Tellus
SINCE 1948

COLLECTION:
TELLUS 75TH JUBILEE SPECIAL COLLECTION: CELEBRATING LEGACY AND SHAPING THE FUTURE OF EARTH SYSTEM SCIENCE

PERSPECTIVE

Perspective Paper on Lorenz (1982): "Atmospheric Predictability Experiments with a Large Numerical Model"

LINUS MAGNUSSON

STOCKHOLM UNIVERSITY PRESS

ABSTRACT

In 1982, Edward Lorenz published a study evaluating the growth of forecast errors in 10-day European Centre for Medium-Range Weather Forecast (ECMWF) forecasts. This was one of the first articles to examine practical predictability in an operational global forecasting system and to estimate the limit of atmospheric predictability. The method of fitting a second-order polynomial to error growth as a function of error magnitude motivated numerous subsequent studies. The present article reviews some of this later work and uses recent ECMWF forecasts to illustrate how the method can be applied today. The error-growth model is also used to discuss recent developments in machine-learning-based forecasting. Although forecast quality has improved substantially since 1982, considerable errors remain for near-surface variables and still need to be addressed.

CORRESPONDING AUTHOR:
Linus Magnusson
European Centre for Medium-Range Weather Forecast (ECMWF), Department of Meteorology, Stockholm University, Sweden
linus.magnusson@ecmwf.int

KEYWORDS:
Numerical Weather Prediction (NWP); forecast verification

TO CITE THIS ARTICLE:
Magnusson, L. (2026). Perspective Paper on Lorenz (1982): "Atmospheric Predictability Experiments with a Large Numerical Model". *Tellus, 79*(1), 204–215. https://doi.org/10.16993/tellus.4141

1. INTRODUCTION

How far into the future can weather be predicted? This question has engaged scientists since the advent of numerical weather prediction (NWP). From the simplified model of Lorenz (1963), it became clear that, in a chaotic system, there exists a forecast horizon beyond which little useful information about the future state remains. This behaviour became known as deterministic chaos: a system governed by deterministic equations but highly sensitive to small perturbations. Although the model was not directly applicable to the real atmosphere, subsequent studies examined error growth in what were, at the time, state-of-the-art NWP models (e.g. Charney et al., 1966). Lorenz (1969a) theoretically described the growth of small-scale perturbations into larger scales and estimated the limit of predictability. In Lorenz (1969b), a large set of atmospheric analyses was used to identify atmospheric states that were relatively close to one another (i.e. analogues) and to compare how the differences between them evolved. That study also introduced the average maximum separation of two atmospheric states as an upper bound on forecast error.

In 1981, Edward Lorenz was on sabbatical from MIT at the European Centre for Medium-Range Weather Forecast (ECMWF). ECMWF had begun operational global forecasting in 1979, and Lorenz took the opportunity to investigate predictability in these forecasts. The results were published in Lorenz (1982), which is the focus of this review and is hereafter referred to as L82. L82 approached this question by attempting to estimate upper and lower bounds on forecast error. The upper bound represented what could be achieved by the forecast system at the time, as measured by the skill of a set of forecasts from winter 1980/81; this is what we would now refer to as practical predictability. The lower bound was estimated by comparing two successive forecasts, taking one of them as a proxy for the truth. In modern medium-range forecasting terminology, this is commonly referred to as potential predictability.

Forecast errors in the current forecasting system, which determine practical predictability, arise from both initial-condition errors and imperfections in the forecast model, leading to systematic (bias) and non-systematic errors. Systematic errors may be present almost from the start of the forecast and may also be embedded in the initial conditions because the forecast model is used in the data-assimilation cycle. If such biases grow during the forecast integration, this is often referred to as model drift. Systematic errors may also depend on the prevailing weather regime. Errors of this kind may appear random when verification is aggregated over all situations, but systematic when verification is conditioned on specific regimes. When forecast skill is evaluated against observations, differences in representation between observations (often point measurements) and model output (typically grid-box averages) must also be considered. Representation errors may be systematic (e.g. an anemometer located in an unusually windy position) or random (e.g. precipitation from a local shower). Because this variability occurs within the model grid box, it should be treated as uncertainty rather than as forecast error when observations are compared with model fields (Saetra et al., 2004).

