes, it is essential to review and quantify services provided, whether for moral or economic reasons. Especially the latter is very important, as it may ensure that either money is saved or even earned. Such review is termed verification and is, of course, also used in meteorology. Due to the fact that such verification is often influenced subjectively, an exact procedure is required if the verification should yield useful results. These results may then be used to trace possible errors, and accordingly increase performance or achieve an optimisation. Especially in times of automated forecasts and increasing requirements, such verification gains in importance.

This document verifies and interprets students' performance at the Vienna University's Institute of Meteorology and Geodynamics. Besides conventional parameters being used (such as the general error), forecasts will be compared to suitable reference forecasts. Particular attention will be paid to MOS forecasts - the so-called automated point forecasts (in this document for Vienna) - in order to finally attempt to answer the all-important question: man or machine?

Inhaltsverzeichnis

Zusammenfassung	II
Abstract	III
1 Einleitung und Motivation	3
2 Verifikation.....	7
2.1 Kontinuierliche Gütemaße.....	8
2.1.1 Mittlerer Fehler (<i>Mean Error</i> , ME).....	8
2.1.2 Mittlerer absoluter Fehler (<i>Mean Absolute Error</i> , MAE)	9
2.1.3 Wurzel des mittleren quadratischen Fehlers (<i>Root Mean Square Error</i> , RMSE) ...	9
2.1.4 Skill Score (SS)	10
2.1.5 Reduktion der Varianzen (<i>Reduction of Variance</i> , RV)	11
2.2 Kategorische Gütemaße.....	11
2.2.1 Trefferrate (<i>Proportion Correct</i> , PC).....	14
2.2.2 BIAS (<i>Frequency Bias Index</i>)	14
2.2.3 Treffersatz (T)	15
2.2.4 Erfolgsanteil (<i>Post Agreement</i> , PA).....	15
2.2.5 Entdeckungswahrscheinlichkeit (<i>Probability of Detection</i> , POD).....	16
2.2.6 Anteil des Falschen Alarms (<i>False Alarm Ratio</i> , FAR).....	17
2.2.7 Entdeckungswahrscheinlichkeit des Falschen Alarms (<i>False Alarm Rate</i> , FR)	17
2.2.8 Kritischer Erfolgsindex (<i>Critical Success Index</i> , CSI)	18
2.2.9 Hanssen-Kuipers Diskriminante (<i>True Skill Statistics</i> , TSS)	18
2.2.10 Heidke Skill Score (HSS)	19
2.3 Korrelationskoeffizient (<i>Correlation Coefficient</i> , r)	20
2.4 Streudiagramme (<i>Scatter Plots</i>).....	20
2.5 Referenzprognosen.....	22
3 Daten	27
3.1 Studentische Prognosen.....	27
3.2 Persistenzprognosen.....	28
3.3 Klimaprognosen	28
3.4 Model Output Statistics (MOS)	29
4 Ergebnisse und Interpretation.....	29
4.1 Subjektive Verifikation	29
4.2 Die Temperatur als kontinuierlicher Parameter	31

4.2.1 Verifikation der Minimumtemperatur (T_{\min}).....	32
4.2.2 Verifikation der 6 UTC-Temperatur	41
4.2.3 Verifikation der 12 UTC-Temperatur	49
4.2.4 Verifikation der Maximumtemperatur (T_{\max}).....	56
4.3 Die Temperatur als kategorischer Parameter	63
4.3.1 Kategorische Verifikation des PETRUS	65
4.3.2 Kategorische Verifikation des MOS	71
4.4 Zusammenfassung der Ergebnisse.....	74
4.5 Rückblick und Ausblick	77
Anhang	79
Abbildungsverzeichnis	85
Tabellenverzeichnis	87
Abkürzungsverzeichnis	89
Literaturverzeichnis	91
Danksagung	93
Curriculum Vitae	95

1 Einleitung und Motivation

Seit jeher beschäftigt es den Mensch, stets den Wert seiner Leistung bestätigt zu wissen bzw. die Qualität der selbigen zu kennen. Auch unter Meteorologen in der Wettervorhersage ist dies nicht anders, wobei gerade in der Meteorologie diese Bestätigung nicht nur zum Wohlwollen des einzelnen Prognostikers von Bedeutung ist, sondern auch – oder gerade deshalb – weil die Ergebnisse eines Meteorologen, also am Ende die Wetterprognose, einer breiteren Öffentlichkeit (also dem Konsumenten) von Nutzen sein sollen. Um diese Qualität gegenüber dem „Endabnehmer“ stets auf hohem Niveau halten zu können, bedient man sich nur allzu gern gewissen Tricks, die es einem Meteorologen ermöglichen, so gut wie immer vermeintlich „richtige“ Prognosen zu erstellen: mit verbalen Beschreibungen wie zum Beispiel „örtlich“ oder „vereinzelt“, aber auch „wechselhaft“ oder ganz simpel „kalt“ entzieht man sich zumeist der Möglichkeit der mathematischen bzw. statistischen Betrachtungsweise dieser Aussagen (Prognosen) und somit der Bestätigung einer möglichen erbrachten Leistung. Aus diesem Grund sollten einer jeden Prognose Zahlenwerte anstatt verbal formulierter Texte der Vorzug gegeben werden. (An dieser Stelle ist natürlich angemerkt, dass dieser statistische Zugang, Prognosen auf Zahlenwerte zu reduzieren, im klassischen, medial aufbereiteten Wetterbericht nur schwer umzusetzen und vermutlich auch nur wenig sinnvoll ist.)

Ein weiteres Problem der „richtigen“ Prognosen ist die Frage, wie „richtig“ definiert ist. Welchen Spielraum gewährt man sich bei der Vorhersage, so dass die Prognose noch „richtig“ ist? Liegt der erlaubte Fehler bei einer Temperaturprognose zum Beispiel bei $\pm 2\text{K}$, so würden heutzutage in Mitteleuropa in etwa 66% aller Prognosen eintreffen. Erweitert man das Fehlerintervall auf $\pm 4\text{K}$, so wären dies schon 92% [DORNINGER, 2002]. An diesem Beispiel ist gut ersichtlich, dass Prozentangaben allein oft gar keine Aussage über die Qualität einer Prognose erbringen und zumeist nur der Willkür des Einzelnen unterworfen sind.

Sind nun also Prognosen statistisch bzw. mathematisch zugänglich, so kann man diese (abseits von Prozentangaben) mit verschiedenen Gütemaßen auf diverse Eigenschaften hin untersuchen: Der Klassiker dabei ist die „Genauigkeit“ einer Vorhersage (*accuracy*), aber auch andere Eigenschaften wie die „Schärfe“ (*sharpness*, wie steht es um die Fähigkeit, Extremereignisse wie zum Beispiel Gewitter richtig zu prognostizieren?) können ermittelt werden. Diese Untersuchung bezeichnet man als *Verifikation von Wetterprognosen* und ist aus drei Gründen notwendig [BROWN, ET AL., WWRP]:

1. Um die Qualität einer Prognose festzuhalten und die Genauigkeit zu überprüfen (wie

- hoch ist der Prognosefehler?),
2. um die Vorhersagequalität zu verbessern (der erste Schritt besser zu werden ist zu wissen, was falsch gelaufen ist) und
 3. um die Qualität verschiedener Vorhersagen vergleichen zu können (inwieweit ist eine Prognose besser als eine andere und was macht diese besser?).

Allan Murphy, ein Pionier im Gebiet der Vorhersageverifikation, führt auf die Frage „Was eine Vorhersage gut macht?“ drei Punkte an [MURPHY, 1993]:

1. Die Beschaffenheit einer Prognose (*consistency*), also den Zusammenhang zwischen einer Vorhersage und dem Wissen des Prognostikers, der die Vorhersage macht;
2. die Leistung bzw. Qualität einer Prognose (*quality*), also den Zusammenhang zwischen einer Vorhersage und der Beobachtung und
3. den Nutzen bzw. Wert einer Prognose (*value*), also der Vorteil einer Vorhersage für den Konsumenten.

Um diese drei Punkte einer Prognose zu bestimmen, bedient man sich der Verifikation, wobei Zweiterer (Leistung bzw. Qualität, „wie gut ist nun die Prognose tatsächlich?“) anhand verschiedener Verifikationsmaße leicht zu bestimmen ist (soweit die Prognosen, wie oben beschrieben, statistisch betrachtet werden können). Der Nutzen bzw. Wert einer Prognose lässt sich dann in Folge dieser Ergebnisse bewerten und ermitteln. Die Beschaffenheit der Prognose hingegen lässt sich in den meisten Fällen nicht mehr bestimmen, da Auskünfte über den Wissensstand eines Prognostikers, zum Zeitpunkt der Prognose, zumeist nicht verfügbar sind.

Neben der Evaluierung der menschlichen Leistung ist natürlich auch die Beurteilung der Qualität diverser Wettervorhersagemodelle oder (automatisierter) Punktprognosen heutzutage von immenser Bedeutung, schließlich kommt die moderne Wettervorhersage heute kaum mehr ohne diese Produkte aus. Dabei stellt sich meist nicht nur mehr die Frage, wie „gut“ das Produkt ist, sondern viel mehr, ob das Produkt „gut genug“ ist. Henry R. Stanski, ein weiterer Pionier in der Vorhersageverifikation, unterscheidet aus diesem Gesichtspunkt zwei Arten der Verifikation [STANSKI ET AL., 1989], wobei der wirtschaftliche Nutzen bzw. der Nutzen für den Konsumenten ignoriert wird:

1. Verifikation aus administrativen Gründen (*administrative verification*): um gegenüber dem Geldgeber (zum Beispiel einem Ministerium) den Unterhalt des Wetterdienstes, die Anschaffung neuer Ausstattung (zum Beispiel bessere Computer) oder neue Forschungsgründe (zum Beispiel etwaige Verbesserungen im Vorhersagemodell) zu rechtfertigen. Dafür sind zumeist Verifikationsstatistiken über

einen größeren Datensatz notwendig.

2. Verifikation aus wissenschaftlichen Gründen (*scientific verification*): um die Stärken und Schwächen des jeweiligen Vorhersagemodells (aber auch eines Prognostikers) zu identifizieren und bei Bedarf Verbesserungen durchzuführen.

Warum auch immer die Qualität einer Prognose von Interesse ist, um repräsentative Ergebnisse zu bekommen, ist stets eine systematische Herangehensweise erforderlich, um jedwede Willkür auszuschließen. Weiters muss, wie schon oben angedeutet, die Prognose quantifizierbar, also statistisch überprüfbar, und die Nachprüfbarkeit sämtlicher Daten, aber auch die der verwendeten Qualitäts- bzw. Verifikationsmaße, gesichert sein.

In dieser Arbeit wird die Leistung bzw. die Qualität (*quality*) der Prognosen der Studenten am Institut für Meteorologie und Geodynamik (IMGW) der Universität Wien erhoben (*scientific verification*), wobei nicht die Leistung Einzelner, sondern stets die Gesamtheit aller zur Überprüfung stehen. Dabei soll der Beschaffenheit der Prognose (*consistency*) keine Bedeutung beigemessen werden, alle Prognosen werden also unabhängig des Wissensgrades der einzelnen Studenten stets zusammen betrachtet. Neben der Evaluierung gewisser Standardparameter wie dem Prognosefehler und der Genauigkeit einer Prognose wird die Leistung der studentischen Vorhersagen auch mit geeigneten Vergleichsprognosen (Referenzprognosen, siehe Kapitel 2.4) überprüft, die sowohl einfacher (Persistenzprognose) als auch statistischer Natur (Klimamittel, automatisierte Punktprognosen) sein können. Daraus kann wiederum, wenn man will, der Wert bzw. Nutzen (*value*) für den Konsumenten abgeleitet werden (ist der Mensch dem Zufall bzw. der Maschine überlegen?).

Im folgenden Kapitel 2 soll nun zuerst auf die Verifikation an sich eingegangen und die in dieser Arbeit verwendeten Verifikationsmaße sowie die Referenzprognosen näher erläutert werden. In Kapitel 3 werden die verwendeten Daten und schließlich in Kapitel 4 die Ergebnisse vorgestellt und interpretiert.

2 Verifikation

Verifikation, die; lat. *veritas* (Wahrheit) und *facere* (machen)

Verifikation ist die Bewertung und Quantifizierung des Zusammenhanges zwischen einem angepassten Satz von Vorhersagen (*forecasts*) und Beobachtungen (*observations*). Zur Überprüfung der Qualität einer Vorhersage werden verschiedene Verifikationsmaße herangezogen, die die folgenden Eigenschaften untersuchen [MURPHY, 1993], [STANSKI ET AL., 1989]:

- den systematischen Prognosefehler (*BIAS*).
- den linearen Zusammenhang zwischen Vorhersage und Beobachtung (*association*).
- die Genauigkeit (*accuracy*). Entspricht der Differenz (dem Fehler) zwischen Vorhersage und Beobachtung; je kleiner die Differenz (der Fehler), desto größer die Genauigkeit.
- die Verbesserung gegenüber einer Standard- oder Referenzprognose (*skill*). Wettervorhersagen können genau sein, weil das Wetter einfach zu prognostizieren ist (z.B. Omega-Lage im Sommer). Der *skill* berücksichtigt dies.
- die Zuverlässigkeit (*reliability*). Sie ist äquivalent dem *BIAS*. Man unterscheidet zwischen der *overall reliability* (*unconditional BIAS*; alle Vorhersagen werden zusammen betrachtet) und der *reliability* (*conditional BIAS*; die Vorhersagen werden in vordefinierte Kategorien (*subsets*) unterteilt).
- die Auflösung (*resolution*). Entspricht der Fähigkeit, verschiedene Ergebnisse zu trennen. Zum Beispiel muss sich die Verteilung der Beobachtungen, wenn der Wert „X“ vorhergesagt wird, deutlich von der Verteilung der Beobachtungen, wenn der Wert „Y“ vorhergesagt wird, unterscheiden. In diesem Fall spricht man von hoher Auflösung, ungeachtet ob die Vorhersage falsch oder richtig ist.
- die Schärfe (*sharpness*), also die Fähigkeit, Extremereignisse zu prognostizieren. Eine Vorhersage kann Schärfe besitzen, auch wenn diese falsch ist. Als Beispiel für eine Prognose ohne jede Schärfe sei die Klimatologie genannt.
- die Unsicherheit (*uncertainty*). Je größer die Unsicherheit, desto größer die Variabilität der Beobachtungen und desto schwieriger die Prognose.

Dabei gibt es eine Reihe von anwendbaren Gütemaßen, ein universelles Gütemaß hingegen nicht: Jedes Verifikationsmaß gibt nur einen oder einige, aber nicht alle Aspekte der

Prognosegüte wieder. Es ist daher ratsam, die Prognosegüte stets anhand mehrerer Verifikationsmaße zu betrachten ([DORNINGER, 2002], [STANSKI ET AL., 1989], [NURMI, 2003]), wobei zu beachten ist, dass nicht jedes Verifikationsmaß zur Beurteilung jeder Vorhersagegröße verwendet werden kann, da grundsätzlich zwei Arten von Verifikationsmaßen unterschieden werden:

- Gütemaße für kontinuierliche Vorhersageelemente
- Gütemaße für kategorische Vorhersageelemente

Im Folgenden sollen die wichtigsten kontinuierlichen und kategorischen Verifikationsmaße vorgestellt werden. Als Quelle dient dabei hauptsächlich, wenn nicht anders angegeben, [STANSKI ET AL., SURVEY OF COMMON VERIFICATION METHODS IN METEOROLOGY, 1989] und [WILKS, D.S.: STATISTICAL METHODS IN THE ATMOSPHERIC SCIENCES, 2006].

2.1 Kontinuierliche Gütemaße

Kontinuierliche Gütemaße untersuchen ob und um wie viel eine Vorhersage von der tatsächlichen Beobachtung abweicht. Sie entsprechen somit den klassischen Fehlermaßen. Der gebräuchlichste, zu untersuchende, meteorologische Parameter in dieser Kategorie ist wohl die Temperatur (typischerweise T_{\min} und T_{\max} bzw. die Temperatur zu festgelegten Zeiten wie 6 und 12 UTC). Aber auch die Windrichtung bzw. Windgeschwindigkeit, die Feuchte, die Niederschlagsmenge oder der Bewölkungsgrad können als kontinuierliche Parameter angesehen werden.

2.1.1 Mittlerer Fehler (*Mean Error*, ME)

Der mittlere Fehler entspricht dem *BIAS* und beschreibt die mittlere Abweichung zwischen einer Vorhersage (F_i) und einer Beobachtung (O_i). Er stellt somit ein generelles (lineares) Maß für die Zuverlässigkeit (*reliability*) einer Prognose dar. Als sehr einfaches Fehlermaß kann der ME bei unbegrenzten Parametern (wie der Temperatur) Werte zwischen $-\infty$ und $+\infty$ annehmen. Sind dabei die vorhergesagten Werte (F_i) im Mittel höher als die Beobachtungen (O_i), so ist der ME positiv; umgekehrt negativ. Im Idealfall (*perfect score*) ist der ME gleich Null.

$$ME = \frac{1}{n} \sum_{i=1}^n (F_i - O_i) \quad (2.1)$$

n ist dabei die Anzahl der Vorhersagen (F_i). Der große Nachteil beim ME ist, dass sich (große) Fehler unterschiedlicher Vorzeichen gegenseitig aufheben und so ein vermeintlich gutes Ergebnis vortäuschen können. Der ME gibt somit keinerlei Auskunft über die Magnitude des Fehlers. Es ist also ratsam, den ME immer zusammen mit dem mittleren absoluten Fehler (MAE; vgl. Kapitel 2.1.2) zu betrachten [NURMI, 2003].

2.1.2 Mittlerer absoluter Fehler (*Mean Absolute Error*, MAE)

Der mittlere absolute Fehler ist das Maß für die Genauigkeit einer Prognose (*accuracy*). Wie der ME (vgl. Kapitel 2.1.1) ist auch der MAE ein lineares Maß, gibt aber im Gegensatz zum ME die mittlere Magnitude eines Fehlers an (nicht jedoch dessen Richtung). Der MAE kann Werte zwischen Null und $+\infty$ annehmen (bei unbegrenzten Parametern); je größer der mittlere absolute Fehler ist, desto schlechter ist die Vorhersage.

$$MAE = \frac{1}{n} \sum_{i=1}^n |F_i - O_i| \quad (2.2)$$

2.1.3 Wurzel des mittleren quadratischen Fehlers (*Root Mean Square Error*, RMSE)

Der RMSE ist ein quadratisches Maß und beschreibt wie der MAE (vgl. Kapitel 2.1.2) die Genauigkeit (*accuracy*) einer Prognose, wichtet jedoch große Fehler stärker. Das Maß verleitet somit zu einer eher konservativen (Klimatologie-ähnlichen) Prognose [NURMI, 2003].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2} \quad (2.3)$$

Auch der RMSE kann bei unbegrenzten Parametern Werte zwischen Null und $+\infty$ annehmen, im Idealfall (*perfect score*) ist der RMSE gleich Null.

MAE und RMSE liefern eine Abschätzung für die Fehlervarianz. Dabei gilt:

RMSE > MAE: hohe Fehlervarianz

RMSE ~ MAE: geringe Fehlervarianz

RMSE = MAE: keine Fehlervarianz (alle Fehler haben die gleiche Magnitude)

Der RMSE kann nicht kleiner als der MAE sein. Bei (zu) kleinen Datensätzen ist der MAE dem RMSE vorzuziehen.

Der Vollständigkeit halber ist an dieser Stelle auch noch der **mittlere quadratische Fehler** (*Mean Square Error*, MSE), der bei der Reduktion der Varianzen (vgl. Kapitel 2.1.5) zum Einsatz kommt, genannt:

$$MSE = \frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2 \quad (2.4)$$

2.1.4 Skill Score (SS)

Oft ist die Aussage über den Fehler einer Vorhersage allein nicht ausreichend, um die Prognosegüte einer Vorhersage zu beschreiben. Schließlich kann auch eine vermeintlich „schlechte“ Prognose noch immer „gut“ sein (und umgekehrt), wenn eine Verbesserung gegenüber einer Referenzprognose (*reference forecast*) vorliegt. Diesen Vergleich von Vorhersage und Referenzprognose bezeichnet man als *Skill Score*. Als Referenzprognose zieht man dabei typischerweise die Klimatologie oder die Persistenz heran (siehe Kapitel 2.5). Dabei gilt: je besser die Referenzprognose, desto schwieriger ist es, einen guten *Skill Score* zu erzielen.

Der gebräuchlichste *Skill Score* ist der sogenannte *Heidke Skill Score* (vgl. Kapitel 2.2.10), der bei kategorischen Gütemaßen zum Einsatz kommt.

Die allgemeine Form des SS sieht wie folgt aus:

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}} \quad (2.5)$$

wobei A für ein beliebiges Gütemaß der zu untersuchenden Prognose, A_{ref} für das Gütemaß der Referenzprognose (*reference forecast*) und A_{perf} für das der perfekten Prognose (*perfect forecast*) steht.

Setzt man für A den MAE ein, so ergibt sich (MAE_{perf} ist per Definition Null):

$$SS_{MAE} = 1 - \frac{MAE}{MAE_{ref}} = 1 - \frac{\sum |F_i - O_i|}{\sum |F_{i,ref} - O_i|} \quad (2.6)$$

Der SS_{MAE} kann Werte zwischen $-\infty$ und Eins annehmen (bei unbegrenzten Parametern); Eins entspricht dabei der perfekten Prognose. Ist der SS_{MAE} gleich Null, so ist keine Verbesserung zur Referenz erzielt werden (*no skill*). Bei negativen Werten ist die Prognose schlechter als die Referenz.

2.1.5 Reduktion der Varianzen (*Reduction of Variance, RV*)

Setzt man in Gleichung (2.5) für A den mittleren quadratischen Fehler (MSE, vgl. Kapitel 2.1.3) ein, so erhält man die *Reduction of Variance*.

$$RV = 1 - \frac{MSE}{MSE_{ref}} = 1 - \frac{\sum (F_i - O_i)^2}{\sum (F_{i,ref} - O_i)^2} \quad (2.7)$$

Wie beim SS_{MAE} (vgl. Kapitel 2.1.3) liegt auch bei der RV der Wertebereich bei unbegrenzten Parametern zwischen $-\infty$ und Eins, mit Eins als perfekter Prognose und Null mit *no skill*.

Bei zu geringer Stichprobe sind *Skill Scores* im Allgemeinen, speziell aber die RV, nur wenig aussagekräftig [NURMI, 2003], da falsch prognostizierte Extremereignisse voll durchschlagen.

2.2 Kategorische Gütemaße

Im Gegensatz zu den kontinuierlichen Gütemaßen (vgl. Kapitel 2.1), die den Vorhersagefehler als solchen untersuchen, verifizieren kategorische Gütemaße, ob und in welcher Häufigkeit bestimmte vorhergesagte Ereignisse (*forecasted events*) überhaupt eingetroffen sind oder nicht. Typische Parameter in dieser Kategorie sind Regen (eingetroffen/nicht eingetroffen bzw. ja/nein), Nebel (ja/nein), aber auch beispielsweise Frost (ja/nein). Viele meteorologische Parameter können somit sowohl als kontinuierlich als auch kategorisch angesehen werden (vgl. am Beispiel Temperatur Kapitel 2.1). Um diese ja/nein Parameter zu verifizieren, bedient man sich im einfachsten Fall einer 2x2 Kontingenztafel (siehe TABELLE

2-1 und TABELLE 2-2), in der die Häufigkeiten der einzelnen *events* dargestellt werden [BROWN ET AL., WWRP]:

- Treffer (*hit*): ein Ereignis ist eingetroffen und wurde auch vorhergesagt
- Negativer Treffer (*correct negative*): ein Ereignis ist nicht eingetroffen und wurde auch nicht vorhergesagt
- Falscher Alarm¹⁾ (*false alarm*): ein Ereignis wurde vorhergesagt, aber nicht beobachtet; das Ereignis wurde somit fälschlicherweise vorhergesagt
- Überraschung (*miss*): ein Ereignis wurde nicht vorhergesagt, aber beobachtet; das Ereignis wurde somit versäumt

Im besten Fall sind die Überraschungen (*misses*) und die Falschen Alarme (*false alarms*) gleich Null. Mit Hilfe dieser Kontingenztabelle können in Folge eine Vielzahl von geeigneten Verifikationsmaßen (Genauigkeitsmaße wie auch *Skill Scores*) berechnet werden. An dieser Stelle sei nochmals angemerkt, dass die Anwendung eines einzelnen Maßes in keinem Fall ausreichend ist [STANSKI ET AL., 1989], [NURMI, 2003].

Die Einteilung in eingetroffene bzw. nicht eingetroffene *events* und die Anwendung einer 2x2 Kontingenztabelle ist die einfachste Form einer solchen. Für manche Parameter sind mehrfach kategoriale Kontingenztabelle zu definieren (siehe TABELLE 2-3). So macht es zum Beispiel Sinn, den Niederschlag als einfach kategorisches Element (zum Beispiel Regen ist eingetroffen/ist nicht eingetroffen) wie auch als ein mehrfach kategorisches Element (wenn Regen eingetroffen ist, in welcher Stärke?) anzusehen. Das Definieren solcher Klassen erfolgt nach jeweiligem Nutzen. Bei der Analyse spaltet man eine solche Kontingenztabelle – unabhängig der Anzahl der Klassen – wieder in mehrere 2x2 Kontingenztabelle auf und berechnet so die einzelnen Verifikationsmaße (TABELLE 2-4).

		Beobachtung	
		ja	nein
Vorhersage	ja	Treffer (<i>hit</i>)	Falscher Alarm (<i>false alarm</i>)
	nein	Überraschung (<i>miss</i>)	Negativer Treffer (<i>correct negative</i>)

TABELLE 2-1: Darstellung der einzelnen events in einer 2x2 Kontingenztabelle.

		Beobachtung		Total
Vorhersage		ja	nein	
	ja	a	b	a + b
	nein	c	d	c + d
Total		a + c	b + d	a + b + c + d = n

TABELLE 2-2: Allgemeine Form einer 2x2 Kontingenztabelle.

Niederschlag in mm/24h		Beobachtung				Total
		0,0	≥0,1	≥1	≥10	Total
Vorhersage	0,0	a	b	c	d	Σ
	≥0,1	e	f	g	h	Σ
	≥1	i	j	k	l	Σ
	≥10	m	n	o	p	Σ
Total		Σ	Σ	Σ	Σ	Σ

TABELLE 2-3: Mehrfach kategoriale Kontingenztabelle am Beispiel Niederschlag.

Niederschlag in mm/24h		Beobachtung			Total	
		0,0	≥0,1	≥1	≥10	
Vorhersage	0,0	a	b		Σ	
	≥0,1				Σ	
	≥1	c	d			
	≥10					
Total		Σ	Σ		Σ	

TABELLE 2-4: Auflösung einer mehrfach kategorialen Kontingenztabelle am Beispiel Niederschlag.

2.2.1 Trefferrate (*Proportion Correct*, PC)

Das einfachste Verifikationsmaß bei der Bestimmung der Genauigkeit (*accuracy*) von Prognosen ist die Trefferrate. Sie beschreibt das Verhältnis der korrekten Vorhersagen (*hits + correct negatives*) zur Gesamtanzahl (*hits + misses + false alarms + correct negatives*).

$$PC = \frac{a + d}{n} \quad (2.8)$$

PC kann Werte zwischen Null und Eins annehmen, die perfekte Prognose liegt bei Eins. Als recht anschauliches Maß kann der PC leicht eine hohe Vorhersagegüte vortäuschen, indem Extremereignisse nicht vorhergesagt werden und somit stets die Kategorie, die zumeist nicht von Interesse ist, einen hohen Einfluss auf das Maß ausübt (oft die *correct negatives*).

Multipliziert man die Trefferrate mit 100%, so erhält man die **Trefferquote** (*Percent Correct*, **P**), also den Prozentsatz der richtigen Vorhersagen:

$$P = 100 \times \frac{a + d}{n} \quad (2.9)$$

2.2.2 BIAS (*Frequency Bias Index*, FBI)

Der *Frequency Bias Index* gibt das Verhältnis der Häufigkeiten von vorhergesagten Ereignissen (*hits + false alarms*) und den positiven Beobachtungen (*hits + misses*) an, ohne eine Aussage über die Genauigkeit (*accuracy*) einer Prognose zur urteilen.

$$FBI = \frac{a + b}{a + c} \quad (2.10)$$

Der Wertebereich liegt zwischen Null und $+\infty$, dabei gilt:

FBI = 1, keine Abweichung im Mittel (jedoch keine Aussage über die Prognosegüte)

FBI > 1, die Kategorie wurde öfters prognostiziert als beobachtet (*over-forecasted*)

FBI < 1, die Kategorie wurde öfters beobachtet als prognostiziert (*under-forecasted*)

2.2.3 Treffersatz (T)

Im Gegensatz zur Trefferrate (vgl. Kapitel 2.2.1), nimmt der Treffersatz mehr Rücksicht auf eventuelle Extremereignisse (zum Beispiel Hagel), gibt aber sonst nur wenig Auskunft [HOFMAN ET AL., 1974]. Der Wertebereich des Treffersatzes liegt zwischen minus Eins und Eins, mit Eins als perfekter Prognose.

$$T = 1 - \frac{2}{k} \sum_{i=1}^k \alpha_i \quad (2.11)$$

mit

$$\alpha_i = \frac{e_i - n_{ij}}{e_i} \quad (2.12)$$

Dabei gilt:

k = Anzahl der einzelnen Kategorien (2 bei einer 2x2 Kontingenztabelle)

e_i = Summe der Beobachtungen der einzelnen Kategorien (a + b bzw. c + d) geteilt durch die Gesamtsumme (n)

n_{ij} = Anzahl der korrekten Prognosen in der jeweiligen Kategorie (a bzw. d) geteilt durch die Gesamtsumme (n)

2.2.4 Erfolgsanteil (*Post Agreement*, PA)

Der Erfolgsanteil gibt nur jenen Anteil an richtigen Prognosen an, der auch vorhergesagt wurde (also den Anteil an prognostizierten Ereignissen, die auch eingetroffen, d.h. beobachtet worden sind). Es stellt somit die Treffer (*hits*) der Gesamtzahl der positiven Vorhersagen (*hits* + *false alarms*) gegenüber und ist somit ein Maß, welches nur auf prognostizierte, aber nicht eingetretene, Ereignisse (*false alarms*) sensibel reagiert und den nicht prognostizierten, aber eingetretenen, *events* (*misses*) keine Bedeutung beimisst.

$$PA = \frac{a}{a + b} \quad (2.13)$$

Der Wertebereich des PA liegt zwischen Null und Eins, mit Eins als perfekter Prognose.

Analog kann zum *Post Agreement* für die positiven Vorhersagen („ja“-Kategorie; Gleichung (2.13)) auch ein PA für die negativen Vorhersagen („nein“-Kategorie) definiert werden. Dieses stellt die negativen Treffer (*correct negatives*) der Gesamtzahl der negativen Vorhersagen (*misses* + *correct negatives*) gegenüber und überprüft somit, wie viele der nicht vorhergesagten *events* dann doch eingetroffen, dh. beobachtet worden sind.

2.2.5 Entdeckungswahrscheinlichkeit (*Probability of Detection, POD*)

Die Entdeckungswahrscheinlichkeit beschreibt das Verhältnis aus Treffern (*hits*) bzw. Negativen Treffern (*correct negatives*) zur Anzahl der Beobachtungen (*hits* + *misses* bzw. *false alarms* + *correct negatives*) und ist somit ein Maß dafür, wie viele der Beobachtungen vom Prognostiker auch korrekt vorhergesagt wurden. Die Entdeckungswahrscheinlichkeit für die „ja“-Kategorie wird gelegentlich auch als **Hit Rate (HR)**¹⁾ bezeichnet.

$$POD = HR = \frac{a}{a + c} \quad (2.14)$$

Analog die POD für die „nein“-Kategorie:

$$POD = \frac{d}{b + d} \quad (2.15)$$

POD (bzw. HR) kann dabei Werte zwischen Null und Eins annehmen; die perfekte Prognose liegt bei Eins. Da bei der *Hit Rate* der Anteil der *false alarms* keine Rolle spielt, ist das Maß leicht zu beeinflussen [NURMI, 2003]: der Prognostiker erstellt mehr „ja“ - Vorhersagen um seine Treffer zu erhöhen (das dabei auftretende Mehr an Falschen Alarmen ist bei der HR ohne Bedeutung) bzw. umgekehrt für die „nein“-Kategorie. Um eine mögliche Fehlinterpretation zu vermeiden, wird daher zusätzlich zur *Hit Rate* die *False Alarm Ratio* (vgl. Kapitel 2.2.6) eingesetzt.

Das Pendant zur HR ist die negativ orientierte *Miss Rate* (MR):

$$MR = 1 - HR = \frac{c}{a + c} \quad (2.16)$$

2.2.6 Anteil des Falschen Alarms (*False Alarm Ratio*, FAR)

Die *False Alarm Ratio* untersucht, wie oft ein vorhergesagtes Ereignis nicht eingetroffen ist oder anders ausgedrückt: wie oft falsch gewarnt wurde. Die FAR ist somit das Verhältnis zwischen den Falschen Alarmen (*false alarms*) und den „ja“- Vorhersagen (*hits* + *false alarms*). Das Maß reagiert somit sensibel auf die Anzahl der Falschen Alarme, lässt aber dafür die Anzahl der Überraschungen (*misses*) außer Acht.

$$FAR = \frac{b}{a + b} \quad (2.17)$$

Die FAR kann Werte zwischen Null und Eins annehmen, ist aber negativ orientiert. Die perfekte Prognose liegt somit bei Null.

Wie auch schon die *Probability of Detection* (vgl. Kapitel 2.2.5) kann die FAR leicht beeinflusst werden [NURMI, 2003]: der Prognostiker reduziert einfach die Zahl der *false alarms*, indem er spezielle *events* nicht vorhersagt (der gleichzeitige Anstieg der *misses* spielt bei der FAR keine Rolle). Aus diesem Grund sind POD und FAR stets gemeinsam zu betrachten. Sowohl POD als auch FAR sind im Übrigen sehr sensibel gegenüber der klimatologischen Häufigkeit eines Ereignisses.

2.2.7 Entdeckungswahrscheinlichkeit des Falschen Alarms (*False Alarm Rate*, FR)

Während die *False Alarm Ratio* (vgl. Kapitel 2.2.6) das Verhältnis zwischen den Falschen Alarmen und den „ja“-Vorhersagen betrachtet, stellt die *False Alarm Rate* die Falschen Alarme (*false alarms*) den Negativen Beobachtungen (*false alarms* + *correct negatives*) gegenüber. FR untersucht somit, wie oft ein nicht eingetroffenes *event* fälschlicherweise vorhergesagt wurde. Die Anzahl der Überraschungen (*misses*) spielen dabei keine Rolle. Die Entdeckungswahrscheinlichkeit des Falschen Alarms wird in der Literatur oft auch als *Probability of False Detection* (POFD) angeführt.

$$FR = POFD = \frac{b}{b + d} \quad (2.18)$$

Der Wertebereich der *False Alarm Rate* liegt zwischen Null und Eins, die perfekte Prognose liegt bei Null (ist demnach negativ orientiert).

