Die Tücken des Root Mean Square Error

Warum AI-Vorhersagemodelle nicht so gut sind wie behauptet

Dietmar Thaler

dietmar.thaler@posteo.at

Admont, 9. Dez. 2023

Im Jahr 2023 wurden von Huawei und Google jeweils ein neuartiges Wettervorhersagemodell (*Pangu-Weather* bzw. *GraphCast*) mit Artikeln in *Nature* (Bi u. a., 2023) *bzw. Science* (Lam u. a., 2023) vorgestellt. Es sind in beiden Fällen AI[1]-basierte Modelle, welche auf ECMWF ERA5-Reanalysedaten (Hersbach u. a., 2020) trainiert wurden. Nach dem *root mean square error*[2] (RMSE) gemessen, übertrafen sie das operationelle hochaufgelöste Modell des ECMWF. Ein Zitat aus Lam u. a. (2023) stimmt jedoch nachdenklich: .. *GraphCast's MSE training objective encourages it to spatially blur its predictions in the presence of uncertainty,* .. Pangu-Weather neigt ebenfalls dazu, die Prognosefelder zu glätten (Magnusson, 2023). Hier wird nun in starker Vereinfachung der Frage nachgegangen, wie sehr Glättung den RMSE beeinflusst.

1 Grundlagen und Methodik

Der RMSE ist definiert als

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}}^{\mathrm{n}} \left(\Phi_{\mathrm{A,i}} - \Phi_{\mathrm{F,i}}\right)^2} \,. \tag{1}$$

Φ_F ist das Vorhersagefeld und Φ_A das Analysefeld für den Zeitpunkt, mit dem das Vorhersagefeld verglichen wird. Der Index i bezeichnet dabei die einzelnen Gitterpunkte, n ist die Anzahl der Gitterpunkte. Im RMSE erhalten große Abweichungen von der Analyse wegen der inhärenten Nichtlinearität überproportionales Gewicht.

Prototypisch wird eine eindimensionale Welle[3] als Funktion der Raumkoordinate x betrachtet, z.B.:

$$\Phi(x) = \Phi_0 \cos(x - \delta) \,. \tag{2}$$

$\Phi_0 > 0$ sei die Amplitude, d.h. die maximale positive Elongation der Welle. $0 \leq \delta \leq 2\pi$ bezeichnet die Phasenverschiebung der Welle. Der Bereich für x sei auf $0 \leq x \leq 2\pi$ beschränkt. Abbildung (1) zeigt modellhaft die *Analyse* (rot-liniert) und die *Vorhersage* (blau-gestrichelt) mit *Amplitudenfehler und Phasenfehler.*

Wir beschränken uns auf *abnehmende* Amplituden im Vorhersagefeld, da dies dem Fall des *spatial blur* der AI-Prognosen entspricht. Aus den Beziehungen (1) und (2) ergibt sich der RMSE in diesem speziellen Fall als Funktion des Phasenfehler δ und des positiven Amplitudenfehler: $\Delta\Phi_0 = \Phi_{0,A} - \Phi_{0,F} \geq 0$, der eine *abnehmende* Amplitude in der Vorhersage bedeutet. Der Einfachheit halber setzen wir die Analyseamplitude auf $\Phi_{0,A} = 1$. Damit lautet der Amplitudenfehler $\Delta\Phi_0 = 1 - \Phi_{0,F}$. Eine Verallgemeinerung auf beliebige Amplitudenfehler ist trivial durchführbar.

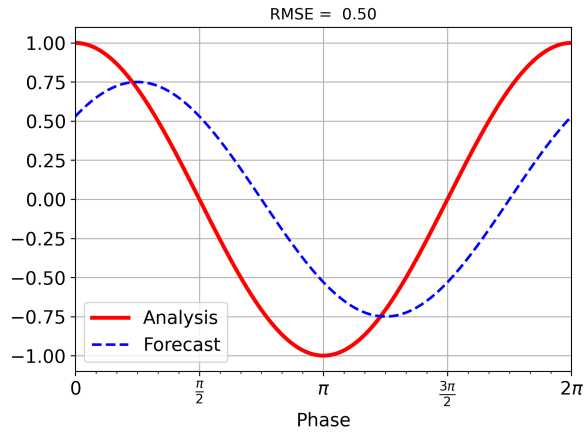

Abbildung 1: Exemplarische Cosinus-Wellen.