Estimating the lower bound of forecast error can be approached by assuming that the model is perfect and by imposing a lower bound on the initial-condition error. This can be tested in a modelling framework by comparing two model simulations. However, this method relies on the model being able to simulate the intrinsic growth of errors correctly; otherwise, the estimated limit of predictability may be misleading.

L82 was one of the first articles to explore practical predictability in operational forecasts and to propose an approach to potential predictability. This article reviews several key aspects of L82 and relates them to subsequent research. The outline of this paper is as follows. Section 2 reviews the error-growth framework introduced in L82 and its later refinements; Section 3 examines how different mechanisms operating across scales shape error growth from small perturbations to long-range predictability; Section 4 assesses how far the forecast improvements envisaged by Lorenz have been realised in ECMWF forecasts since 1981; Section 5 considers what machine-learning forecasts reveal about the Lorenz framework and its interpretation; Section 6 discusses what L82 implies about the remaining scope for further gains in forecast skill; Section 7 shows how predictability and error growth depend on the variable being forecast; and Section 8 places L82 in a broader context by discussing related developments not addressed directly in the original paper.

2. ERROR GROWTH MODEL FROM L82 AND FURTHER DEVELOPMENTS

This section provides a brief review of the development of the parametric error-growth model used in L82. As background, Figure 1a shows the global root-mean-square error (RMSE) of 500 hPa geopotential height (z500) from a recent forecasting system. The forecasts considered here are from the control (unperturbed) members of the ECMWF sub-seasonal forecasting system based on the Integrated Forecast System (IFS) model, run with a horizontal grid spacing of 36 km for the period from 1 April 2025 to 31 March 2026 and covering lead times of up to 45 days. During this period, the ECMWF forecasting system used model version IFS 49r1. The forecasts are verified against the 12-hour-window 4D-Var analysis from the 9 km version of the same IFS model. Although the forecasts were initialised from the 8-hour-window analysis, the two analyses share a substantial part of the analysis error.

In L82, predictability was assessed by calculating the difference between two forecasts initialised X days apart but valid at the same time, Y days into the older forecast. If X = Y, the latter forecast is at day 0, corresponding to verification against the analysis of that day. This provides a measure of

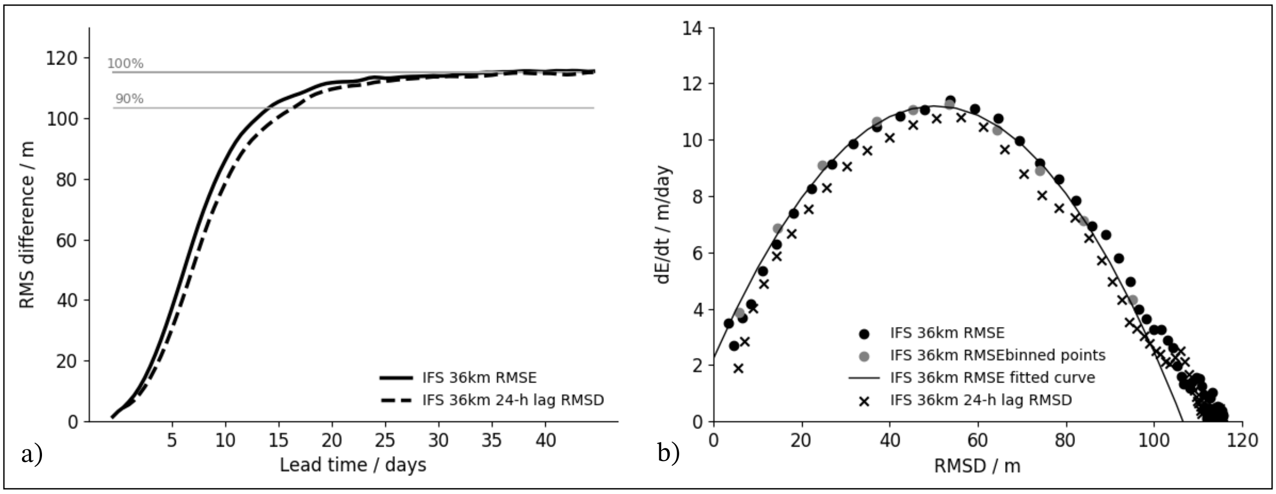

Figure 1 (a) Global z500 RMSE (solid) and RMS difference between forecasts initialised 24 hours apart (dashed), based on the control member of the ECMWF sub-seasonal forecast system for the period 1 April 2025 to 31 March 2026 and lead times up to 45 days. The 100% (thick dashed) and 90% (thin dashed) levels of the mean RMSE over days 30–45 are also shown. **(b)** dE/dt as a function of E for the same data as in panel (a), shown for RMSE (crosses) and lagged-forecast RMS difference (black dots). The solid line shows a second-order polynomial fitted to the RMSE data, and the binned RMSE values are shown as grey dots.