2.2.8 Kritischer Erfolgsindex (*Critical Success Index*, CSI)

Der kritische Erfolgsindex ist ein gutes Maß für Extremereignisse (*rare events*), das sowohl die Treffer (*hits*) wie auch die Falschen Alarme (*false alarms*) und die Überraschungen (*misses*) berücksichtigt, den Negativen Treffern (*correct negatives*) hingegen keinerlei Bedeutung beimisst. Der CSI entspricht somit der Trefferrate (vgl. Kapitel 2.2.1) ohne den Negativen Treffern. Der Kritische Erfolgsindex ist auch bekannt als **Threat Score (TS)**.

$$CSI = TS = \frac{a}{a + b + c} \quad (2.19)$$

Der CSI kann Werte zwischen Null und Eins annehmen, die perfekte Prognose liegt bei Eins (und bedeutet, dass alle „ja“-Vorhersagen auch beobachtet wurden). Er ist sehr stark abhängig von der klimatologischen Häufigkeit eines Ereignisses (schlechtere Werte für seltenere *events*).

2.2.9 Hanssen-Kuipers Diskriminante (*True Skill Statistics*, TSS)

Die Hanssen-Kuipers Diskriminante untersucht, wie gut eine Vorhersage beobachtete Ereignisse von den nicht beobachteten Ereignissen unterscheidet. Anders ausgedrückt wird untersucht, ob die Vorhersage eines (speziellen) *events* öfter zu einem Falschen Alarm führt, oder nicht. Hierfür werden alle Elemente der Kontingenztabelle (vgl. TABELLE 2.1 und 2.2) verwendet:

$$TSS = \frac{(ad - bc)}{(a + c)(b + d)} \quad (2.20)$$

Die TSS entspricht somit der Differenz von Entdeckungswahrscheinlichkeit (POD bzw. *Hit Rate* HR; vgl. Kapitel 2.2.5) und der Entdeckungswahrscheinlichkeit des Falschen Alarms (POFD bzw. *False Alarm Rate* FR; vgl. Kapitel 2.2.7):

$$TSS = \frac{a}{(a+c)} - \frac{b}{(b+d)} \quad (2.21)$$

$$= POD - POFD \equiv HR - FR$$

Die TSS gilt als *Skill Score*, auch wenn es nicht dessen klassischem Aufbau entspricht. Dementsprechend liegt der Wertebereich zwischen minus Eins und Eins; die perfekte Prognose liegt bei Eins (POD = 1, POFD = 0). Bei Null (POD = POFD) entspricht die Vorhersage einer Zufallsprognose und hat daher *no skill*. Die TSS ist für seltene Ereignisse sehr gut geeignet, da sie von den Treffern (*hits*) besonders abhängig ist. Würde man also ein seltenes Ereignis nie (richtig) vorhersagen, so kann das Ergebnis im besten Fall nur Null sein – abhängig von der Anzahl der Negativen Treffer (*correct negatives*) und der Falschen Alarme (*false alarms*).

2.2.10 Heidke Skill Score (HSS)

Der Heidke Skill Score ist der am häufigsten verwendete *Skill Score* (vgl. Kapitel 2.1.4). Wie jeder *Skill Score* bedient er sich einer Referenzprognose (vgl. Kapitel 2.4). Ist diese Referenz der Zufall (*events by chance*), so ergibt sich (mit Z als Zufall):

$$Z = \frac{(a+b)(a+c) + (b+d)(c+d)}{n} \quad (2.22)$$

Der HSS folgt nun der allgemeinen Form des *Skill Scores* (vgl. Gleichung 2.5), als Verifikationsmaß (A) dient die Trefferrate (vgl. Kapitel 2.2.1), als Verifikationsmaß der Referenzprognose (A_{ref}) die Trefferrate durch Zufall:

$$HSS = \frac{\frac{a+d}{n} - \frac{Z}{n}}{1 - \frac{Z}{n}} = \frac{a+d-Z}{n-Z} \quad (2.23)$$

Durch Einfügen und Vereinfachung der Gleichungen (2.22) und (2.23) erhält man:

$$HSS = \frac{2(ad - bc)}{[(a + c)(c + d) + (a + b)(b + d)]} \quad (2.24)$$

Der HSS gibt somit den Anteil der korrekten Vorhersagen wieder, eliminiert dabei aber den Anteil der durch Zufall richtigen Prognosen. Auch wenn beim HSS in der Regel der Zufall als Referenz genutzt wird (aufgrund der Einfachheit), so kann auch die Persistenz oder die Klimatologie herangezogen werden (wie ursprünglich von Heidke auch präferiert [STANSKI ET AL., 1989]; siehe Kapitel 2.5). Der Wertebereich des HSS liegt zwischen $-\infty$ und Eins, mit Eins als perfekter Prognose. Bei Null ist keine Verbesserung gegenüber der Referenz erzielt worden (*no skill*), bei negativen Werten ist die Referenz besser (*negative skill*).

2.3 Korrelationskoeffizient (*Correlation Coefficient, r*)

Der Korrelationskoeffizient untersucht, wie gut die Vorhersagen (F_i) mit den Beobachtungen (O_i) zusammenhängen (korrelieren). Visuell ausgedrückt, wie nah einzelne Wertepaare der 45° Linie (= perfekte Prognose) sind (vgl. ABBILDUNG 2-1). Der Vorhersagefehler (*BIAS*) als solcher wird dabei nicht berücksichtigt; somit können auch Vorhersagen mit vermeintlich großen Fehlern einen guten Korrelationskoeffizienten vorweisen [BROWN, ET AL., WWRP]. Der Korrelationskoeffizient ist sensibel gegenüber Ausreißer.

$$r = \frac{\sum(F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum(F_i - \bar{F})^2} \sqrt{\sum(O_i - \bar{O})^2}} \quad (2.25)$$

Der Wertebereich des Korrelationskoeffizienten liegt zwischen minus Eins und Eins; die perfekte Prognose liegt bei Eins (alle Wertepaare liegen in einem Streudiagramm auf der 45° Linie).

2.4 Streudiagramme (*Scatter Plots*)

Das Streudiagramm ist ein verhältnismäßig einfaches, nichts desto trotz recht aussagekräftiges Verifikationsmaß für kontinuierliche Parameter (wie der Temperatur oder dem Wind; siehe ABBILDUNG 2-1). Es berücksichtigt alle Vorhersagen eines Datensatzes und stellt diese gegen die Beobachtungen (als Wertepaar) graphisch in einem kartesischen Koordinatensystem dar (*forecast vs. observation*). Durch die daraus resultierende

Punktwolke können Informationen der Abhängigkeit zwischen prognostizierten und beobachteten Werten gesammelt werden. Die perfekte Prognose (*perfect prognosis*) liegt auf der 45°-Achse (Vorhersage gleich Beobachtung), sofern die Abszisse (x-Achse) und die Ordinate (y-Achse) denselben Maßstab besitzen. Durch eine Regressionsanalyse kann in der Punktwolke eine Trendlinie bestimmt werden, die im besten Fall der 45°-Achse entspricht. Diese Trendlinie gibt (visuell) Auskunft über die relative Qualität der Vorhersagen: Abweichungen der Linie von dieser 45°-Achse entsprechen dem *BIAS* und sind somit ein Maß für die Verlässlichkeit (*reliability*). Ist diese Trendlinie horizontal, so haben die Vorhersagen *no skill*. Aus dem Grad der Neigung dieser Trendlinie kann zudem auch der „Mut-Faktor“ interpretiert werden, also ob und in welchem Ausmaß auch Extremwerte (ungeachtet deren Eintreffens) prognostiziert werden.

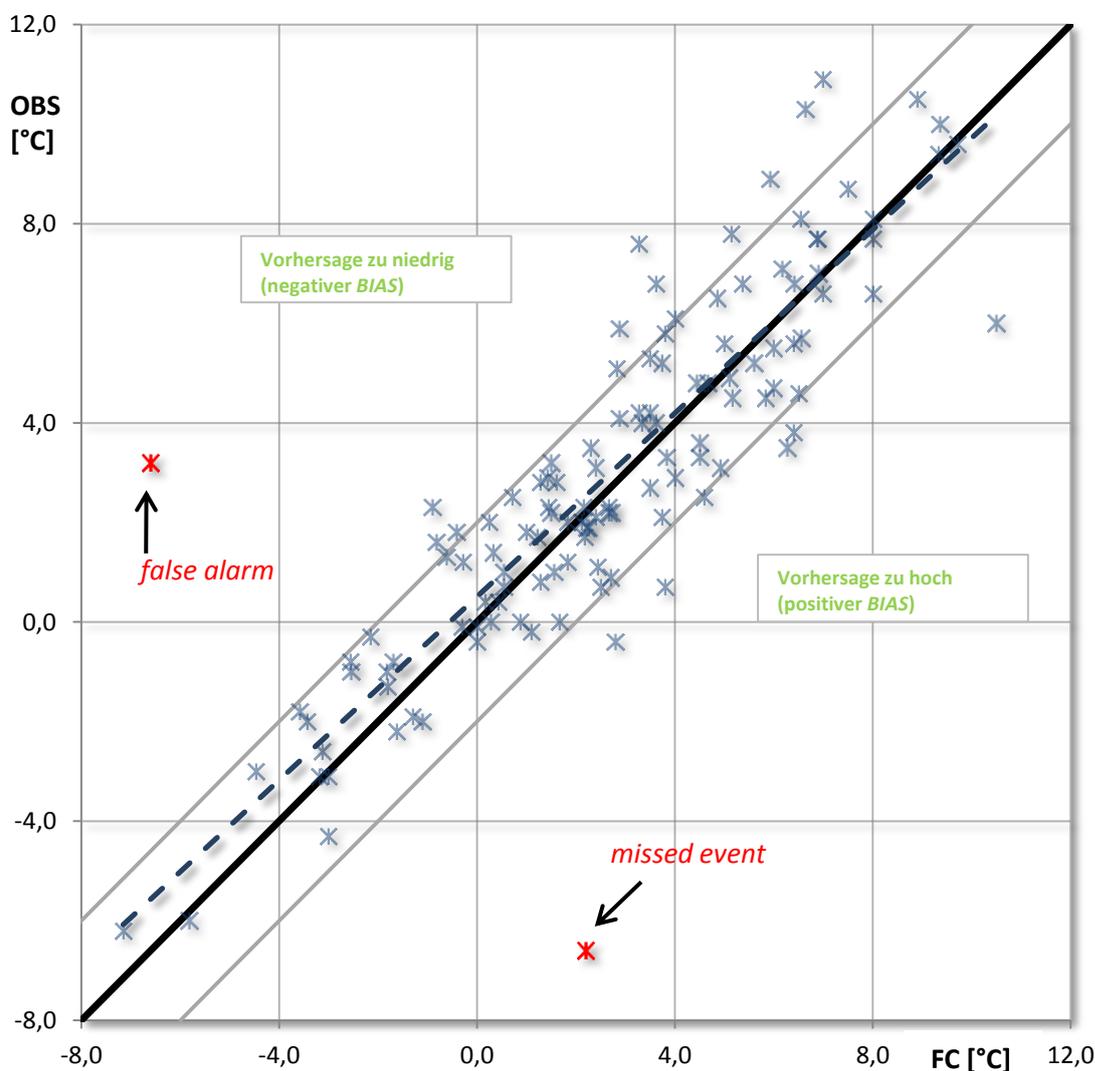


ABBILDUNG 2-1: Beispiel eines Streudiagramms nach [STANSKI ET AL., 1989]; die Daten (blaue Sterne) entsprechen den Vorhersagen und den dazugehörigen Beobachtungen der Minimumtemperatur für den Monat März der Jahre 2000-2010 für Wien. Die Wertepaare false alarm und missed event (rote Sterne) wurden zur Veranschaulichung hinzugefügt. Die blaue (gestrichelte) Linie ist die Trendlinie der Daten, die schwarze Linie entspricht der perfekten Prognose, die grauen Linien stehen für ein (mögliches erlaubtes) Fehlerintervall von $\pm 2^{\circ}\text{C}$. In Summe wurde die Minimumtemperatur zu niedrig vorhergesagt, vor allem bei Frost.

2.5 Referenzprognosen

Referenz, die; lat. *referre* (berichten, in Beziehung setzen)

Wie schon im Kapitel 2.1.4 angedeutet, hängt der Erfolg oder Misserfolg einer Prognose nicht allein vom Prognosefehler der selbigen ab. Schließlich kann eine vermeintlich „gute“ Prognose noch immer als „schlecht“ angesehen werden, nämlich dann, wenn keinerlei Verbesserung im Vergleich mit anderen Prognosen erzielt worden ist (*no skill*). Diese Vergleichsprognosen bezeichnet man als Referenzprognosen (*reference forecasts*), die gängigsten sind ([NURMI, 2003], [STANSKI ET AL., 1989]):

- Zufallsprognosen
Die Zufallsprognose basiert auf dem klassischen 50/50-Prinzip – ein Ereignis tritt ein oder nicht (das beste Beispiel hierfür ist das Werfen einer Münze). Als schlechteste aller Referenzprognosen (niedrige Verlässlichkeit (*reliability*) und Auflösung (*resolution*), Trefferchance 50%) ist jedem Meteorologen nahe zu legen, stets besser als der Zufall zu sein.
- Persistenzprognosen
Unter Persistenz versteht man das langfristige Fortbestehen einer Sache. Im Fall der Persistenzprognose bedeutet dies, dass der beobachtete Wetterzustand zum Zeitpunkt der Prognose gleich der selbigen entspricht. Oder anders ausgedrückt: Das Wetter gestern ist gleich dem Wetter heute, das Wetter heute ist gleich dem Wetter morgen. Die Trefferchance einer Persistenzprognose ist höher als die der Zufallsprognose und liegt in Mitteleuropa bei etwa 66%, was auf die Langlebigkeit von Wetterlagen zurückzuführen ist. Kurzfrist-Persistenzprognosen haben zudem eine gute Verlässlichkeit (*reliability*) und eine zumeist hohe Auflösung (*resolution*).
- Prognosen basierend auf der Klimatologie
Eine Vorhersage basierend auf der Klimatologie greift auf die jeweiligen Klimanormalwerte des Vorhersagezeitraums bzw. des Vorhersageortes zurück. Die Klimaprognose hat zwar eine höhere Trefferrate als der Zufall, ist aber der Persistenz doch unterlegen, da Klimanormalwerte stets als ein Mittel zu verstehen sind. Klimaprognosen haben eine hohe Verlässlichkeit (*reliability*), diese leidet aber unter der fehlenden Fähigkeit, Extremereignisse zu erkennen (*sharpness*). Zudem besitzen diese Prognosen auch eine schwache Auflösung (*resolution*).

In dieser Arbeit werden als Referenzprognose die Persistenz und die Klimatologie herangezogen, auf die Zufallsprognose wird verzichtet. Um eine zusätzliche Vergleichsmöglichkeit zu bekommen, wird die MOS-Prognose als dritte Referenzprognose definiert.

- MOS-Prognosen

MOS steht für *Model Output Statistics*. MOS interpretiert und korrigiert numerische Modellvorhersagen durch die Anwendung statistischer Beziehungen zwischen früheren Vorhersagen dieser Modelle und den tatsächlichen Beobachtungen [DWD, ERKLÄRUNGEN ZU MOS UND MOS-GUIDANCE]. Als statistische Vorhersagemethode werden dabei Beziehungen zwischen einem sogenannten Prädiktand (jener Größe, die vorhergesagt werden soll; auch als abhängige Variable bezeichnet) und einem oder mehreren Prädiktoren (jener Größe, die den Prädiktanden maßgeblich beeinflusst und sich aus synoptischen Beobachtungen oder numerischen Modellprognosen zusammensetzt; auch als abhängige Variable bezeichnet) abgeleitet:

$$y = f(x_1, x_2, x_3, \dots, x_n) \quad (2.26)$$

Dabei ist y die abhängige (vorherzusagende) Variable (oder der Prädiktand) und x die unabhängige(n) Variable(n) (die Prädiktoren). Als anschauliches Beispiel sei die Minimumtemperatur (hier der Prädiktand) erwähnt, die von einem oder mehreren Prädiktoren (unter anderem dem Bewölkungsgrad, dem Wind, etc.) abhängig sein kann.

Mit Hilfe einer multiplen linearen Regression werden daraus Vorhersagegleichungen definiert, die der folgenden Form entsprechen:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (2.27)$$

mit a_n als Regressionskoeffizienten (für eine genauere Beschreibung bzw. Herleitung von MOS-Prognosen und anderen statistischen Vorhersagemethoden siehe ua. [WILKS, 2006]).

Im Unterschied zu anderen statistischen Vorhersagemethoden (wie der sogenannten „klassischen Methode“ oder der „Perfect Prog Methode“) werden beim MOS die unabhängigen Variablen bei der Ableitung dieser Vorhersagegleichungen (Prädiktoren) aus einem Satz von in der Vergangenheit prognostizierten Werten eines numerischen Vorhersagemodells entnommen.

$$y_t = f_{MOS}(x_t) \quad (2.28)$$

Der Index t soll trotz der Verwendung eines historischen Datensatzes auf einen gleichen aber zukünftigen Zeitpunkt hindeuten. MOS vergleicht somit das Bodenwetter direkt mit den numerischen Modellprognosen und kann als statistische Interpretation eines Vorhersagemodells in Hinblick auf das lokale Wetter am Boden angesehen werden [HERMANN, 1998].

MOS-Vorhersagegleichungen sind für verschiedene Vorhersagetermine („warme“ und „kalte“ Jahreszeit) und verschieden Vorhersageorte stets eigens zu erstellen und sind im Vergleich zu den anderen statistischen Vorhersagemethoden modellabhängig. Aufgrund der hohen Verlässlichkeit (*reliability*) und Auflösung (*resolution*) sind MOS-Prognosen heute dennoch weltweit im Gebrauch, haben aber, wie jede statistische Prognose, eine niedrige Schärfe (*sharpness*).

3 Daten

Ein 10-jähriger Datensatz, bestehend aus subjektiv erstellten Prognosen (studentische Prognosen), liegt dieser Arbeit zu Grunde. Neben diesen Vorhersagen wurden vier Referenzprognosen (siehe Kapitel 2.4) definiert, um die Vorhersagegüte (den *Skill*) der studentischen Prognosen zu bestimmen: eine dieser Referenzprognosen basiert auf der Persistenz, zwei auf der Klimatologie und die vierte auf einem statistischen (automatisierten) Vorhersageverfahren (MOS). In diesem Kapitel sollen nun die in dieser Arbeit verwendeten Daten näher vorgestellt werden.

3.1 Studentische Prognosen

Seit dem Jahr 1992 wird am Institut für Meteorologie und Geophysik (IMGW) der Universität Wien die Lehrveranstaltung „Besprechung des aktuellen Wettergeschehens“ mit den Schwerpunkten „advektive Prozesse“ (in den Wintersemestern) und „konvektive Prozesse“ (in den Sommersemestern) abgehalten, in der die Studenten das in anderen Lehrveranstaltungen angeeignete meteorologische Wissen in die Praxis umsetzen und lernen, ihre eigenen Wetterprognosen zu erstellen. Die so gesammelten studentischen Prognosen entsprechen einer klassischen Punktprognose für die Stadt Wien, im Speziellen für die Hohe Warte im 19. Wiener Gemeindebezirk. Im Laufe der Jahre wurden diese Prognosen mehrmals adaptiert und angepasst. Folgende Parameter werden derzeit (2010) prognostiziert:

- Minimumtemperatur der Vorhersagenacht in Grad Celsius [T_{\min}]
- Maximumtemperatur des Vorhersagetages in Grad Celsius [T_{\max}]
- Temperatur um 6 UTC des Vorhersagetages in Grad Celsius [T_6]
- Temperatur um 12 UTC des Vorhersagetages in Grad Celsius [T_{12}]
- Wetterzustand in vordefinierten Zeitintervallen (1. & 2. Nachthälfte, Vormittag, Nachmittag); die Verschlüsselung entspricht der der allgemeinen SYNOP-Verschlüsselung (4 für Nebel, 5 für Nieselregen, 6 für Regen, 7 für Schneefall, 8 für Schauer und 9 für Gewitter) [W_{18-00} , W_{00-06} , W_{06-12} , W_{12-18}]
- 24h-Niederschlagsmenge des Vorhersagezeitraums in mm [RR_{18-18}]
- relative Sonnenscheindauer des Vorhersagezeitraums in % [Sd_{06-18}]
- Bedeckungsgrad um 12 UTC des Vorhersagetages in Achtel [N_{12}]

- Windrichtung um 12 UTC des Vorhersagetages in Grad [dd_{12}]
- Windgeschwindigkeit um 12 UTC des Vorhersagetages in Meter pro Sekunde [ff_{12}]

In dieser Arbeit werden die Temperaturen (T_{\min} , T_{\max} , T_6 , T_{12}) der studentischen Prognosen verifiziert, der Datensatz erstreckt sich über die Jahre 2000-2010 und beinhaltet jeweils 8138 (T_{\min} , T_{\max}) bzw. 8076 (T_6 , T_{12}) Daten pro Parameter (die kleine Diskrepanz zwischen den Parametern beruht darauf, dass in den ersten Tagen des Datensatzes die 6 UTC und 12 UTC Temperatur noch kein Vorhersageelement war). An dieser Stelle ist angemerkt, dass die Lehrveranstaltung nur montags bis donnerstags stattfindet. An Freitagen, sowie an vorlesungsfreien Tagen (Feiertagen, Wochenenden) und in den vorlesungsfreien Monaten (Februar, Juli-September) liegen keine Prognosen vor. Die Prognosen der Studenten sind stets in ganzen Zahlen gegeben, etwaige Dezimalprognosen wurden adäquat gerundet.

Für viele Verifikationsmaße werden nicht die studentischen Prognosen im Einzelnen, sondern wird stets das Mittel aller verwendet. Bei diesem Mittel handelt es sich um ein arithmetisches Mittel aller am jeweiligen Tag abgegebenen Prognosen. Dies wird in dieser Arbeit als „**PETRUS**“ bezeichnet. PETRUS ist stets auf eine Kommazahl gerundet. Der so gewonnene Datensatz beinhaltet 1040 (bei T_{\min} und T_{\max}) bzw. 1035 Daten (T_6 und T_{12}).

3.2 Persistenzprognosen

Wenn die Prognose zum Zeitpunkt ihrer Erstellung gleich dem beobachteten Wetterzustand ist, so spricht man von Persistenz (siehe Kapitel 2.5). Die Persistenzprognosen entsprechen somit dem Wetterzustand des jeweiligen Vortages und wurden aus den offiziellen SYNOP-Meldungen der Zentralanstalt für Meteorologie und Geodynamik (ZAMG) entnommen. Der Umfang des so rekonstruierten Datensatzes für die Jahre 2000-2010, sowie die Prognosetermine und der Prognoseort entsprechen denen des PETRUS (siehe Kapitel 3.1). Die Persistenzprognosen sind auf jeweils eine Kommastelle genau gegeben und in dieser Arbeit stets als „**PERS**“ gekennzeichnet.

3.3 Klimaprognosen

Die Klimaprognosen greifen auf die jeweiligen Klimanormalwerte für den Vorhersageort Wien Hohe Warte zurück, welche von der Zentralanstalt für Meteorologie und Geodynamik (ZAMG) zur Verfügung gestellt wurden. Dabei wurden zwei verschiedene Klimaprognosen definiert, mit gleichem Datenumfang wie der des PETRUS (siehe Kapitel 3.1):

- das 30-jährige Klima-Monatsmittel aus den Jahren 1971-2000, auf eine Kommastelle genau, in dieser Arbeit als „**KLIMA**“ bezeichnet
- ein 10-jähriges Klimamittel aus den Jahren 1991-2000, für den jeweiligen Tag berechnet, auf eine Kommastelle genau, in dieser Arbeit als „**KLITA**“ bezeichnet

3.4 Model Output Statistics (MOS)

MOS leitet zwischen den Vorhersagen eines numerischen Vorhersagemodells und dem beobachteten Bodenwetter statistische Beziehungen ab, die in Vorhersagegleichungen formuliert und prognostisch angewendet werden ([HERMANN, 1998], siehe Kapitel 2.5).

Die in dieser Arbeit verwendeten MOS-Prognosen entsprechen dem sogenannten AUSTROMOS, der österreichischen Variante des statistischen Vorhersageverfahrens *Model Output Statistics*, welche von der Zentralanstalt für Meteorologie und Geodynamik (ZAMG) in Wien für die Jahre 2000-2010 dankenswerter Weise rekonstruiert und zur Verfügung gestellt wurden (einzig das Jahr 2003, der November 2004 und der März 2007 sowie einige wenige Tage konnten aufgrund einer Systemumstellung nicht nachgerechnet werden). Grundlage dafür ist das vom Europäischen Zentrum für mittelfristige Wettervorhersagen EZMW (*European Centre for Medium-Range Weather Forecasts, ECMWF*) in Reading (Großbritannien) gerechnete numerische Wettervorhersagemodell (EZ-Modell), an dessen Gitterpunktprognosen die statistischen Interpretationen für über 112 österreichische (und für 60 mitteleuropäische) Wetterstationen durchgeführt werden.

AUSTROMOS wurde zwischen 1990 und 1993 von der ZAMG entwickelt und ist seit 1994 im österreichischen Wetterdienst in Gebrauch. Nach einer Weiterentwicklung bzw. Adaptierung 1998 werden nun folgende Prognoseelemente für fünf Prognosestage und 18 Prognosetermine vorhergesagt:

- die Lufttemperatur in Grad Celsius zu ausgewählten Terminen (0, 6, 12, 18 UTC) sowie das Tagesminimum und das Tagesmaximum
- die Taupunktdifferenz (daraus werden der Taupunkt (in Grad Celsius) und die relative Luftfeuchtigkeit (in Prozent) berechnet)
- der Bodenwind in Stärke [m/s] und Richtung [°]
- der 6-stündige Niederschlag in Wahrscheinlichkeit (Prozent) und Menge (mm)
- die Bewölkung und der Anteil der tiefen Bewölkung in Wahrscheinlichkeit (Prozent) und Bedeckungsgrad (Achtel)
- die Wolkenuntergrenze in Wahrscheinlichkeit (Prozent) für 3 Kategorien
- Wettererscheinungen (Nebel, Konvektion)

- die Sichtweite (in Kilometer)

Der erste Prognosetermin (1.Prognosetag, „heute“) ist 18 Stunden nach der Ausgangslage (12 UTC des Vortages) fertig und reicht im 6-Stunden Intervall bis 120 Stunden in die Zukunft (5.Prognosetag, „in vier Tagen“).

Auch die AUSTROMOS-Vorhersagen sind stets auf eine Kommastelle gegeben und werden in dieser Arbeit stets als **MOS** bezeichnet. Der Datenumfang ist etwas geringer als der der studentischen Prognosen und beläuft sich auf 911 (T_{\min} und T_{\max}) bzw. 906 (T_6 und T_{12}) Daten.

4 Ergebnisse und Interpretation

4.1 Subjektive Verifikation

Subjektivität, die; lat. (*persönlich, aus eigener Sicht*)

Wie in Kapitel 1 bereits angedeutet, stellt sich bei richtigen Prognosen die Frage, wie „richtig“ definiert ist. In der Lehrveranstaltung „Besprechung des aktuellen Wettergeschehens“ des Instituts für Meteorologie und Geodynamik (IMGW) der Universität Wien wird eine Temperaturvorhersage dann als brauchbar angesehen, wenn diese in einem Intervall von $\pm 1^\circ\text{C}$ liegt. Im Folgenden soll nun diese Definition beibehalten werden.

Mit über 8000 einzelnen Prognosen stellen die Studenten den in dieser Arbeit größtmöglichen Datensatz eines Vorhersageelementes. Mehr als die Hälfte davon entfallen auf Prognosen aus den Wintersemestern (4416 Prognosen), demgegenüber stehen 3722 (bzw. 3660) Prognosen aus den Sommersemestern. Die größtmögliche Anzahl an „richtigen“ Treffern (mit, wie oben beschrieben, einem Intervall von $\pm 1^\circ\text{C}$) liefert dabei das T_{\min} aus den Wintersemestern mit einer Trefferchance von **50,1%**. Da auch die der Sommersemester mit **49,3%** knapp an der Hälfte der Prognosen liegt, ist die Minimumtemperatur das am besten vorhergesagte Prognoseelement (gesamt **49,7%**). Generell am Schlechtesten abgeschnitten hat die 12 UTC-Temperatur mit gesamt **40,8%** richtigen Prognosen. Hierbei zeigt sich, dass eine Änderung des Fehlerintervalls deutliche Auswirkungen hat: Bei einem tolerierten Fehler von $\pm 2^\circ\text{C}$ steigt die Anzahl der richtigen Prognosen auf knapp 80% (T_{\min}), wobei auch hier die T_{12} -Prognosen mit **69,6%** im Gesamten am schlechtesten abschneiden.

Noch deutlicher fällt das Ergebnis von PETRUS aus: bei einer Gesamtzahl von 1040 (bzw. 1035) Prognosen sind alle Elemente um bis zu 5 Prozentpunkte besser, wobei auch hier das T_{\min} der Wintersemester mit einem Spitzenwert von **54,9%** am besten abschneidet und so von allen den höchsten Prozentsatz an richtigen Prognosen aufstellt (dieser Wert lässt sich mit einem Fehlerintervall von $\pm 2^\circ\text{C}$ auf über 80% steigern). Auch bei PETRUS sind die wenigsten richtigen Prognosen mit gesamt **43,2%** bei den Vorhersagen der 12 UTC-Temperatur zu finden, aufgespalten in Winter- und Sommersemester zeigt sich hierbei aber ein deutlicher, Jahreszeit-bedingter Unterschied: Während in den Wintersemestern (insgesamt 548 Prognosen) der Prozentsatz bei **40,7%** liegt, sind es in den Sommersemestern (insgesamt 492 bzw. 487 Prognosen) immerhin **46,0%**. Eine mögliche Erklärung für diese Differenz ist wahrscheinlich in der Anfälligkeit für Hochnebel im Vorhersagegebiet zu suchen, der in den Wintersemestern wiederholt eine

Temperaturprognose über den Haufen werfen kann (was bei PETRUS als Mittel aller studentischen Prognosen stärker zum Tragen kommt als bei den Studenten im Einzelnen). Bei einer Erweiterung des Fehlerintervalls auf $\pm 2^{\circ}\text{C}$ gleicht sich diese Differenz hingegen wieder aus.

Die Persistenz (PERS, gleiche Anzahl wie PETRUS) steigt ebenso mit der Vorhersage der Minimumtemperatur am besten aus, der Unterschied zu PETRUS ist aber bereits deutlich: **33,6%** aller T_{\min} -Vorhersagen (im Gesamten) waren richtig und damit um mehr als 20% weniger als bei PETRUS. Geringfügig höher (mit **34,5%**) liegt die Maximumtemperatur der Wintersemester, wohingegen die der Sommersemester mit **21,3%** richtigen Prognosen deutlich schlechter abschneiden. Diese Diskrepanz zwischen WS und SS kann auf die fehlende *reliability* der Persistenzprognosen zurückgeführt werden (zum Beispiel können Gewitterlagen im Sommer schon mal ein Grund für deutliche Temperatursprünge sein, während in den Wintersemestern die advektiven Wetterlagen dominieren).

Die Klimaprognosen (KLIMA und KLITA, gleiche Anzahl wie PETRUS) schneiden im Vergleich generell am Schlechtesten ab. Den höchsten Prozentsatz an richtigen Prognosen liefert dabei die Minimumtemperatur der Sommersemester mit **22,8%**, sonst liegen die Werte bei oder deutlich unter 20% mit nur **14,0%** richtigen Prognosen (T_{\max} der Sommersemester) als absoluten Tiefstwert. Selbst eine Erhöhung des Fehlerintervalls auf $\pm 2^{\circ}\text{C}$ lässt diese Werte auf maximal **42,9%** (T_{\min} der Sommersemester) ansteigen, was in etwa der Hälfte der PETRUS-Werte entspricht. Im Allgemeinen schneiden die KLITA-Prognosen geringfügig besser als die KLIMA-Prognosen ab.

Den in dieser Arbeit geringsten Datensatz liefern die MOS-Prognosen, da, wie im Kapitel 3.4 beschrieben, einzelne Tage, aber auch ganze Monate sowie ein Jahr fehlen. Der Datensatz beinhaltet somit 911 (bzw. 906) Prognosen, aufgeteilt in 475 Prognosen aus den Wintersemestern und 436 (bzw. 431) aus den Sommersemestern. Die Anzahl der richtigen Prognosen aller Vorhersageelemente beim MOS liegt (mit einem Fehlerintervall von $\pm 1^{\circ}\text{C}$) bei knapp 50%, auch hier liefern die Prognosen der Minimumtemperatur die besten Ergebnisse mit **47,3%** im Gesamten. Anders als bei den menschlichen Prognosen liegt aber hier der Spitzenwert mit **50,5%** bei den Minimumtemperaturen der Sommersemester, die der Wintersemester liefern hingegen nur **44,4%** (tiefe Temperaturen werden vom MOS zumeist überschätzt). Besonders schwach sind die MOS-Vorhersagen bei den Prognosen der 12 UTC-Temperatur mit lediglich **33,3%** in den Wintersemestern. Dieser Prozentsatz wird vom deutlich besseren Abschneiden in den Sommersemestern (**43,9%**) auf einen Gesamtwert von **38,3%** gehoben. Generell scheint das MOS mehr Probleme in den Wintermonaten zu haben, da alle Vorhersageelemente in den Wintersemestern niedrigere Prozentsätze (auf hohem Niveau) aufweisen als in den Sommersemestern – selbst bei Erhöhung des Fehlerintervalls auf $\pm 2^{\circ}\text{C}$. Eine mögliche Erklärung hierfür könnten die häufigeren Inversionswetterlagen in den WS sein. Herausragend dabei die T_{\max} -Prognosen: Lediglich **34,7%** richtige Vorhersagen in den Wintersemestern (und damit nahezu ident mit

PERS!) stehen **46,1%** richtigen Prognosen in den Sommersemestern gegenüber – ein Wert, der sogar den des PETRUS (mit **42,9%**) in den Schatten stellt (diese Verbesserung bleibt übrigens auch bei einer Erhöhung des erlaubten Fehlers auf $\pm 2^\circ\text{C}$ erhalten!). Mit Ausnahme der T_{\max} -Vorhersagen in den Sommersemestern liegen die Ergebnisse aller anderen Vorhersageparameter ansonsten unter denen des PETRUS – eine höhere Leistung gegenüber dem automatisierten Vorhersageverfahren MOS ist also in dieser ersten, subjektiven Überprüfung gegeben.

