

**Ein Verfahren zur Extrapolation
von Daten mit Erhaltungsneigung und
dessen Anwendung für die Kurzfristprognose
von meteorologischen Einpunktmessungen**

Autoren:

***C. Temme
B. Thees***

GKSS 2004/16

**Ein Verfahren zur Extrapolation
von Daten mit Erhaltungsneigung und
dessen Anwendung für die Kurzfristprognose
von meteorologischen Einpunktmessungen**

Autoren:

C. Temme

*(GKSS, Institut für
Küstenforschung)*

B. Thees

*(Umweltbundesamt,
Messstelle Zingst)*

Die Berichte der GKSS werden kostenlos abgegeben.
The delivery of the GKSS reports is free of charge.

Anforderungen/Requests:

GKSS-Forschungszentrum Geesthacht GmbH
Bibliothek/Library
Postfach 11 60
D-21494 Geesthacht
Germany
Fax.: (49) 04152/871717

Als Manuskript vervielfältigt.
Für diesen Bericht behalten wir uns alle Rechte vor.

ISSN 0344-9629

GKSS-Forschungszentrum Geesthacht GmbH · Telefon (04152)87-0
Max-Planck-Straße · D-21502 Geesthacht / Postfach 11 60 · D-21494 Geesthacht

Ein Verfahren zur Extrapolation von Daten mit Erhaltungsneigung und dessen Anwendung für die Kurzfristprognose von meteorologischen Einpunktmessungen

Christian Temme und Barnim Thees

26 Seiten mit 5 Abbildungen

Zusammenfassung

Es wurde ein Computerprogramm entwickelt, das mit Hilfe von Einpunktmessungen unterschiedlichster Umweltparameter den zukünftigen Verlauf einzelner Parameter für diesen Ort berechnen kann. Es sind Kurzzeitprognosen (bis zu 10 Stunden) für den Temperaturverlauf und die Globalstrahlung an der Messstelle des Umweltbundesamtes in Zingst berechnet und im Detail untersucht worden. Als Grundlage dienten die Monatsdatensätze der jeweiligen Halbstundenwerte. Die Konfigurationseinstellungen des Computerprogramms werden aufgezeigt. Die Prognosegüte wird durch das 80-Perzentil der Restabweichungen (80 %ige Eintreffwahrscheinlichkeit) bewertet, und durch Anwendung eines multiplen Regressionsmodells können Veränderungen der Eigenschaften des Prozesses „gelernt“ und entsprechend angepasst werden. Gleichzeitig kann entschieden werden, ob und bis zu welchem Prognoseschritt das Verfahren brauchbare Ergebnisse liefert. Das beschriebene Prognoseverfahren kann prinzipiell auf alle Daten mit Erhaltungsneigung angewendet werden. Der verwendete „selbstlernende“ Algorithmus sichert eine automatische Anpassung für alle möglichen Datenkollektive mit sich verändernden Eigenschaften ab, so dass keine langjährigen Messreihen für Kurzzeitprognosen mit hoher zeitlicher Auflösung notwendig sind. Potentielle Anwendungsgebiete werden diskutiert.

A procedure for the extrapolation of data from an autoregressive process and the application to short-term predictions of meteorological single-point measurements

Abstract

Software was developed which calculates future developments of an individual parameter with the help of single-point measurements of different environmental parameters. Short-term predictions (up to 10 hours) for temperature and global radiation were calculated and investigated in detail for the measurement location Zingst of the German Environmental Protection Agency. The monthly time-series of the half-hour averages were used for calculation. The configuration of the software is

shown. The quality of the predictions is evaluated by the 80-percentile of the residual variance (80 % probability). Short-term changes of the process can be learned and adjusted by a multiple regression model. At the same time it helps to decide if and up to which prediction step the procedure produces useful results. The described procedure can be applied in principle to all data from autoregressive processes. The here used “selftrained” algorithm ensures an automatic adjustment for all possible datasets with changing properties; therefore, using long-term datasets for short-term predictions in high time resolution is not necessary. Potential fields of application are discussed.

Inhaltsverzeichnis

1	EINLEITUNG	7
2	THEORETISCHER HINTERGRUND	8
3	DAS PROGNOSEMODELL	9
3.1	Prognosegleichung	9
3.2	Der Eintreffwahrscheinlichkeitsbereich (Prognosegüte)	10
4	ERSTE ERGEBNISSE VON PROGNOSEBERECHNUNGEN	10
4.1	Temperaturprognose	15
4.2	Globalstrahlungsprognose	17
5	SCHLUSSBEMERKUNGEN / AUSBLICK	20
	Literatur	23
	Anhang	25

1 EINLEITUNG

Geophysikalische Messwerte besitzen die Eigenschaft der **Erhaltungsneigung**.

Bei im genügend kleinen zeitlichen (oder auch räumlichen) Abstand gemessenen Daten ist der nachfolgende Messwert nicht unabhängig vom Vorgängermesswert. In den Datenaufzeichnungen ist eine zeitliche (oder auch räumliche) Struktur zu erkennen, wobei die Änderung der Messwerte mehr oder weniger stetig erfolgt. Dies dürfte das Vorhandensein von permanent existierenden (stetigen) Prozessabläufen in der natürlichen Umwelt widerspiegeln. Die registrierten Datenstrukturen besitzen „quasi ein *Gedächtnis*“. Beispielsweise „weiß“ die tägliche Temperaturentwicklung, dass es einen je nach Messstandort unterschiedlich ausgeprägten Tagesgang geben wird.

Weiterhin ist zu beachten, dass die Messwerte den Zustand eines bestimmten (willkürlichen) Teils der Umwelt beschreiben, z.B. die Temperatur den Wärmeinhalt eines Luftteilchens. In der Änderung der Temperatur kommt somit die Wärmeabgabe /Wärmezufuhr zum Ausdruck. Dies könnte u.a. eine *Folge* von veränderten Strahlungsbedingungen sein. Wird nun als weiterer Parameter z.B. die Globalstrahlung gemessen, so ist klar, dass die willkürlich gewählten Parameter zur Beschreibung der natürlichen Umwelt zusätzlich noch untereinander in Beziehung stehen können. Dabei laufen diese Prozesse in der Regel nicht gleichzeitig ab. Die zukünftige zeitliche Änderung eines Parameters kann sich in der Kombination der zeitlichen Strukturen anderer „sozusagen ankündigen“. Beispielsweise kündigt ein stärkerer Luftdruckfall häufig das Ende einer sonnigen Witterungsperiode an.