the practical predictability, that is, the forecast skill of the system in 1980/81. By contrast, comparing two forecasts initialised 1 day apart (X = 1 day) provides an estimate of potential predictability, under the assumption of a perfect model and with the initial-condition error taken to be equal to the 24-hour forecast error. In Figure 1a, the RMS curves for X = Y (forecast error) and X = 1 day (hereafter referred to as the 24-hour lagged forecast) are shown; these correspond to the uppermost and lowermost curves in Figure 1 of L82.[1]

A key element of L82 was to plot the growth rate of the forecast error (dE/dt) as a function of the error magnitude (E) itself. Figure 1b shows this relationship for the error data in Figure 1a. The method is based on the assumption that, in a chaotic system, the growth rate of the error depends on the magnitude of the error, leading to exponential error growth:

$$\frac{dE}{dt} = \alpha \mathrm{E} \qquad \text{Eq. 1}$$

where α is related to the effective Lyapunov exponent of the dynamical system and determines the intrinsic chaotic growth of errors (Boffetta et al., 1998, and references therein). Because the maximum error is constrained by the dynamical range of atmospheric variables and therefore cannot grow without bound, a negative quadratic term was added so that large errors grow more slowly and eventually approach an asymptotic level, as discussed in Lorenz (1969b). On this basis, Lorenz fitted the following function:

$$\frac{dE}{dt} = \mathrm{aE} + \mathrm{bE}^2 \qquad \text{Eq. 2}$$

The basic model of L82 can be reformulated slightly to better isolate the processes controlling error growth on different time scales. Dalcher and Kalnay (1987), hereafter DK87, used the following formulation:

$$\frac{dE}{dt} = (\alpha \mathrm{E} + \mathrm{S})\left(1 - \frac{E}{E_\infty}\right) \qquad \text{Eq. 3}$$

Here, E is taken to be the standard deviation of the error, that is, the error after bias correction. DK87 introduced a term *S*, representing a linear error-growth contribution that has its strongest influence when E is small. The addition of such a linear term had already been discussed in the lecture notes from the ECMWF Annual Seminar 1981 by C. Leith, published in Burridge and Källén (1984), and was probably introduced earlier in Leith (1978). By rewriting the quadratic term in Eq. 2, DK87 obtained the parameter E_∞ as the asymptotic limit of the error. Based on the decomposition of RMSE (Murphy, 1988), this asymptotic limit E_∞ is given by:

$$E_\infty = \sqrt{\overline{(F-C)^2} + \overline{(A-C)^2}} \qquad \text{Eq. 4}$$

where F was the forecast value valid at the verification time, A is the analysis, and C is the climatological value of the verified variable. The overbar denotes an average over many grid points and forecast cases. In the case of lagged-forecast differences, the variances of the two forecasts may be assumed to be equal.

Integrating the DK87 equation shows that the bias-corrected forecast error as a function of lead time is determined by the initial error (E_0), the intrinsic chaotic error growth of the system (α), the random model error term (S), and the variability of both the atmosphere and the model, which together determine E_∞.

One interpretation of *S* is that it represents the growth of errors in the absence of initial-condition errors, arising from model imperfections. In the case of lagged-forecast differences, both forecasts are produced with the same model, and S should therefore be zero, which appears to be a reasonable assumption from inspection of Figure 1b. However, the interpretation of S as a pure model-error term is debatable. Zhang et al. (2019) and Sun and Zhang (2020), for example, interpreted S as representing rapid error growth associated with convective-scale processes (see discussion in Section 3).

Magnusson and Källén (2013) used the DK87 equation to examine improvements in ECMWF forecast skill from 1986 to 2011, with the aim of separating the effects of reduced initial-condition error and reduced model error by analysing the evolution of the individual terms in Eq. 3. The study concluded that, in particular, the introduction of 4D-Var data assimilation (Rabier et al., 2000) in 1997, together with the subsequent improved use of satellite observations, contributed substantially to forecast improvement by reducing the initial-condition error and, consequently, the error growth associated with α.