		Ges.	WS	SS	(-1 ≤ Fi-Oi ≤ 1)						(-2 ≤ Fi-Oi ≤ 2)					
		N	N	N	N	N	WS	SS	%	%	%	N	N	WS	SS	%
STUDENTEN	T_{\min}	8138	4416	3722	4048	2212	1836	49,7	50,1	49,3	6372	3419	2953	78,3	77,4	79,3
	T_{6UTC}	8076	4416	3660	3662	1951	1711	45,3	44,2	46,7	5983	3163	2820	74,1	71,6	77,0
	T_{12UTC}	8076	4416	3660	3297	1796	1501	40,8	40,7	41,0	5622	3066	2556	69,6	69,4	69,8
	T_{\max}	4104	4416	3722	1708	1913	1444	41,6	43,3	38,8	2853	3169	2573	69,5	71,8	69,1
PETRUS	T_{\min}	1040	548	492	562	301	265	54,0	54,9	53,9	872	455	417	83,8	83,0	84,8
	T_{6UTC}	1035	548	487	515	269	246	49,8	49,1	50,5	826	423	403	79,8	77,2	82,8
	T_{12UTC}	1035	548	487	447	223	224	43,2	40,7	46,0	759	401	358	73,3	73,2	73,5
	T_{\max}	1040	548	492	457	246	211	43,9	44,9	42,9	779	416	363	74,9	75,9	73,8
PERSISTENZ	T_{\min}	1040	548	492	349	186	163	33,6	33,9	33,1	601	335	266	57,8	61,1	54,1
	T_{6UTC}	1035	548	487	328	185	143	31,7	33,8	29,4	610	314	296	58,9	57,3	60,8
	T_{12UTC}	1035	548	487	285	165	120	27,5	30,1	24,6	511	294	217	49,4	53,6	44,6
	T_{\max}	1040	548	492	294	189	105	28,3	34,5	21,3	514	303	211	49,4	55,3	42,9
KLIMA	T_{\min}	1040	548	492	204	92	112	19,6	16,8	22,8	414	203	211	39,8	37,0	42,9
	T_{6UTC}	1035	548	487	205	110	95	19,8	20,1	19,5	381	195	186	36,8	35,6	38,2
	T_{12UTC}	1035	548	487	170	93	77	16,4	17,0	15,8	315	176	139	30,4	32,1	28,5
	T_{\max}	1040	548	492	162	93	69	15,6	17,0	14,0	316	176	140	30,4	32,1	28,5
KLITA	T_{\min}	1040	548	492	223	106	117	21,4	19,3	23,8	406	184	222	39,0	33,6	45,1
	T_{\max}	1040	548	492	169	99	70	16,3	18,1	14,2	321	179	142	30,9	32,7	28,9
MOS	T_{\min}	911	475	436	431	211	220	47,3	44,4	50,5	714	351	363	78,4	73,9	83,3
	T_{6UTC}	906	475	431	414	206	208	45,7	43,4	48,3	655	318	337	72,3	66,9	78,2
	T_{12UTC}	906	475	431	347	158	189	38,3	33,3	43,9	594	287	307	65,6	60,4	71,2
	T_{\max}	911	475	436	366	165	201	40,2	34,7	46,1	635	303	332	69,7	63,8	76,1

TABELLE 4-1: Anzahl (N) und Prozentsatz (%) der „richtigen“ Prognosen mit einem Fehlerintervall von $\pm 1^\circ\text{C}$ und $\pm 2^\circ\text{C}$.

4.2 Die Temperatur als kontinuierlicher Parameter

Nach der subjektiven Betrachtungsweise aus Kapitel 4.1 widmen wir uns jetzt der in Kapitel 2 definierten Art und Weise der Verifikation von Punktprognosen. Die Temperatur kann demnach sowohl als kontinuierlicher (bei einer Vorhersage von Null Grad wurde welche Temperatur beobachtet?) wie auch kategorischer Parameter (ist Frost eingetreten oder nicht?) angesehen werden. In diesem Kapitel werden die kontinuierlichen Eigenschaften der

Temperatur-Vorhersage näher betrachtet; um die Übersicht zu bewahren, werden dabei die Ergebnisse der Vorhersageelemente (Minimumtemperatur, Maximumtemperatur, 6 UTC-Temperatur und 12 UTC-Temperatur) in einzelnen Kapiteln präsentiert.

4.2.1 Verifikation der Minimumtemperatur (T_{\min})

Die Minimumtemperatur ist die tiefste erreichte Temperatur während eines gewissen Zeitraumes. In unserem Fall erstreckt sich dieser zwischen 18 UTC und 6 UTC und entspricht somit dem Tiefstwert der Vorhersagenacht. T_{\min} wird stets am Morgen um 6 UTC mit Hilfe eines Minimumthermometers, einem speziellen Ausdehnungsthermometer, bei dem Alkohol anstelle von Quecksilber als Messflüssigkeit verwendet wird, bestimmt. Aufgrund der Kapillarität kann ein in der Flüssigkeit enthaltenes Glasstäbchen zwar abwärts, aber nicht wieder aufwärts gleiten [LILJEQUIST, 2006].

Gleich zu Beginn ist zu sagen, dass die Minimumtemperatur im Mittel zu niedrig prognostiziert wurde – mit Ausnahme des MOS, dessen Prognosewerte als einzige im Schnitt deutlich zu hoch waren, besonders in den Wintersemestern (siehe TABELLE 4-2). Der BIAS (mittlerer Fehler, ME; vgl. Kapitel 2.1.1) der studentischen Prognosen liegt bei **-0,31**. PETRUS als Mittel aller studentischen Prognosen verbessert diesen Wert auf **-0,23**. PERS zeigt im Mittel keine nennenswerte Abweichung (**-0,03**) – ein überraschend gutes Abschneiden, welches spätestens mit dem mittleren absoluten Fehler (MAE) relativiert wird. Bei den Klimaprognosen steigt das Tagesmittel KLITA mit **-0,65** ein wenig besser aus als das Monatsmittel KLIMA, dessen Fehler bei rund einem Grad liegt (**-0,87**). Sowohl bei KLIMA als auch bei KLITA kann daraus auf den Klimawandel und dem daraus resultierenden Temperaturanstieg geschlossen werden, müsste doch der BIAS gerade bei den Klimaprognosen annähernd null sein. Wie schon erwähnt sind die Prognosen des MOS tendenziell zu hoch (**0,56**), in den Wintersemestern hat das MOS sogar die größte Abweichung von allen (**0,85**).

ME			
	GESAMT	WS	SS
STUDENTEN	-0,31	-0,41	-0,19
PETRUS	-0,23	-0,34	-0,12
PERS	-0,03	0,08	-0,16
KLIMA	-0,87	-0,81	-0,94
KLITA	-0,65	-0,66	-0,63
MOS	0,56	0,85	0,24

TABELLE 4-2: mittlerer Fehler (ME) der T_{\min} -Prognosen

Auch die graphische Auswertung der studentischen Prognosen (siehe ABBILDUNG 4-1) zeigt, dass die Minimumtemperatur tendenziell zu niedrig vorhergesagt wurde, vor allem bei negativen Temperaturen (Frost), wobei diese Unterschätzung unabhängig von Winter- und Sommersemester ist. Bei höheren Temperaturen ist hingegen ein leichter Umkehrtrend zu erkennen.

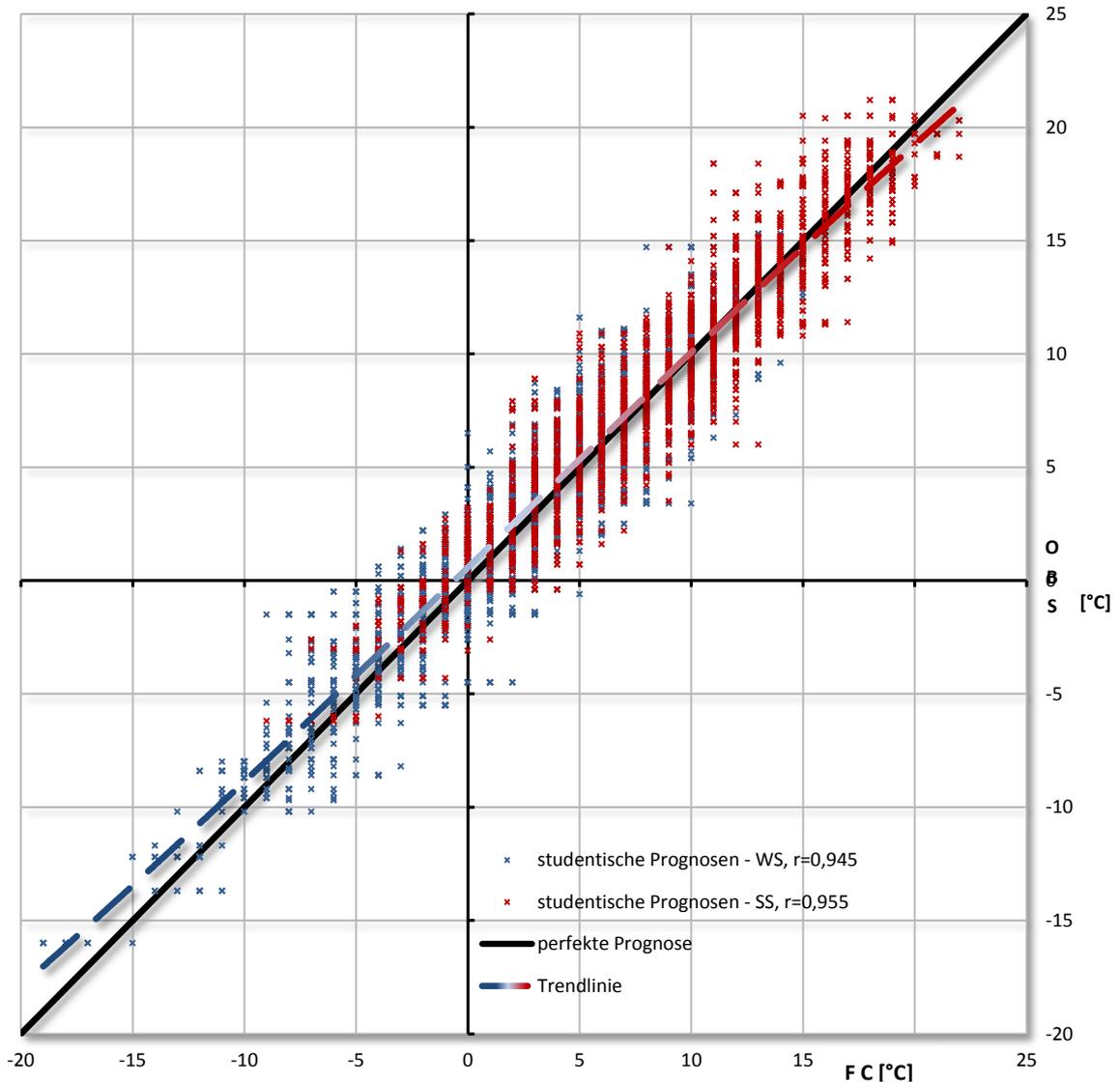


ABBILDUNG 4-1: Streudiagramm der studentischen T_{min} -Prognosen mit Trendlinie und Korrelationskoeffizient r . Negative Temperaturen wurden tendenziell zu niedrig, Temperaturen ab 15 Grad zu hoch vorhergesagt. Der Temperaturbereich um 10 Grad entspricht dem verlässlichsten Prognosebereich.

Betrachtet man die Verteilung der Referenzprognosen PERS und KLITA (ABBILDUNG 4-2), so zeigt sich folgendes Bild: die Streuung dieser Prognosen um die 45°-Achse (perfekte Prognose) ist größer, zusätzlich fällt das Fehlen von hohen und tiefen Werten (Extremwerten) bei der KLITA sofort ins Auge (fehlende *sharpness*). Die große Streuung der

PERS-Prognosen in beide Richtungen resultiert im Mittel zu dem vermeintlich guten BIAS-Wert.

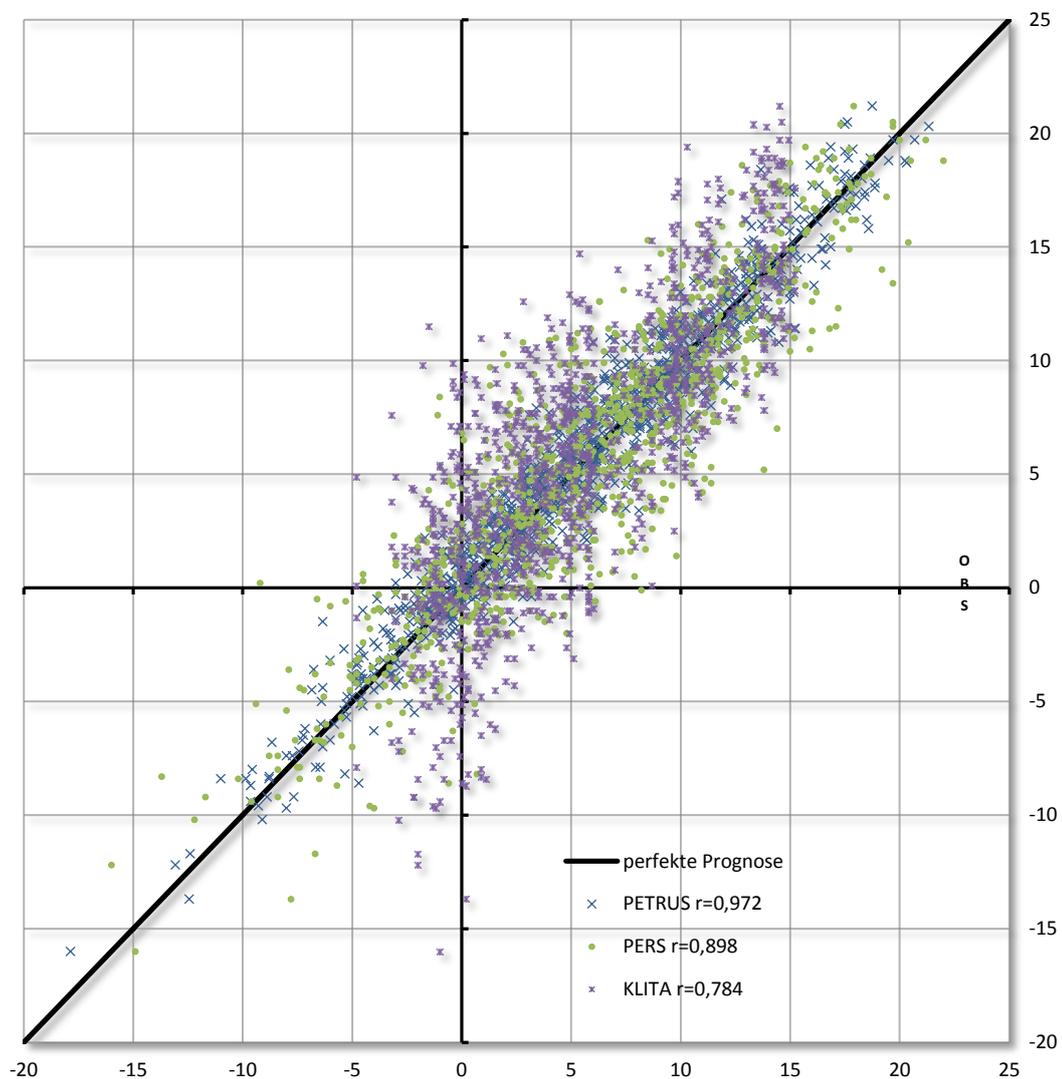


ABBILDUNG 4-2: Streudiagramm der T_{min} - Prognosen von PETRUS, PERS und KLITA mit Korrelationskoeffizient r . Sowohl PERS als auch KLITA liefern eine deutliche Streuung um die perfekte Prognose.

Anders der Vergleich mit MOS (siehe ABBILDUNG 4-3): visuell scheinen die Prognosen annähernd gleich wie jene des PETRUS zu sein, lediglich der fehlende Mut-Faktor im negativen Bereich ist gut zu erkennen und zeigt somit die Schwäche der MOS-Prognosen.

Widmet sich der *BIAS* der Verlässlichkeit einer Prognose, so steht der mittlere absolute Fehler (MAE; vgl. Kapitel 2.1.2, siehe TABELLE 4-3) für die Genauigkeit der Vorhersagen und zeigt im Gegensatz zum mittleren Fehler auch die mittlere Magnitude eines Vorhersagefehlers. Speziell bei der PERS wird nun ersichtlich, wie wichtig es ist, sich nicht auf einzelne Verifikationsmaße zu stützen, sondern diese stets zusammen mit anderen zu

betrachten: lag der mittlere Fehler der PERS bei fast Null, so zeigt der mittlere absolute Fehler eine deutliche Abweichung von **2,13** (die Ursache hierfür liegt, wie bereits oben beschrieben, beim gegenseitigen Aufheben der einzelnen Fehler unterschiedlicher Vorzeichen).

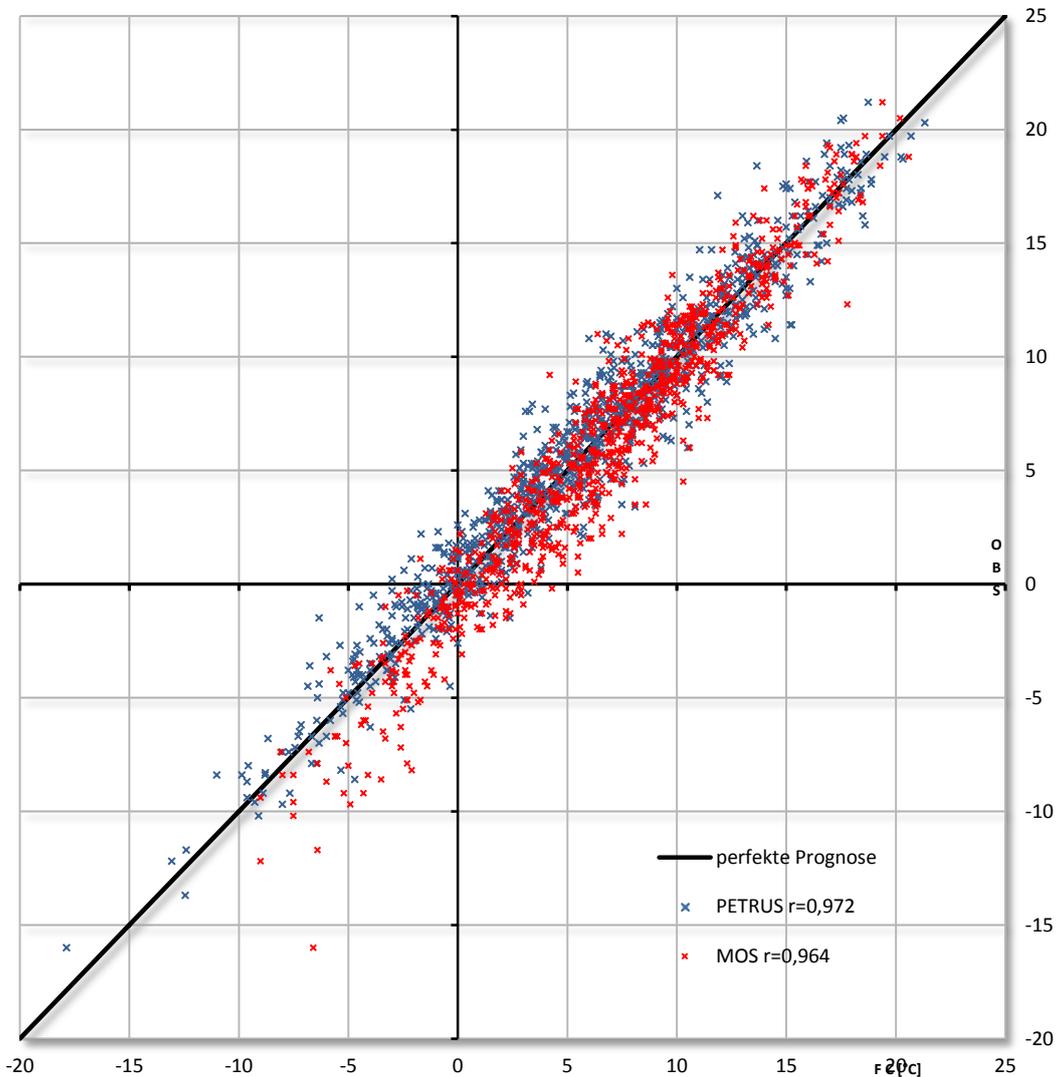


ABBILDUNG 4-3: Streudiagramm der T_{min} -Prognosen von PETRUS und MOS mit Korrelationskoeffizienten r . Die Vorhersagen des MOS sind im Mittel zu hoch, deutlich zu sehen bei Temperaturen $< 0^{\circ}\text{C}$

MAE	SS_{MAE}					
	GESAMT	WS	SS	GESAMT	WS	SS
STUDENTEN	1,33	1,36	1,31	---	---	---
PETRUS	1,15	1,18	1,12	---	---	---
PERS	2,13	2,11	2,15	0,46	0,44	0,48
KLIMA	3,12	3,43	2,78	0,63	0,66	0,60
KLITA	3,10	3,50	2,64	0,63	0,66	0,58
MOS	1,37	1,52	1,22	0,16	0,22	0,08

TABELLE 4-3: mittlerer absoluter Fehler (MAE) und der dazugehörige Skill Score (SS_{MAE}) der T_{min} -Prognosen. SS_{MAE} gibt die Verbesserung des PETRUS wieder.

Die größten Abweichungen finden sich auch hier bei den Klimaprognosen, wobei kein Unterschied mehr zwischen KLIMA und KLITA auszumachen ist. Mit einer Abweichung von gut einem Grad (**1,15**) verfehlt PETRUS nur knapp das aus Kapitel 4.1 erlaubte Fehlerintervall der subjektiven Verifikation. Der absolute Fehler von MOS liegt mit **1,37** etwas höher.

Um diesen Prognosefehler auch graphisch darstellen zu können, eignen sich sogenannte Differenzplots. In ABBILDUNG 4-4 und ABBILDUNG 4-5 sind diese Fehler ersichtlich, wobei auch hier wieder die deutlich positive Abweichung des MOS hervorsticht. Das Fehlerintervall liegt dabei zumeist bei $\pm 5^{\circ}\text{C}$, mit einzelnen Ausreißern bei MOS nach oben. Bemerkenswert, wenn auch nicht anders zu erwarten, die Verteilungen von PERS und KLITA mit einem doch deutlich größeren Fehlerintervall von $\pm 10^{\circ}\text{C}$.

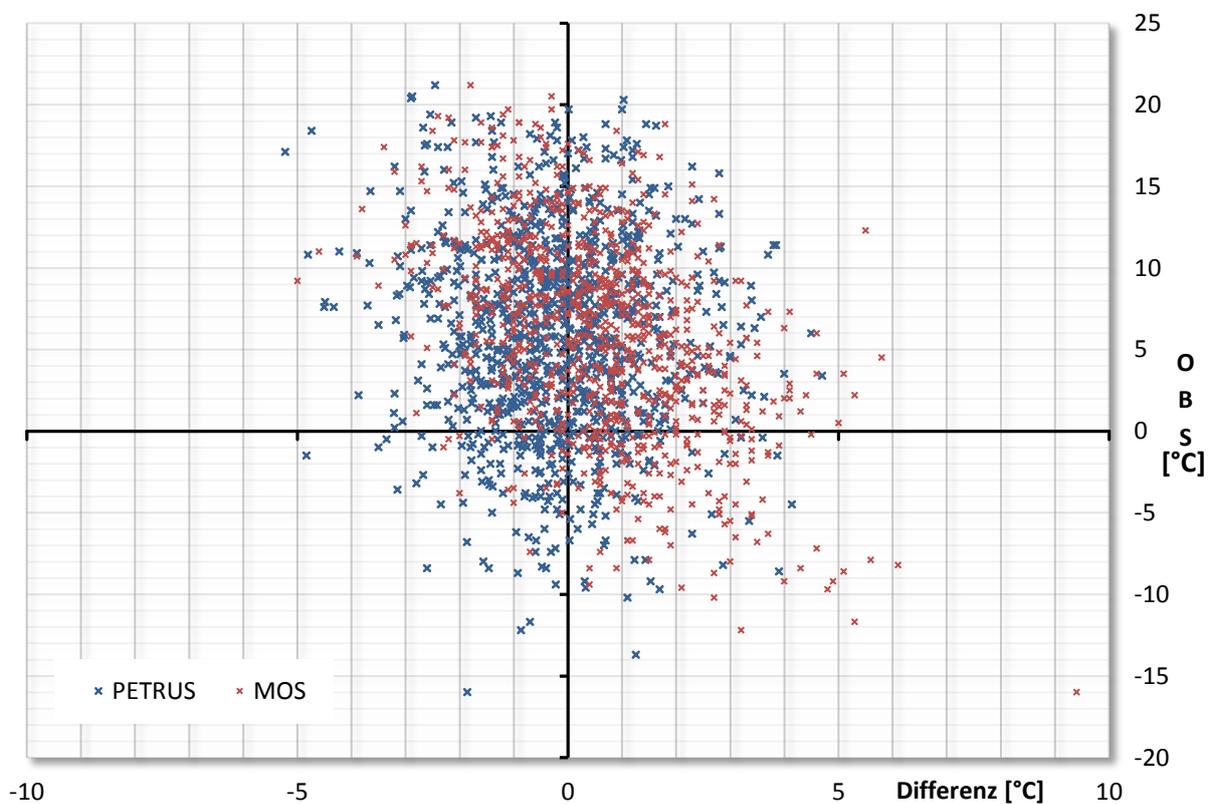


ABBILDUNG 4-4: Differenzplot T_{min} ; aufgetragen sind die Vorhersagefehler vs. FC von PETRUS & MOS. Der Fehlerbereich liegt bei $\pm 5^{\circ}\text{C}$; die Tendenz zu positiven Differenzen des MOS ist deutlich.

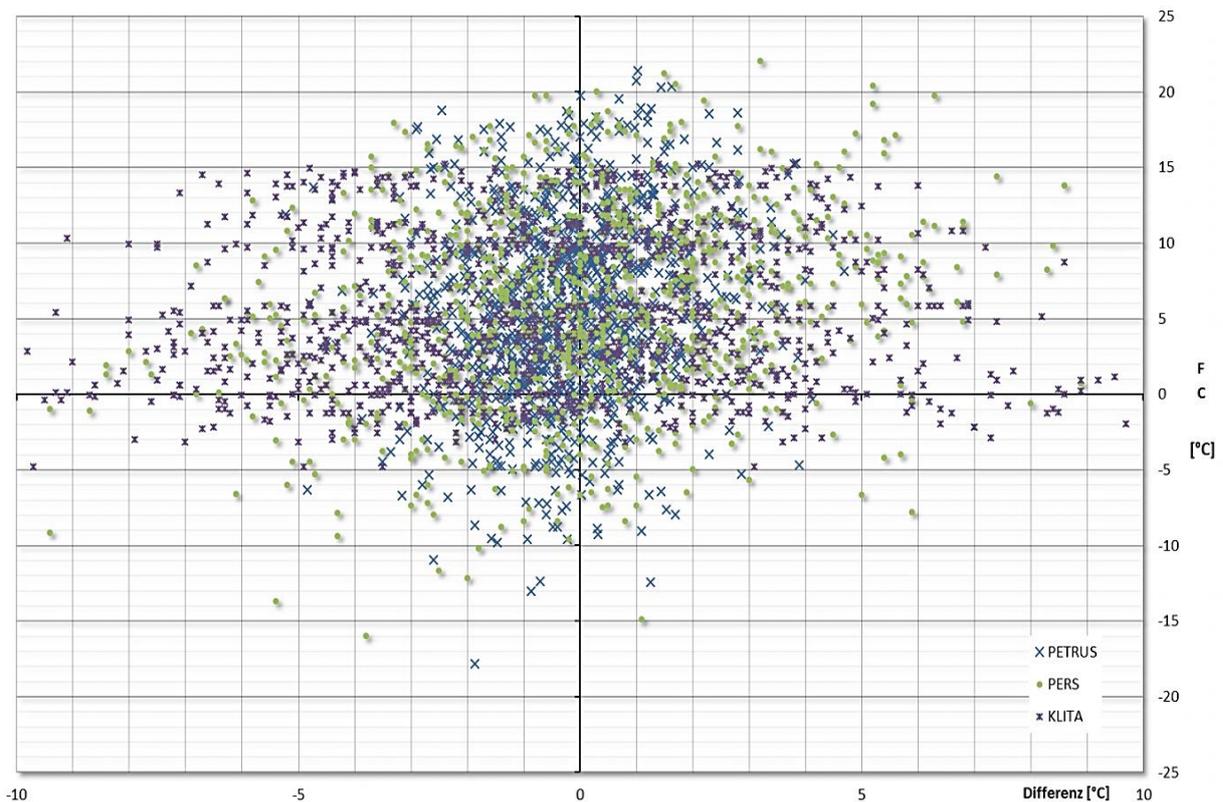


ABBILDUNG 4-5: Differenzplot T_{min} ; aufgetragen sind die Vorhersagefehler vs. FC von MOS, PERS & KLITA. Im Vergleich zu PETRUS liefern sowohl PERS als auch KLITA eine Verdoppelung des Fehlerintervalls auf $\pm 10^\circ\text{C}$. Das Fehlen jeglicher Extremwerte bei den KLITA-Vorhersagen ist ersichtlich.

Beim mittleren absoluten Fehler macht es bereits Sinn, sich den daraus resultierenden *Skill Score* (SS_{MAE} ; siehe Kapitel 2.1.4) zu berechnen, also ob und um wie viel eine Verbesserung gegenüber den Referenzprognosen erzielt worden ist. Die Ergebnisse zeigen diese Verbesserung der Studenten gegenüber allen Vergleichsprognosen (siehe TABELLE 4-3), am größten gegenüber den beiden Klimavorhersagen (KLIMA, KLITA). Beim MOS kommt die Verbesserung nur noch durch die Minimumtemperaturvorhersage der Wintersemester zu tragen (wie bereits weiter oben beschrieben hat das MOS hier die größten Schwierigkeiten), die Verbesserung gegenüber die der Sommersemester geht hingegen gegen Null.

Nach der Verlässlichkeit und Genauigkeit der Prognosen stellt sich noch die Frage, wie hoch die Fehlervarianz der einzelnen Prognosen ist. Ein Maß dafür ist die Wurzel des mittleren quadratischen Fehlers (RMSE; vgl. Kapitel 2.1.3). Alle Prognosen weisen dabei eine gewisse Fehlervarianz auf, mit **1,48** schneidet PETRUS am besten ab (siehe TABELLE 4-4). MOS liegt etwas darüber (**1,76**), wobei auch hier wieder ein deutlicher Unterschied zwischen den Jahreszeiten zu erkennen ist. Eine höhere Varianz weisen die anderen Referenzprognosen auf, was aber auch nicht anders zu erwarten war.

RMSE			
	GESAMT	WS	SS
STUDENTEN	1,73	1,77	1,68
PETRUS	1,48	1,51	1,44
PERS	2,76	2,79	2,73
KLIMA	3,89	4,26	3,43
KLITA	3,86	4,32	3,27
MOS	1,76	1,95	1,52

TABELLE 4-4: Wurzel des mittleren quadratischen Fehlers (RMSE) der T_{min} -Prognosen.

Zieht man beim RMSE nicht die Wurzel, erhält man den MSE (mittlerer quadratischer Fehler; siehe Kapitel 2.1.3), bei dem sich, wie schon beim MAE, ein *Skill Score* berechnen lässt (Reduktion der Varianzen, RV; vgl. Kapitel 2.1.5). Auch die RV gibt die Verbesserung der Studenten gegenüber den Referenzprognosen wieder, wobei die perfekte Prognose bei Eins liegt. Die Verbesserung gegenüber den klassischen Referenzprognosen (PERS, KLIMA, KLITA) ist mit einer RV von bis zu **0,86** (KLIMA) deutlich (siehe TABELLE 4-5). Auch gegenüber MOS ist mit **0,29** eine Verbesserung gegeben, wobei sich auch hier wieder der Unterschied zwischen Winter- und Sommersemester bemerkbar macht.

RV			
	GESAMT	WS	SS
STUDENTEN	---	---	---
PETRUS	---	---	---
PERS	0,71	0,71	0,72
KLIMA	0,86	0,87	0,82
KLITA	0,85	0,88	0,80
MOS	0,29	0,40	0,10

TABELLE 4-5: Reduktion der Varianzen (RV) der T_{min} -Prognosen. Die Werte geben die Verbesserung des PETRUS gegenüber den Referenzprognosen wieder.

Abschließend stellt sich noch die Frage, ob und inwiefern ein Unterschied der Prognosegüte für die einzelnen Monate (Jänner, März, April, Mai, Juni, Oktober, November und Dezember) besteht. Dabei bestätigt sich der schon zuvor festgestellte Trend, dass die Prognosen des MOS eher zu hoch, wohingegen die des PETRUS tendenziell zu niedrig lagen. Auffallend dabei, dass dieser Trend nur in den kühleren bzw. kalten Monaten zwischen Oktober und März auftritt, wohingegen in den wärmeren Monaten April, Mai und Juni die Abweichungen gegen Null gehen (siehe ABBILDUNG 4-6).

Der mittlere absolute Fehler des PETRUS liegt dabei in allen Monaten um die 1°C (siehe ABBILDUNG 4-7) – am schlechtesten schneiden die Übergangsmonte Oktober und November ab (beide inversions-freundliche Monate). Bei MOS zeigt sich wieder der große Unterschied zwischen den warmen und kalten Monaten, mit der größten Abweichung im Wintermonat Jänner. Auffallend dabei, dass im Monat Juni sowohl der mittlere absolute Fehler als auch die Wurzel des mittleren quadratischen Fehlers (ABBILDUNG 4-8) deutlich unter dem des PETRUS liegen und auch die Monate Mai und November in punkto Vorhersagefehler und Fehlervarianz an die Studenten aufschließen können.

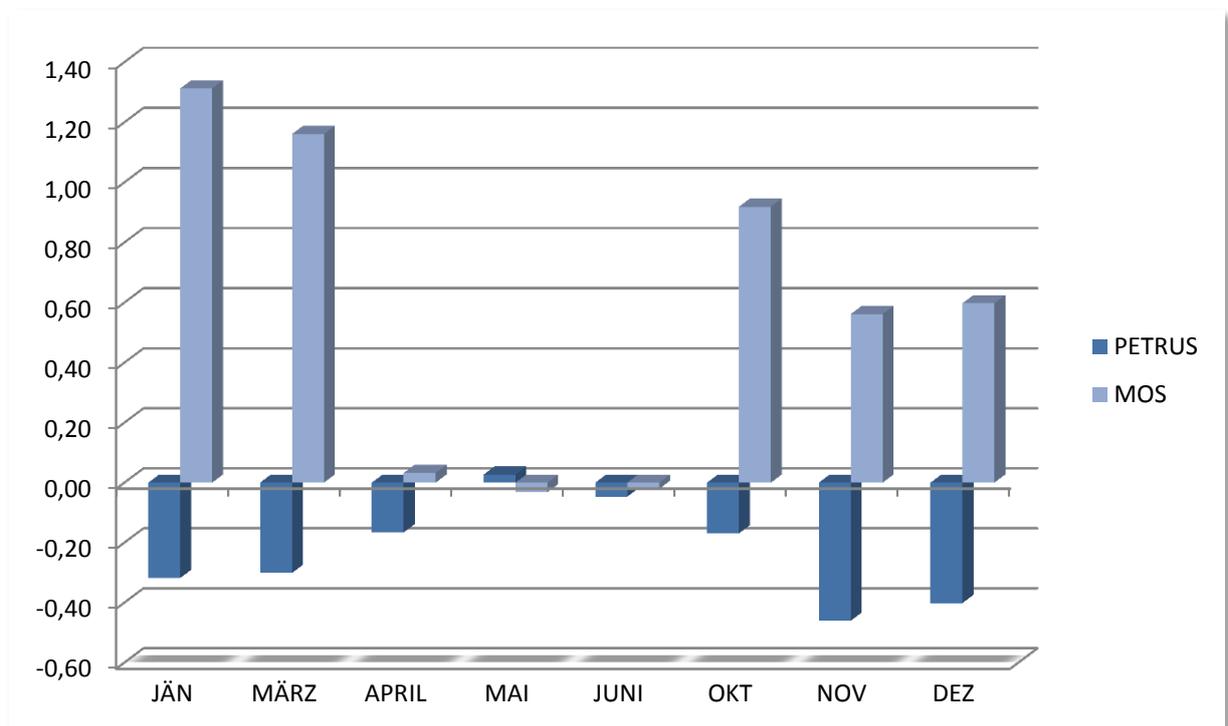


ABBILDUNG 4-6: Mittlerer Fehler (ME) der Minimumtemperatur von PETRUS & MOS, aufgespaltet auf die einzelnen Vorhersagemonate. Während die „warmen“ Monate kaum Abweichungen liefern, sind die Fehler in den „kalten“ Abweichungen teils deutlich, vor allem beim MOS.