→ Die zeitliche (oder auch räumliche) Änderung von Messdaten ist ein **multiples Phänomen**.

Diese Eigenschaften wurden durch den Menschen schon immer (intuitiv) benutzt, z.B.: Vorhersage des Wetters durch Schäfer/Fischer für ihren Lebens-/Arbeitsort; gelernt durch permanente (multiple) Beobachtung des zeitlichen Ablaufes verschiedenster Umweltparameter; „erzwungen“ durch die große Abhängigkeit des Erfolges ihrer Arbeit von diesen Wetterabläufen.

In dieser Arbeit wird versucht, diese Methodik mit Hilfe mathematisch statistischer Verfahren zu objektivieren.

Das Ziel war:

ein Computerprogramm zu entwickeln, das nur mit Hilfe von Einpunktmessungen unterschiedlichster Umweltparameter den wahrscheinlichsten zukünftigen Verlauf einzelner Parameter für diesen Ort berechnen kann. Dabei sollte gleichzeitig die Güte dieser „Prognose“ permanent aus den Fehlern dieses Verfahrens abgeschätzt werden (d.h. selbstlernend; im Sinne neuronaler Netze).

2 THEORETISCHER HINTERGRUND

Zeitreihen mit mehr oder weniger ausgeprägter Erhaltungsneigung (autoregressive Prozesse) lassen sich immer wie folgt darstellen:

$$y(t) = c_1*y(t-1) + c_2*y(t-2) + \dots + c_p*y(t-p) + e_t \quad \text{für alle } t \quad (1)$$

[Hartung S. 678]

mit e_t : Restfehler in Form eines weißen Rauschens mit dem Erwartungswert 0

Das multiple lineare Regressionsverfahren ist somit in der Lage die Parameter dieser Gleichung zu bestimmen, indem dazu die entsprechenden zeitlichen Verschiebungen der Ordnung p durchgeführt werden.

Die meisten Fluktuationsprozesse in der Meteorologie und Geophysik haben die Form eines roten Rauschens [hohe Amplituden (Energie) bei niedrigen Frequenzen (hohen Perioden); exponentielles Abnehmen auf niedrige Amplituden (Energie) bei hohen Frequenzen (niedrigen Perioden)] mit häufig einem hervortretendem Frequenzbereich (auch farbiges Rauschen genannt)].

[Taubenheim: S 233]

Das rote Rauschen vom exponentiellen Typ hat die Eigenschaft, dass jeder Messwert nur vom Vorgängermesswert aber nicht von in weiterer Vergangenheit liegenden Werten abhängt.

[Taubenheim S.232]

→ Notwendigkeit der gleitenden Verschiebung mit kleinen p -Werten.

Beim farbigen Rauschen gibt es zusätzlich evtl. mehrere hervortretende Frequenzbereiche, wobei die feste Verschiebung entsprechend der hervortretenden Perioden angepasst werden muss.

→ harmonische Analyse des Prozesses mit Ermittlung dieser Perioden.

[Bronstein – Semendjajev S.480]

Zur Ermittlung dieser hervortretenden Frequenzen ist folgendes zu beachten

[Taubenheim S.292]:

die maximale (interpretierbare) Verschiebung sollte nicht mehr als etwa $1/10$ der Beobachtungslänge betragen (wenn der mittlere quadratische Fehler (MSE) bei der Bestimmung der Spektrallinie nicht mehr als etwa 33 % des wahren Energiewertes betragen soll). Das bedeutet z.B.: bei 1488 Halbstundendaten (Monatsdatei) knapp 150, d.h. ca. 3 Tage.

→ Es können nur die Amplitudenpeaks mit Perioden von < 150 aus dem mittels harmonischer Analyse ermittelten Spektrum benutzt werden.

3 DAS PROGNOSEMODELL

Unter der **Bedingung eines angenähert stationären Prozesses** können diese Eigenschaften für die zeitliche Extrapolation (Prognose) verwendet werden. Über obiges multiples Regressionsmodell, als selbstlernendes Verfahren, können Veränderungen der Eigenschaften des Prozesses "gelernt" und entsprechend angepasst werden.

Gleichzeitig könnten evtl. vorhandene "Querkorrelationen" zu anderen Messgrößen (z), die ebenfalls wieder autoregressive Prozesse darstellen, zur Modellerweiterung benutzt werden:

$$y(t) = c_1*y(t-1) + c_2*y(t-2) + \dots + c_p*y(t-p) + (c_{p+1})*z(t-1) + (c_{p+2})*z(t-2) + \dots + (c_{p+q})*z(t-q) + e_t \quad (2)$$

Hieraus lässt sich die folgende multiple Regressionsgleichung aufstellen:

$$y = a_0 + c_1*x_1 + c_2*x_2 + \dots + c_n*x_n \quad (3)$$

$$\begin{aligned} \text{mit: } y &= y(t) & x_1 &= y(t-p) & x_2 &= y(t-p-pgl) & x_3 &= y(t-pfest1) & x_4 &= y(t-pfest2) \\ & & x_5 &= z_1(t-p-pgl) & x_6 &= z_1(t-pfest1) & x_7 &= z_1(t-pfest2) \\ & & x_8 &= z_2(t-p-pgl) & x_9 &= z_2(t-pfest1) & x_{10} &= z_2(t-pfest2) \\ & & \dots & & & & & \end{aligned}$$

mit: p = Prognoseverschiebung; pgl = gleitende Verschiebung; pfest = feste Verschiebung bzgl. t der jeweils einzelnen Regressoren y, z1, z2,...

Der ermittelte **Zusammenhang gelte nun auch in $t=t+p$** , so dass aus der obigen Regressionsgleichung (3) durch Ersetzen von t durch t+p die folgende Prognosegleichung (4) folgt:

3.1 Prognosegleichung

$$y(t+p) = a_0 + c_1*y(t) + c_2*y(t-pgl) + c_3*y(t+p-pfest1) + c_4*y(t+p-pfest2) + \dots \quad (4)$$

für einen frei vorgebbaren Prognosebereich von p = pa ... pe in Prognoseschritten dp

(pa = Prognoseanfang; pe = Prognoseende; dp = Prognoseschritte).