Die Berechnung des RMSE im idealisierten Fall kann auf analytischem Weg erfolgen, indem man folgendes Integral löst:

$$\mathrm{RMSE} = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} \left[\Phi_{0,A} \cos(x) - \Phi_{0,F} \cos(x - \delta_F)\right]^2 dx} \,.$$

Hier jedoch erfolgt die Berechnung nach hinreichend feiner Diskretisierung der analytischen Funktionen[4] gemäß Gleichung (1) rein numerisch.

2 Ergebnisse

Betrachten wir zwei Spezialfälle.

Reiner Phasenfehler Der Amplitudenfehler sei 0 (Abb. 2 auf der nächsten Seite, oben): Man kann erkennen, dass der RMSE mit wachsender Phasenverschiebung rasch ansteigt und für die Verschiebung π, wenn die Vorhersagewelle gegenphasig zur Analysewellen steht, mit RMSE=1,41 maximal wird. Bei weiter bis 2π zunehmendem Phasenfehler sinkt der RMSE gegen 0.

[1] *artificial intelligence*, künstliche Intelligenz (KI)

[2] Wurzel des mittleren quadratischen Fehlers

[3] Eine Erweiterung auf allgemeinere Funktionen in höheren Dimension ist möglich.

[4] mittels 201 Datenpunkten im Intervall 0 bis 2π

Reiner Amplitudenfehler Der Phasenfehler sei 0 (Abb. 2, unten): Der RMSE nimmt mit wachsendem Amplitudenfehler (sinkender Amplitude) zu und erreicht mit dem Erreichen des *klimatologischen Mittelwerts* (Amplitude 0, Amplitudenfehler 1) den maximalen RMSE von 0,71.

Das heißt, dass bereits geringe Phasenfehler relativ große RMSE bedeuten, während vergleichsweise große Amplitudenfehler sich nur moderat auf den RMSE auswirken.

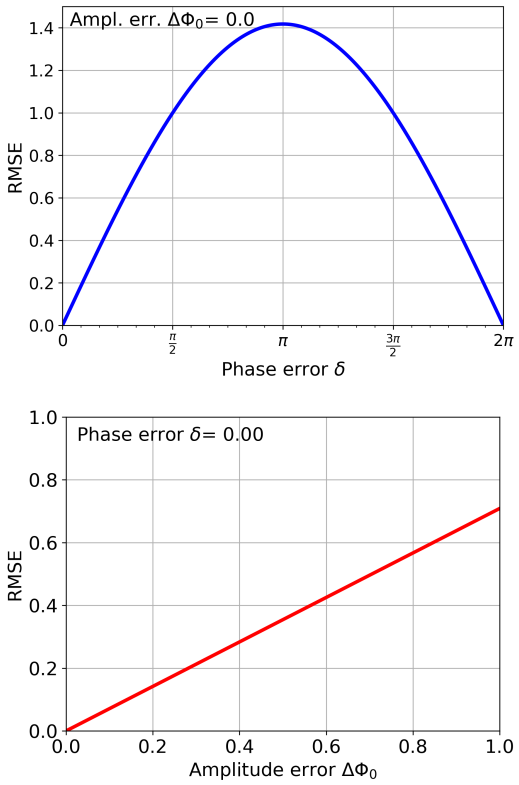

Abbildung 2: Oben: RMSE als Funktion des Phasenfehler bei exakter Amplitude. Unten: RMSE als Funktion der sinkenden Amplitude bei exakter Phase.