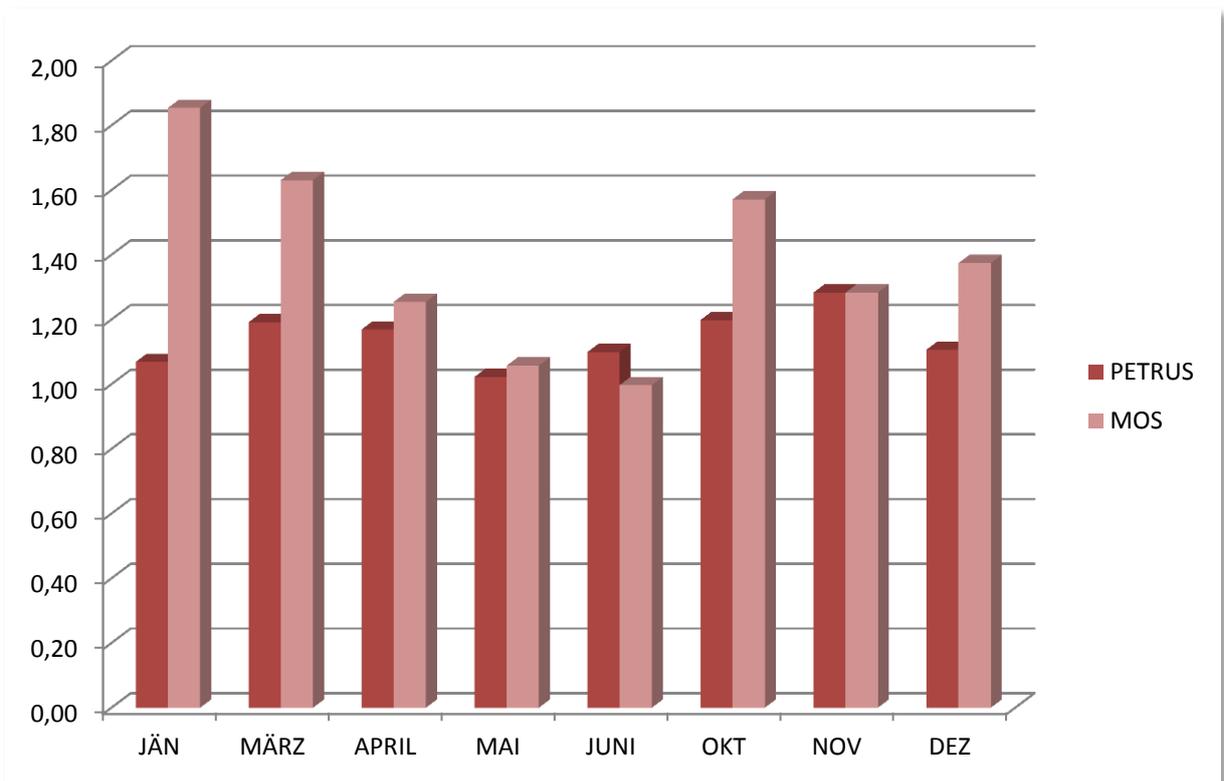


ABBILDUNG 4-7: Mittlerer absoluter Fehler (MAE) der Minimumtemperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate. PETRUS schwankt um $\pm 1^\circ\text{C}$, beim MOS liegen die absoluten Fehler hingegen in den „kalten“ Monaten deutlich darüber. Lediglich im Juni liefert das MOS ein besseres Ergebnis als PETRUS.

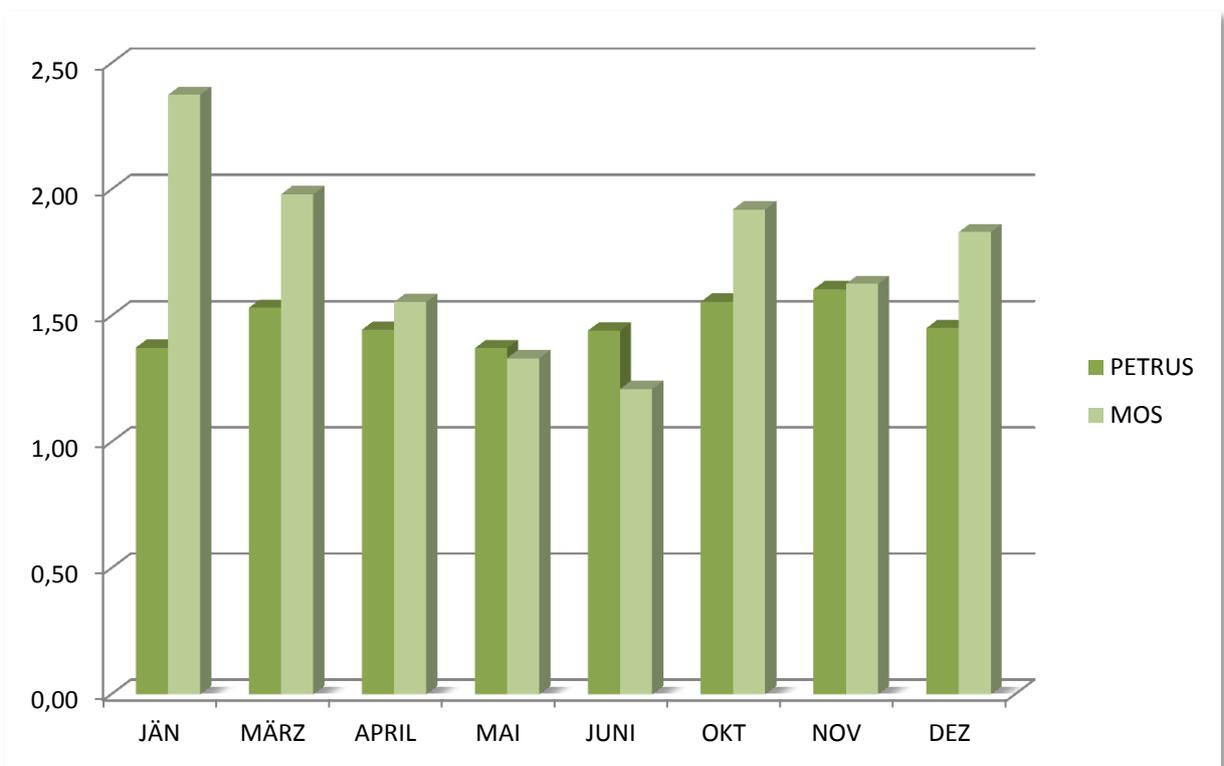


ABBILDUNG 4-8: Wurzel des mittleren quadratischen Fehler (RSME) der Minimumtemperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate. Die Verteilung ist ähnlich der des MAE (vgl. ABBILDUNG 4-6) mit Juni als bestem Monat des MOS.

Zusammenfassend ist zu sagen, dass, wie auch schon die subjektive Verifikation im vorangegangenen Kapitel gezeigt hat, die Minimumtemperatur derjenige Parameter ist, der sowohl von PETRUS als auch MOS am besten vorhergesagt wird. Dabei zeigt sich bei beiden eine bessere Vorhersagegüte in den Sommersemestern, wohingegen in den Wintersemestern vor allem das MOS mit doch deutlichen Abweichungen zu kämpfen hat und in punkto Vorhersagegüte sogar unter die anderen Referenzprognosen (PERS, KLIMA, KLITA) fällt. Eine Erklärung hierfür kann das nicht Berücksichtigen des MOS (als statistische Vorhersagemethode) von möglichen Strahlungsnächten (wolkenlose, windschwache Nächte, womöglich noch mit einer bestehenden Schneedecke) sein. Auch bei PETRUS schneiden die Wintersemester ein wenig schlechter ab, wobei die Tendenz zu zu niedrigen Vorhersagen überwiegt; ein Umkehrtrend ist nur bei hohen Minimumtemperaturen (ab etwa 15°C) erkennbar. Der tendenziell verlässlichste Vorhersagebereich liegt bei etwa 10°C. Eine Verbesserung des PETRUS gegenüber allen Referenzprognosen, auch gegenüber MOS, ist gegeben – lediglich im Monat Juni übertrifft die Leistung des MOS die der Studenten und auch in den Monaten Mai und November kann die des MOS mithalten.

4.2.2 Verifikation der 6 UTC-Temperatur

Während die Minimumtemperatur aus dem vorangegangenen Kapitel die tiefste erreichte Temperatur während eines gewissen Zeitraumes (zumeist die Nacht) darstellt, ist die 6 UTC-Temperatur diejenige Temperatur, die um 6 UTC (7 Uhr MEZ, 8 Uhr MESZ) mithilfe eines Thermometers gemessen wird. Im Winter entspricht dieser Wert nicht selten der Minimumtemperatur selbst (oder ist dieser zumindest nahe daran), während im Sommer die Temperatur am Morgen schon stark vom Tiefstwert abweichen kann.

Trotz dieses Umstandes, überrascht das Ergebnis doch etwas. Es zeigt nämlich, dass die Abweichungen (mittlerer Fehler, vgl. Kapitel 2.1.1) von PETRUS und MOS sich genau konträr dazu verhalten: im Mittel sind diese nämlich in den Wintersemestern höher (und nicht wie angenommen werden könnte, tiefer) als in den Sommersemestern (siehe TABELLE 4-6). Dabei liegen die Prognosen, wie auch schon bei der Minimumtemperatur, generell zu niedrig, lediglich das MOS prognostiziert deutlich zu hoch: Während PETRUS auf gesamt **-0,56** kommt, weist das MOS mit **0,91** fast eine Verdoppelung des T_{\min} -Ergebnisses auf. Mit **1,02** in den WS liefert das MOS zudem die größte Abweichung von allen. Das vermeintlich gute Abschneiden der PERS (gesamt **-0,12**) ist auch hier nicht über zu interpretieren, da sich erneut zeigt, dass sich große Fehler unterschiedlicher Vorzeichen einander aufheben. Eine Verbesserung gegenüber der Minimumtemperatur kann lediglich das KLIMA erzielen, mit gesamt **-0,76**.

ME			
	GESAMT	WS	SS
STUDENTEN	-0,59	-0,70	-0,47
PETRUS	-0,56	-0,66	-0,46
PERS	-0,12	0,04	-0,29
KLIMA	-0,76	-0,56	-0,99
MOS	0,91	1,02	0,78

TABELLE 4-6: mittlerer Fehler (ME) der T_6 -Prognosen.

Auch die graphische Auswertung zeigt, dass die 6 UTC-Temperatur im Mittel, deutlicher allerdings in den Wintersemestern, zu tief prognostiziert wurde (siehe ABBILDUNG 4-9). Eine mögliche Erklärung hierfür könnte bereits das Abschneiden bei den Minimumtemperaturen sein, da das T_{\min} speziell in den Wintersemestern das T_6 maßgeblich beeinflusst (und T_{\min} vor allem im unteren Temperaturbereich ebenso zu tief vorhergesagt wurde, siehe Kapitel 4.2.1). Ab etwa 20°C wurden zu hoch vorhergesagt, was aber aufgrund der deutlich geringen Anzahl an Wertepaaren nicht über zu interpretieren ist.

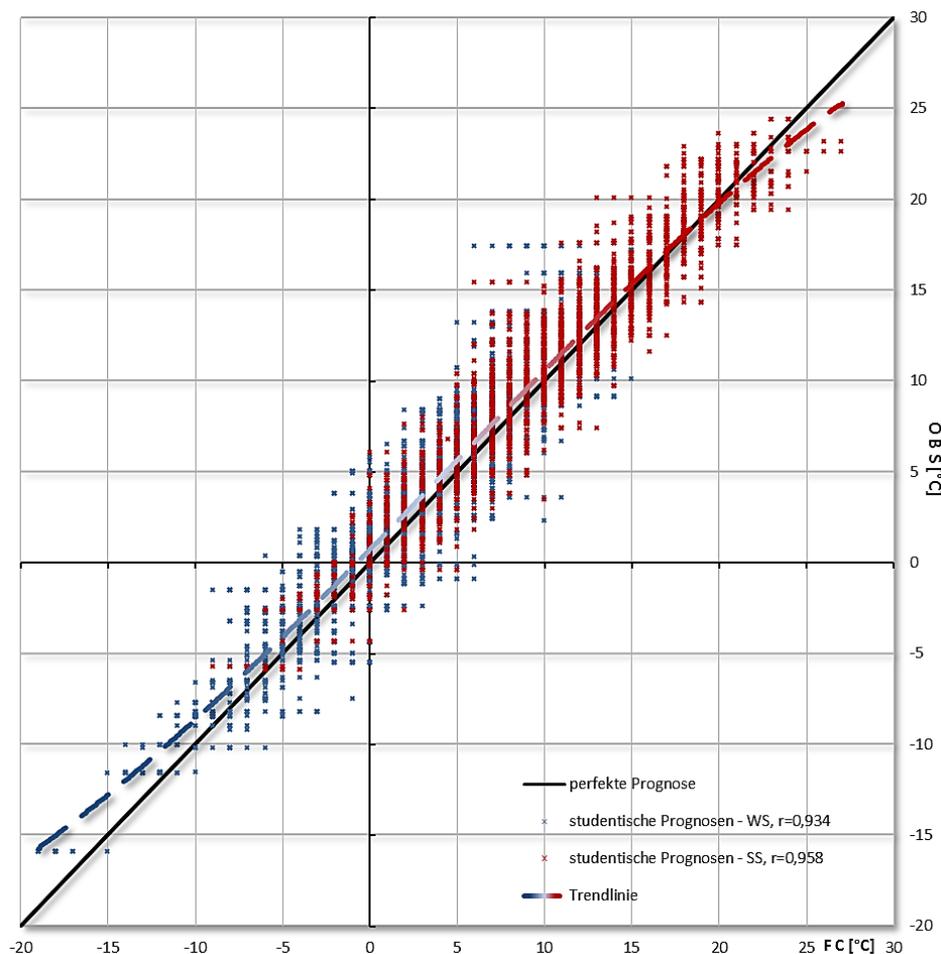


ABBILDUNG 4-9: Streudiagramm der studentischen T_6 -Prognosen mit (polynomischer) Trendlinie und Korrelationskoeffizient r . Die Prognosen liegen fortlaufend zu niedrig, erst ab etwa 18°C liegen die Prognosen zu hoch.

Bei der graphischen Auswertung der 6 UTC-Temperatur für PETRUS und MOS ist die (fast schon) überdeutliche Trennung der beiden Vorhersagen bemerkenswert (siehe ABBILDUNG 4-10 und ABBILDUNG 4-11). Einmal mehr zeigt sich wieder der fehlende Mut-Faktor des MOS im Frostbereich. Bezüglich des Vorhersagefehlers ist vor allem bei MOS eine deutliche Streuung innerhalb des Fehlerintervalls von $\pm 5^{\circ}\text{C}$ zu erkennen, wohingegen PETRUS, von den deutlich mehr Ausreißern abgesehen, doch ein ähnliches Fehlerintervall wie das der T_{\min} vorweisen kann (vgl. ABBILDUNG 4-4). Die PERS-Prognosen zeigen auch bei der 6UTC – Temperatur den großen Streuungsbereich um die perfekte Prognose, der zu dem vermeintlich guten mittleren Fehler führt. Auf eine graphische Auswertung von KLIMA wurde auf Grund der zu erwartenden Verteilung verzichtet (ABBILDUNG 4-12).

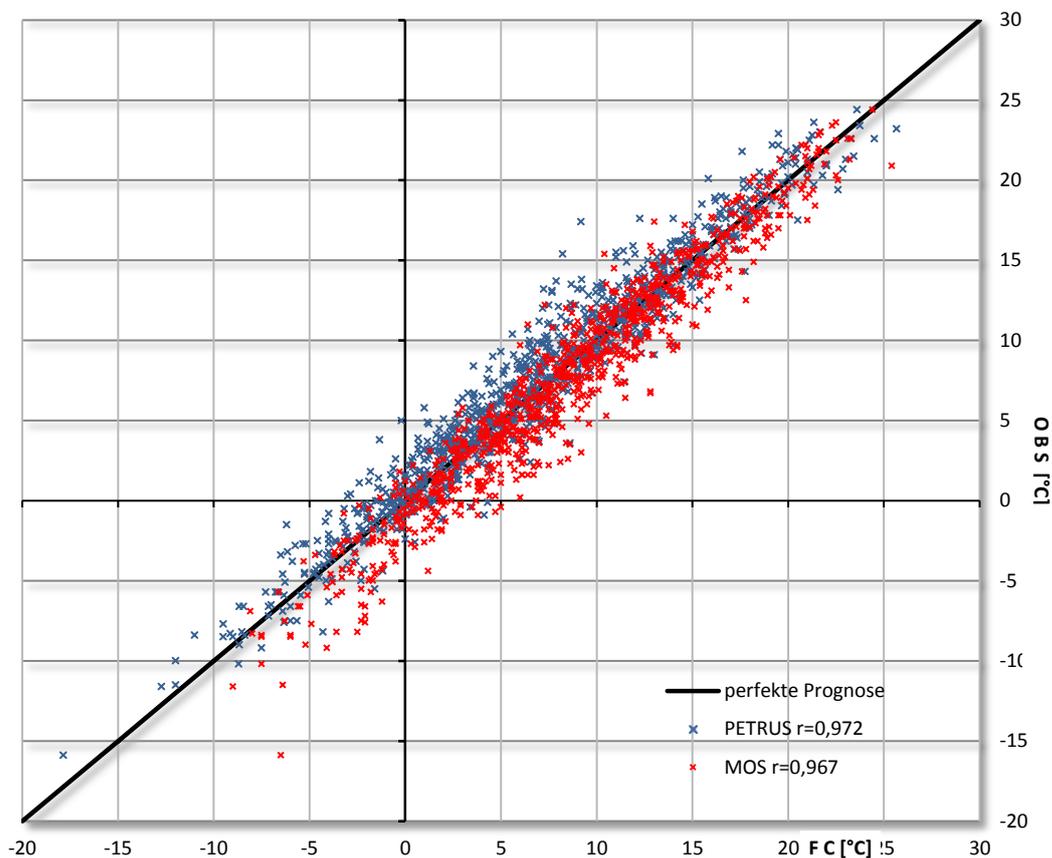


ABBILDUNG 4-10: Streudiagramm der T_6 -Prognosen von PETRUS und MOS mit Korrelationskoeffizient r . Die deutliche Trennung der PETRUS- und der MOS-Vorhersagen ist gut ersichtlich.

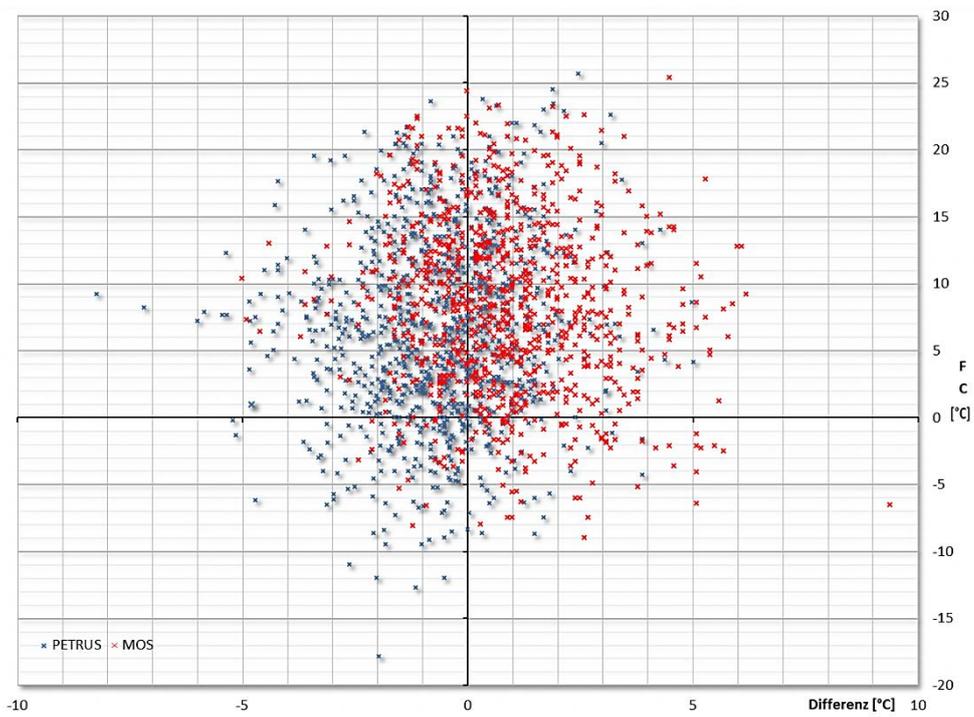


ABBILDUNG 4-11: Differenzplot der T_6 -Prognosen; aufgetragen sind die Vorhersagefehler vs. FC von PETRUS & MOS. Der Fehlerbereich liegt $\pm 5^\circ\text{C}$, mit einzelnen Ausreißern nach oben. Auch hier zeigt sich wieder, dass die MOS-Abweichungen deutlich im positiven Bereich liegen.

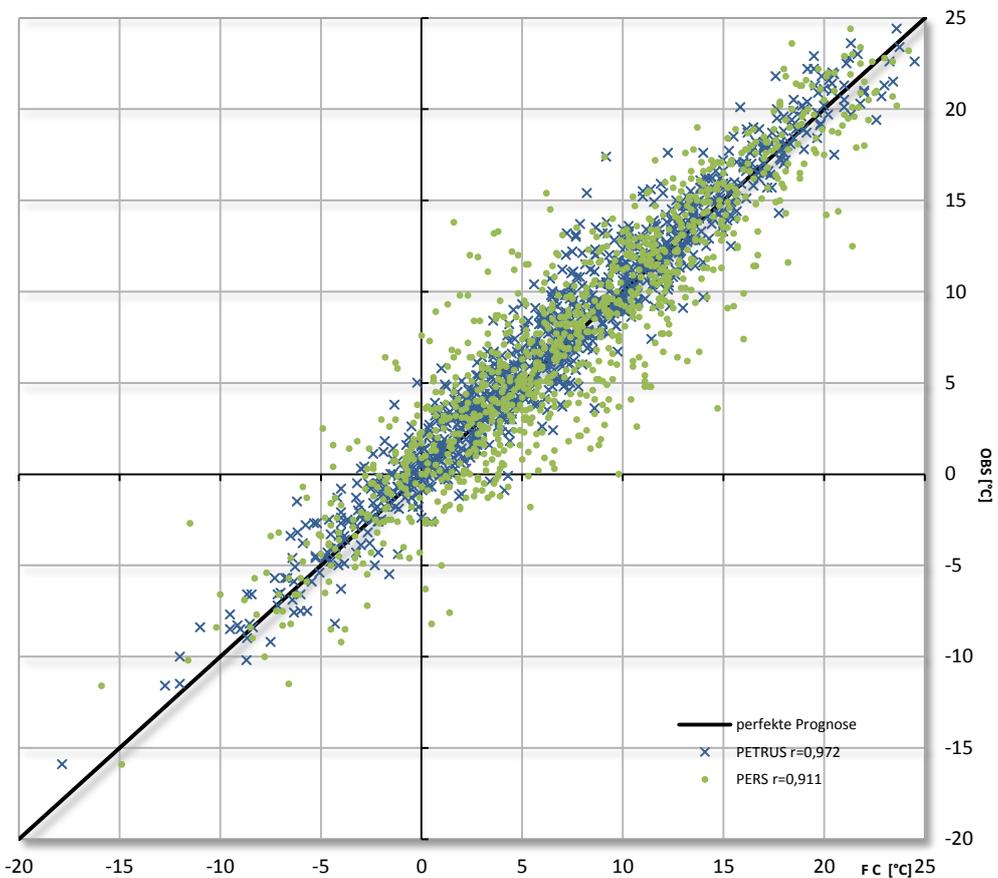


ABBILDUNG 4-12: Streudiagramm T_6 der Prognosen von PETRUS und PERS mit Korrelationskoeffizient r . Die größere Streuung der PERS um die perfekte Prognose ist gut zu erkennen.

Der mittlere absolute Fehler (MAE; vgl. Kapitel 2.1.2) aller 6 UTC-Prognosen liegt ein wenig höher als der der Minimumtemperaturprognosen (siehe TABELLE 4-7). Die Verbesserung gegenüber KLIMA (*Skill Score*; siehe Kapitel 2.1.4) bleibt dabei deutlich (gesamt **0,60**), wohingegen die Verbesserung gegenüber PERS leicht nachlässt (gesamt **0,40**). Gegenüber MOS ist keine Änderung festzustellen, eine leichte Verbesserung ist auch hier gegeben (gesamt **0,14**). Die Verbesserung in den Sommersemestern (wie auch schon bei T_{\min} ; vgl. TABELLE 4-3) geht allerdings erneut gegen Null (**0,10**).

Der RMSE als Maß für die Fehlervarianz (mittlerer quadratischer Fehler, siehe Kapitel 2.1.3) der 6 UTC-Prognosen zeigt ebenso, ganz dem Trend folgend, höhere Werte als bei den Prognosen für die Minimumtemperatur (siehe TABELLE 4-8). Die Verbesserung des PETRUS gegenüber allen Referenzprognosen ist aber auch hier wieder teils deutlich (mit **0,82** bei KLIMA) bzw. gegeben (0,26 bei MOS, vgl. TABELLE 4-9).

MAE	SS _{MAE}					
	GESAMT	WS	SS	GESAMT	WS	SS
STUDENTEN	1,52	1,57	1,45	---	---	---
PETRUS	1,31	1,37	1,25	---	---	---
PERS	2,20	2,27	2,11	0,40	0,40	0,41
KLIMA	3,29	3,51	3,04	0,60	0,61	0,59
KLITA	---	---	---	---	---	---
MOS	1,53	1,67	1,38	0,14	0,18	0,10

TABELLE 4-7: mittlerer absoluter Fehler (MAE) und der dazugehörige Skill Score (SS_{MAE}) der T_6 -Prognosen. SS_{MAE} gibt die Verbesserung des PETRUS wieder.

RMSE	GESAMT	WS	SS
	STUDENTEN	1,99	2,08
PETRUS	1,73	1,80	1,63
PERS	2,90	3,04	2,74
KLIMA	4,11	4,43	3,72
KLITA	---	---	---
MOS	2,00	2,18	1,79

TABELLE 4-8: Wurzel des mittleren quadratischen Fehlers (RMSE) der T_6 -Prognosen.

RV			
	GESAMT	WS	SS
STUDENTEN	---	---	---
PETRUS	---	---	---
PERS	0,65	0,65	0,65
KLIMA	0,82	0,83	0,81
KLITA	---	---	---
MOS	0,26	0,32	0,16

TABELLE 4-9: Reduktion der Varianzen (RV) der T_6 -Prognosen.
Die Werte geben die Verbesserung des PETRUS gegenüber den Referenzprognosen wieder.

Die Aufspaltung auf die einzelnen Vorhersagemonate zeigt noch einmal den Unterschied der Vergleichsprognosen (vgl. ABBILDUNG 4-13). Während die 6 UTC-Temperatur-Prognosen des PETRUS in allen Vorhersagemonaten zu niedrig liegen (wobei die Wintermonate (November – Jänner) dabei geringfügig schlechter liegen als die restlichen), zeigt das MOS eine deutlich positive Abweichung, wobei auch hier die „kalten“ Monate, tendenziell schlechter abschneiden.

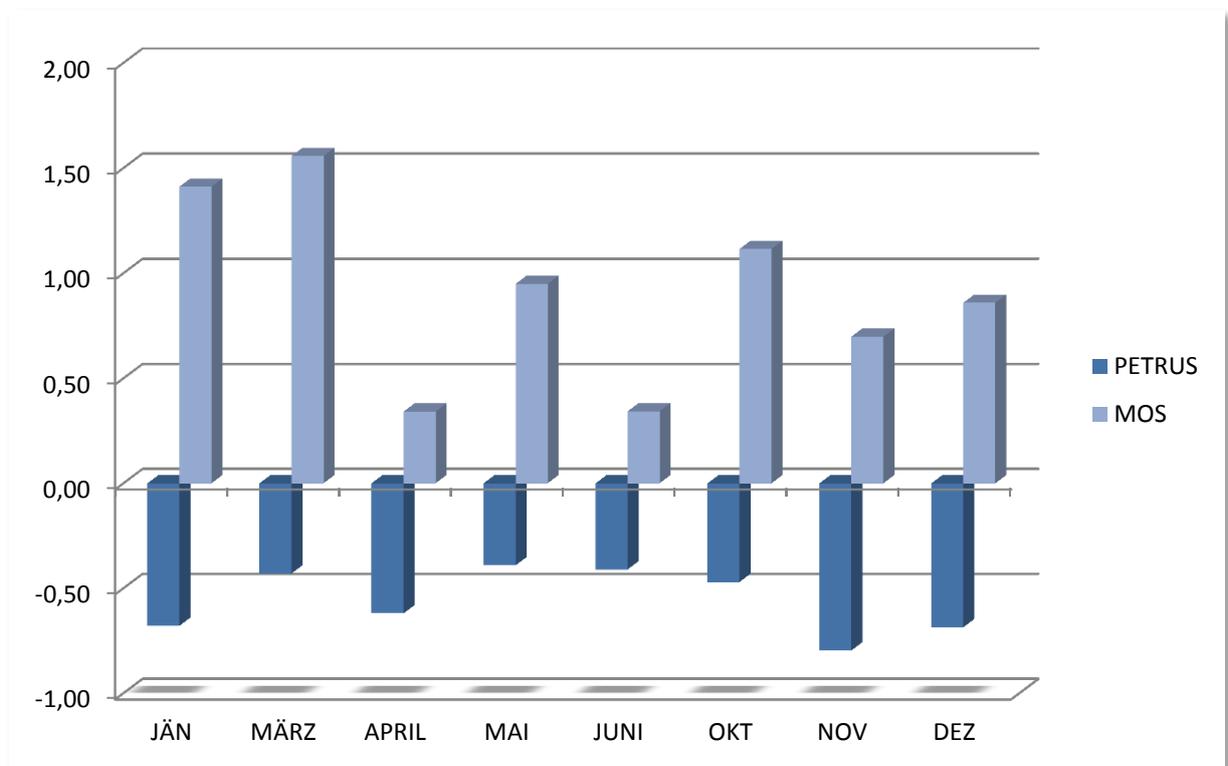


ABBILDUNG 4-13: Mittlerer Fehler (ME) der 6 UTC-Temperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate. Anders als bei der Minimumtemperatur ist ein deutlicher mittlerer Fehler auch in den „warmen“ Monaten erkennbar. Das tendenziell schlechtere Abschneiden in den „kalten“ Monaten bleibt jedoch bestehen.

Die Verteilung des mittleren absoluten Fehlers (ABBILDUNG 4-14) bzw. des RMSE (ABBILDUNG 4-15) ist der der Minimumtemperaturen recht ähnlich (vgl. ABBILDUNG 4-7 und ABBILDUNG 4-8): wieder kristallisiert sich das deutlich schlechtere Abschneiden der MOS Prognosen in den Wintersemestern heraus, mit einem Maximum im Jänner und März. PETRUS hingegen tendiert zwar zu leicht größeren Abweichungen im Oktober und November, bleibt aber im Großen und Ganzen über alle Vorhersagemonate stabil. Auffallend wieder der Monat Juni, der sich auch bei der 6 UTC-Temperatur als stärkster Vorhersagemonat des MOS erweist. Auch im November, bei der Minimumvorhersage noch gleich stark, schafft es das MOS besser abzuschneiden als PETRUS. Neu hinzugekommen ist auch eine leichte Überlegenheit im April, wohingegen im Mai die Leistung des PETRUS wieder über die des MOS steigt.

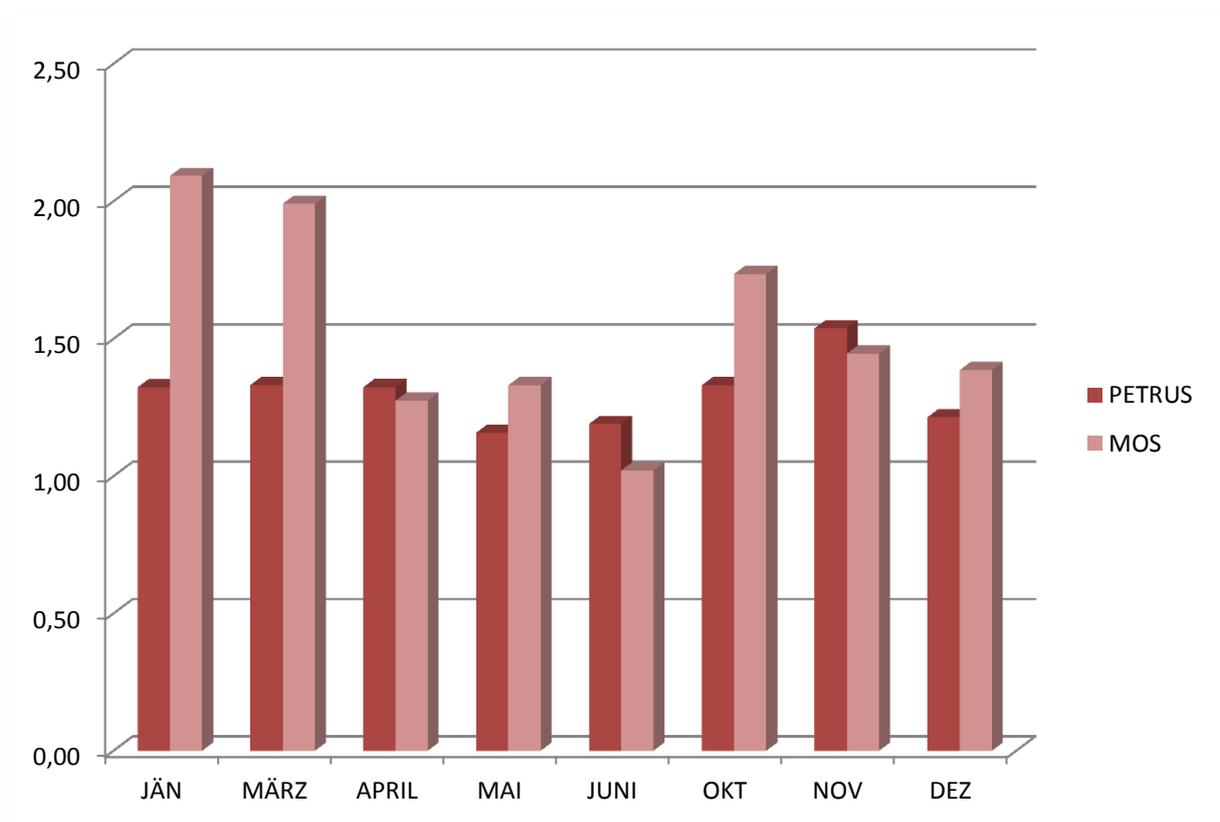


ABBILDUNG 4-14: Mittlerer absoluter Fehler (MAE) der 6 UTC-Temperatur von PETRUS & MOS, aufgespaltet auf die einzelnen Vorhersagemonate. PETRUS weist dabei annähernd konstante MAEs auf, das MOS schwankt hingegen stark. Auch bei der 6 UTC-Temperatur liegt im Monat Juni die Qualität der MOS-Vorhersagen über der des PETRUS.