Damit die Gleichungen (3) und (4) immer berechnet werden können, müssen die folgenden Bedingungen eingehalten werden:

- (a) $p_{fest} \geq p_e$ für jeweils jeden einzelnen Regressoren: y, z1, z2, ...
- (b) $p + p_{gl} < p_{fest}$ für jeweils jeden einzelnen Regressoren: y, z1, z2, ...

Zu (a): Im Falle $p_{fest} < p_e$ folgt aus Gleichung (4): $p - p_{fest} > 0$, d.h. es müsste ein y aus der „Zukunft“ bekannt sein.

Zu (b): Bei $p + p_{gl} = p_{fest}$ folgt aus Gleichung (3): Doppeldeutigkeit der Regressionsgleichung durch zwei identische x2 und x3 (Gleichung (3) ist nicht lösbar)

Welche zusätzlichen Messparameter und mit welchem zeitlichen Versatz zu einer besseren Prognoseabbildung führen, muss der „Intuition“ (dem „Ausprobieren“) überlassen bleiben. Beispielhaft sollen nun die verwendeten Konfigurationseinstellungen für die Temperatur- und Globalstrahlungsprognose aufgeführt werden:

Für die **Temperaturprognose** wurde die folgende gleitende Verschiebung gewählt:

14p01 Temperatur mit zeitlichem Versatz von 1 Zeitschritt (halbe Stunde) (pgl)

Es wurden folgende weitere Messparameter benutzt:

37p02 Globalstrahlung mit zeitlichem Versatz von 2 Zeitschritten (1 Stunde) (pgl)
37p48 Globalstrahlung mit zeitlichem Versatz von 48 Zeitschritten (pfest)
15p12 relative Feuchte mit zeitlichem Versatz von 12 Zeitschritten (pgl)
15p48 relative Feuchte mit zeitlichem Versatz von 48 Zeitschritten (pfest)
15p72 relative Feuchte mit zeitlichem Versatz von 72 Zeitschritten (pfest)
19p00 Windböe mit zeitlichem Versatz von 0 Zeitschritten (pgl)
19p48 Windböe mit zeitlichem Versatz von 48 Zeitschritten (pfest)
12p24 Luftdruck mit zeitlichem Versatz von 24 Zeitschritten (pgl)
91p00 Nord- Süd- Windkomponente mit zeitlichem Versatz von 0 Zeitschritten (pgl)
91p48 Nord- Süd- Windkomponente mit zeitlichem Versatz von 48 Zeitschritten
 (pfest)
90p00 West- Ost- Windkomponente mit zeitlichem Versatz von 0 Zeitschritten (pgl)
90p48 West- Ost- Windkomponente mit zeitlichem Versatz von 48 Zeitschritten
 (pfest)

Für die Prognose der **Globalstrahlung** wurde die folgende gleitende Verschiebung benutzt:

37p02 Globalstrahlung mit zeitlichem Versatz von 2 Zeitschritten (pgl)

Es wurden folgende weitere Messparameter benutzt:

37p48 Globalstrahlung mit zeitlichen Versatz von 48 Zeitschritten (pgl)
18p00 Taupunkt mit zeitlichen Versatz von 0 Zeitschritten (pgl)
18p48 Taupunkt mit zeitlichem Versatz von 48 Zeitschritten (pgl)
12p00 Luftdruck mit zeitlichem Versatz von 0 Zeitschritten (pgl)
12p06 Luftdruck mit zeitlichem Versatz von 6 Zeitschritten (pgl)

In der Abbildung 1 sind die entsprechenden Prognosen vom 06.06.2003 zum Zeitpunkt 11:19 MEZ mit Hilfe des Visualisierungsprogramms BILDFOLGE5.EXE in Form der Konfiguration PROGNOSECOMP.CFG dargestellt. Es wurden jeweils die Prognosebereiche mit einer „gelernten“ Eintreffwahrscheinlichkeit von 80 % (80-Perzentil) für +1 Stunde, +2 Stunden, +4 Stunden, +6 Stunden, +8 Stunden und +10 Stunden angezeigt.

Wetterkurzfristprognose für Zingst vom : **06.06.2003** **11:19 MEZ**

Der Messstandort :



auf dem Gelände der Mutter -
Kind - Klinik in Zingst (Ost)

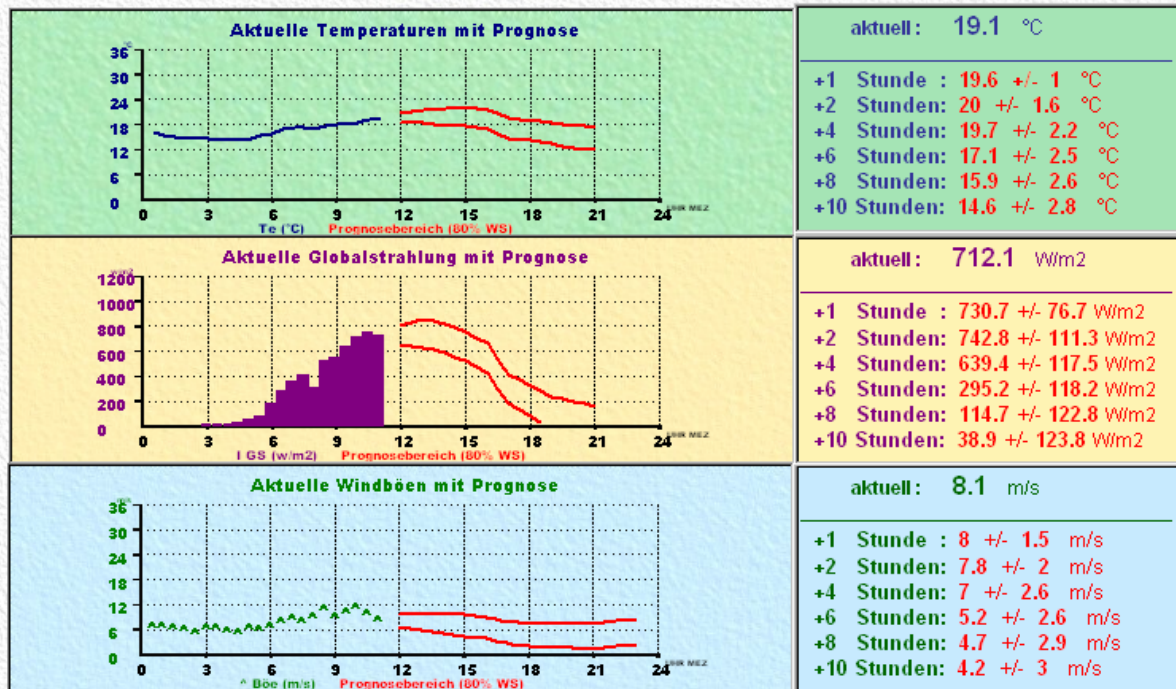


Abb. 1: Beispiel von Prognoseberechnungen (Temperatur, Globalstrahlung, Maximale Windböen). Dieses Beispiel ist mit Hilfe des Programms BILDFOLGE5.EXE und der Konfiguration PROGNOSECOMP.CFG visualisiert worden.