One source of uncertainty when applying Eq. 3 to estimate the contributing terms is the calculation of the time derivative dE/dt. To address this difficulty, Žagar et al. (2017) introduced a curve-fitting approach applied directly to the error itself, rather than to dE/dt, while retaining parameters that can still be related to the terms in Eq. 3.

3. ERROR GROWTH FROM SMALL ERRORS TO LONG-RANGE PREDICTABILITY: IMPLICATIONS FOR THE SHAPE OF THE PARABOLA

Although the first data point in the L82 version of the figure (Figure 2 in L82) was still relatively far from E = 0, the points nevertheless resembled a parabola. In Figure 1b of the present paper, a second-order polynomial has been fitted to the data following the approach of L82, after first binning the points. The binned values are shown in Figure 1b as grey dots. Here, 10 equally sized bins were used, but only RMSE values below 100 m were included in the fit in order to avoid the influence of long-range forecasts (see below). The current data follow the parabolic form reasonably well, although some deviations are evident for very small and very large errors.

Turning first to the growth of small errors, L82 states in the introduction:

> *However, the models which have indicated a doubling time of several days do not explicitly contain smaller-scale features ranging in size from squall lines and thunderstorms to dust whirls, whose amplitudes should double in hours or minutes or less. The effects of these features upon the larger scales appear in the models in parameterized form, but the uncertainties in these features do not.*

A central point in this quotation is that the atmosphere contains processes with intrinsically rapid error growth on small spatial scales, and that these errors can project onto larger scales through upscale transfer (Lorenz, 1969a). The mechanism of upscale error growth from convective to synoptic scales has been, and remains, an area of active research. Zhang et al. (2007) proposed a three-stage framework consisting of rapid error growth associated with moist convective instability, a transition in which unbalanced perturbations adjust towards balanced structures through geostrophic adjustment, and subsequent growth through baroclinic instability. A similar conceptual model was diagnosed by Baumgart et al. (2018).

Several studies using the L82-style dE/dt versus E diagnostic reported apparently "super-exponential" growth for the smallest errors (e.g. Nicolis, Vannitsem, & Royer, 1995), consistent with the idea of rapid convective- or mesoscale error growth. However, what appears to be very rapid initial growth may also arise from the changing relationship between analysis errors and forecast errors: at short lead times, forecast errors are partially correlated with the analysis error, and as that correlation decays the apparent growth rate is inflated (Simmons & Hollingsworth, 2002). In Figure 1b, this effect is most evident for the first cross in the RMSE data; beyond that point, there is no clear evidence of sustained super-exponential behaviour.

At the other end of the parabola, Lorenz acknowledged already in the introduction that "*We shall not be concerned with climatic or other long-range predictability, which should*