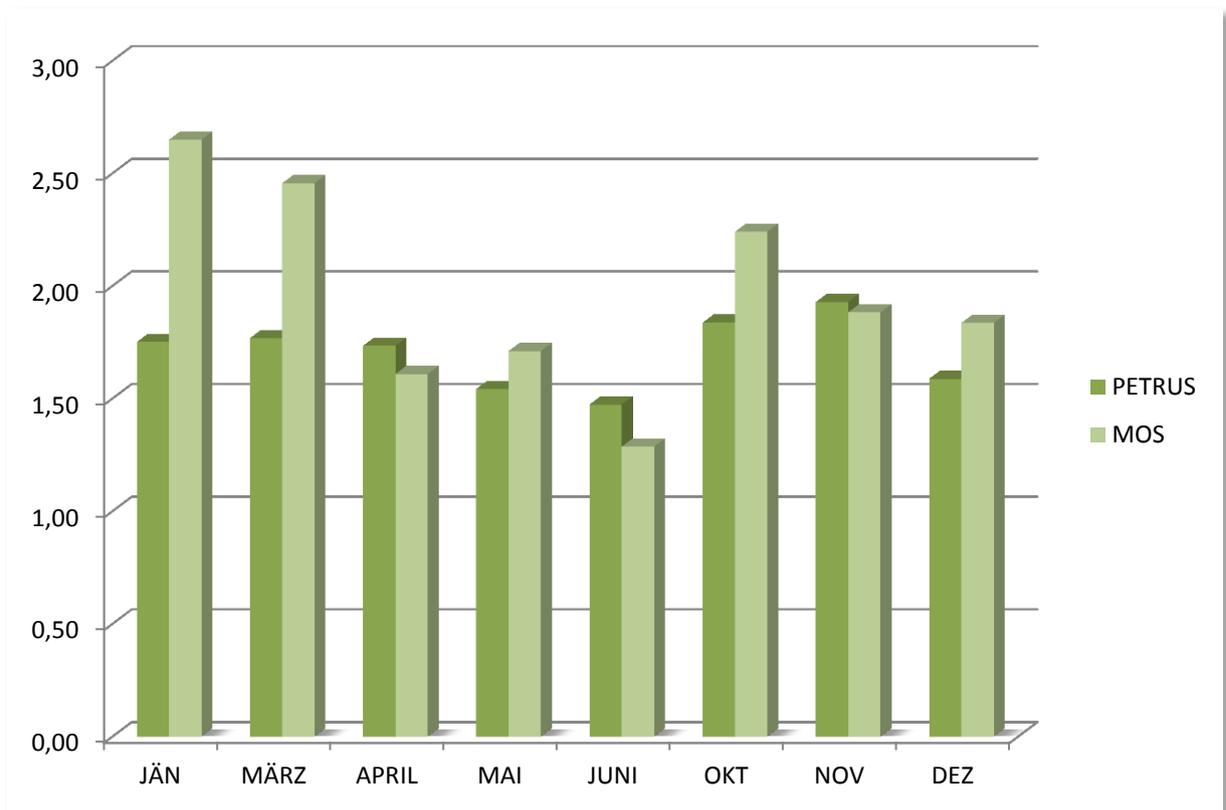


ABBILDUNG 4-15: Wurzel des mittleren quadratischen Fehler (RSME) der 6 UTC-Temperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate.

Zusammenfassend kann bei der Prognose der 6 UTC-Temperatur eine leichte Verschlechterung der Prognosegüte festgestellt werden, wobei die Sommersemester tendenziell besser abschneiden als die Wintersemester. Vor allem das MOS hat in der kalten Jahreszeit nicht nur wieder das Nachsehen, sondern schneidet bei der mittleren Abweichung auch am Schlechtesten ab (wenngleich MOS beim absoluten Fehler bzw. bei der Fehlervarianz dann doch wieder der PERS und KLIMA überlegen ist). Allgemein ist zu sagen, dass sich der Trend zu zu tiefen Vorhersagen- (PETRUS) bzw. zu zu hohen (MOS), wie auch schon bei der Minimumtemperatur, in allen untersuchten Monaten fortsetzt; die Umkehr zu zu hoch prognostizierten 6 UTC-Temperaturen des PETRUS ab etwa 20 Grad ist auf Grund des zu geringen Datensatzes in diesem Bereich nicht aussagekräftig genug. In punkto Qualität der Vorhersagen können sich zum für das MOS traditionell guten Vorhersagemonat Juni auch die Monate April (knapp aber doch) und November (der November ist auch das Monat mit der schlechtesten studentischen Leistung) reihen, während in den anderen Monaten PETRUS, teils deutlich, besser abschneidet. Gegenüber den beiden anderen Referenzprognosen (PERS, KLIMA) ist auch bei der 6 UTC-Temperatur die Verbesserung des PETRUS eindeutig.

4.2.3 Verifikation der 12 UTC-Temperatur

Die 12 UTC-Temperatur entspricht derjenigen Temperatur, die um 12 UTC (13 MEZ, 14 MESZ) mit Hilfe eines Thermometers ermittelt wird. Der Mittagstermin ist einer von vier Haupt-Messzeitpunkten neben der Morgenmessung um 6 UTC (vgl. Kapitel 4.2.2), dem Abendtermin um 18 UTC, sowie dem Nachttermin um 00 UTC. Sie entspricht im Wintersemester (respektive in den "kalten" Monaten) nicht selten dem Höchstwert des Tages oder ist diesem zumindest nahe dran. Im Sommersemester hingegen (also in den "warmen" Monaten) weicht diese zumeist stark vom Tageshöchstwert ab.

Eine erste Auswertung der 12 UTC-Temperatur zeigt ein gutes Abschneiden der studentischen Leistung mit einem mittleren Fehler (ME; vgl. Kapitel 2.1.1) von gesamt **-0,16** bzw. **-0,15** bei PETRUS. Im Vergleich zu den anderen Vorhersageparametern (T_{\min} und T_6) weist die 12 UTC-Temperatur sogar das bislang beste Ergebnis auf (siehe TABELLE 4-10). Bei der statistisch-automatisierten Prognose MOS ist der mittlere Fehler hingegen wieder eindeutig, mit gesamt **0,77** bzw. **0,96** in den Wintersemestern weist das MOS erneut das schlechteste Ergebnis aller Vergleichsprognosen auf. Im Mittel keine Abweichung (gesamt **-0,02**) bietet erneut die PERS-Vorhersage, und auch die KLIMA-Vorhersage schneidet mit gesamt **-0,60** im Vergleich recht gut ab.

ME			
	GESAMT	WS	SS
STUDENTEN	-0,16	-0,10	-0,23
PETRUS	-0,15	-0,13	-0,17
PERS	-0,02	0,12	-0,18
KLIMA	-0,60	-0,42	-0,81
KLITA	---	---	---
MOS	0,77	0,96	0,55

TABELLE 4-10: mittlerer Fehler (ME) der T_{12} -Prognosen.

Das gute Abschneiden der Studenten zeigt auch die graphische Auswertung, man beachte aber die doch recht breite Streuung der Vorhersagen sowohl in positiver als auch negativer Richtung (siehe ABBILDUNG 4-16). Bei tieferen 12 UTC-Temperaturen ($< 0^{\circ}\text{C}$) liegen die Prognosen zu niedrig, bei höheren ($> 25^{\circ}\text{C}$) hingegen zu hoch. Bei strengem Frost (Temperaturen unter -5°C), aber auch bei Hitze (Temperaturen über 30°C) handelt es sich allerdings um einzelne Extremwerte, die nur selten auftreten bzw. aufgetreten sind. Für den größten Anteil des Temperaturbereichs zwischen 0°C und 25°C ist keinerlei Abweichung festzustellen.

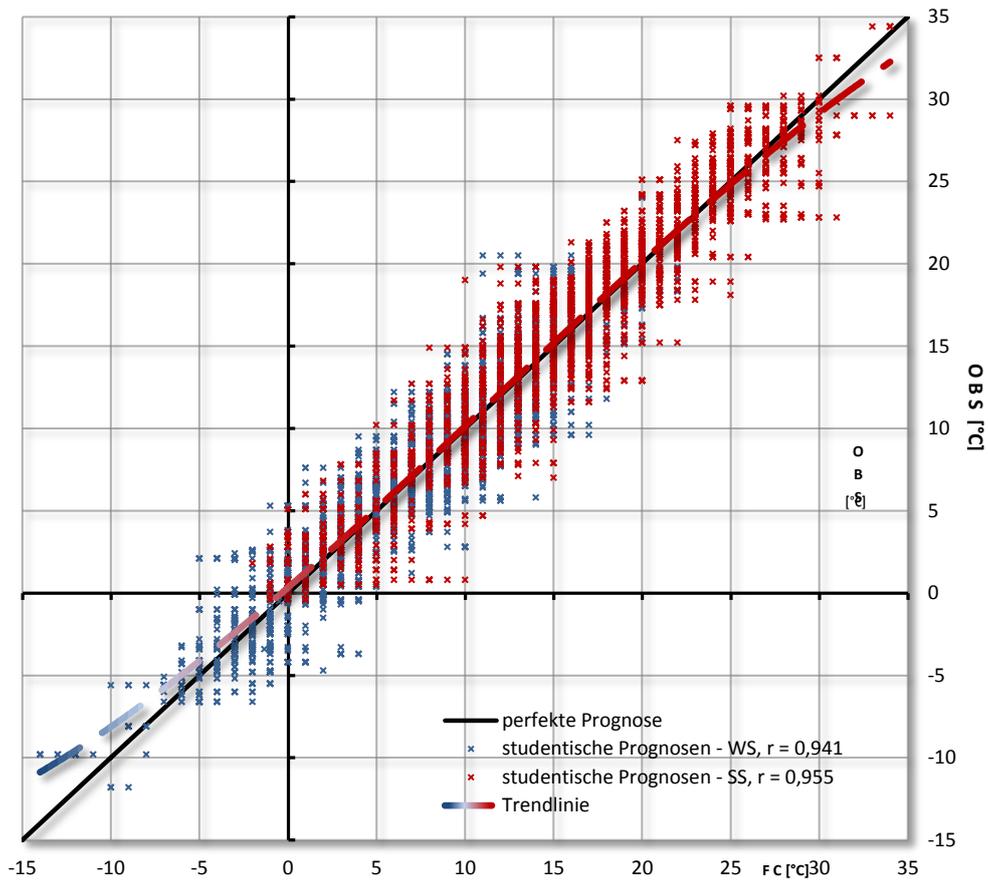


ABBILDUNG 4-16: Streudiagramm der studentischen T_{12} -Prognosen mit (polynomischer) Trendlinie und Korrelationskoeffizient r . Mit Ausnahme der Extremwerte ($<0^{\circ}\text{C}$ und $>25^{\circ}\text{C}$) ist keinerlei Trend zu erkennen.

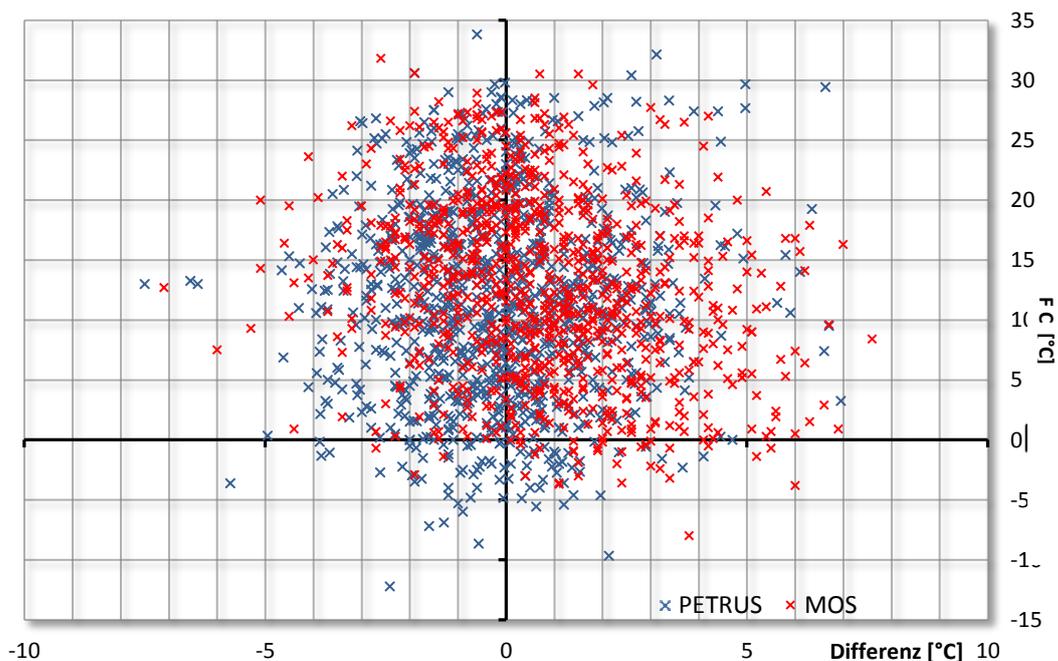


ABBILDUNG 4-17: Differenzplot der T_{12} -Prognosen; aufgetragen sind die Vorhersagefehler vs. FC von PETRUS & MOS. Der Streubereich des MOS ist auf positiver Seite fast doppelt so groß, wie auf negativer Seite. PETRUS hingegen zeigt eine Drängung auf negativer Seite.

ABBILDUNG 4-18 zeigt die Verteilung der PETRUS und MOS-Vorhersagen. Das tendenzielle, bei Frost ständige, zu-tief-vorhersagen des MOS, tritt dabei deutlich in den Vordergrund - ab etwa 20°C scheint sich dies auszugleichen. Von einzelnen Ausreißern abgesehen, ist dabei der Streubereich auf positiver Seite fast doppelt so groß ($\pm 8^\circ\text{C}$) wie der der negativen Seite ($\pm 4^\circ\text{C}$; siehe ABBILDUNG 4-17). Die breite Streuung des PETRUS hingegen wird im positiven Differenzbereich durch die dichtere Drängung im negativen Bereich ausgeglichen und resultiert in einem in Summe negativen mittleren Fehler.

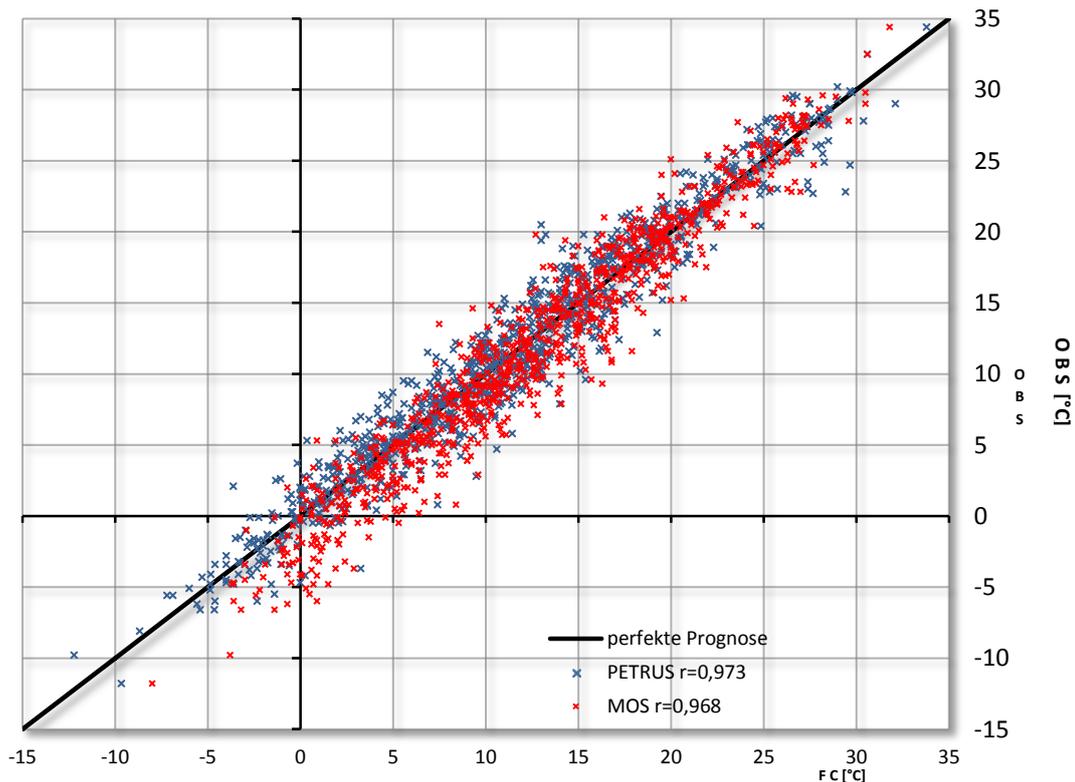


ABBILDUNG 4-18: Streudiagramm der T_{12} -Prognosen von PETRUS und MOS mit Korrelationskoeffizient r . MOS neigt bei Temperaturen $<15^\circ\text{C}$ zu zu tiefen Vorhersagen, besonders bei Frost.

So gut die Studenten beim mittleren Fehler auch abschneiden mögen, generell ist zu sagen, dass die 12 UTC-Temperatur derjenige Parameter ist, der der größten Schwankung unterlegen ist. Schon die subjektive Verifikation aus Kapitel 4.1 hat gezeigt, dass die 12 UTC Temperatur Vorhersage die niedrigste Anzahl an „richtigen“ Prognosen aufweist. Die Gründe hierfür liegen auf der Hand: als Punktprognose nicht nur lokal sondern auch temporal, unterliegt die T_{12} Vorhersage besonderen Herausforderungen. Die Hilfestellung, die die 6 UTC-Temperatur genießt (der Nähe zum Temperaturminimum zum Beispiel), ist der 12 UTC-Temperatur nicht oder nur zum Teil gegeben. Kein Wunder also, dass schon der mittlere absolute Fehler (MAE; vgl. Kapitel 2.1.2 und siehe TABELLE 4-11) ein etwas anderes Ergebnis aufweist: bei den STUDENTEN (1,67), PETRUS (1,49), PERS (2,65) und MOS

(1,78) zeigt der MAE den größten Wert aller der vier zu untersuchenden Parameter (T_6 , T_{12} , T_{\min} , und T_{\max}). Die Verbesserung gegenüber den anderen Vorhersagen ist trotzdem teils deutlich (SS_{MAE} ; vgl. Kapitel 2.1.2), lediglich die Verbesserung der menschlichen Leistung gegenüber MOS geht in den Sommersemestern - wie auch schon bei T_{\min} und T_6 - wieder gegen Null (0,09). Bei den Ergebnissen des Quadratischen Fehlers (bzw. dessen Wurzel, RMSE; vgl. Kapitel 2.1.3) und der Verbesserung davon (Reduktion der Varianzen, RV; vgl. Kapitel 2.1.5) zeigt sich einmal mehr das deutlich schlechtere Abschneiden der MOS-Vorhersagen (siehe TABELLE 4-12 und TABELLE 4-13) in den WS (2,45) gegenüber den SS (2,13), wohingegen PERS in den WS deutlich besser abschneidet (3,20) als in den SS (3,74). Eine mögliche Erklärung hierfür könnte im advektiven Charakter der Wetterlagen in den Herbst- und Wintermonaten liegen. Die Studenten bzw. PETRUS weisen keinerlei nennenswerte Unterscheidung zwischen den Semestern, KLIMA hingegen mit einer Abweichung von bis zu 5,08 das bis dato schlechteste Ergebnis auf.

MAE	SSMAE					
	GESAMT	WS	SS	GESAMT	WS	SS
STUDENTEN	1,67	1,66	1,68	---	---	---
PETRUS	1,49	1,50	1,48	---	---	---
PERS	2,65	2,39	2,94	0,44	0,37	0,43
KLIMA	4,00	3,79	4,23	0,63	0,60	0,65
KLITA	---	---	---	---	---	---
MOS	1,78	1,93	1,62	0,16	0,22	0,09

TABELLE 4-11: mittlerer absoluter Fehler (MAE) und der dazugehörige Skill Score (SS_{MAE}) der T_{12} -Prognosen. SS_{MAE} gibt die Verbesserung des PETRUS wieder.

RMSE	GESAMT	WS	SS
	STUDENTEN	2,14	2,13
PETRUS	1,92	1,92	1,91
PERS	3,46	3,20	3,74
KLIMA	4,88	4,69	5,08
KLITA	---	---	---
MOS	2,30	2,45	2,13

TABELLE 4-12: Wurzel des mittleren quadratischen Fehlers (RMSE) der T_{12} -Prognosen.

RV			
	GESAMT	WS	SS
STUDENTEN	---	---	---
PETRUS	---	---	---
PERS	0,69	0,64	0,74
KLIMA	0,85	0,83	0,86
KLITA	---	---	---
MOS	0,31	0,38	0,20

TABELLE 4-13: Reduktion der Varianzen (RV) der T_{12} -Prognosen. Die Werte geben die Verbesserung des PETRUS gegenüber den Referenzprognosen wieder.

Auf die einzelnen Monate aufgespalten, bestätigt sich erneut das tendenzielle zu niedrig- (bei PETRUS) bzw. zu hoch (bei MOS), prognostizieren wobei die Abweichungen bei PETRUS im überschaubar niedrigem Rahmen bleiben (mit Oktober als besten Monat; siehe ABBILDUNG 4-19). Auffällig dabei der Mai, der ein leicht positives Ergebnis aufweist. Anders bei MOS: einmal mehr weisen besonders die Wintermonate Dezember und Jänner und der Übergangsmonat März eine doch recht eindeutige Abweichung auf, wohingegen der Juni, als einziger Sommermonat in dieser Verifikation, keine Abweichung zeigt.

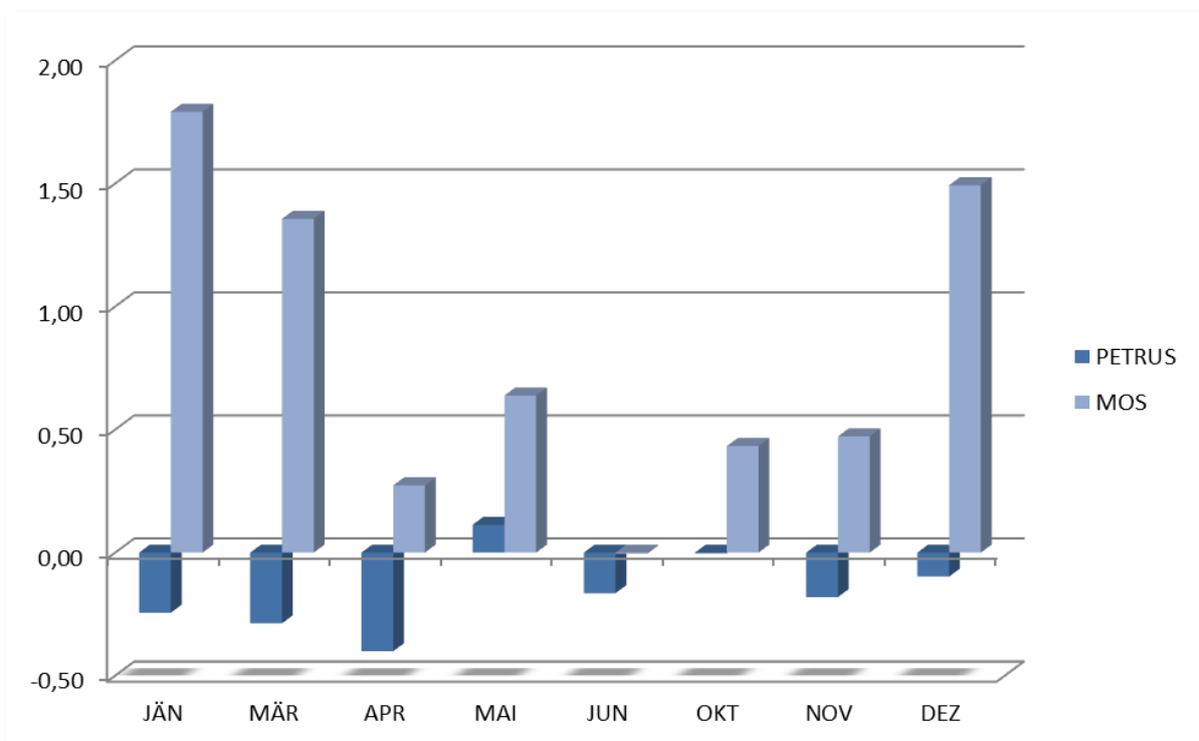


ABBILDUNG 4-19: Mittlerer Fehler (ME) der 12 UTC-Temperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate. PETRUS schneidet allgemein gut ab, auffällig der Mai, der leicht positiv abschneidet (Prognosen im Mittel zu hoch). Die Abweichungen des MOS sind teils deutlich, der Juni weist hingegen keine Abweichung auf.

Bei der Monatsaufspaltung des mittleren absoluten Fehlers (siehe ABBILDUNG 4-20) ist bei PETRUS zwar keinerlei Trend zwischen den Monaten zu erkennen, im Vergleich zur Minimumtemperatur (siehe ABBILDUNG 4-7) liegt jedoch das allgemeine Niveau des MAE um ein gutes halbes Grad höher. Diese allgemeine Verschlechterung ist auch für MOS auszumachen, wobei hier wieder die „kalten“ Monate Dezember und Jänner sowie März deutlich herausstechen. Selbiges gilt auch für die Ergebnisse des RMSE (siehe ABBILDUNG 4-21). Wie auch schon bei den vorangegangenen Parametern kann sich auch bei der 12 UTC-Temperatur das MOS beim Vorhersagemonat Juni gegenüber PETRUS behaupten.

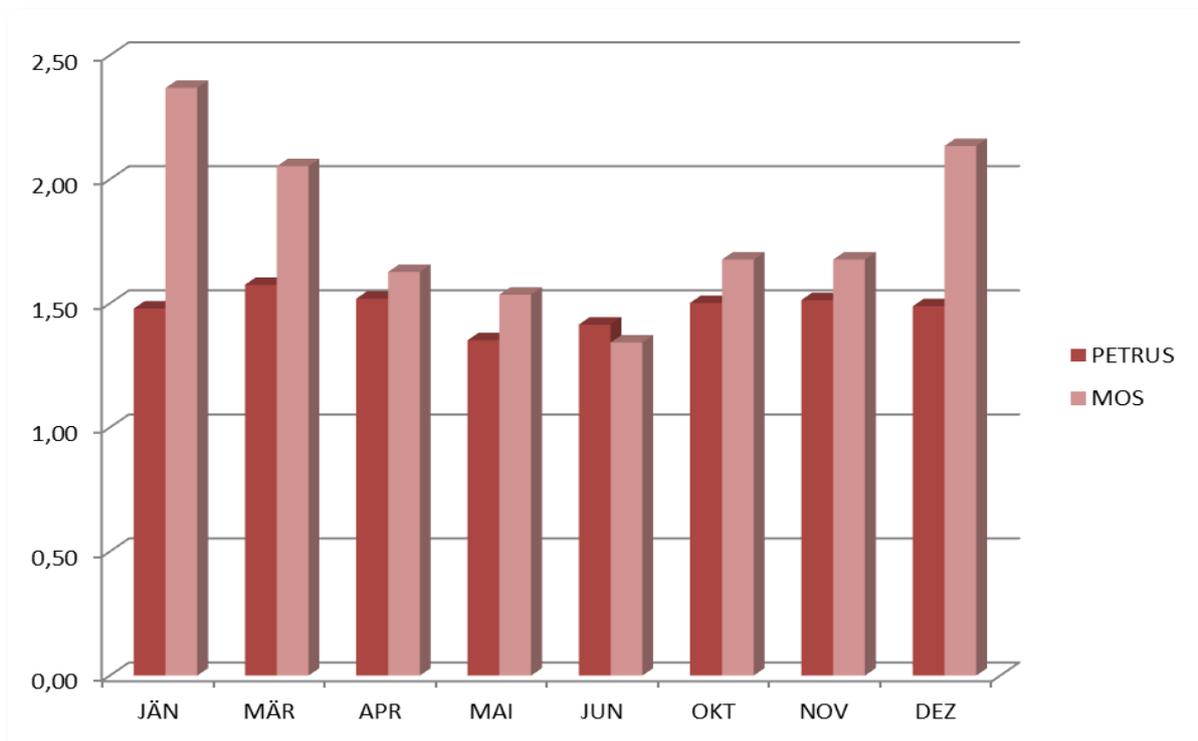


ABBILDUNG 4-20: Mittlerer absoluter Fehler (MAE) der 12 UTC-Temperatur von PETRUS & MOS, aufgespaltet auf die einzelnen Vorhersagemonate. Dabei zeigt PETRUS zwar konstante, aber im Vergleich mit T_{min} und T_6 allgemein schlechtere Ergebnisse. Die Werte des MOS liegen, mit Ausnahme des Juni, teils deutlich darüber.

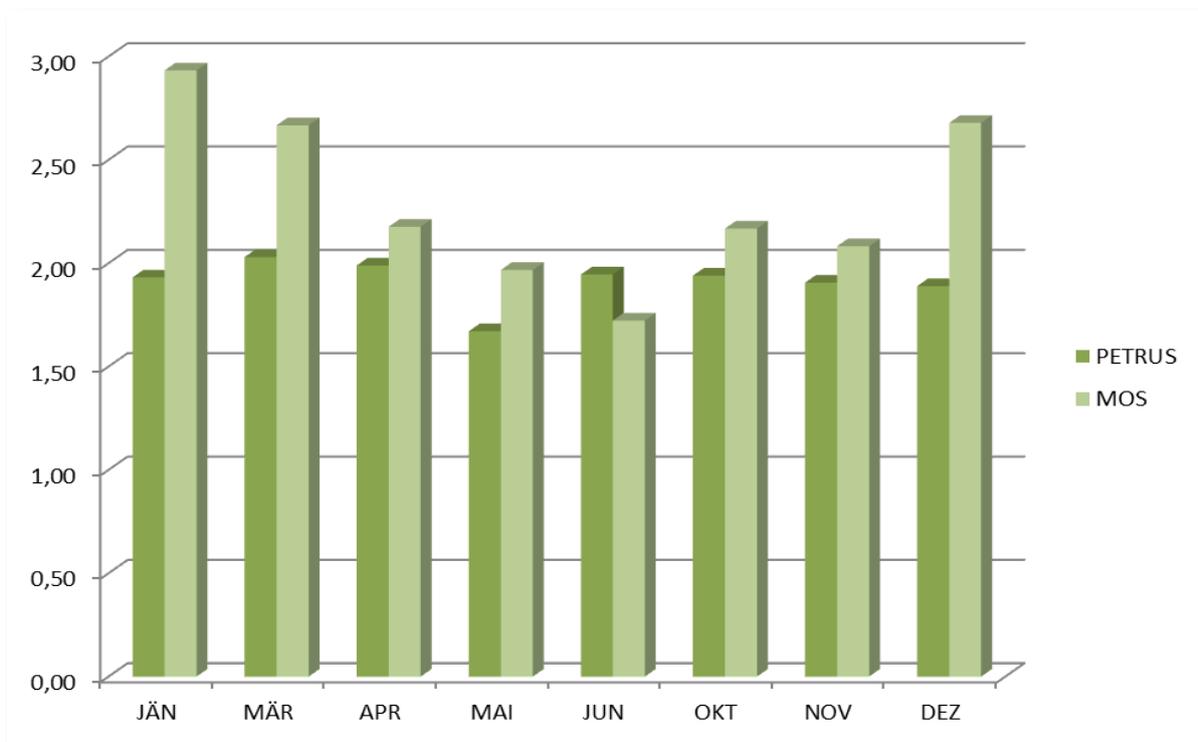


ABBILDUNG 4-21: Wurzel des mittleren quadratischen Fehlers (RSME) der 12 UTC-Temperatur von PETRUS & MOS, aufgesplittet auf die einzelnen Vorhersagemonate. Sowohl PETRUS als auch MOS weisen dabei eine im Vergleich mit T_{min} und T_6 höhere Fehlervarianz auf.

Zusammenfassend ist zu sagen, dass die 12 UTC- Temperatur, trotz des im Mittel geringen Vorhersagefehlers, die größten Probleme bei der Vorhersage aufweist, wie die durchwegs hohen Werte beim MAE und RMSE zeigen – und zwar unabhängig des Monats oder des Semesters. Vom Monat Mai abgesehen, haben die Studenten im Mittel zu niedrig prognostiziert, wobei der absolute Vorhersagefehler zumeist unter $1,5^{\circ}\text{C}$ liegt. Anders die Situation bei MOS, das durchgehend und teils recht deutlich zu hohe Prognosen liefert, besonders in den Wintermonaten, in denen der absolute Fehler bei oder sogar über $2,0^{\circ}\text{C}$ liegt. Erst ab einer Temperatur von etwa 20°C liefert das MOS deutlich bessere Vorhersagen. Eine besseres Abschneiden der menschlichen Leistung gegenüber MOS ist somit gegeben, mit einer Ausnahme: im Monat Juni überzeugt auch bei der 12 UTC-Temperatur das MOS mit im Mittel keiner Abweichung und dem niedrigsten absoluten Fehler im Vergleich. Wenngleich die beiden anderen Referenzprognosen (PERS und KLIMA) auch bei der 12 UTC-Temperatur eine nur untergeordnete Rolle spielen, ist dennoch eine Änderung festzustellen: waren bei beiden Vergleichsprognosen die höheren Fehler sowohl bei T_{min} als auch T_6 in der „kalten“ Jahreszeit (respektive den WS) anzutreffen, so sind es bei der 12 UTC-Temperatur mehr die „wärmeren“ Monate (respektive die SS), in denen die größeren Abweichungen vorliegen. Die Ursache hierfür liegt in den für die Sommerzeit typischen konvektiven Großwetterlagen, die die Temperaturprognose tagsüber zusätzlich erschweren können, wohingegen bei advektiven Wetterlagen in der Herbst- und Winterzeit das

Temperaturniveau über Tage hinweg gleich bleiben kann (nicht selten nahe dem langjährigen Mittel).

4.2.4 Verifikation der Maximumtemperatur (T_{max})

Analog zur Minimumtemperatur aus Kapitel 4.2.1 ist die Maximumtemperatur die höchste Temperatur, die während eines vordefinierten Zeitraumes erreicht wird. Dieser Zeitraum erstreckt sich in unserem Fall von 6 bis 18 UTC und entspricht somit dem Höchstwert des jeweiligen Vorhersagetages. Die Ablesung erfolgt stets um 18 UTC und wird mit einem sogenannten Maximumthermometer (einem Quecksilberthermometer, welcher ähnlich einem Fieberthermometer konstruiert ist) ermittelt. Die Kapillare ist dabei unmittelbar nach dem Thermometergefäß verengt. Die Messflüssigkeit kann sich zwar bei steigender Temperatur ungehindert durch diese Verengung ausdehnen, bei sinkender Temperatur jedoch reißt der Quecksilberfaden an der Engstelle [LILJEQUIST, 2006].