Die Temperaturprognosen für +3 Stunden (+3H) und +9 Stunden (+9H) und die Globalstrahlungsprognosen für +1 Stunde (+1H) und +3 Stunden (+3H) des gesamten Monats Juni sind im „Detail“ untersucht worden und die Ergebnisse dieser Untersuchungen sind in den Abbildungen 2 bis 5 dargestellt.

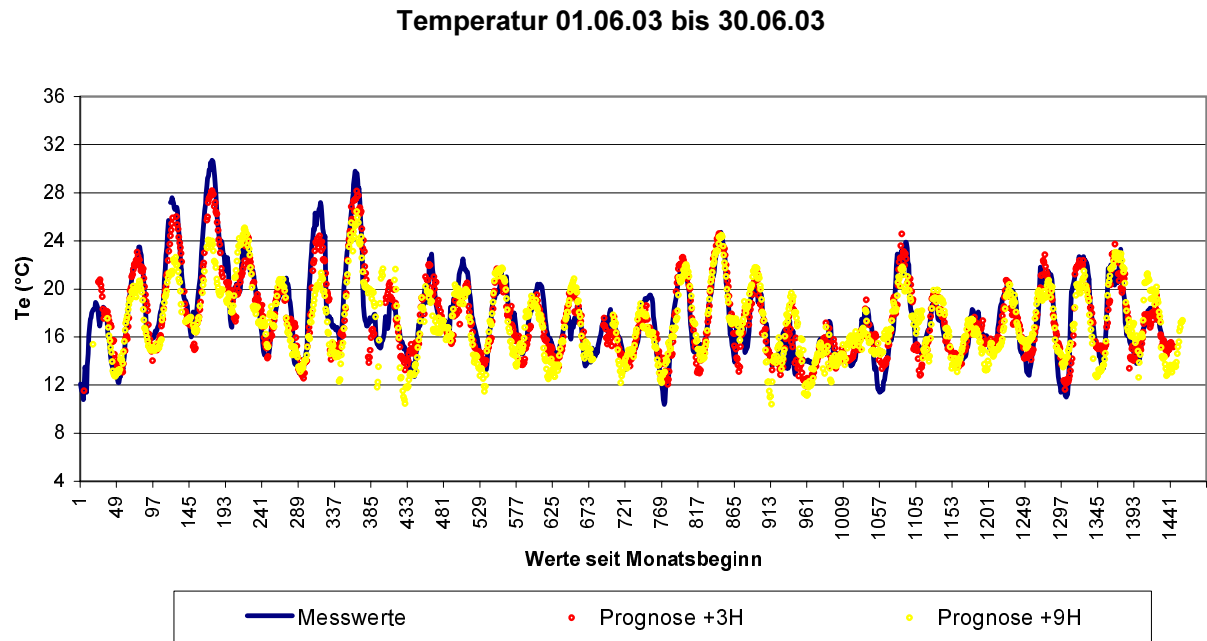


Abb. 2: Temperaturmesswerte und Prognosen für den Zeitraum 01. bis 30. Juni 2003.

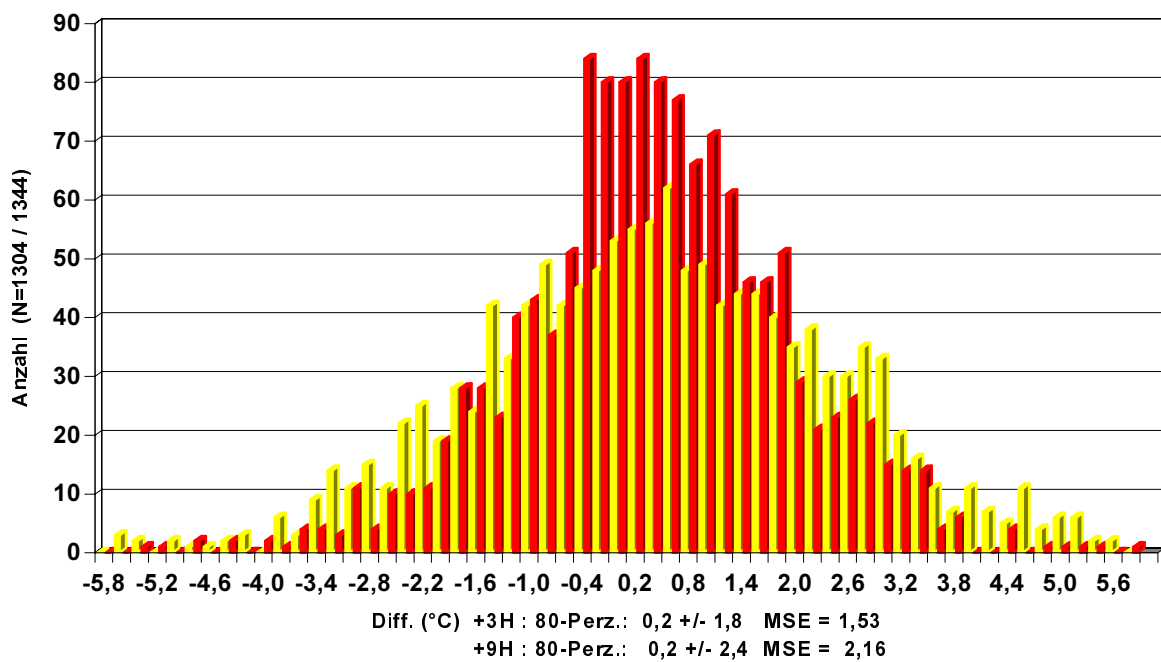


Abb. 3: Absolute Häufigkeiten der Temperaturprognoseabweichungen für +3H (rot) und +9H (gelb) im Juni 2003.