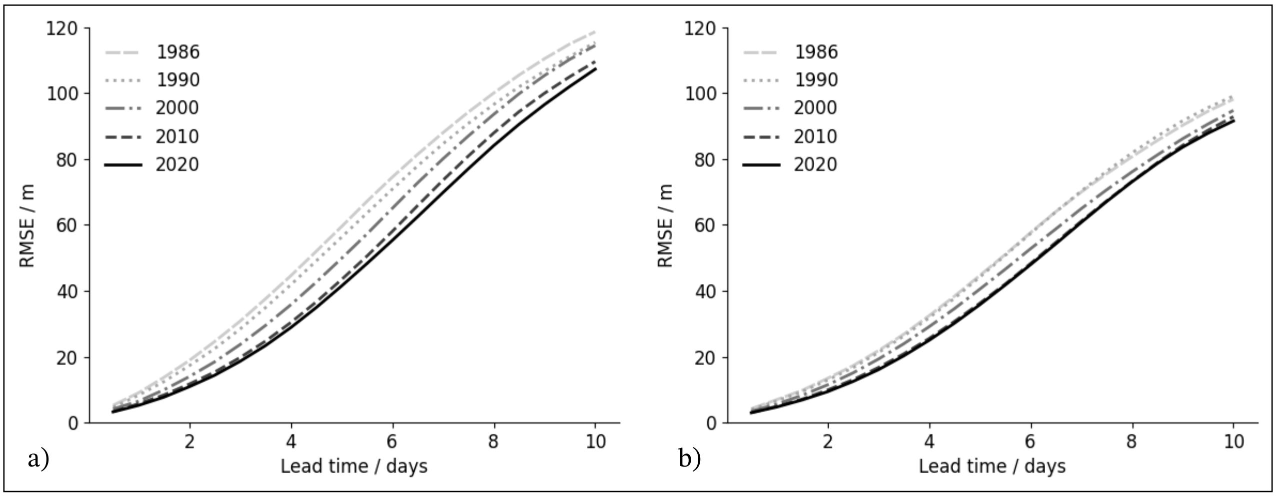

Figure 2 (a) RMSE of z500 for operational ECMWF forecasts over the Northern Hemisphere for selected years (1986, 1990, 2000, 2010, and 2020). **(b)** As in panel **(a)**, but for forecasts produced within the ERA5 system.

exist if certain features of the atmosphere can still be predicted when most of the atmosphere cannot". Since L82, extensive research has examined forms of predictability that do not arise primarily from rapid, internally generated growth of small errors, but instead from relatively slowly varying components of the Earth system that act as boundary conditions or external forcings for the mid-latitude atmosphere (e.g. Stan et al., 2017). Examples include tropical forcing associated with slow variations in tropical heating from organised convection (e.g Hoskins & Karoly, 1981) such as Madden-Julian Oscillation (MJO)-related variability, lower-boundary influences from sea-surface temperature, terrestrial influences from soil moisture, snow and sea-ice, and upper-boundary influences from the stratosphere, such as the Quasi-Biennial Oscillation (QBO) and variations in the polar vortex.

Within the framework of the L82/DK87 error-growth model, these slowly varying influences may be interpreted as "slow modes" with weaker intrinsic growth and a longer time to approach saturation (Palmer, 1993; Peña & Kalnay, 2004). When rapid, chaotic atmospheric error growth is combined with slow, externally forced variability, the total error need not approach its asymptotic level at a single rate. Instead, rapid initial growth may be followed by a more gradual increase as the remaining predictable signal becomes increasingly tied to slowly evolving boundary or forcing anomalies. This mixed behaviour is consistent with the tendency in Figure 1b for the largest-error points to lie to the right of the fitted parabola.

Departures from a perfect parabola in the dE/dt versus E plot can therefore reflect the superposition of several growth mechanisms with different characteristic time scales. For very small errors, convective or mesoscale processes, and/or the rapid decorrelation of forecast and analysis errors, can produce an initially steep slope, making the left-hand side of the curve appear too high relative to a simple quadratic fit. At larger errors, the remaining predictable component of the flow may be increasingly influenced by slowly varying, externally forced components, implying weaker effective growth and hence a slower approach to saturation. The resulting curve may therefore be skewed rather than symmetric. Stroe and Royer (1993) examined a more general functional form for the growth law, later revised by Simmons et al. (1995), in which the dependence on normalised error (E/E_∞) is not constrained to be linear, as in DK87. They found that the effective exponent varies with season and is often lower than 1, consistent with a slower-than-quadratic reduction in growth as saturation is approached. This skewness has practical implications when diagnosing the different components in Eq. 3, because parameter estimates will depend on whether the fitted data primarily sample the left- or right-hand side of the curve.

4. HOW HAS THE SKILL OF ECMWF FORECASTS EVOLVED SINCE 1981?

In L82, the potential for forecast improvement was summarised as follows:

> *Assuming that we have correctly estimated the doubling time, we find that, even without further improvement in one-day forecasting, we may eventually make ten-day forecasts as good as present seven-day forecasts, and 13.5-day forecasts as good as present ten-day forecasts. Cutting the one-day RMSE error in half should add another two days to the range of predictability; possibly we may cut this error in half once more.*

It is not entirely clear how Lorenz derived the statement that "ten-day forecasts as good as present seven-day forecasts". One possible interpretation is that it was based on (1) correcting the forecast mean state and variability over the Northern Hemisphere (20°N–90°N), and (2) assuming that forecast skill over the Southern Hemisphere (20°S–90°S) could eventually match that over the Northern Hemisphere. Alternatively, the statement may have been based on comparing the actual forecast error with the difference between two successive forecasts initialised 1 day apart, as shown in Figure 1 of L82. The latter interpretation would assume that the forecast model can become a perfect representation of the atmosphere, so that the remaining errors arise entirely from the initial conditions. However, if this were the intended method, the implication would be closer to a ten-day forecast performing as well as a present-day 6-day forecast, based on Figure 1 of L82.