Wie schon bei den vorangegangenen Parametern (T_{min} , T_6 , T_{12}) liegen auch bei der Maximumtemperatur die Prognosen, mit Ausnahme denen des MOS, im Mittel unterhalb der Beobachtungen, die Maximumtemperatur wurde also zu niedrig vorhergesagt. PETRUS weist dabei einen mittleren Fehler (ME; vgl. Kapitel 2.1.1, siehe TABELLE 4-14) von gesamt **-0,39** auf, PERS **-0,08** und die beiden Klima-Prognosen **-0,64** bzw. **-0,44**. Eine im Gegensatz dazu deutliche Abweichung findet sich hingegen beim MOS, mit gesamt **0,70** bzw. **1,04** in den WS liefert die Prognose sogar die (vermeintlich) schlechtesten Ergebnisse. In den SS hingegen holt das MOS mit **0,34** deutlich auf.

ME				
	GESAMT	WS	SS	
STUDENTEN	-0,36	-0,18	-0,57	
PETRUS	-0,39	-0,24	-0,55	
PERS	-0,08	0,08	-0,27	
KLIMA	-0,64	-0,24	-1,09	
KLITA	-0,44	-0,32	-0,57	
MOS	0,70	1,04	0,34	

TABELLE 4-14: mittlerer Fehler (ME) der T_{max} -Prognosen.

Ein Blick auf die graphische Auswertung der Studenten zeigt nahezu keine Abweichungen in einem großen Teil des vorherzusagenden Temperaturbereiches. Lediglich bei ganz tiefen

Maxima (Eistage; Tage mit einem Maximum $< 0^{\circ}\text{C}$) sind negative Abweichungen auszumachen, bei Frostereignissen wird also tendenziell zu niedrig prognostiziert (siehe **ABBILDUNG 4-22**). Bei höheren Temperaturen (Tropentage; Tage mit einem Maximum $\geq 30^{\circ}$) hingegen liegen die Prognosen eher über den Beobachtungen. Der Stichprobenumfang mag zwar bei Maxima $>30^{\circ}$ schon deutlich geringer sein, spiegelt aber im Grunde die Erkenntnisse aus der Untersuchung der 12 UTC-Temperatur (vgl. Kapitel 4.2.3) wider, die im höheren Temperaturbereich ($>30^{\circ}\text{C}$) ebenfalls zu hohe Prognosen im Mittel gezeigt haben (bei noch geringerem Stichprobenumfang).

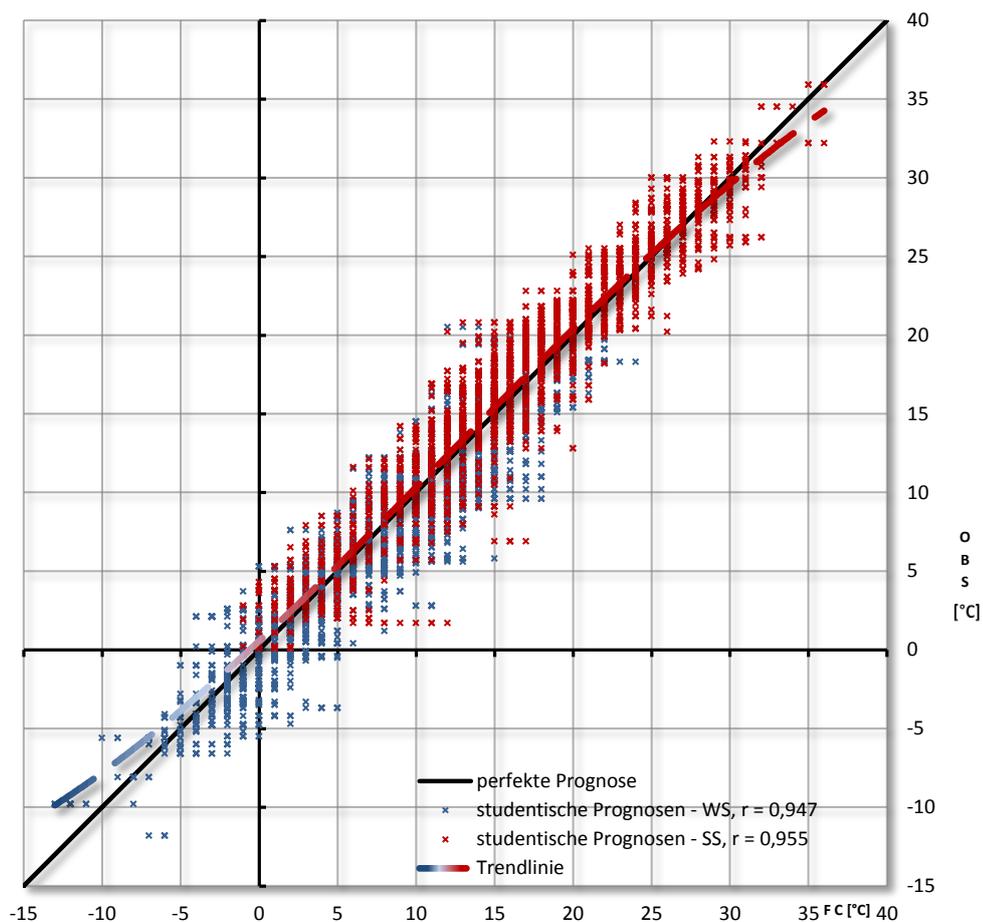


ABBILDUNG 4-22: Streudiagramm der studentischen T_{max} -Prognosen mit (polynomischer) Trendlinie und Korrelationskoeffizient r . Im Falle eines Eistages ($< 0^{\circ}\text{C}$) liegen die Vorherzusagen im Mittel zu niedrig, bei heißen Tagen ($> 30^{\circ}\text{C}$) im Mittel zu hoch.

ABBILDUNG 4-23 zeigt das Abschneiden des MOS, wobei die Streuung beider Vergleichsprognosen (also PETRUS und MOS) ein ähnliches Muster wie das der 12 UTC-Temperatur aufweist (vgl. **ABBILDUNG 4-18**). Bei MOS liegen die Vorhersagen teils deutlich zu hoch, bei Temperaturen $>20^{\circ}\text{C}$ hingegen schneidet das MOS wesentlich besser ab (siehe dazu auch **ABBILDUNG 4-24**). Die beiden anderen Vergleichsprognosen (PERS und KLITA) sind in **ABBILDUNG 4-25** aufgetragen – sowohl der Grund für das vermeintlich gute Abschneiden der

PERS als auch das Fehlen jeglicher Extremwerte der KLITA-Vorhersage sind wieder deutlich auszumachen.

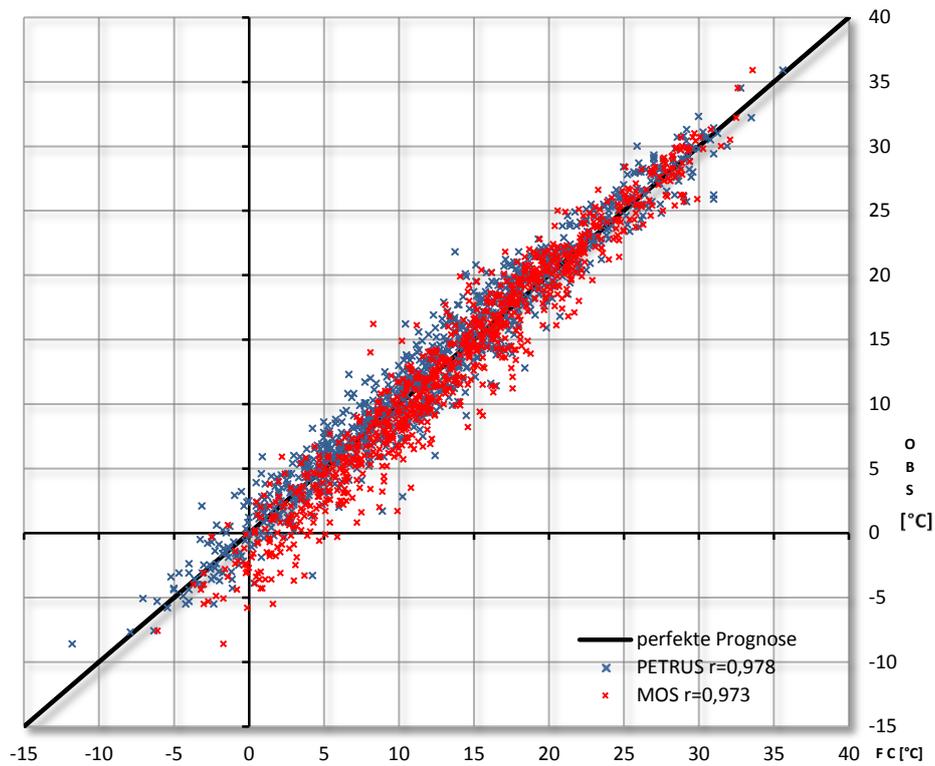


ABBILDUNG 4-23: Streudiagramm der T_{max} -Prognosen von PETRUS und MOS mit Korrelationskoeffizienten r . Das MOS zeigt einerseits eine deutlich positive Abweichungen im unteren Vorhersagebereich ($<20^{\circ}\text{C}$), darüber andererseits keinen Trend.

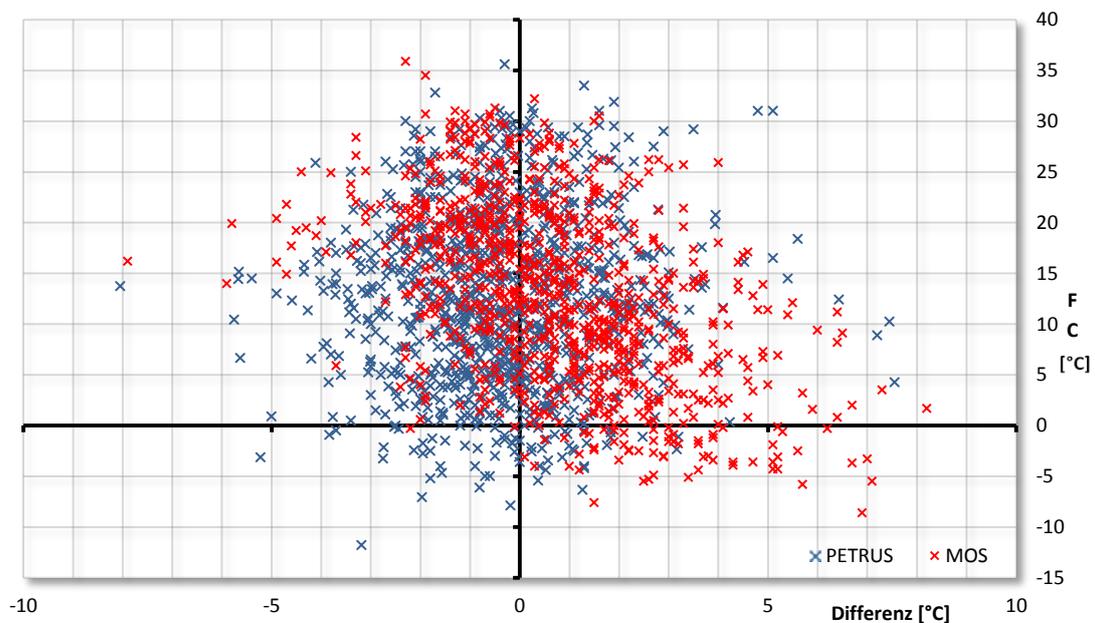


ABBILDUNG 4-24: Differenzplot der T_{max} -Prognosen; aufgetragen sind die Vorhersagefehler vs. FC von PETRUS & MOS. PETRUS weist mit $\pm 5^{\circ}\text{C}$ (von einzelnen Ausreißern abgesehen) eine gleichmäßige Verteilung auf. Das MOS hingegen hat im unteren Temperaturbereich ($<10^{\circ}\text{C}$) eine Schwankung von -2 bis +8, im oberen ($>15^{\circ}\text{C}$) von -8 bis +4.

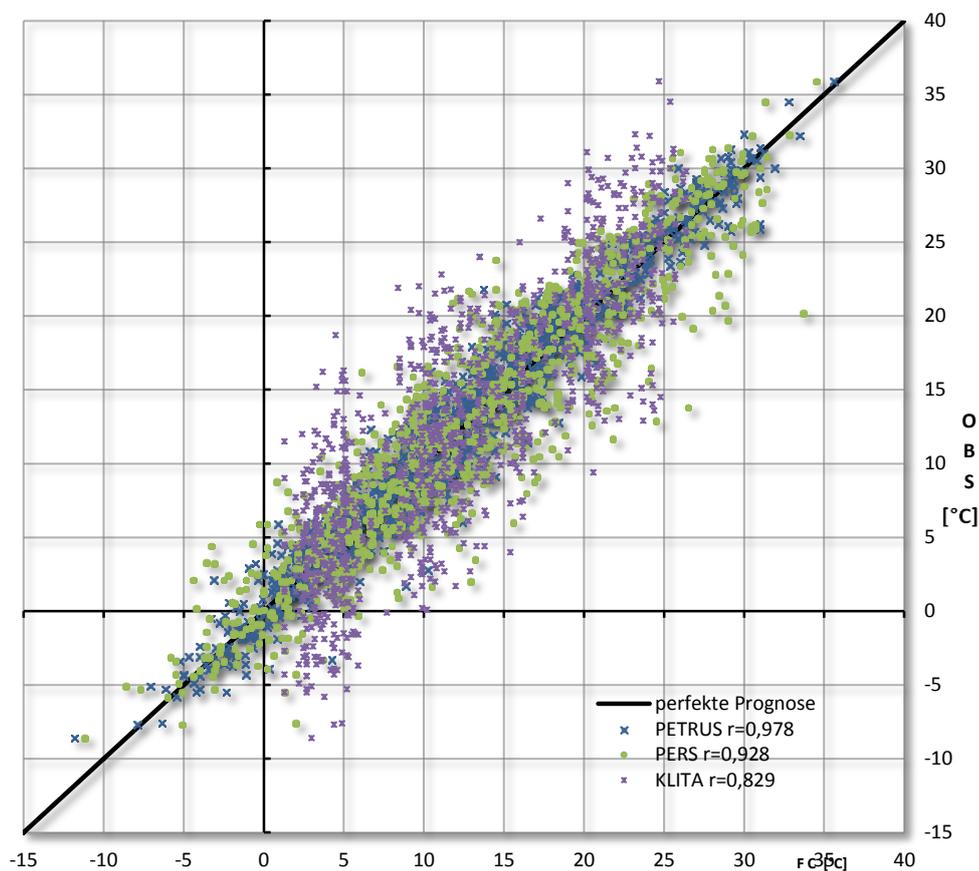


ABBILDUNG 4-25: Streudiagramm der T_{max} -Prognosen von PETRUS, PERS und KLITA mit Korrelationskoeffizient r . Die breite Streuung der PERS sowie die fehlende Schärfe der KLITA sind auch bei der Maximumtemperatur ersichtlich.

Die Ergebnisse des mittleren absoluten Fehlers (MAE; vgl. Kapitel 2.1.2, siehe TABELLE 4-15), liegen durchaus im Bereich dessen, was aufgrund der Ergebnisse der vorangegangenen Parameter zu erwarten war: Mit einem mittleren absoluten Fehler von gesamt **1,44** bei PETRUS bzw. **1,62** bei den STUDENTEN zeigen die Ergebnisse zwar eine leichte Verbesserung gegenüber den Ergebnissen der 12 UTC-Temperatur (mit Ausnahme der Klima-Prognosen), doch wird auch hier das „erlaubte“ Fehlerintervall aus Kapitel 4.1 von $\pm 1^\circ\text{C}$ um rund eines halben Grades überschritten. Als Prognose über einen gewissen Zeitraum muss hier allerdings die deutliche Verschlechterung gegenüber den Ergebnissen der ebenfalls über einen gewissen Zeitraum prognostizierten Minimumtemperaturen angemerkt werden (vgl. TABELLE 4-3). Das Gleiche gilt auch für die Ergebnisse des MOS (gesamt **1,69**), und für die der beiden anderen Referenzprognosen – wenngleich die Verschlechterung der Klima-Prognosen sich doch recht deutlich auf die Tageszeit bezieht: Die Ergebnisse sind, wie auch schon die der T_{12} (vgl. TABELLE 4-11), um bis zu 1,5 Grad höher als die der T_{min} bzw. T_6 . Speziell die Klima-Prognosen schneiden somit tagsüber noch schlechter ab, besonders in den Sommersemestern – ein Umstand, der durch häufigere Auftreten von „Extrema“ tagsüber leicht zu erklären ist (angemerkt sei hier nochmals, dass, bezogen auf eine Klimaprognose, jede Abweichung bereits als „Extremwert“ angesehen

werden kann). Zusammenfassend ist eine Verbesserung des PETRUS (SS_{MAE} ; vgl. Kapitel 2.1.2) gegenüber allen Referenzprognosen gegeben; hervorzuheben sind allerdings die MOS-Vorhersagen in den SS, in denen diese Verbesserung gegen Null (**0,03**) geht – das beste Ergebnis des MOS.

	MAE			SS_{MAE}		
	GESAMT	WS	SS	GESAMT	WS	SS
STUDENTEN	1,62	1,59	1,65	---	---	---
PETRUS	1,44	1,42	1,45	---	---	---
PERS	2,52	2,28	2,78	0,43	0,38	0,48
KLIMA	4,05	3,84	4,28	0,65	0,63	0,66
KLITA	3,94	3,83	4,06	0,63	0,63	0,64
MOS	1,69	1,88	1,49	0,15	0,24	0,03

TABELLE 4-15: mittlerer absoluter Fehler (MAE) und der dazugehörige Skill Score (SS_{MAE}) der T_{max} -Prognosen. SS_{MAE} gibt die Verbesserung des PETRUS gegenüber den Referenzprognosen wieder.

Ein Blick auf die Wurzel des mittleren quadratischen Fehlers (RMSE; vgl. Kapitel 2.1.3, siehe TABELLE 4-16) zeigt noch einmal deutlich die Probleme der Klimaprognosen tagsüber. Mit **5,17** in den SS verbucht dabei KLIMA das schlechteste Ergebnis aller zu untersuchenden Parameter und ist einmal mehr ein Indiz dafür, dass Prognosen basierend auf der Klimatologie im Allgemeinen als Vorhersagen nur bedingt zu gebrauchen sind. Die anderen Prognosen schneiden etwas besser ab als noch die 12 UTC-Prognosen, wengleich die Ergebnisse an die der Minimumtemperatur (vgl. TABELLE 4-4) nicht herankommen. Mit gesamt **1,84** liegen dabei die „Ausreißer“ bei PETRUS nach wie vor unter 2°C . Eine Verbesserung der menschlichen Leistung gegenüber den Referenzprognosen ist somit in allen Fällen gegeben und orientiert sich in etwa an den Ergebnissen der vorangegangenen Parametern (Reduktion der Varianzen, RV; vgl. Kapitel 2.1.5).

	RMSE		
	GESAMT	WS	SS
STUDENTEN	2,09	2,07	2,10
PETRUS	1,84	1,85	1,83
PERS	3,26	3,05	3,47
KLIMA	4,95	4,75	5,17
KLITA	4,84	4,77	4,93
MOS	2,20	2,39	1,97

TABELLE 4-16: Wurzel des mittleren quadratischen Fehlers (RMSE) der T_{max} -Prognosen.

RV			
	GESAMT	WS	SS
STUDENTEN	---	---	---
PETRUS	---	---	---
PERS	0,68	0,63	0,72
KLIMA	0,86	0,85	0,87
KLITA	0,85	0,85	0,86
MOS	0,30	0,40	0,14

TABELLE 4-17: Reduktion der Varianzen (RV) der T_{max} -Prognosen. Die Werte geben die Verbesserung des PETRUS gegenüber den Referenzprognosen wieder.

Bei der Untersuchung der einzelnen Vorhersagemonate bestätigt sich einmal mehr das tendenzielle zu tief- bei PETRUS bzw. zu hoch-vorhersagen bei MOS - allerdings sticht der April in zweierlei Hinsicht deutlich hervor (siehe ABBILDUNG 4-26): Mit einem Fehler von fast -1°C bei PETRUS ist die Abweichung bei keinem anderen Parameter in keinem anderen Monat so deutlich und auch das MOS unterliegt zum ersten Mal einer deutlichen Unterschätzung. Eine mögliche Erklärung für diesen doch eindeutigen Ausreißer könnte klimatologische Ursachen haben: Mit einer deutlichen Erhöhung des durchschnittlichen Temperaturmaximums im April von 14 auf 18 Grad an der Messstelle Wien Hohe Warte (ZAMG) innerhalb der letzten 30 Jahre, zeigt gerade das Frühjahr, und hier insbesondere der April, eine deutliche Zunahme der monatlichen Durchschnittsmaxima. Seit dem Jahre 2000 findet sich in der Statistik auch kein April, der im Mittel zu kühl temperiert war [ZAMG, MONATSRÜCKBLICKE]. Mit Ausnahme des März, dessen Abweichung bei PETRUS ähnlicher Natur sein dürfte, zeigen ansonsten alle übrigen Monate sowohl bei PETRUS als auch bei MOS eine ähnliche Verteilung wie bei den vorangegangenen Vorhersageparametern, wobei abermals das deutlich schlechtere Abschneiden der MOS-Vorhersagen in den Wintermonaten herauszuheben ist.

Die Monatsverteilung des MAE, wie auch die des RMSE (siehe ABBILDUNG 4-28) zeigen eine fast idente Verteilung wie die der 12 UTC-Temperatur (vgl. ABBILDUNG 4-20 und ABBILDUNG 4-21), wengleich eine leichte Verbesserung auszumachen ist. Somit schneidet bei PETRUS einmal mehr der Mai am besten ab, und kann auch im Juni zumindest mit MOS gleichziehen, das wiederum seine Stärke in diesem Sommermonat unter Beweis stellt. Dafür ist der Fehler vor allem in den Wintermonaten Dezember und Jänner erheblich.

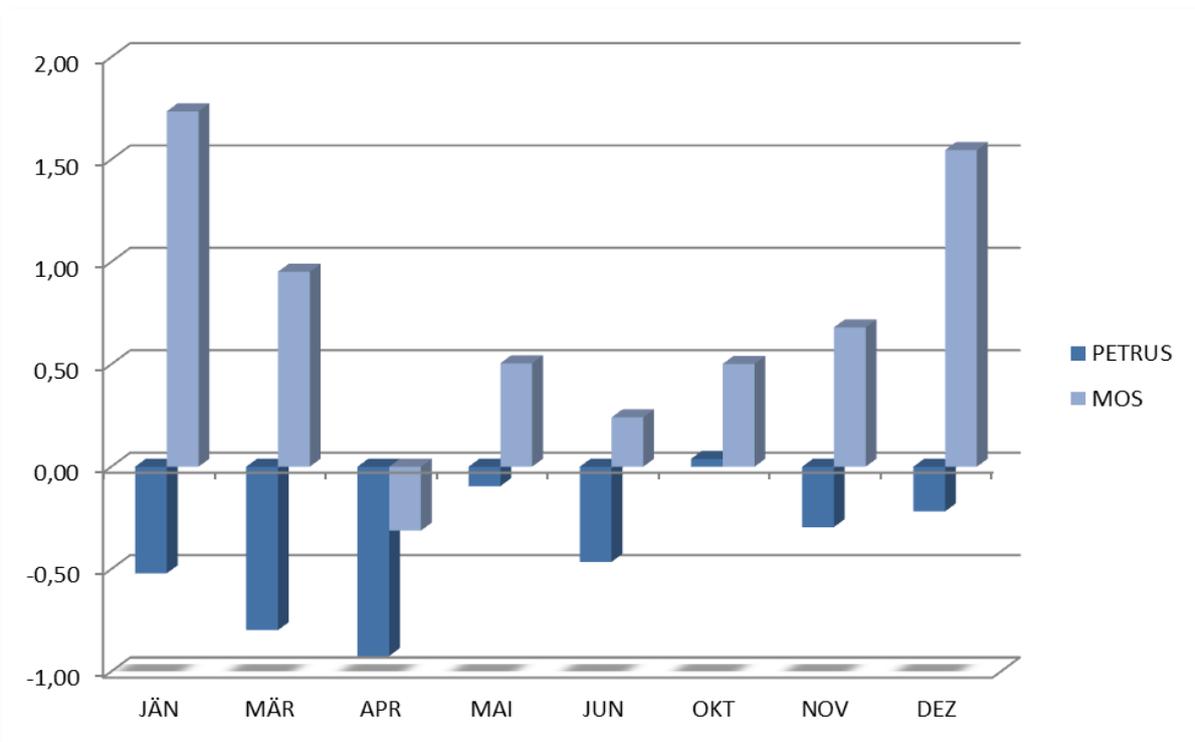


ABBILDUNG 4-26: Mittlerer Fehler (ME) der Maximumtemperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate. Der April sticht als besonderes Monat hervor: in keinem anderen Monat ist der mittlere Fehler so deutlich (bei PETRUS) bzw. negativ (bei MOS). Generell finden sich bei PETRUS die größten Abweichungen in den Frühlingsmonaten März und April, wohingegen die Wintermonate besser abschneiden. Bei MOS sind, mit Ausnahme des April, die größeren Abweichungen einmal mehr in den „kalten“ Monaten zu finden.

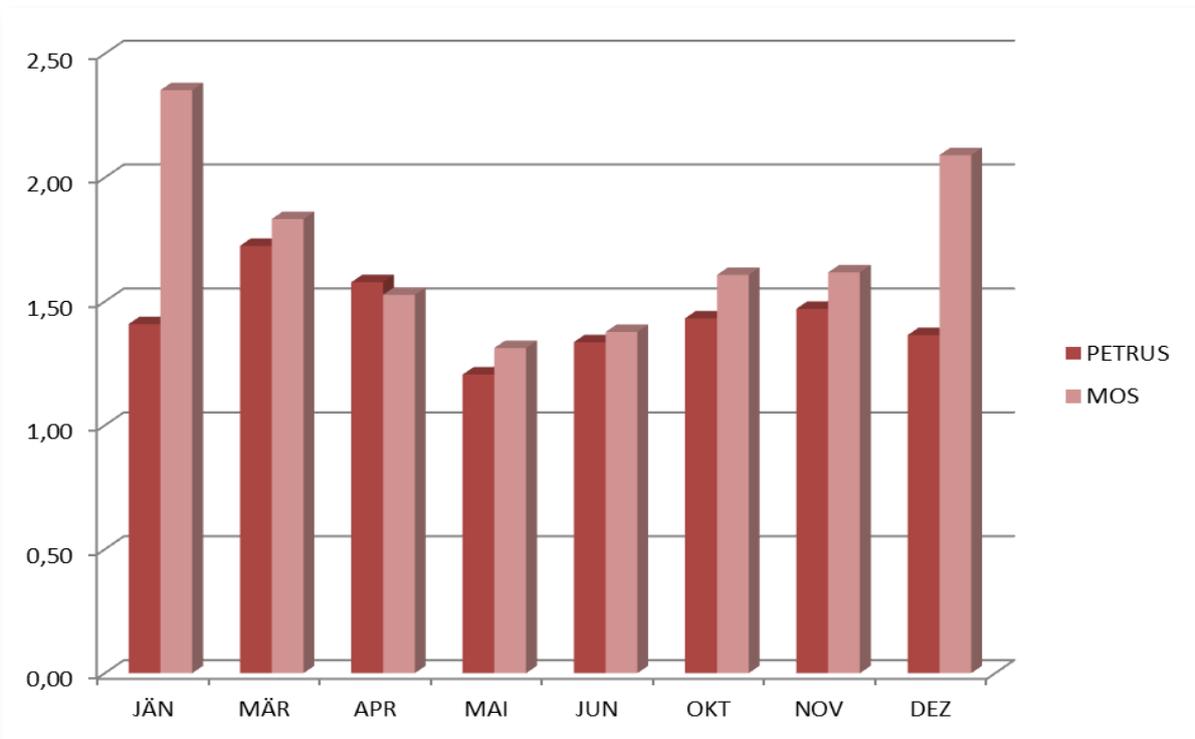


ABBILDUNG 4-27: Mittlerer absoluter Fehler (MAE) der Maximumtemperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate.

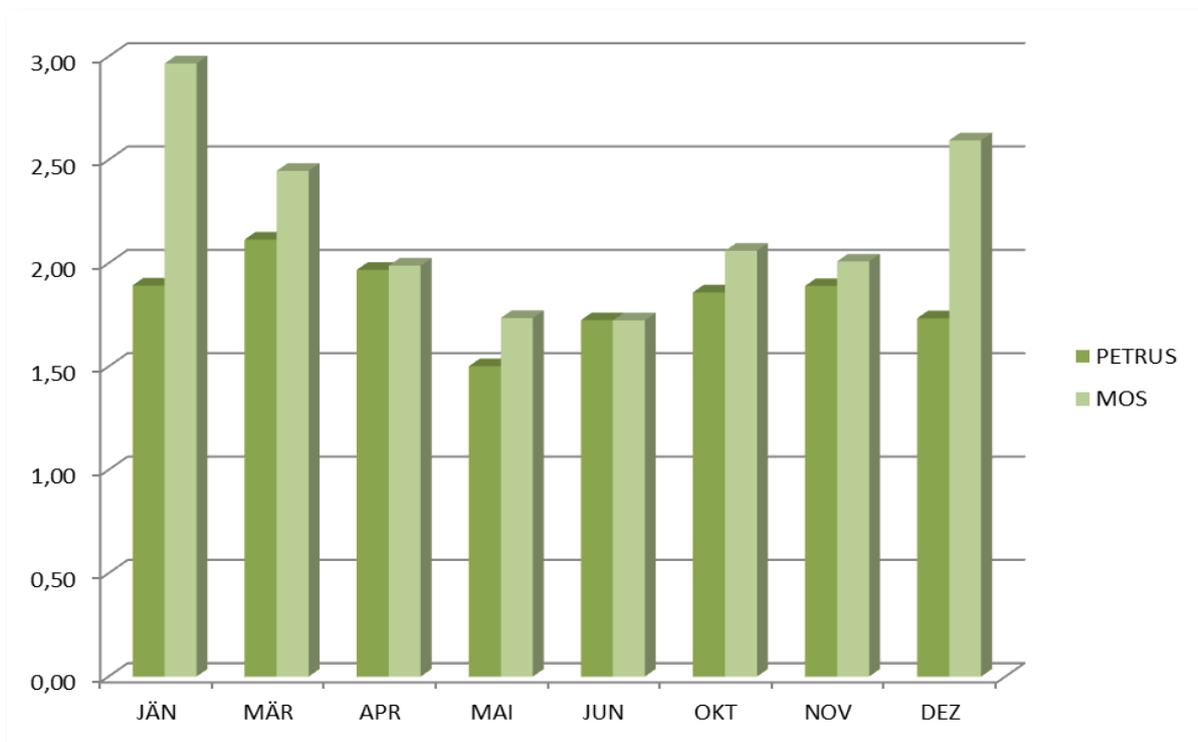


ABBILDUNG 4-28: Wurzel des mittleren absoluten Fehlers (RMSE) der Maximumtemperatur von PETRUS & MOS, aufgespalten auf die einzelnen Vorhersagemonate.

Zusammenfassend mag zwar die Vorhersage der Maximumtemperatur nicht den gleichen Schwierigkeiten wie die der 12 UTC-Temperatur unterliegen, wird aber von diesen maßgeblich beeinflusst. Somit ist es nur wenig verwunderlich, dass bei T_{\max} die MAE und RMSE Ergebnisse eine nahezu gleiche Verteilung wie bei T_{12} aufweisen, wobei sich der Monat Mai bei PETRUS als bester Vorhersagemonat bestätigt hat. Am Schlechtesten schneiden die Frühjahrsmonate März und April ab, der März weist sogar die größte mittlere, absolute Abweichung aller Parameter auf und der April sticht mit seiner doch recht deutlichen Unterschätzung hervor. Das MOS hingegen kann sich einmal mehr im Monat Juni behaupten, auch im Mai liefert es gute Ergebnisse. Die Wintermonate weisen hingegen auch bei der Maximumtemperatur die größten Probleme auf. Beim tendenziellen „zu niedrig vorhersagen“ des MOS hat sich gezeigt, dass sich dieses im höheren Temperaturbereich ($>20^{\circ}\text{C}$) etwas ausgleicht. Auch hier liegt die Vermutung einer klimatologischen Ursache nahe, weist nämlich die Maximumtemperatur des MOS im April als einzigem Vorhersagemonat eine negative Abweichung auf.

4.3 Die Temperatur als kategorischer Parameter

Bisher haben wir die Temperatur als fortlaufenden (kontinuierlichen) Parameter angesehen, was in Bezug auf die zu erwartende Aussage (wie genau ist meine Prognose? bzw. wie hoch ist

mein Prognosefehler?) durchaus angebracht ist. Gerade aber auch bei der Temperatur macht es Sinn, diese in spezielle Kategorien einzuteilen, um so zu erkennen, in wieweit ein prognostiziertes Ereignis (*event*) eingetroffen ist, oder nicht. Diese Kategorisierung mag zwar für den „normalen“ Konsumenten unnötig erscheinen, macht aber zum Beispiel in der Wirtschaft, bei Behörden, etc. durchaus Sinn. Im Straßendienst etwa ist es von entscheidender Bedeutung (und nicht zuletzt eine Kostenfrage), die Streufahrzeuge rechtzeitig in Bereitschaft bzw. Dienst zu versetzen, wenn Frost (Temperaturen $< 0^{\circ}\text{C}$) vorhergesagt ist. Tritt dieser Frost als vorhergesagtes *event* dann nicht ein, spricht man von einem Falschen Alarm (*false alarm*; vgl. Kapitel 2.2). Was im Einzelfall passieren wird und kann, sollte natürlich kein Dauerzustand sein bzw. werden – schon allein aus wirtschaftlicher Sicht.

Dieses Kapitel orientiert sich ganz an der klimatologischen Einteilung von Tagen und definiert folgende Kategorien (für die einzelnen Definitionen siehe auch [DWD, WETTERLEXIKON]):

- Frosttag: ein Tag, an dem die Minimumtemperatur unter 0°C sinkt ($T_{\min} < 0^{\circ}\text{C}$).
- Eistag: ein Tag, an dem sowohl Minimum- als auch Maximumtemperatur unter 0°C liegen ($T_{\min} \& T_{\max} < 0^{\circ}\text{C}$), das heißt, es herrscht 24 Stunden Dauerfrost.
- Sommertag: ein Tag, an dem das Temperaturmaximum bei oder über 25°C liegt ($T_{\max} \geq 25^{\circ}\text{C}$).
- Tropentag (Heißer Tag): ein Tag, an dem das Temperaturmaximum bei oder über 30°C liegt ($T_{\max} \geq 30^{\circ}\text{C}$).
- Tropennacht: eine Nacht, in der das Temperaturminimum nicht unter 20°C sinkt ($T_{\min} \geq 20^{\circ}\text{C}$).

Zusätzlich seien noch folgende Parameter definiert, die keiner klimatologischen Definition entsprechen:

- Strenger Frost: ein Tag, an dem die Minimumtemperatur unter -5°C sinkt ($T_{\min} < -5^{\circ}\text{C}$).
- Frühlingstag: ein Tag, an dem die Maximumtemperatur bei oder über 15°C liegt ($T_{\max} \geq 15^{\circ}\text{C}$).