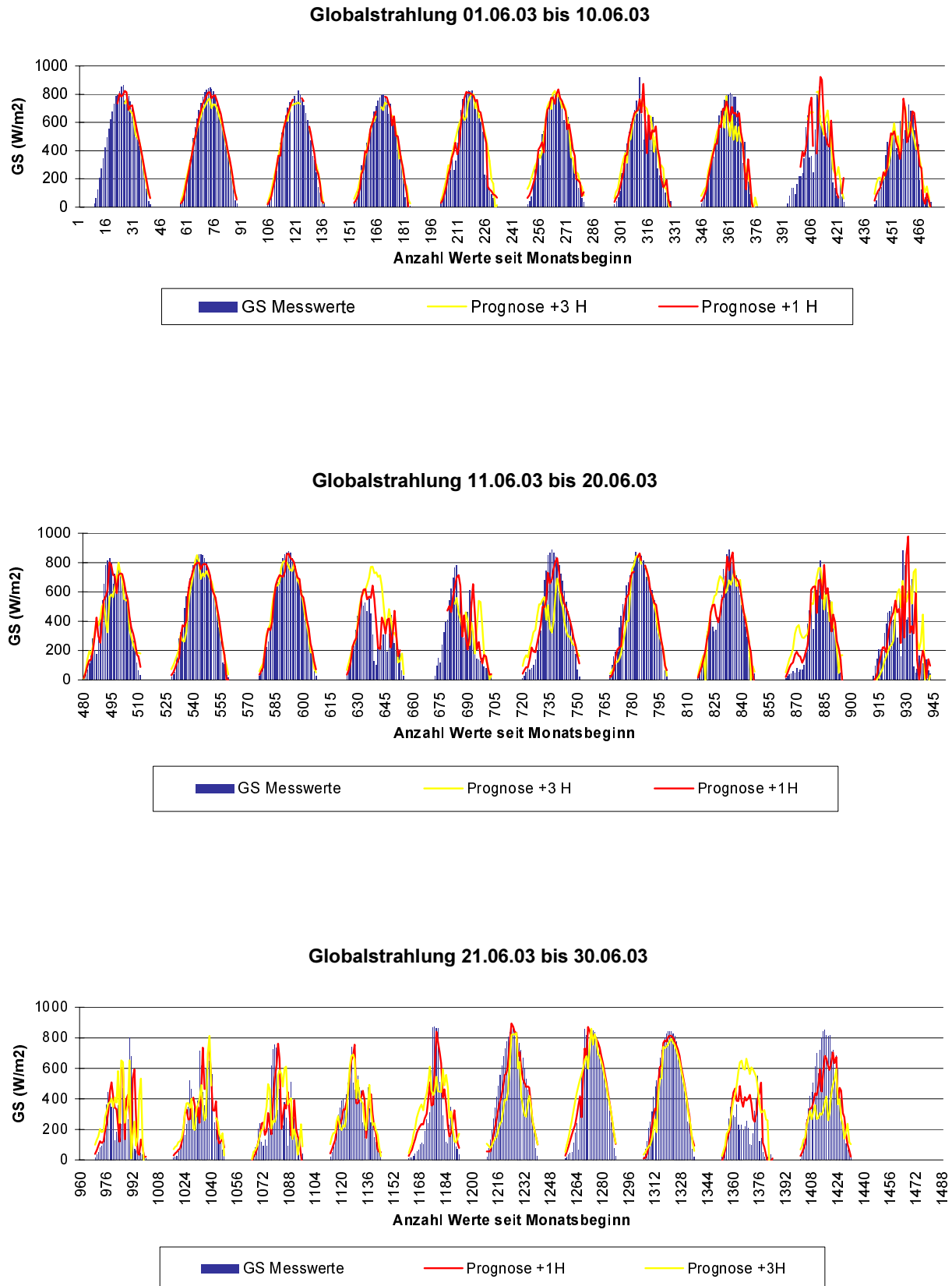


Abb. 4: Globalstrahlungsmesswerte (GS-Messwerte) und Prognosen für den Zeitraum 01. bis 30. Juni 2003.

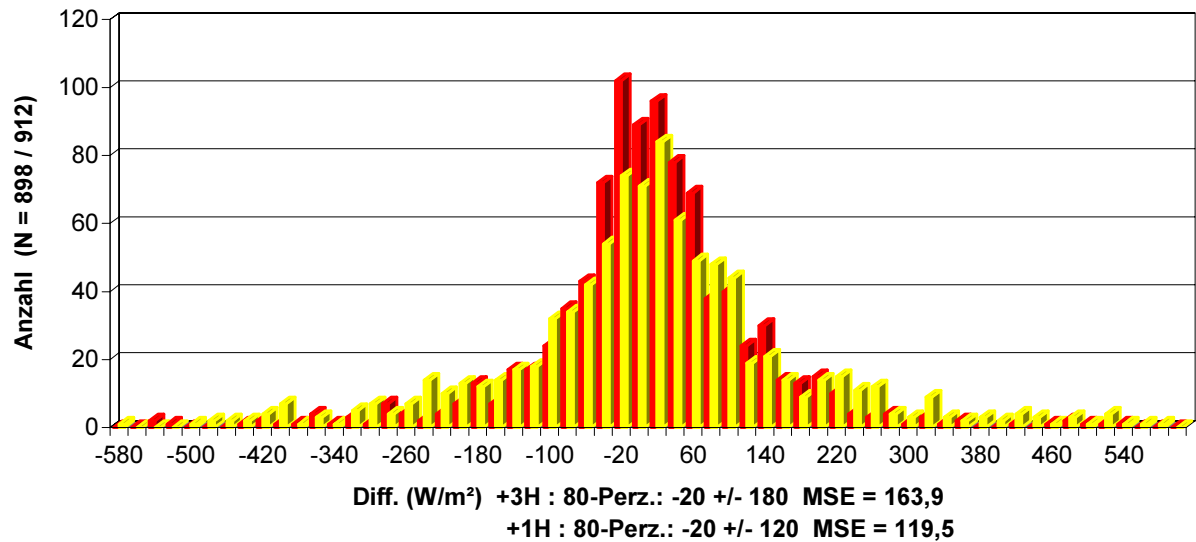


Abb. 5: Absolute Häufigkeiten der Globalstrahlungsprognoseabweichungen für +3H (gelb) und +1H (rot) im Juni 2003 (4–20 Uhr MEZ).