Abbildung (3) zeigt in einer zweidimensionalen Darstellung die Isolinien des RMSE als Funktion von Phasenfehler und Amplitudenfehler. Dabei zeigen sich folgende Merkmale:

- RMSE = 0: kein Phasen- und Amplitudenfehler (exakte Prognose)[5]
- Bei maximalem Amplitudenfehler (Amplitude des Prognosefeldes verschwindet, d.h. klimatologisches Mittel wird angenommen) ist RMSE $\approx 0,7$ unabhängig vom Phasenfehler von mittlerer Größenordnung.
- Bei konstanter Amplitude nimmt der RMSE bei wachsendem Phasenfehler zunächst zu und erreicht bei $\delta = \pi$ ein Maximum (RMSE $\approx 1,4$), wird also doppelt so groß wie der reine Klimafall.
- Bei eher geringem Phasenfehler (z.B. $\delta \approx \pi/4$) *sinkt* der RMSE mit wachsendem Alpitudenfehler (bis zum Amplitudenfehler von ca. 0,6) und steigt erst bei weiterer Annäherung zum klimatologischen Mittel.
- Der RMSE reagiert auf Phasenfehler deutlich empfindlicher als auf Amplitudenfehler.

[5]Unwahrscheinlicher Glücksfall: RMSE = 0, wenn ohne Amplitudenfehler der Phasenfehler genau 2π beträgt

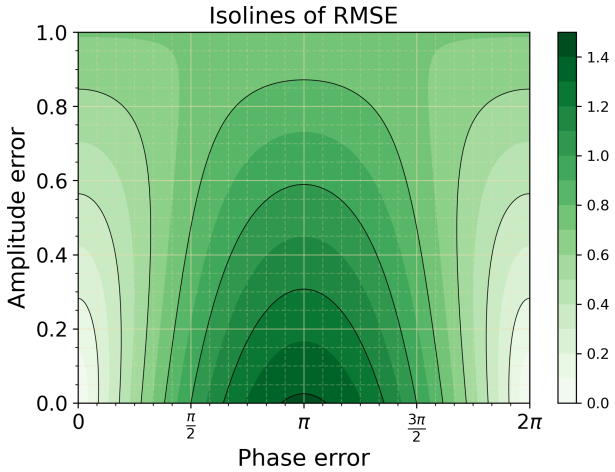

Abbildung 3: RMSE als Funktion von Phasenfehler und Amplitudenfehler

3 Zusammenfassung

Die im Vergleich zum operationellen ECMWF-Lauf günstigere Fehlerentwicklung der AI-Modelle hat ihre Ursache teilweise in der Unschärfe der Prognosefelder. Das ECMWF-Modell, nahe an der realen Physik der Atmosphäre, simuliert die Variablität mit großer Auflösung. Damit besteht besonders im atmosphärischen Kurzwellenbereich die Gefahr von Phasenfehlern, die sich entsprechend ungünstig auf den RMSE auswirken. Die AI-Modelle, ganz ohne Physik, neigen im Bereich der Kurzwellentröge zur Dämpfung der Amplituden, welche sich günstig auf den RMSE auswirkt. Allerdings sind die AI-Modelle schlechter in der Prognose von Extremereignissen. *Die AI-Modelle scheuen das Risiko.*

Selbstverständlich sind diese Umstände allen beteiligten Forschern bekannt, insbesondere die überproportionale Wirkung von Phasenfehlern auf den RMSE. Den zahlreichen fachfremden Lesern der Artikel in *Nature* und *Science* wird das hochwahrscheinlich nicht bewusst sein. Noch weniger bewusst ist es jenen, die ihre Information aus den Wissenschaftsformaten der Medien beziehen und sich dabei auf das Lesen von Schlagzeilen beschränken.

Unabhängig von der vorgetragenen Kritik ist die Qualität der AI-Prognosefelder beeindruckend. Mit Gewissheit werden AI-Methoden den weiteren Weg der Wettervorhersage auf vielfältige Weise verändern.

Literatur

Bi, Kaifeng u. a. (2023). "Accurate medium-range global weather forecasting with 3D neural networks". In: *Nature* 619, S. 533–538. DOI: https://doi.org/10.1038/s41586-023-06185-3.

Hersbach, Hans u. a. (2020). "The ERA5 global reanalysis". In: *Q J R Meteorol Soc.* 146, 1999–2049. DOI: https://doi.org/10.1002/qj.3803.

Lam, Ramy u. a. (2023). “Learning skillful medium-range global weather forecasting”. In: *Science*. DOI: https://doi.org/10.1126/science.adi2336.

Magnusson, Linus (2023). *Exploring machine-learning forecasts of extreme weather*. ECMWF. URL: https://www.ecmwf.int/en/newsletter/176/news/exploring-machine-learning-forecasts-extreme-weather.