Die Untersuchung der jeweiligen Kategorien erfolgt jeweils für die Winter- und Sommersemester. Da bei der kategorischen Betrachtungsweise Ereignisse, welche nicht vorhergesagt und auch nicht beobachtet wurden, als sogenannte Negative Treffer (*correct negatives*; vgl. Kapitel 2.2) in die Verifikation mit einfließen können, macht aber noch eine zusätzliche zeitliche Beschränkung Sinn. So kommen beispielsweise Tropennächte oder heiße Tage in den Wintermonaten nur selten bis gar nicht vor, umgekehrt aber wäre ein

Frühlingstag in den Sommermonaten wahrscheinlich. Diese Beschränkung ist willkürlich gesetzt und entspricht dem Auftreten des jeweiligen Ereignisses in dem zu untersuchenden Datensatz:

- Frosttag: Oktober – April
- Eistag: November – Jänner
- Sommertag: April – Juni
- Tropentag: Mai – Juni
- Tropennacht: Mai – Juni
- Strenger Frost: Oktober – April
- Frühlingstag: November – März

4.3.1 Kategorische Verifikation des PETRUS

In erster Linie ist natürlich auch bei der kategorischen Betrachtungsweise die Genauigkeit der Prognosen (*accuracy*) von Bedeutung. Anders als bei der kontinuierlichen Verifikation in den vorangegangenen Kapiteln, bei der diese Genauigkeit quantitativ festgestellt wurde (wie hoch ist der Prognosefehler?), wird nun anhand der einzelnen Kategorien überprüft, in wie weit diese korrekt prognostiziert wurden. Ein erstes, recht anschauliches, Maß hierfür ist die Trefferrate (PC; vgl. Kapitel 2.2.1) und daraus resultierend die Trefferquote (P). Mit Werten stets über 0,90 bzw. 90% liefern beide Maße gute Ergebnisse (siehe TABELLE 4-18).

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
PC	0,94	0,93	0,98	0,98	0,97	1,00	0,94	0,96	1,00	0,94	1,00	0,95	0,96	1,00	0,98	0,95	0,93	0,92	0,94	1,00	0,99
P	93,9	93,2	97,8	98,1	97,3	100,0	94,4	96,0	100,0	93,5	99,8	95,1	96,4	100,0	98,2	95,0	93,4	92,5	94,4	100,0	98,8

TABELLE 4-18: Trefferrate (PC) und Trefferquote (P) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Der Wertebereich liegt zwischen Null und Eins (Hundert), mit Eins (Hundert) als perfekter Prognose.

Allerdings sind Trefferrate und Trefferquote allein nicht über zu interpretieren. Beispielsweise liefern die Ergebnisse der Kategorie „Tropennacht“ eine fast perfekte Trefferquote - in beiden Semestern. Fast jede Prognose hat also dem Anschein nach gestimmt. Man darf jedoch davon ausgehen, dass diese Werte lediglich durch die „nein“-Vorhersagen zustande kommen, da in unseren Breiten Tropennächte im Winter eher unwahrscheinlich und auch im Sommer mehr die Ausnahme denn die Regel sind. Das Interesse gilt nun also einzig den Fällen, in denen entweder eine Tropennacht prognostiziert

oder beobachtet - im besten Fall prognostiziert und beobachtet - wurde. Die nicht vorhergesagt/nicht beobachtet-Ereignisse (die sogenannten Negativen Treffer oder *correct negatives*) müssen also so gut es geht eliminiert werden um im Falle der Tropennacht auch brauchbare Ergebnisse zu bekommen. Anders ausgedrückt: um die Prognosegüte von seltenen Ereignissen (*rare event*) oder gar Extremereignissen feststellen zu können, benötigen man andere Qualitätsmaße, wie den Treffersatz (T; siehe Kapitel 2.2.3). Sofort wird ersichtlich, dass speziell die Tropennacht mit einer Trefferquote von **99%** in den Sommersemestern und **94%** in den Beobachtungsmonaten Mai und Juni beim Treffersatz nur mehr bescheiden abschneiden: mit **0,22** bzw. **0,24** sinkt die Anzahl der Treffer auf nur mehr rund 20% (siehe TABELLE 4-19). Mit Werten um die **0,80** sind Frostereignisse (Frosttag, strenger Frost, Eistag) noch vorne mit dabei, das Erkennen von Frühlings- und Sommertagen gelingt ebenso weitgehend zufriedenstellend mit einem Spitzenwert von **0,88** (Sommertag, SS), und die Prognosen für einen heißen Tag kommen auf knapp über **0,70**.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
T	0,82	0,83	0,78	0,79	0,77	1,00	0,80	0,81	---	0,86	---	0,88	0,71	---	0,72	0,86	0,74	0,81	0,22	---	0,24

TABELLE 4-19: Treffersatz (T) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich liegt zwischen Null und Eins, mit Eins als perfekter Prognose.

Neben der richtigen Prognose eines Ereignisses sind auch die falschen Prognosen von Bedeutung, da auch diese aussagekräftige Informationen enthalten können. Wird ein Ereignis gerne übersehen oder öfter nicht erkannt, handelt es sich um eine Überraschung oder ein *missed event*. Den anderen Fall, ein Ereignis öfter vorherzusagen als es eintritt, bezeichnet man als Falschen Alarm (*false alarm*). Eine erste Information über darüber liefert der *Frequency Bias Index* (FBI; siehe Kapitel 2.2.2; vgl. TABELLE 4-20).

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
FBI	1,03	1,03	1,04	1,07	1,08	---	1,00	1,00	---	0,90	0,00	0,90	0,94	---	0,94	0,63	0,95	0,96	1,00	---	1,00

TABELLE 4-20: Frequency Bias Index (FBI) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich liegt zwischen Null und $+\infty$, wobei die Ergebnisse keine Aussage über die Prognosegüte treffen, sondern lediglich, ob ein Ereignis öfter prognostiziert als beobachtet (>1) oder öfter beobachtet als prognostiziert (<1) wurde.

Dabei zeigt sich, dass Frostereignisse tendenziell einem *overforecasting* unterliegen, also öfter prognostiziert als beobachtet werden. Frühlings-, Sommer- und heiße Tage hingegen werden öfter beobachtet als prognostiziert (*underforecasted*). Auffällig dabei die Prognose des Frühlingstages, die speziell in den dafür interessanten Monaten November bis März lediglich auf **0,63** kommt. Die Diskrepanz mit den Werten der WS (**0,95**) bzw. der SS (**0,96**) ist aufgrund des höheren Auftretens von Tagen $\geq 15^{\circ}\text{C}$ leicht erklärt (in den WS sind auch die Monate September und Oktober enthalten!), nicht jedoch das *underforecasting*, welches per Definition nur durch eine häufigere Anzahl von *missed events* zustande kommen kann. Bei genauerer Betrachtung der Monate November bis März zeigt sich, dass gerade im November eine große Anzahl von *missed events* in dieser Kategorie zu finden ist, welches auf eine zu hohe Orientierung an das November-Klimamittel (Wien Hohe Warte, $7,5^{\circ}\text{C}$, 1971-2000) vermuten lässt. Eistage und Tropennächte weisen keine Abweichung auf, eine Aussage über die Prognosegüte lässt dies aber nicht zu.

Getätigte Prognosen lassen sich einerseits mit dem *Post Agreement* (PA; siehe Kapitel 2.2.4), aber auch mit der *False Alarm Ratio* (FAR; siehe Kapitel 2.2.6) untersuchen. Beide Maße geben Auskunft über den Anteil der Ereignisse wieder, die sowohl prognostiziert als auch beobachtet wurden (die sogenannten Treffer, *hits*; TABELLE 4-21).

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
PA	0,85	0,87	0,79	0,80	0,79	---	0,84	0,84	---	0,92	---	0,92	0,73	---	0,73	0,91	0,78	0,96	0,25	---	0,25
FAR	0,15	0,13	0,21	0,20	0,21	---	0,16	0,16	---	0,08	---	0,08	0,27	---	0,27	0,09	0,22	0,04	0,75	---	0,75

TABELLE 4-21: *Post Agreement* (PA) und Anteil des Falschen Alarms (FAR) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich des PA liegt zwischen Null und Eins, mit Eins als perfekter Prognose. Der Wertebereich der FAR liegt ebenfalls zwischen Null und Eins, die perfekte Prognose liegt aber bei Null.

Dabei zeigt sich, dass mit Ausnahme der Tropennacht, die Werte des PA durchaus im hohen Bereich liegen. Anders ausgedrückt: wenn die jeweilige Kategorie vorhergesagt wurde, so trifft diese mit einer Wahrscheinlichkeit zwischen 79% (**0,79** PA, Frost, SS) und 96% (**0,96** PA, Frühlingstag, SS) auch ein (die Differenzen auf die 100% geben den jeweiligen Anteil an Falschen Alarmen wieder). Frühlingstage wurden demnach in den Monaten November bis März zwar öfter beobachtet als prognostiziert (BIAS von **0,63**), wenn diese Kategorie jedoch vorhergesagt wurde, so liegt der Anteil der richtigen Prognosen bei 91% (**0,91** PA). Lediglich die Prognose von heißen Tagen liegt mit **0,73** PA etwas darunter – rund ein Viertel aller Vorhersagen (**0,27** FAR) sind somit ein Falscher Alarm. Deutlicher hingegen das Ergebnis der Prognosen von Tropennächten: mit einem PA von **0,25** und einer FAR von **0,75** tritt dieses Ereignis in vier Fällen (Prognosen) nur einmal ein. Hier muss aber auch die Sensibilität gegenüber der klimatologischen Häufigkeit dieses *events* angemerkt werden.

Neben den Informationen über fälschlicherweise prognostizierte Ereignisse (Falsche Alarmer), kann auch das Gegenpart, also nicht prognostizierte aber eingetroffene Ereignisse (Überraschungen oder *misses*), von Interesse sein. Die Entdeckungswahrscheinlichkeit (POD, siehe Kapitel 2.2.5) gibt darüber Auskunft. Dabei untersucht die sogenannte *Hit Rate* (die Entdeckungswahrscheinlichkeit der „ja“-Kategorie), wie oft ein beobachtetes *event* auch vorhergesagt wurde und analog die Entdeckungswahrscheinlichkeit der „nein“-Kategorie, wie oft ein nicht beobachtetes *event* auch nicht vorhergesagt wurde. Das Pendant zur *Hit Rate* ist die *Miss Rate*, also der Anteil an Ereignissen, die zwar beobachtet aber fälschlicherweise nicht vorhergesagt wurden.

Die Ergebnisse (vgl. TABELLE 4-22) zeigen einerseits eine geringfügig besseres Abschneiden im Vergleich zu den Falschen Alarmen bei Frostereignissen (zwischen **0,81** und **0,89** HR), andererseits eine geringfügig schlechtere bei Sommer- und heißen Tagen (zwischen **0,69** und **0,83** HR). Unverändert schneiden Eistage mit **0,84** HR und Tropennächte mit **0,25** HR ab. Bei den Frühlingstagen hingegen werden nur 57% (**0,57** HR, INT) der beobachteten Ereignisse auch prognostiziert, 43% (**0,43** MR, INT) hingegen nicht. Dieses Ergebnis deckt sich auch mit den Werten des BIAS. Die Diskrepanz zwischen WS und SS ist erneut mit der Monatsaufteilung zu erklären. Die Ergebnisse der „nein“-Kategorie (POD) liegen allesamt mit deutlich über 90% - was aufgrund der Einteilung der Kategorien in die speziellen klimatologisch-relevanten Tage auch nicht anders zu erwarten war (seltene Ereignisse produzieren standardmäßig eine große Anzahl an Negativen Treffern!).

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
HR	0,88	0,89	0,81	0,85	0,85	---	0,84	0,84	---	0,83	0,00	0,83	0,69	---	0,69	0,57	0,74	0,93	0,25	---	0,25
POD	0,96	0,95	0,99	0,99	0,98	1,00	0,97	0,98	1,00	0,97	1,00	0,98	0,98	1,00	0,99	0,99	0,97	0,92	0,97	1,00	0,99
MR	0,12	0,11	0,19	0,15	0,15	---	0,16	0,16	---	0,17	1,00	0,17	0,31	---	0,31	0,43	0,26	0,07	0,75	---	0,75

TABELLE 4-22: Die Entdeckungswahrscheinlichkeit der „ja“-Kategorie (HR) und der „nein“-Kategorie (POD) sowie die Miss Rate (MR) des PETRUS für die einzelnen Kategorien, das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich der HR und der POD liegt zwischen Null und Eins, mit Eins als perfekter Prognose. Der Wertebereich der MR liegt zwischen Null und Eins, mit Null als perfekter Prognose.

Während die *False Alarm Ratio* (FAR) das Verhältnis von Falschen Alarmen und Treffern beschreibt, untersucht die *False Alarm Rate* (FR, siehe Kapitel 2.2.7) das Verhältnis von Ereignissen, die nicht beobachtet und nicht vorhergesagt wurden (Negative Treffer, *correct negatives*) zu denen, die nicht beobachtet, sehr wohl aber vorhergesagt wurden (Falsche Alarmer). Es untersucht somit, wie oft ein nicht beobachtetes Ereignis fälschlicherweise vorhergesagt wurde und entspricht somit dem Pendant der POD der „nein“-Kategorie. Wie

schon POD liefert auch FR durchwegs gute Ergebnisse (vgl. TABELLE 4-23), die aufgrund der schon erwähnten Gründe allesamt keiner näheren Betrachtung bedürfen.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
FR	0,04	0,05	0,01	0,01	0,02	0,00	0,03	0,02	0,00	0,03	0,00	0,02	0,02	0,00	0,01	0,01	0,03	0,08	0,03	0,00	0,01

TABELLE 4-23: Die Entdeckungswahrscheinlichkeit des Falschen Alarms (FR) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Der Wertebereich der FR liegt zwischen Null und Eins, mit Null als perfekter Prognose.

Wie schon erwähnt können die Ergebnisse einiger Parameter eine Leistung vortäuschen, die de facto nicht existiert. Die Ursache liegt unter anderem in der Definition der Kategorien, die als solche eigentlich als selten und somit als Extremereignisse (*rare events*) anzusehen sind. Da seltene Ereignisse ganz dem Namen nach auch selten vorkommen, kann sich eine Prognose derselbigen als einfach erweisen, solange das jeweilige *event* nicht eintrifft. Von Interesse ist also nicht nur, wie oft ein Ereignis richtig prognostiziert wurde, sondern auch - und gerade ob - ein *beobachtetes* Ereignis richtig prognostiziert wurde. Diesen Umstand untersucht der Kritische Erfolgsindex (CSI; siehe Kapitel 2.2.8). Er entspricht der Trefferrate allerdings ohne Berücksichtigung der Negativen Treffern.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
CSI	0,77	0,79	0,67	0,70	0,69	---	0,72	0,72	---	0,78	0,00	0,78	0,55	---	0,55	0,54	0,61	0,90	0,14	---	0,14

TABELLE 4-24: Der Kritische Erfolgsindex (CSI) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich des CSI liegt zwischen Null und Eins, mit Eins als perfekter Prognose.

Die Ergebnisse bieten ein leicht differenzierteres Bild als bisher. Sie zeigen nämlich, dass das bisherige Abschneiden der einzelnen Kategorien maßgeblich durch eben diese Negativen Treffer beeinflusst wurde. Ohne diese, liegt der Prozentsatz an beobachteten Ereignissen, die korrekt prognostiziert wurden, in etwa zwischen 70 und 80% für Frosttage, Tage mit strengem Frost, Eis- und Sommertage. Für heiße Tage sinkt die Anzahl an richtigen Prognosen auf die Hälfte (**0,55**) und damit ein deutlicher Verlust im Vergleich zur bisherigen Trefferrate von **0,96** (bzw. **0,98** SS). Auch die Frühlingstage kommen im vordefinierten Zeitintervall von November bis März auf nur **0,54** – verglichen mit der Trefferrate von **0,95** ebenfalls eine Halbierung. Betrachtet man in diesem Fall die Winter- und Sommersemester, so steigt dieser Wert in den WS geringfügig auf **0,61** bzw. in den SS deutlich auf **0,90** (was in den Monaten mit einem täglichen Durchschnittsmaximum von >15°C ebenfalls keine Leistung darstellt). Die Tropennacht schließlich kommt gar nur auf **0,14**, nur eine von sieben

wurde demnach auch prognostiziert. Angesichts dieses Wertes stellt sich die Frage, ob hier überhaupt noch eine Leistung vorliegt, oder nicht doch beispielsweise eine Zufallsprognose mehr Sinn machen würde. Dies untersuchen die sogenannten *Skill Scores* (siehe Kapitel 2.1.4). Mit der Hanssen-Kuipers Diskriminante (TSS; siehe Kapitel 2.2.9) und dem Heidke Skill Score (HSS; siehe Kapitel 2.2.10) wollen wir abschließend diesen Umstand untersuchen. Beide *Skill Scores* zeigen dabei ähnliche Ergebnisse. Die Vorhersagegüte bei Frostereignissen sowie bei Sommertagen liegt bei rund **0,80** – eine deutliche Leistung gegenüber einer Zufallsprognose ist also gegeben. Bei der Prognose des heißen Tages sinkt dieser Wert auf etwas unter **0,70**, eine Leistung ist aber auch hier erkennbar. Bei Frühlingstagen unterscheiden sich die zwei *Skill Scores* etwas, zumindest in ihrem Beobachtungsintervall: TSS liefert **0,57** während HSS einen Wert von **0,68** gibt. Die Ursache hierfür liegt in der großen Anzahl von Überraschungen (*misses*) in dieser Kategorie, die beim TSS stärker zum Tragen kommen. In den WS, und deutlicher in den SS, erhöht sich automatisch wieder die Anzahl der Treffer (*hits*) und die *Skill Scores* steigen. Wie schon die vorangegangenen Maße vermuten haben lassen, sind die größten Prognose-Schwierigkeiten bei der Tropennacht zu finden. Mit **0,22** bzw. **0,24** liegt die Leistung hier nur knapp über der einer Zufallsprognose.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
TSS	0,84	0,84	0,80	0,84	0,83	---	0,80	0,81	---	0,81	0,00	0,82	0,67	---	0,68	0,57	0,70	0,85	0,22	---	0,24
HSS	0,83	0,83	0,79	0,81	0,80	---	0,80	0,81	---	0,83	0,00	0,85	0,69	---	0,70	0,68	0,72	0,82	0,22	---	0,24

TABELLE 4-25: Die Hanssen-Kuipers Diskriminante (TSS) und der Heidke Skill Score (HSS) des PETRUS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich der TSS und des HSS liegt zwischen minus Eins und Eins, mit Eins als perfekter Prognose und mit Null als no skill.

Zusammenfassend lässt sich also sagen, dass speziell Frostereignisse, also Frosttage, das Auftreten von strengem Frost oder Eistage, in dem untersuchten Datensatz gut erkannt und dementsprechend auch prognostiziert wurden. Selbst nach der Eliminierung der Negativen Treffer (*correct negatives*), welche bei seltenen Ereignissen traditionell hoch sind, liegt die Anzahl der richtigen Prognosen zwischen 70% und 80%; die Verbesserung gegenüber dem Zufall liegt bei rund 80%. In dieser Größenordnung sind auch die Ergebnisse der Kategorie „Sommertag“. Bei heißen Tagen differenziert sich das Gesamtbild etwas, die an sich guten Ergebnisse der allgemeinen Trefferrate von über 90% sinken ohne Beachtung der Negativen Treffer um etwa die Hälfte auf 55%. Die bei heißen Tagen vergleichsweise hohe Anzahl an Falschen Alarmen (*false alarms*) und Überraschungen (*misses*) kann ohne diese Negativen Treffer nicht mehr ausgeglichen werden. Die Leistung gegenüber dem Zufall ist mit 70% dennoch als brauchbar anzusehen. Selbiges gilt für die Kategorie des Frühlingstages in dem für uns relevanten Zeitintervall von November bis März – wenngleich hier der Wert des CSI (also der Trefferrate ohne Negative Treffer) mit 54% stark von der hohen Anzahl der

Überraschungen beeinflusst wird (MR von 0,43 und FAR von 0,09). Die Leistungsverbesserung gegenüber dem Zufall liegt ebenfalls bei rund 70% (in den SS hingegen „nur“ bei 80%). Die große Herausforderung stellt die Kategorie der Tropennächte dar. Von der Anzahl der beobachteten Fälle ein Extremereignis, wurde die ursprüngliche allgemeine Trefferrate von fast 100% (in den SS) nach Eliminierung der Negativen Treffer auf 14% deutlich relativiert und kann mit 24% auch nur eine schwache Leistungssteigerung gegenüber dem Zufall vorweisen. Hier beeinflusst die klimatologische Häufigkeit sowohl Prognostiker als auch Verifikation. An dieser Stelle sei nochmals darauf hingewiesen, dass die Untersuchung mit kategorischen Maßen keinerlei Auskunft über die quantitative Genauigkeit der Prognosen liefert. Es zeigt lediglich den Zusammenhang von richtigen und falschen Prognosen.

4.3.2 Kategorische Verifikation des MOS

Nach der Verifikation der studentischen Prognosen im vorangegangenen Kapitel wollen wir nun abschließend noch die des MOS näher betrachten. Trefferrate (PC) und –quote (P; vgl. Kapitel 2.2.1, siehe TABELLE 4-26) zeigen auf hohem Niveau ähnliche Werte wie PETRUS, allerdings schneidet das MOS bei dem für Extremereignissen wichtigeren Treffersatz (T; vgl. Kapitel 2.2.3) gleich in mehreren Punkten besser ab: So ist das Erkennen von Frost in den SS mit **0,97** nahezu perfekt, PETRUS kommt hier auf „nur“ **0,78**. Des Weiteren ist auch beim generellen Erkennen von strengem Frost und in der Kategorie der heißen Tage eine Verbesserung von rund 10% gegeben. Spannend auch das Erkennen von Tropennächten: mit **0,49** erkennt zwar auch das MOS nur jedes zweite Ereignis, dies ist dennoch eine Verdoppelung der Leistung von PETRUS (**0,22**).

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
PC	0,92	0,91	0,97	0,97	0,96	1,00	0,88	0,92	1,00	0,95	1,00	0,96	0,97	1,00	0,99	0,97	0,94	0,92	0,98	1,00	1,00
P	92,0	90,9	97,5	96,8	95,8	99,5	88,2	91,6	100,0	94,7	99,8	95,9	97,4	100,0	98,6	96,8	93,9	92,4	97,8	100,0	99,5
T	0,83	0,81	0,97	0,87	0,87	---	0,84	0,87	---	0,87	---	0,88	0,85	---	0,86	0,84	0,75	0,85	0,49	---	0,50

TABELLE 4-26: Trefferrate (PC) und Trefferquote (P), sowie Treffersatz (T) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Der Wertebereich liegt zwischen Null und Eins (Hundert), mit Eins (Hundert) als perfekter Prognose.

Ein Blick auf den BIAS (vgl. Kapitel 2.2.2, siehe TABELLE 4-27) zeigt allerdings, dass das MOS, mit Ausnahme des Frühlingstages und der Tropennacht, in allen Kategorien einem deutlichen *underforecasting* unterliegt. Negativer Spitzenreiter hierbei ist die Kategorie des Eistages mit **0,41**; speziell der Eistag wurde demnach öfter beobachtet als prognostiziert, was

wiederum eine höhere Anzahl an *missed events* (Überraschungen) impliziert. Nähere Auskunft darüber gibt die *Hit Rate* (Entdeckungswahrscheinlichkeit, HR; vgl. Kapitel 2.2.5, siehe TABELLE 4-29) und dessen Pendant die Miss Rate (MR). Anhand dieser Ergebnisse zeigt sich deutlich, dass das Zustandekommen der guten Ergebnisse beim Treffersatz nur durch die geringe Anzahl an Falschen Alarmen zustande kommen kann (siehe dazu auch die Ergebnisse des *Post Agreements*, PA, und des Anteil des Falschen Alarms, FAR; TABELLE 4-28). So wurde beispielsweise bei Frostereignissen in den SS (Treffersatz von **0,97!**) mit einer HR und MR von **0,50** lediglich jedes zweite beobachtete *event* richtig vorhergesagt (zum Vergleich PETRUS: HR **0,81** und MR **0,19**). Selbiges gilt auch für die anderen Kategorien, die beim Treffersatz gut abgeschnitten haben (strenger Frost und heißer Tag), lediglich die Verbesserung in der Kategorie „Tropennacht“ bestätigt sich.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
FBI	0,79	0,84	0,50	0,54	0,57	0,00	0,41	0,41	---	0,98	0,00	0,98	0,67	---	0,67	0,90	1,07	1,05	1,00	---	1,00

TABELLE 4-27: *Frequency Bias Index (FBI) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich liegt zwischen Null und $+\infty$, wobei die Ergebnisse keine Aussage über die Prognosegüte treffen, sondern lediglich, ob ein Ereignis öfter prognostiziert als beobachtet (>1) oder öfter beobachtet als prognostiziert (<1) wurde.*

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
PA	0,91	0,90	1,00	0,90	0,90	---	0,96	0,96	---	0,91	---	0,91	0,88	---	0,88	0,86	0,78	0,92	0,50	---	0,50
FAR	0,09	0,10	0,00	0,10	0,10	---	0,04	0,04	---	0,09	---	0,09	0,13	---	0,13	0,14	0,22	0,08	0,50	---	0,50

TABELLE 4-28: *Post Agreement (PA) und Anteil des Falschen Alarms (FAR) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich des PA liegt zwischen Null und Eins, mit Eins als perfekter Prognose. Der Wertebereich der FAR liegt ebenfalls zwischen Null und Eins, die perfekte Prognose liegt aber bei Null.*

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
HR	0,72	0,76	0,50	0,49	0,51	0,00	0,39	0,39	---	0,89	0,00	0,89	0,58	---	0,58	0,78	0,83	0,97	0,50	---	0,50
POD	0,98	0,97	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,97	1,00	0,98	1,00	1,00	1,00	0,99	0,96	0,80	0,99	1,00	1,00
MR	0,28	0,24	0,50	0,51	0,49	1,00	0,61	0,61	---	0,11	1,00	0,11	0,42	---	0,42	0,23	0,17	0,03	0,50	---	0,50

TABELLE 4-29: *Die Entdeckungswahrscheinlichkeit der „ja“-Kategorie (HR) und der „nein“-Kategorie (POD) sowie die Miss Rate (MR) des MOS für die einzelnen Kategorien, das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich der HR und der POD liegt zwischen Null und Eins, mit Eins als perfekter Prognose. Der Wertebereich der MR liegt zwischen Null und Eins, mit Null als perfekter Prognose.*

Durch die geringe Anzahl an Falschen Alarmen schneidet das MOS bei der Entdeckungswahrscheinlichkeit des Falschen Alarms (FR; vgl. Kapitel 2.2.7, siehe TABELLE

4-30), das das Verhältnis zwischen den Negativen Treffern (*correct negatives*) und den Falschen Alarmen (*false alarms*) untersucht, ausgesprochen gut ab. Allerdings liefert auch PETRUS hier gute Ergebnisse.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
FR	0,02	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,02	0,00	0,00	0,00	0,01	0,04	0,20	0,01	0,00	0,00

TABELLE 4-30: Die Entdeckungswahrscheinlichkeit des Falschen Alarms (FR) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Der Wertebereich der FR liegt zwischen Null und Eins, mit Null als perfekter Prognose.

Abschließend noch die Ergebnisse des CSI (*Critical Success Index*; vgl. Kapitel 2.2.8, siehe TABELLE 4-31), der die Negativen Treffer unberücksichtigt lässt, sowie der *Skill Scores* nach Hanssen-Kuipers (TSS; vgl. Kapitel 2.2.9, siehe TABELLE 4-32) und Heidke (HSS; vgl. Kapitel 2.2.10), die die Verbesserung und damit die Leistungssteigerung gegenüber dem Zufall angeben. Im Vergleich mit den Ergebnissen des PETRUS zeigt sich auch hier wieder das schlechtere Abschneiden bei Frostereignissen, wobei die Kategorie des Eistages mit einem CSI von **0,38** und einem TSS von **0,39** besonders schlecht abschneidet. In der Kategorie der Sommer- und der heißen Tage kann die Leistung des MOS mit der des PETRUS mithalten. Bei der Prognose des Frühlingstages und der der Tropennacht hingegen übertrifft die Leistung des MOS die des PETRUS.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
CSI	0,67	0,70	0,50	0,46	0,49	0,00	0,38	0,38	---	0,81	0,00	0,81	0,54	---	0,54	0,69	0,67	0,90	0,33	---	0,33

TABELLE 4-31: Der Kritische Erfolgsindex (CSI) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Die kursiven Werte sind lediglich der Vollständigkeit angeführt und spiegeln keinerlei Leistung wider. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich des CSI liegt zwischen Null und Eins, mit Eins als perfekter Prognose.

	FROST			STRENGER FROST			EISTAG			SOMMERTAG			HEISSER TAG			FRÜHLINGSTAG			TROPENNACHT		
	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS	INT	WS	SS
TSS	0,70	0,72	0,50	0,48	0,51	0,00	0,39	0,39	---	0,85	0,00	0,86	0,58	---	0,58	0,76	0,79	0,77	0,49	---	0,50
HSS	0,75	0,76	0,66	0,62	0,63	0,00	0,50	0,52	---	0,86	0,00	0,87	0,69	---	0,69	0,80	0,77	0,80	0,49	---	0,50

TABELLE 4-32: Die Hanssen-Kuipers Diskriminante (TSS) und der Heidke Skill Score (HSS) des MOS für die einzelnen Kategorien, sowie für das jeweilige Beobachtungsintervall, das Winter- und das Sommersemester. Fehlende Werte resultieren aus einem undefinierten Divisor. Der Wertebereich der TSS und des HSS liegt zwischen minus Eins und Eins, mit Eins als perfekter Prognose und mit Null als no skill.

4.4 Zusammenfassung der Ergebnisse

Die Ergebnisse dieser Arbeit zeigen und bestätigen in erster Linie das, was bereits in Kapitel 2 angedeutet und von mehreren Autoren zu diesem Thema (u.a. MURPHY A., STANSKI H.R.) erläutert wurde: die Ergebnisse einzelner Verifikationsmaße allein sind nur von beschränkter Aussagekraft. So mag die Persistenzprognose zwar eine geringe mittlere Abweichung haben, im Hinblick auf deren absoluten Fehler und die daraus resultierende Fehlervarianz ist dies jedoch mit Vorsicht zu genießen. Oder das beeindruckende Abschneiden des MOS bei Frostereignissen in den SS: mit einem Treffersatz von 97% wäre dieses Ergebnis knapp an der Perfektion, verliert jedoch bei Berücksichtigung anderer Qualitätsmaße, wie der *Hit Rate*, deutlich an Glanz, zeigt es doch, dass mit einer HR von 0,50 nur mehr jedes zweite beobachtete Ereignis auch ein Treffer war. Eine einzelne Prozentzahl allein sagt also zumeist wenig aus.

Die Ergebnisse zeigen, dass Prognosen der Studenten (und demnach auch des PETRUS) den klassischen Referenzprognosen (PERS, KLIMA, KLITA) deutlich überlegen sind. Die Gründe hierfür liegen auf der Hand: Eine gewöhnliche Persistenzprognose mag zwar bei weitgehend stationären Wetterlagen (zum Beispiel bei Hochdruckeinfluss oder bei Omega-Wetterlagen) ihre Stärke voll ausspielen können (da bei solche Wetterlagen über Tage oder auch Wochen hinweg mit einem ähnlichen Wetter zu rechnen ist; hohe *reliability*), hat aber bei häufigen Wetterwechseln das Nachsehen, da die Persistenz solche Änderungen erst mit Verspätung berücksichtigen kann. Die Klimaprognose hingegen hat das Problem, Extremereignisse schlecht oder überhaupt nicht erkennen zu können (keine *sharpness*), wobei ein Ereignis dabei gar nicht so „extrem“ sein muss: Praktisch alles, was vom Klima abweicht, kann als mehr oder weniger „extrem“ angesehen werden. Ein Fehler kann so rasch sehr groß werden.

Zwei Beispiele sollen dies im Folgenden veranschaulichen: Für die Zuverlässigkeit einer Prognose (*reliability*) betrachten wir den Witterungsverlauf vom 13.-15. Mai 2003 [ZAMG, 2003; MONATSRÜCKBLICK MAI 2003]: Die Tage zuvor bestimmt Hochdruckeinfluss mit Warmluftadvektion aus dem westlichen Mittelmeerraum das Wetter in Österreich mit Höchsttemperaturen bis zu 25 Grad. Bevor am 13. Mai eine kräftige Kaltfront mit schweren Gewittern und heftigem Regen den Osten Österreichs überquert, wird in Wien (Hohe Warte) noch einmal ein Tagesmaximum (T_{\max}) von 24,5°C registriert. Die Tage darauf liegt Österreich im Einflussbereich eines Höhentrog, mit der deutlich kühlere Luftmassen in den Alpenraum gelangen. So liegt der Höchstwert (T_{\max}) am 14. Mai in Wien bei nur mehr 12,9°C. Dieser Temperatursturz wird von den studentischen Prognosen und demnach auch von PETRUS gut wiedergegeben (siehe TABELLE 4-33), ungeachtet des doch größeren Vorhersagefehlers (mittlerer Fehler, ME; siehe Kapitel 2.1.1). Die Persistenz hingegen hat bei solch einem Wettersturz das Nachsehen, kann diese den Wetterumschwung erst um einen

Tag verspätet berücksichtigen: ist die PERS-Vorhersage am 12. Mai auf Grund der zuvor andauernden stabilen Wetterlage noch ganz brauchbar (Vorhersagefehler von unter einem Grad), so mehr als verzehnfacht sich dieser Fehler tags darauf. Der daraus resultierende mittlere quadratische Fehler (MSE; siehe Kapitel 2.1.3) zeigt die angesprochene Schwäche dieser Vorhersagemethode deutlich. Auch die Klimaprognosen (KLIMA und KLITA) sind in der Regel mit solchen Wetterstürzen heillos überfordert, der Fehler liegt aber in diesem Fall unter dem der PERS. Die MOS-Prognose fehlt für diesen Zeitraum, sie kann aber im Allgemeinen mit Wetteränderungen deutlich besser umgehen als PERS, KLIMA und KLITA.

Das zweite Beispiel soll die Schärfe einer Prognose (*sharpness*) veranschaulichen: Ende Jänner 2002 gelangen mit einer stürmischen Westströmung ausgesprochen milde Luftmassen nach Österreich, mit sehr warmen Nächten und Tagesmaxima bis teils über 20 Grad [ZAMG, 2002; MONATSRÜCKBLICK JÄNNER 2002]. Am 29. Jänner 2002 wurde dabei in Wien (Hohe Warte) ein Höchstwert (T_{max}) von 18,7°C gemessen (siehe TABELLE 4-34). Alle Vorhersagen haben dieses Ereignis unterschätzt, wobei die PERS-Prognose (die dem PETRUS gleicht) erstaunlich gut abgeschnitten hat (da die stürmische Westwetterlage auch in den Tagen vor dem 29. Jänner dominant war). Das MOS ist geringfügig schlechter, die deutlichste Abweichung haben allerdings die Klimaprognosen, da sowohl KLIMA als auch KLITA den Monat Jänner standardmäßig als einen „kalten“ Monat sehen. Mit rund 16 Grad Abweichung (mittlerer Fehler, ME) und dem daraus resultierenden mittleren quadratischen Fehler (MSE) entspricht diese KLIMA-Prognose der schlechtesten Prognose im gesamten, in dieser Arbeit verwendeten, 10-jährigen Datensatz.