4.1 Temperaturprognose

Der Vergleich sämtlicher Prognosen des Monats Juni 2003 mit den anschließend gemessenen Temperaturen für die Prognoseschritte +3 Stunden und +9 Stunden ergab folgende Ergebnisse:

Plus 3 Stunden:

Gesamtzahl der berechneten Prognosen: Nges = 1344
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 1,5 °C
80-Perzentil in 0,2 °C – Klassen zwischen +6 und –6 °C: +0,2 +/- 1,8 °C
Reduktion der Varianz zur „simplen“ Erhaltungsneigung: RV = + 61 %

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt: 30.06. 12 MEZ

Gesamtzahl der regressierten Werte: Nges = 1383
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 1,4 °C
(für die „simple“ Erhaltungsneigung: RMSE = 2,4 °C (Nges=1476))

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt: 30.06. 12 MEZ

ohne zusätzliche Messparameter

(d.h. nur Temperatur – Erhaltungsneigung)

Gesamtzahl der regressierten Werte:	Nges = 1400
Wurzel aus mittlerer quadratischer Abweichung:	RMSE = 1,8 °C

Plus 9 Stunden:

Gesamtzahl der berechneten Prognosen:	Nges = 1304
Wurzel aus mittlerer quadratischer Abweichung:	RMSE = 2,2 °C
80-Perzentil in 0,2 °C – Klassen zwischen +6 und –6 °C:	+0,2 +/- 2,4 °C
Reduktion der Varianz zur „simplen“ Erhaltungsneigung:	RV = + 82 %

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt: 30.06. 12 MEZ

Gesamtzahl der regressierten Werte:	Nges = 1383
Wurzel aus mittlerer quadratischer Abweichung:	RMSE = 1,8 °C
(für die „simple“ Erhaltungsneigung:	RMSE = 5,2 °C (Nges=1464))

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt: 30.06. 12 MEZ

ohne zusätzliche Messparameter

(d.h. nur Temperatur – Erhaltungsneigung)

Gesamtzahl der regressierten Werte:	Nges = 1400
Wurzel aus mittlerer quadratischer Abweichung:	RMSE = 2,4 °C

→ Die Wurzel des mittleren quadratischen Fehlers (RMSE) der Regressionswerte im Vergleich zu den Messwerten der letzten 1488 Daten zum Zeitpunkt: 30.06. 12 MEZ lag mit 1,4 °C für die 3-Stundenprognose und mit 1,8 °C für die 9-Stundenprognose deutlich unter den RMSE-Werten für die „simple“ Erhaltungsneigung (2,4 bzw. 5,2 °C; siehe obige Liste). Bei der „simplen“ Erhaltungsneigung würde sich der Messwert nicht ändern, d.h. der Prognosewert in 3 Stunden bzw. in 9 Stunden ist mit dem aktuellen Messwert identisch. Diese deutlich niedrigeren RMSE-Werte sind natürlich die Grundvoraussetzung für ein sinnvolles Prognoseverfahren.

→ Die 80 %-Eintreffwahrscheinlichkeit (bezogen auf den Monat Juni 2003) der berechneten Prognosen lag bei der 3-Stundenprognose bei etwa $\pm 1,8$ °C (RMSE = 1,5) und verschlechterte sich für die 9-Stundenprognose auf etwa $\pm 2,4$ °C (RMSE = 2,2). Dies spiegelt die abnehmende Erhaltungsneigung der erkannten zeitlichen Strukturen und natürlich die Strukturbrüche durch Verletzung der Modellvoraussetzung wider, nach der die vorliegenden Prozesse angenähert stationär sein sollen (siehe 2. Prognosemodell).

→ Die Reduktion der Varianz (RV) zur simplen Erhaltungsneigung betrug bei der 3-Stundenprognose + 61 % und bei der 9-Stundenprognose + 82 %. Beide Werte liegen damit deutlich oberhalb der minimalen Reduktion der Varianz (für $N = 1300$: $RV(MIN) \approx 10$ % für das 5 % Niveau der Irrtumswahrscheinlichkeit). Das verwendete Prognoseverfahren liefert somit signifikant (wesentlich) bessere Ergebnisse als die „simple“ Erhaltungsneigung (Persistenz).

Da natürlich sämtliche Besonderheiten des Messstandortes mit ihren Auswirkungen auf die Temperaturstruktur der Messwerte enthalten sind, dürften diese Prognoseergebnisse für die etwa 10 folgenden Stunden schon relativ gut sein. Erst recht, wenn man bedenkt, dass z.B. kleinräumige Strukturen im relativ großen Gitternetz der Prognosemodelle des Wetterdienstes kaum berücksichtigt werden können.

→ Der Beitrag der „querkorrelierten“ Messparameter auf den RMSE-Wert liegt bei der 3-Stundenprognose bei 21,9 % (Erniedrigung von 1,8 (nur Temperatur) auf 1,4) und bei der 9-Stundenprognose bei 26,7 % (Erniedrigung von 2,4 (nur Temperatur) auf 1,8)

Das bedeutet:

je niedriger der Erhaltungsneigungseffekt des zu prognostizierenden Parameters auf die Prognosegüte wird, umso größer wird somit der Querkorrelationseffekt der zusätzlichen Modellparameter.

Die willkürlich ausgewählten zusätzlichen Modellparameter haben schon einen relativ großen Einfluss auf die Prognosegüte. Das „Optimum“ wäre aber bestimmt noch zu finden.

4.2 Globalstrahlungsprognose

Es wurden die Globalstrahlungsprognosen der Prognoseschritte: +1 Stunde und +3 Stunden des Monats mit den tatsächlichen Messwerten verglichen. Berechnet wird jeweils auch der RMSE – „Root Mean Squared Error“ (Wurzel aus mittlerem quadratischen Abweichung).