Datum	Prognose	FC [°C]	OBS [°C]	ME	MAE	MSE
12.05.2003	PETRUS	23,1	24,5	-1,4	1,4	2,0
	PERS	25,3	24,5	0,8	0,8	0,6
	KLIMA	20,5	24,5	-4,0	4,0	16,0
	KLITA	20,5	24,5	-4,0	4,0	16,0
13.05.2003	PETRUS	14,6	12,9	1,7	1,7	2,9
	PERS	24,5	12,9	11,6	11,6	134,6
	KLIMA	20,5	12,9	7,6	7,6	57,8
	KLITA	21,3	12,9	8,4	8,4	70,6
14.05.2003	PETRUS	15,1	15,4	-0,3	0,3	0,1
	PERS	12,9	15,4	-2,5	2,5	6,3
	KLIMA	20,5	15,4	5,1	5,1	26,0
	KLITA	21,1	15,4	5,7	5,7	32,5

TABELLE 4-33: T_{max} - Prognosen und deren Fehler vom 12.-14. Mai 2003

29.01.2002	FC [°C]	OBS [°C]	ME	MAE	MSE
PETRUS	15,0	18,7	-3,7	3,7	13,7
PERS	15,0	18,7	-3,7	3,7	13,7
KLIMA	2,9	18,7	-15,8	15,8	249,6
KLITA	4,5	18,7	-14,2	14,2	201,6
MOS	14,6	18,7	-4,1	4,1	16,8

TABELLE 4-34: T_{max} - Prognosen und deren Fehler vom 29. Jänner 2002

Im Gegensatz zur Persistenz und den beiden Klimaprognosen erweist sich der Vergleich zwischen den studentischen Prognosen (bzw. PETRUS) und der statistisch-automatisierten Prognose MOS weitaus spannender, denn hier zeigt sich deutlich, dass die Leistung an sich und die Verbesserung der menschlichen Leistung gegenüber diesen MOS-Prognosen stark schwankt und in einigen Fällen die MOS-Prognosen denen der Studenten (knapp aber doch) überlegen sind.

PETRUS liefert für alle Parameter (T_{min} , T_{max} , T_6 und T_{12}) im Mittel zu tiefe Prognosen, während das MOS im Mittel zu hoch liegt. Diesen Umstand in Zukunft zu nutzen, könnte die Prognosequalität deutlich steigern. Generell haben alle Parameter mit einer gemittelten absoluten Abweichung von $\pm 1,35^\circ$ bei PETRUS und von $\pm 1,60^\circ$ beim MOS das Ziel der subjektiven Verifikation aus Kapitel 4.1, den Fehler auf $\pm 1^\circ$ zu beschränken, klar verfehlt. Auch wenn die studentische Leistung deutlich ist, so zeigt es klar, dass das MOS in Bezug auf die Vorhersagequalität keineswegs zu unterschätzen ist. So zeigte sich bei der Aufspaltung der einzelnen Monate, dass das MOS in Monat Juni bei allen Parametern eine höhere Prognosegüte aufweist, als die Studenten. Klar geschlagen geben muss sich hingegen das MOS in der kalten Jahreszeit, praktisch alle Ergebnisse liegen hier teils deutlich unter denen der Studenten.

Bezogen auf die einzelnen Parameter hat die Minimumtemperatur die höchste Vorhersagegüte – sowohl bei PETRUS als auch beim MOS. Hier wird gerade die Nähe zwischen Prognose- und Beobachtungszeitpunkt eine maßgebliche Rolle spielen, zumindest auf menschlicher Seite. Synoptisch kann man ja die Minimumtemperatur der kommenden Nacht bereits im Laufe des Nachmittags allein anhand der aktuellen Beobachtungen grob abschätzen und dies anhand aktueller Wetterinformationen weiter verfeinern. Die größten Unsicherheiten liefert hingegen die 12 UTC-Temperatur – wieder sowohl bei PETRUS als auch beim MOS. Als Prognose für einen Zeitpunkt unterliegt sie einerseits besonderen Schwierigkeiten gegenüber den anderen Parametern, andererseits genießt sie auch nicht die Hilfestellung, die etwa die 6 UTC-Temperatur genießt (dem Zusammenhang zum Temperaturminimum zum Beispiel). Mit einem absoluten Fehler von $\pm 1,5^\circ$ (PETRUS) und

1,8° (MOS) und einer Fehlervarianz von $>2^\circ$ liefert die 12 UTC-Temperatur die schlechtesten Ergebnisse.

Bei der kategorischen Verifikation bestätigen sich nochmals die teils deutlichen Schwierigkeiten, die das MOS in Bezug auf Frostereignisse (vornehmliches Auftreten in den WS) hat. Wobei hier deutlich wird, dass das MOS recht konservativ (Klimatologie-ähnlich) agiert: Einer großen Anzahl an Überraschungen (*misses*) steht eine nur geringe Anzahl an Falschen Alarmen (*false alarms*) gegenüber. Das MOS neigt demnach, Frostereignisse nicht nur zu unterschätzen (niedrige BIAS-Werte!; nicht zu verwechseln mit der allgemeinen Abweichung, bei dem das MOS zu positiven BIAS-Werten neigt), sondern diese häufig auch nicht zu erkennen (niedrige HR, hohe MR). Eine Leistungssteigerung gegenüber dem Zufall ist aber mit 50%-80% (je nach Parameter) gegeben. Bei Frühlingstagen hingegen liegt das Erkennen der solchen in dem dafür relevanten Zeitraum von November bis März über der studentischen Leistung. Hier agieren wiederum die Studenten als zu konservativ, eine Beeinflussung durch die Klimatologie ist wahrscheinlich. In der Kategorie der Tropennacht hat ebenfalls das MOS die Nase vorn, da das Auftreten sich aber auf Einzelereignisse beschränkt, ist eine Aussage hier nur bedingt möglich.

Zusammenfassend kann also die Anfangsfrage „Mensch oder Maschine“ in dieser Arbeit mit „Mensch“ beantwortet werden, wenngleich die Ergebnisse des MOS (= der Maschine) teils beachtlich sind. Diese hohe Qualität mag mit ein Grund sein, warum diese statistisch-automatisierte Vorhersage aus dem operationellen Vorhersagebetrieb nicht mehr weg zu denken ist.

4.5 Rückblick und Ausblick

Der in dieser Arbeit verwendete, 10-jährige Datensatz basiert auf den Vorhersagen der Studenten des Instituts für Meteorologie und Geodynamik (IMGW) der Universität Wien (siehe Kapitel 3) – und hat somit einen großen „Nachteil“: es liegen in der vorlesungsfreien Zeit keine Prognosen vor. Dadurch fehlen einerseits der wichtige Wintermonat Februar, andererseits aber auch die Sommermonate Juli und August – beides wäre in Bezug auf Extremwerte interessant: sowohl die Anzahl der Eis- (Tage mit $T_{\max} < 0^\circ\text{C}$) als auch die der Tropentage ($T_{\max} \geq 30^\circ\text{C}$) und –nächte ($T_{\min} \geq 20^\circ\text{C}$) wäre signifikant höher und die Ergebnisse dahingehend aussagekräftiger. So fällt man auch um die Möglichkeit, das Abschneiden des MOS als Vergleichsprognose in Bezug auf Tropentage als Gegen-Extremum zu den Frostereignissen, bei denen das MOS standardmäßig zu hoch prognostiziert (also überschätzt), zu verifizieren und damit zu zeigen, dass das MOS bei höheren Temperaturen zur Unterschätzung neigt.

Da das Vorhersageverfahren des MOS ständig weiterentwickelt und verfeinert wird, soll der Nutzen dieser Arbeit (Ergebnisse zusammengefasst in Kapitel 4.4) in erster Linie dem Prognoseverhalten der Studenten dienen. So könnte der Umstand, dass PETRUS im Mittel - unabhängig des Vorhersageparameters - stets zu niedrig prognostiziert, in Zukunft die Prognosegüte steigern – eine gute Hilfe hierfür liefern auch die Ergebnisse der einzelnen Monate des jeweiligen Parameters (siehe die jeweiligen Kapitel). In Bezug auf das MOS soll diese Arbeit eine (vorläufige) Interpretationshilfe stellen – so kann es zum Beispiel von Vorteil sein, zu wissen, dass Frostereignisse egal welcher Art (Frosttage, Eistage, etc.) standardmäßig eher überschätzt werden – auch wenn ein Ergebnis in der anderen Richtung (bei höheren Temperaturen), aufgrund der zu Beginn dieses Kapitels erwähnten Ursache, ausbleibt. Für die anderen Vergleichsprognosen (PERS, KLIMA, KLITA) gilt: wenn man sich deren Stärken, ungeachtet dessen, dass die Schwächen überwiegen, bewusst ist, können auch diese Referenzprognosen, richtig eingesetzt, Hilfe leisten – was de facto im operationellen Vorhersagebetrieb umgesetzt wird: eine Analyse des Vortages und somit das Vorwissen über die vergangene Wetterlage und des Wetterablaufs sowie Kenntnisse über die klimatologischen Möglichkeiten sind zumeist Voraussetzung für brauchbare Prognosen.

Das Hauptaugenmerk dieser Arbeit gilt in erster Linie – neben den einführenden Worten zur Verifikation und deren Anwendung – der Untersuchung des Parameters Temperatur (Minimum, Maximum, 6 UTC und 12 UTC-Temperatur). Selbstverständlich besteht eine Prognose aus weit mehr Vorhersageelementen, deren Abhandlung und Untersuchung den Rahmen dieser Arbeit sprengen würde. Sie soll somit auch als Ausgangslage für die Untersuchung der anderen, noch offenen, Parameter dienen, wie dem Wetterzustand oder der Niederschlagsmenge. Selbstverständlich sind hierfür nicht mehr alle der in dieser Arbeit vorgestellten Verifikationsparameter zu gebrauchen – so macht es zum Beispiel keinen Sinn, kontinuierliche Gütemaße (siehe Kapitel 2.1) für Prognosen des Wetterzustandes anzuwenden, für andere Parameter (wie dem Wind) müssten diese modifiziert werden. Beim Niederschlag hingegen wären neben der klassischen ja/nein Betrachtung (2x2 Kontingenztafel) auch eine mehrdimensionale Herangehensweise sinnvoll (siehe als Beispiel Tabelle 2-3). Hierfür die geeigneten Verifikationsmaße und Techniken zu finden und zu definieren, und damit weitere Parameter zu untersuchen, wäre eine hilfreiche Fortsetzung dieser Arbeit.

Anhang

Kapitel 4.3 zeigt die Ergebnisse einer kategorischen Verifikation von Temperaturprognosen der Studenten des IMGW an der Universität Wien. Für die Kategorien wurden Eigenschaften gewählt, die unter anderem in der Klimastatistik, aber auch in anderen Bereichen (wie der Wirtschaft) eine Rolle spielen. Diese Einteilung ist subjektiv erfolgt, selbstverständlich sind auch andere Definitionen dieser Kategorien zulässig. Im Anhang finden sich nun die (mehrdimensionalen) Kontingenztabelle der Jahre 2000-2010 für die die Minimumtemperatur (T_{\min} ; TABELLE 4-35) und die Maximumtemperatur (T_{\max} ; TABELLE 4-36).

T_{min}		VORHERSAGE [°C]																																						
		<-10	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	>20						
<-10,5	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-10,5≤-9,6	3	7	3	7	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-9,5≤-8,6	4	9	8	4	2	3	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-8,5≤-7,6	10	20	17	9	10	8	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-7,5≤-6,6	0	0	4	14	15	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-6,5≤-5,6	0	0	2	10	20	13	10	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-5,5≤-4,6	0	0	1	1	6	17	26	29	12	12	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-4,5≤-3,6	0	0	0	4	11	10	45	47	17	13	2	4	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-3,5≤-2,6	0	0	0	2	8	16	34	47	38	25	2	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-2,5≤-1,6	0	0	0	0	0	3	15	29	32	37	18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-1,5≤-0,6	0	0	1	4	8	2	6	10	40	99	114	42	9	11	7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-0,5≤0,4	0	0	0	0	0	2	4	9	13	45	81	106	73	37	17	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0,5≤1,4	0	0	0	0	0	0	0	2	9	15	37	107	89	74	33	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,5≤2,4	0	0	0	0	0	0	0	0	0	7	44	73	98	122	73	50	10	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,5≤3,4	0	0	0	0	0	0	0	0	0	0	4	28	63	113	100	49	39	10	4	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,5≤4,4	0	0	0	0	0	0	0	0	0	0	0	0	3	24	70	165	159	89	52	27	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4,5≤5,4	0	0	0	0	0	0	0	0	0	0	0	0	1	3	32	99	117	109	104	47	24	9	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5,5≤6,4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	23	51	105	144	65	22	12	7	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
6,5≤7,4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	7	38	67	140	113	88	41	18	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7,5≤8,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	25	25	33	91	145	145	93	31	15	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8,5≤9,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	16	41	70	118	136	99	29	21	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9,5≤10,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	9	38	47	98	107	52	22	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10,5≤11,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	19	41	66	85	89	69	16	12	10	5	1	0	0	0	0	0	0	0	0	0	0	0
11,5≤12,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	19	48	45	61	59	26	4	0	0	0	0	0	0	0	0	0	0	0	0	0
12,5≤13,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	18	18	54	40	37	13	12	2	0	0	0	0	0	0	0	0	0	
13,5≤14,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6	25	48	50	30	11	3	1	0	0	0	0	0	0	0	0	
14,5≤15,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	6	7	14	23	31	40	19	13	2	2	0	0	0	0	0		
15,5≤16,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	7	12	24	21	7	5	6	0	0	0	0	0	0		
16,5≤17,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	2	2	15	14	23	28	10	1	0	0	0	0	0		
17,5≤18,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	6	13	13	18	21	17	4	0	0	0	0	0	0	
18,5≤19,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	10	13	23	8	2	6	0	0	0		
19,5≤20,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
>20,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		

TABELLE 4-35: mehrdimensionale Kontingenztafel der Minimumtemperatur; Prognosen der Studenten des IMGW der Jahre 2000-2010

Abbildungsverzeichnis

ABBILDUNG 2-1: BEISPIEL EINES STREUDIAGRAMMS NACH [STANSKI ET AL., 1989].....	21
ABBILDUNG 4-1: STREUDIAGRAMM DER STUDENTISCHEN T_{MIN} -PROGNOSEN MIT TRENDLINIE UND KORRELATIONSKOEFFIZIENT ...	33
ABBILDUNG 4-2: STREUDIAGRAMM DER T_{MIN} -PROGNOSEN VON PETRUS, PERS UND KLITA MIT KORRELATIONSKOEFFIZIENT	33
ABBILDUNG 4-3: STREUDIAGRAMM DER T_{MIN} -PROGNOSEN VON PETRUS UND MOS MIT KORRELATIONSKOEFFIZIENTEN	35
ABBILDUNG 4-4: DIFFERENZPLOT T_{MIN} ; AUFGETRAGEN SIND DIE VORHERSAGEFEHLER VS. FC VON PETRUS & MOS.	36
ABBILDUNG 4-5: DIFFERENZPLOT T_{MIN} ; AUFGETRAGEN SIND DIE VORHERSAGEFEHLER VS. FC VON MOS, PERS & KLITA.	37
ABBILDUNG 4-6: MITTLERER FEHLER (ME) DER MINIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	39
ABBILDUNG 4-7: MITTLERER ABSOLUTER FEHLER (MAE) DER MINIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	40
ABBILDUNG 4-8: WURZEL DES MITTLEREN QUADRATISCHEN FEHLER (RSME) DER MINIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.	40
ABBILDUNG 4-9: STREUDIAGRAMM DER STUDENTISCHEN T_6 -PROGNOSEN MIT TRENDLINIE UND KORRELATIONSKOEFFIZIENT	42
ABBILDUNG 4-10: STREUDIAGRAMM DER T_6 -PROGNOSEN VON PETRUS UND MOS MIT KORRELATIONSKOEFFIZIENT	43
ABBILDUNG 4-11: DIFFERENZPLOT DER T_6 -PROGNOSEN; AUFGETRAGEN SIND DIE VORHERSAGEFEHLER VS. FC VON PETRUS & MOS.	44
ABBILDUNG 4-12: STREUDIAGRAMM T_6 DER PROGNOSEN VON PETRUS UND PERS MIT KORRELATIONSKOEFFIZIENT R.	44
ABBILDUNG 4-13: MITTLERER FEHLER (ME) DER 6 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.	46
ABBILDUNG 4-14: MITTLERER ABSOLUTER FEHLER (MAE) DER 6 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.	47
ABBILDUNG 4-15: WURZEL DES MITTLEREN QUADRATISCHEN FEHLER (RSME) DER 6 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.	48
ABBILDUNG 4-16: STREUDIAGRAMM DER STUDENTISCHEN T_{12} -PROGNOSEN MIT TRENDLINIE UND KORRELATIONSKOEFFIZIENT	50
ABBILDUNG 4-17: DIFFERENZPLOT DER T_{12} -PROGNOSEN	50
ABBILDUNG 4-18: STREUDIAGRAMM DER T_{12} -PROGNOSEN VON PETRUS UND MOS MIT KORRELATIONSKOEFFIZIENT R.	51
ABBILDUNG 4-19: MITTLERER FEHLER (ME) DER 12 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	53
ABBILDUNG 4-20: MITTLERER ABSOLUTER FEHLER (MAE) DER 12 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	54
ABBILDUNG 4-21: WURZEL DES MITTLEREN QUADRATISCHEN FEHLERS (RSME) DER 12 UTC-TEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	55
ABBILDUNG 4-22: STREUDIAGRAMM DER STUDENTISCHEN T_{MAX} -PROGNOSEN MIT TRENDLINIE UND KORRELATIONSKOEFFIZIENT	57
ABBILDUNG 4-23: STREUDIAGRAMM DER T_{MAX} -PROGNOSEN VON PETRUS UND MOS MIT KORRELATIONSKOEFFIZIENTEN	58
ABBILDUNG 4-24: DIFFERENZPLOT DER T_{MAX} -PROGNOSEN.....	58
ABBILDUNG 4-25: STREUDIAGRAMM DER T_{MAX} -PROGNOSEN VON PETRUS, PERS UND KLITA MIT KORRELATIONSKOEFFIZIENT	59
ABBILDUNG 4-26: MITTLERER FEHLER (ME) DER MAXIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTET AUF DIE EINZELNEN VORHERSAGEMONATE.....	62
ABBILDUNG 4-27: MITTLERER ABSOLUTER FEHLER (MAE) DER MAXIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTEN AUF DIE EINZELNEN VORHERSAGEMONATE.....	62
ABBILDUNG 4-28: WURZEL DES MITTLEREN ABSOLUTEN FEHLERS (RMSE) DER MAXIMUMTEMPERATUR VON PETRUS & MOS, AUFGESPALTEN AUF DIE EINZELNEN VORHERSAGEMONATE.....	63

Tabellenverzeichnis

TABELLE 2-1: DARSTELLUNG DER EINZELNEN EVENTS IN EINER 2X2 KONTINGENZTABELLE	12
TABELLE 2-2: ALLGEMEINE FORM EINER 2X2 KONTINGENZTABELLE	13
TABELLE 2-3: MEHRFACH KATEGORIALE KONTINGENZTABELLE AM BEISPIEL NIEDERSCHLAG	13
TABELLE 2-4: AUFLÖSUNG EINER MEHRFACH KATEGORIALEN KONTINGENZTABELLE	13
TABELLE 4-1: ANZAHL (N) UND PROZENTSATZ (%) DER „RICHTIGEN“ PROGNOSEN	31
TABELLE 4-2: MITTLERER FEHLER (ME) DER T_{MIN} -PROGNOSEN.....	32
TABELLE 4-3: MITTLERER ABSOLUTER FEHLER (MAE) UND DER DAZUGEHÖRIGE SKILL SCORE (SS_{MAE}) DER T_{MIN} - PROGNOSEN.	35
TABELLE 4-4: WURZEL DES MITTLEREN QUADRATISCHEN FEHLERS (RMSE) DER T_{MIN} - PROGNOSEN.....	38
TABELLE 4-5: REDUKTION DER VARIANZEN (RV) DER T_{MIN} - PROGNOSEN.	38
TABELLE 4-6: MITTLERER FEHLER (ME) DER T_6 -PROGNOSEN.	42
TABELLE 4-7: MITTLERER ABSOLUTER FEHLER (MAE) UND DER DAZUGEHÖRIGE SKILL SCORE (SS_{MAE}) DER T_6 - PROGNOSEN..	45
TABELLE 4-8: WURZEL DES MITTLEREN QUADRATISCHEN FEHLERS (RMSE) DER T_6 - PROGNOSEN.....	45
TABELLE 4-9: REDUKTION DER VARIANZEN (RV) DER T_6 -PROGNOSEN.	46
TABELLE 4-10: MITTLERER FEHLER (ME) DER T_{12} -PROGNOSEN.	49
TABELLE 4-11: MITTLERER ABSOLUTER FEHLER (MAE) UND DER DAZUGEHÖRIGE SKILL SCORE (SS_{MAE}) DER T_{12} -PROGNOSEN.	52
TABELLE 4-12: WURZEL DES MITTLEREN QUADRATISCHEN FEHLERS (RMSE) DER T_{12} -PROGNOSEN.....	52
TABELLE 4-13: REDUKTION DER VARIANZEN (RV) DER T_{12} -PROGNOSEN.	53
TABELLE 4-14: MITTLERER FEHLER (ME) DER T_{MAX} -PROGNOSEN	56
TABELLE 4-15: MITTLERER ABSOLUTER FEHLER (MAE) UND DER DAZUGEHÖRIGE SKILL SCORE (SS_{MAE}) DER T_{MAX} - PROGNOSEN. ...	60
TABELLE 4-16: WURZEL DES MITTLEREN QUADRATISCHEN FEHLERS (RMSE) DER T_{MAX} - PROGNOSEN.	60
TABELLE 4-17: REDUKTION DER VARIANZEN (RV) DER T_{MAX} - PROGNOSEN.	61
TABELLE 4-18: TREFFERRATE (PC) UND TREFFERQUOTE (P) DES PETRUS	65
TABELLE 4-19: TREFFERSATZ (T) DES PETRUS	66
TABELLE 4-20: FREQUENCY BIAS INDEX (BIAS) DES PETRUS	66
TABELLE 4-21: POST AGREEMENT (PA) UND ANTEIL DES FALSCHEN ALARMS (FAR) DES PETRUS	67
TABELLE 4-22: DIE ENTDECKUNGSAHRSCHAENLICHKEIT DER „JA“-KATEGORIE (HR) UND DER „NEIN“-KATEGORIE (POD) SOWIE DIE MISS RATE (MR) DES PETRUS	68
TABELLE 4-23: DIE ENTDECKUNGSAHRSCHAENLICHKEIT DES FALSCHEN ALARMS (FR) DES PETRUS	69
TABELLE 4-24: DER KRITISCHE ERFOLGSINDEX (CSI) DES PETRUS	69
TABELLE 4-25: DIE HANSSEN-KUIPERS DISKRIMINANTE (TSS) UND DER HEIDKE SKILL SCORE (HSS) DES PETRUS	70
TABELLE 4-26: TREFFERRATE (PC) UND TREFFERQUOTE (P), SOWIE TREFFERSATZ (T) DES MOS	71
TABELLE 4-27: FREQUENCY BIAS INDEX (BIAS) DES MOS F	72
TABELLE 4-28: POST AGREEMENT (PA) UND ANTEIL DES FALSCHEN ALARMS (FAR) DES MOS.....	72
TABELLE 4-29: DIE ENTDECKUNGSAHRSCHAENLICHKEIT DER „JA“-KATEGORIE (HR) UND DER „NEIN“-KATEGORIE (POD) SOWIE DIE MISS RATE (MR) DES MOS	72
TABELLE 4-30: DIE ENTDECKUNGSAHRSCHAENLICHKEIT DES FALSCHEN ALARMS (FR) DES MOS.....	73
TABELLE 4-31: DER KRITISCHE ERFOLGSINDEX (CSI) DES MOS.....	73
TABELLE 4-32: DIE HANSSEN-KUIPERS DISKRIMINANTE (TSS) UND DER HEIDKE SKILL SCORE (HSS) DES MOS	73
TABELLE 4-33: T_{MAX} - PROGNOSEN UND DEREN FEHLER VOM 12.-14. MAI 2003.....	75
TABELLE 4-34: T_{MAX} - PROGNOSEN UND DEREN FEHLER VOM 29. JÄNNER 2002	76
TABELLE 4-35: MEHRDIMENSIONALE KONTINGENZTABELLE DER MINIMUMTEMPERATUR; PROGNOSEN DER STUDENTEN DES IMGW DER JAHRE 2000-2010	81

TABELLE 4-36: MEHRDIMENSIONALE KONTINGENZTABELLE DER MAXIMUMTEMPERATUR, PROGNOSEN DER STUDENTEN DES IMGW
DER JAHRE 2000-2010 83

Abkürzungsverzeichnis

BIAS	mittlerer Fehler (kontinuierlich) oder <i>Frequency Bias Index</i> (kategorisch)
CSI	Kritischer Erfolgsindex (<i>Critical Success Index</i>)
DWD	Deutscher Wetterdienst
EUMETCAL	<i>European Meteorological Computer Assisted Learning</i>
FR	Entdeckungswahrscheinlichkeit des Falschen Alarms (<i>False Alarm Rate</i>)
FAR	Anteil des Falschen Alarms (<i>False Alarm Ratio</i>)
FC	Vorhersage (<i>forecast</i>)
HR	<i>Hit Rate</i>
HSS	Heidke Skill Score
IMGW	Institut für Meteorologie und Geodynamik (der Universität Wien)
INT	Intervall
KLIMA	Klimaprognose, monatliches Mittel
KLITA	Klimaprognose, tägliches Mittel
MAE	mittlerer, absoluter Fehler (<i>Mean Absolute Error</i>)
ME	mittlerer Fehler (<i>Mean Error</i>)
MEZ	mitteleuropäische Zeit (UTC +1h)
MESZ	mitteleuropäische Sommerzeit (UTC +2h)
MR	<i>Miss Rate</i>
MSE	mittlerer, quadratischer Fehler (<i>Mean Square Error</i>)
MOS	<i>Model Output Statistics</i>
OBS	Beobachtung (<i>observation</i>)
P	Trefferquote (<i>Percent Correct</i>)
PA	<i>Post Agreement</i>
PC	Trefferrate (<i>Proportion Correct</i>)
PERS	PersistenzPrognose
PETRUS	studentisches Prognose-Mittel
POD	Entdeckungswahrscheinlichkeit (<i>Probability of Detection</i>)
POFD	Entdeckungswahrscheinlichkeit des Falschen Alarms (<i>Probability of False Detection</i>)
r	Korrelationskoeffizient
RMSE	Wurzel des mittleren absoluten Fehlers (<i>Root Mean Square Error</i>)
RV	Reduktion der Varianzen (<i>Reduction of Variance</i>)
SS	<i>Skill Score</i>

SS	Sommersemester
T	Treffersatz
TSS	Hanssen-Kuipers Diskriminante (<i>True Skill Statistics</i>)
UTC	Koordinierte Weltzeit (<i>Universal Time Coordinated</i>)
WMO	Weltorganisation für Meteorologie (<i>World Meteorological Organization</i>)
WS	Wintersemester
WWRP	<i>World Weather Research Program</i>
ZAMG	Zentralanstalt für Meteorologie und Geodynamik

Literaturverzeichnis

Buchliteratur

Liljequist, Gösta H. und Cehak, Konrad: *Allgemeine Meteorologie*; 3.Auflage, Stockholm 1994, Springer Verlag

Schönwiese, Christian-Dietrich: *Praktische Statistik für Meteorologen und Geowissenschaftler*; 4.Auflage, Stuttgart 2006, Gebrüder Borntraeger Verlag

Wilks, Daniel S.: *Statistical Methods in the Atmospheric Sciences*; International Geophysics Series; Second Edition, Burlington 2006, Elsevier Verlag

Wissenschaftliche Artikel und Arbeiten

Dobek, Astrid: *Auswertung von Nebel und Hochnebel, sowie Prognose mittels logistischer Regression für den Flughafen Wien-Schwechat*; Diplomarbeit, Kapitel 6, Wien 2009, Universität Wien

Hermann, Gerhard: *AUSTRAMOS – Ein Modell zur Vorhersage des lokalen Wetters*, Abschlussbericht; Wien 1998, Zentralanstalt für Meteorologie und Geodynamik (ZAMG)

Hofmann, A. et al. : *Vorhersageprüfung*; PROMET, meteorologische Fortbildung; 4.Jahrgang, Heft 4, Bad Homburg, 1974, Herausgeber: DWD

Murphy, Allan H.: *What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting*; Oregon 1993, College of Oceanic and Atmospheric Sciences, Oregon State University; American Meteorological Society: *Weather and Forecasting*, Volume 8, pp. 281-293

Nurmi, Pertti: *Recommendations on the verification of local weather forecasts*, ECMWF Technical Memoranda; Reading 2003, European Centre for Medium Range Weather Forecasts

Sodoudi, Sahar: *Verifikation der EZMW Niederschlagsprognose für den Iran*; Dissertation, Kapitel 6, Berlin 2004, Freie Universität Berlin

Stanski, Henry R., Wilson, Laurence J. und Burrows, William R.: *Survey of common verification methods in meteorology*; Second Edition, Ontario 1989, Environment Canada, Atmospheric Environment Service

Internetquellen

Brown, B., Ghelli, A., Nurmi, P. et al.: *Forecast Verification: Issues, Methods and FAQ*; World Weather Research Programm (WWRP) Joint Working Group on Forecast Verification Research, laufende Updates, <http://www.cawcr.gov.au/projects/verification/>

Deutscher Wetterdienst (DWD): *Erklärungen zu MOS und MOS Guidance*; <http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/TI/TI1/Informationstechnik/Datenverteilung/Broadcasting/DWDSAT/files/MOS-Erklaerungen.templateId=raw.property=publicationFile.pdf/MOS-Erklaerungen.pdf>

Deutscher Wetterdienst (DWD): *Interpretation und Verifikation Numerischer Wettervorhersagen*;

http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=_dwdwww_aufgabenspektrum_vorhersagedienst&_state=maximized&_windowLabel=T19600331211153305221610

Deutscher Wetterdienst (DWD): *Wetterlexikon*; <http://www.dwd.de/lexikon>

World Meteorological Organization (WMO): *Lead Centre for the Long Range Forecast Verifiaction System – Users Guide*, laufende Updates, <http://www.bom.gov.au/wmo/lrfvs/users.shtml>

Wilson, L. und Nurmi, P.: *Forecast Verification*, Learning Module; European Virtual Organisation for Meteorological Training (EUMETCAL);

<http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/courses/msgcrs/index.htm>

Zentralanstalt für Meteorologie und Geodynamik (ZAMG): *Klimatologische Monatsrückblicke*; http://zamg.ac.at/klima/klima_monat/wetterrueckblick/

Weitere Quellen

Dorninger, Manfred: *Verifikation und ökonomischer Nutzen von Wetterprognosen*; Vorlesung am Institut für Meteorologie und Geodynamik an der Universität Wien, 2002

Danksagung

Mein ganzer Dank gebührt all den Menschen, die selbst dann noch an der Fertigstellung dieser Arbeit geglaubt haben, als ich selbst diese Vorstellung schon längst zu Grabe getragen habe. In erster Linie sei hier mein Freund, Kollege und Padawan Mag. Thomas Kumpfmüller erwähnt, der es wie kein anderer verstand, mich zu motivieren. Aber auch die dezenten Hinweise meiner Freunde, allen voran Dr. Gulistan Salaschor, Nicole Rascher, Mag. Michael Tiefgraber, Mag. Andreas Jäger und anderer, sowie das gute gemeinte Erinnern meiner Familie haben maßgebend an der Fertigstellung dieser Arbeit beigetragen.

Für die Betreuung inhaltlicher Natur und für die Geduld möchte ich mich besonders bei meinem Professor und Betreuer Dr. Reinhold Steinacker sowie bei Ass. Prof. Mag. Dr. Manfred Dorninger, für die Beantwortung meiner vielen E-Mail Fragen, bedanken. Auch die Kollegen und –innen am Institut für Meteorologie und Geophysik waren mir (zumindest ab und zu) eine große Hilfe – allen voran Mag. Niko Filipovic und ganz besonders Mag. Stefan Sperka, der mir schon während des Studiums immer zur Seite gestanden ist. Nicht zu vergessen sei an dieser Stelle Mag. Clemens Grohs – der Vollblut-Synoptiker, dessen Diskussionen ich kaum missen möchte. Abseits des Instituts gilt mein Dank der Zentralanstalt für Meteorologie und Geodynamik, hier im Speziellen Mag. Martin Kober, Dr. Alexander Beck und Dr. Gerhard Hermann für das Bereitstellen der MOS-Daten und der Beantwortung der vielen Fragen dahinter.

Danke auch für das Korrekturlesen an Mag. Daniel Landau und für den gescheiterten Versuch darin (immerhin: der Wille zählt!) an all die anderen, insbesondere an Captain Selwyn Rogers.

Ihnen allen – und allen anderen nicht namentlich erwähnten – möchte ich ein herzliches Dankeschön aussprechen! Sofern sie nicht schon der Zunft der Meteorologen angehören, bekommen sie persönliche Wetterberichte auf Lebenszeit! =)

Persönliche Angaben

- Name: Manuel Kelemen
- Geburtsdatum: 03.10.1981
- Geburtsort: Wien
- Eltern: Edeltraud Lengheim, Luka Kelemen
- Familienstand: ledig, keine Kinder
- Staatsbürgerschaft: Österreich

Ausbildung

- 1987 – 1990: Vor- und Volksschule, 1220 Wien
- 1990 – 1992: Volksschule mit Öffentlichkeitsrecht der Wiener Sängerknaben, Augartenpalais, 1020 Wien
- 1992 – 1996: Allgemeinbildende höhere Privatschule der Wiener Sängerknaben, Augartenpalais, 1020 Wien
- 1996 – 2000: Oberstufenrealgymnasium BORG 2, Hegelgasse 14, 1020 Wien
- Juni 2000: Matura mit gutem Erfolg bestanden
- Oktober 2000: Beginn des Studiums der Meteorologie und Geophysik an der Universität Wien
- November 2003: Abschluss der 1.Diplomprüfung (erster Studienabschnitt)
- April 2008: Abschluss des ersten Teils der 2.Diplomprüfung

Berufserfahrung

- Sommer 1999, 2000: Ferialpraktikant bei der Firma ZIELPUNKT Warenhandel GmbH & CoKG
- September 2000-Dezember 2010: Angestellter der Firma ZIELPUNKT Warenhandel in unterschiedlichen Positionen (abhängig vom Studienfortschritt)
- September & Oktober 2005: Praktikant bei ATV Privatfernsehen (vormals ATV+), Wetterredaktion, Aspernbrückengasse 2, 1020 Wien
- November 2005 – Juni 2007: Angestellter bei ATV Privatfernsehen als Assistenz-Meteorologe und stellvertretender Leiter der ATV Wetterredaktion
- seit Juni 2007: Chefmeteorologe bei ATV Privatfernsehen
- Juli 2009 – August 2010: Medienbetreuung an der Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Hohe Warte, 1190 Wien