Plus 1 Stunde:

Gesamtzahl der berechneten Prognosen: Nges = 1358
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 98,7 W/m²
Reduktion der Varianz zur „simplen“ Erhaltungsneigung: RV = + 32 %

Nur Prognosen zwischen 04 und 20 Uhr MEZ

Gesamtzahl der berechneten Prognosen: Nges = 912
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 119,5 W/m²
80-Perzentil in 20 W/m² – Klassen
zwischen +600 und –600 W/m²: -20 +/- 120 W/m²

Regressionsrechnungen der letzten 1488 Daten
zum Zeitpunkt : 30.06. 12 MEZ

Gesamtzahl der regressierten Werte: Nges = 1417
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 94,9 W/m²
(für die „simple“ Erhaltungsneigung: RMSE = 119,3 (Nges=1480))

Regressionsrechnungen der letzten 1488 Daten
zum Zeitpunkt : 30.06. 12 MEZ

ohne zusätzliche Messparameter

(d.h. nur Globalstrahlungs – Erhaltungsneigung)

Gesamtzahl der regressierten Werte: Nges = 1420
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 96,6 W/m²

Plus 3 Stunden:

Gesamtzahl der berechneten Prognosen: Nges = 1341
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 136,6 W/m²
Reduktion der Varianz zur „simplen“ Erhaltungsneigung: RV = + 71 %

Nur Prognosen zwischen 04 und 20 Uhr MEZ

Gesamtzahl der berechneten Prognosen: Nges = 902
Wurzel aus mittlerer quadratischer Abweichung: RMSE = 163,9 W/m²
80-Perzentil in 20 W/m² – Klassen
zwischen +600 und –600 W/m²: -20 +/- 180 W/m²

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt : 30.06. 12 MEZ

Gesamtzahl der regressierten Werte:

$N_{ges} = 1398$

Wurzel aus mittlerer quadratischer Abweichung

$RMSE = 128,9 \text{ W/m}^2$

(für die „simple“ Erhaltungsneigung:

$RMSE = 253,9 \text{ (} N_{ges}=1476 \text{))}$

Regressionsrechnungen der letzten 1488 Daten

zum Zeitpunkt : 30.06. 12 MEZ

ohne zusätzliche Messparameter

(d.h. nur Globalstrahlungs – Erhaltungsneigung)

Gesamtzahl der regressierten Werte:

$N_{ges} = 1415$

Wurzel aus mittlerer quadratischer Abweichung

$RMSE = 132 \text{ W/m}^2$

→ Die Wurzel des mittleren quadratischen Fehlers (RMSE) der Regressionswerte im Vergleich zu den Messwerten der letzten 1488 Daten zum Zeitpunkt: 30.06. 12 MEZ lag mit $94,9 \text{ W/m}^2$ für die 1-Stundenprognose und mit $128,9$ für die 3-Stundenprognose wieder unter den RMSE-Werten für die „simple“ Erhaltungsneigung ($119,3$ bzw. $253,9$, siehe obige Liste). Der Abstand fällt jedoch nicht mehr so deutlich aus wie bei der Temperaturprognose !

→ Die Reduktion der Varianz (RV) zur simplen Erhaltungsneigung betrug bei der 1-Stundenprognose $+32 \%$ (von $RMSE = 119,3$ auf $98,7$) und bei der 3-Stundenprognose $+71 \%$ (von $RMSE = 253,9$ auf $136,6$) . Beide Werte liegen damit deutlich oberhalb der minimalen Reduktion der Varianz (für $N = 1300$: $RV(MIN) \approx 10 \%$ für das 5% Niveau der Irrtumswahrscheinlichkeit). Das verwendete Prognoseverfahren liefert somit auch für die Globalstrahlungsprognose signifikant bessere Ergebnisse als die „simple“ Erhaltungsneigung (Persistenz).

→ Beim Vergleich der berechneten Prognosen mit den Messwerten ist zu beachten, dass die „Nachtprognosen“ simpel sind (keine Sonnenstrahlung). Deshalb wurden nur die berechneten Prognosen für den Zeitraum zwischen 04 Uhr MEZ und 20 Uhr MEZ des Juni 2003 berücksichtigt.

Die 80-%-Eintreffwahrscheinlichkeit dieser berechneten Prognosen lag bei der 1-Stundenprognose bei etwa $\pm 120 \text{ W/m}^2$ und verschlechterte sich für die 3–Stundenprognose auf etwa $\pm 180 \text{ W/m}^2$.

→ Der Beitrag der zusätzlich gewählten Messparameter auf den RMSE-Wert liegt bei der 1-Stundenprognose bei 1,8 % (Erniedrigung von 96,6 (nur Globalstrahlung) auf 94,9) und bei der 3-Stundenprognose bei 2,4 % (Erniedrigung von 132 (nur Globalstrahlung) auf 128,9).

Im Gegensatz zur Temperaturprognose können die hier ausgewählten Parameter nur einen relativ geringen Beitrag zur Prognoseverbesserung leisten. Im Gegensatz zur Temperatur ist es für die Globalstrahlung sicher schwieriger Parameter zu finden, die für eine zukünftige Änderung der Strahlungsbedingungen stehen könnten. Auch wenn die Erfahrung zeigt, dass z.B. fallender Luftdruck bei „Strahlungswetter“ häufig für einen Wetterumschwung zu einem unbeständigeren und bewölkten Wetter steht.

5 SCHLUSSBEMERKUNGEN / AUSBLICK

Das beschriebene Prognoseverfahren kann prinzipiell auf alle Daten mit Erhaltungsneigung angewendet werden. Der verwendete „selbstlernende“ Algorithmus sichert eine automatische Anpassung für alle möglichen Datenkollektive mit sich verändernden Eigenschaften ab. Es sind keine langjährigen Messreihen notwendig. Veränderungen der Messgenauigkeit durch beispielsweise sich ändernde Messverfahren werden automatisch gelernt.

Über die gleichzeitige Ermittlung der Prognosegüte mit Hilfe des 80-Perzentils der Restabweichungen kann entschieden werden, ob und bis zu welchem Prognoseschritt das Verfahren brauchbare Ergebnisse liefert. Beispielsweise ist die Anwendung auf Umweltdaten, wie z.B. auf Ozonwerte, denkbar. Erste Versuche zeigen für eine ca. 10-stündige Ozonprognose ein 80-Perzentil der Restabweichungen von etwa 15 bis 20 $\mu\text{g}/\text{m}^3$.

Weitere potentielle Anwendungsgebiete liegen vor allem dort, wo man am zukünftigen Verlauf von Umweltparametern mit hoher zeitlicher Auflösung für die nächsten Stunden an einem ganz bestimmten Ort interessiert ist: Vorhersage des Windangebotes für Windkraftanlagen, Vorhersage von Überschreitungswahrscheinlichkeiten meteorologischer Größen für den Start- und Landebereich von Flugplätzen bis hin zu Risiko-Nutzenabschätzungen von Umwelteinwirkungen (Umweltreizen) auf den Menschen wie sie im Bereich klimatherapeutischer Maßnahmen von Kureinrichtungen gefordert wären (z.B.: Abschätzung der zu erwartenden natürlichen UV-Dosis auf einen Kurpatienten bei seinem Aufenthalt im Freien. Hierdurch wird eine Abschätzung der Dosierung von Umweltreizen im Rahmen klimatherapeutischer Interventionen möglich.)

Das Prognoseprogramm ist durch Einbeziehung von Messdaten der „Nachbarstationen“ eines möglichen Messnetzes als „querkorrelierte“ Modellparameter auch in der Fläche erweiterbar. Dabei ist mit einer weiteren Verbesserung der Prognosegüte zu rechnen, da damit Strukturänderungen durch Wetterfronten vom Modell besser „gelernt“ werden können als durch beispielsweise vorausseilende Luftdruckänderungen.

Bei Kenntnis von Prognoseergebnissen deterministischer Modelle könnten natürlich auch diese Datenreihen als Modellparameter benutzt werden. In diesem Fall wird aus dem oben beschriebenen Prognoseverfahren ein konfigurierbares adaptives „Model Output Statistics“ (MOS) Verfahren [siehe *K. Mönnich (2000), Kap. 7*]. Die möglichen Vorteile eines solchen MOS-Verfahrens wären:

- Eine hohe Anpassungsfähigkeit an jeden einzelnen Messort (jeder Ort seine eigene, optimierte Konfiguration).
- Höchstmögliche Adaptation an das Datenkollektiv (nach jedem neuen Datensatz erfolgt eine Neubestimmung der Regressionskoeffizienten aus den jeweils letzten 1488 Daten).
- Zu jedem Prognosezeitpunkt erfolgt eine Abschätzung der aktuellen Prognosegüte (mit Hilfe des 80-Perzentils der Restabweichungen).

Das Prognoseprogramm PROGNOSEINC.EXE wurde auf handelsüblichen Rechnern unter Windows 98, NT und XP über einen Zeitraum von mehreren Monaten getestet. Es könnte parallel zur Datenerfassung von meteorologischen und Umweltmessstationen laufen. Die Prognoseergebnisse wären dann ebenso wie die Messdaten durch eine Datenzentrale abrufbar.

Literatur

- Damrath, U. Verifikation von Wettervorhersagen
Promet, Jahrg. 28 Nr. 1 / 2 , S. 8 – 16 , DWD 2002
- Hartung, J. Statistik , Lehr- und Handbuch der angewandten Statistik, 12.
Auflage, R. Oldenbourg Verlag, München Wien, 1999
- Taubenheim, J. Statistische Auswertung geophysikalischer und meteorologischer
Daten, Akademische Verlagsgesellschaft Geest & Portig K.-G.,
Leipzig 1969
- Bronstein, I.N. Taschenbuch der Mathematik für Ingenieure und Studenten der
technischen Hochschulen
BSB B.G. Teubner Verlagsgesellschaft
Leipzig 1973
- Enke, W.
Schneider, F. Verbesserung der Ozonprognose
Schlussbericht zum Umweltforschungsplan des BMU:
FK 201 42 257, Berlin (UBA), August 2003
- Mönnich, K. Vorhersage der Leistungsabgabe netzeinspeisender
Windkraftanlagen zur Unterstützung der
Kraftwerkseinsatzplanung; Dissertation, Fachbereich Physik, Carl
von Ossietzky Universität Oldenburg,
Oldenburg, September 2000

Anhang

Begriffsbestimmungen, Definitionen:

→ RMSE – „Root Mean Squared Error” (Wurzel aus mittlerem quadratischen Fehler)

$$\text{RMSE} = \text{SQR} \left(\left(\sum_{i=1}^N (y(i) - y_p(i))^2 \right) / N \right) \quad [\text{Damrath 2002}]$$

mit: $y(i)$ - i-te Messung

$y_p(i)$ - i-te Prognose

N - Gesamtanzahl der Messungen (Prognosen)

SQR - Quadratwurzel

→ Reduktion der Varianz (RV)

$$\text{RV} = 100 * (1 - \text{RMSE}^2 / \text{RMSE}^2(\text{EEN})) \quad [\text{Enke 2003}]$$

mit: EEN – Einfache Erhaltungsneigung (Persistenz)

$$\text{RV}(\text{Min}) = 186 / (N-2)^{0.415} \text{ für eine Irrtumswahrscheinlichkeit von 5 \%}$$

mit: $\text{RV}(\text{Min})$ – Mindestwert, damit RV signifikant größer Null

→ Zusammenhang zwischen Mittelwert (MW), Streuung (ST) und RMSE

$$\text{MW}(x) = 1/N * \sum_{i=1}^N x(i) \quad \text{ST}(x) = \text{SQR} \left(1/(N-1) * \sum_{i=1}^N (x(i) - \text{MW}(x))^2 \right)$$

$$\text{RMSE}(x) = \text{SQR} \left(\text{ST}^2(x) + \text{MW}^2(x) \right)$$

→ Zusammenhang zwischen ST, RMSE und Perzentil

Bei einer (exakten) Gauss – Normalverteilung einer Stichprobe x gelten die folgenden Beziehungen:

$$80\text{-Perzentil}(x) = MW(x) \pm 1,28 * ST(x) \quad [Taubenheim(S.50)]$$

im Falle $MW(x) = 0$ somit:

$$80\text{-Perzentil}(x) = \pm 1,28 * RMSE(x)$$

mit: 80-Perzentil = 80 % aller x -Werte liegen zwischen $MW(x) - 1,28 * ST(x)$ und $MW(x) + 1,28 * ST(x)$