Since the start of operational forecasting in 1979, the skill of ECMWF forecasts has improved substantially. Figure 2 shows the RMSE of z500 over the Northern Hemisphere for five snapshot years (1986, 1990, 2000, 2010, and 2020). The figure updates Figure 1 of Magnusson and Källén (2013) by including 1986 and 2020; ECMWF does not retain regular-archive forecasts prior to 1986. The forecasts are verified against the 5th generation of ECMWF Reanalysis (ERA5; Hersbach et al., 2020).

With regard to the statement in L82 about reducing the 1-day error, the 1-day RMSE in ECMWF operational forecasts over the Northern Hemisphere decreased from 19.8 m in 1986 to 5.0 m in 2020, while over the Southern Hemisphere it decreased from 36.9 m to 5.8 m. Thus, since 1986, the error has been halved twice in the Northern Hemisphere and reduced to approximately one-sixth in the Southern Hemisphere; the reduction is even larger if the comparison is extended back to 1981, as documented in Simmons et al. (1995). In the operational ECMWF forecasts, the day-6 error in 2020 is comparable to the day-3 error in 1990 (Figure 2a). These improvements in 1-day forecasts reflect a combination of increased observational coverage, especially from satellites, improved data assimilation that makes better use of the observations, and improved forecast models that provide a more accurate first guess for the assimilation system (Bauer, Thorpe, & Brunet, 2015).

To isolate the impact of the observing system, one can instead examine forecast improvement based on forecasts from ERA5 reanalysis which uses the same forecast model and data-assimilation system throughout the full period but with an evolving observing network (Figure 2b). In this case, the 1-day error decreases from 7.0 m to 4.7 m over the Northern Hemisphere and from 9.0 m to 5.2 m over the Southern Hemisphere. The lower errors for the ERA5 forecasts than for the operational ECMWF forecasts in 2020 are partly due to the ERA5 analysis using a longer 12-hour assimilation window, which incorporates observations 4 hours later into the forecast, and partly because ERA5 analyses are also used for verification, which favours the ERA5 forecasts.

5. LORENZ ERROR GROWTH MODEL IN THE LIGHT OF MACHINE-LEARNING-BASED FORECASTS

The emergence of machine-learning weather prediction (MLWP; Bi et al., 2023; Lam et al., 2023; Lang et al., 2024) has brought renewed attention to several of the issues discussed in L82. Figure 3 shows the RMSE and the RMS difference for lagged forecasts from the 9 km version of the physics-based IFS model (the same version as discussed in the previous sections) and from the deterministic ECMWF Artificial Intelligence Forecasting System (AIFS-single), with a horizontal resolution of approximately 25 km (Lang et al., 2024; Moldovan et al., 2026). The AIFS-single are initialised from the same initial conditions as the IFS forecasts. Verification is based on the same region and period as in Figure 1.

AIFS-single exhibits slower error growth than the IFS forecasts. This immediately raises the question of which components of the error-growth model account for the difference. Because both forecast systems are initialised from the same IFS analysis, differences in the initial-condition error (E_0) can be excluded.

A central issue in the discussion of first-generation ML-based forecasting systems has been the role of smoothing. This is not a new concern: in Section 4, L82 discusses the implications of differences in mean state and variance between the forecast fields and the verifying analysis. Lorenz states:

> *On the other hand, when we make the variance and hence the standard deviation of X' equal to those of Y, we may actually increase the mean square of Y-X'; this is especially likely to be so if X has a smaller variance than Y, and if X and Y are not highly correlated. … the possibility of an increased mean square points to a serious shortcoming of the mean-square error, or root-mean-square error, as a general measure of the goodness of a forecasting procedure.*

If a forecast is smoother than the truth, the RMSE can be reduced by avoiding the double-penalty effect, whereby spatially displaced but otherwise realistic structures are penalised twice. In the limiting case in which all skill is lost and forecasts and analyses are uncorrelated, a forecast with the correct variance has an RMSE that is $\sqrt{2}$ times larger than that of a forecast equal to climatology, as follows from Eq. 4. For example, smoothing is obtained in an ensemble-mean forecast, as the ensemble mean approaches climatology then the ensemble members become mutually uncorrelated (Bengtsson, Magnusson, & Källén, 2008).

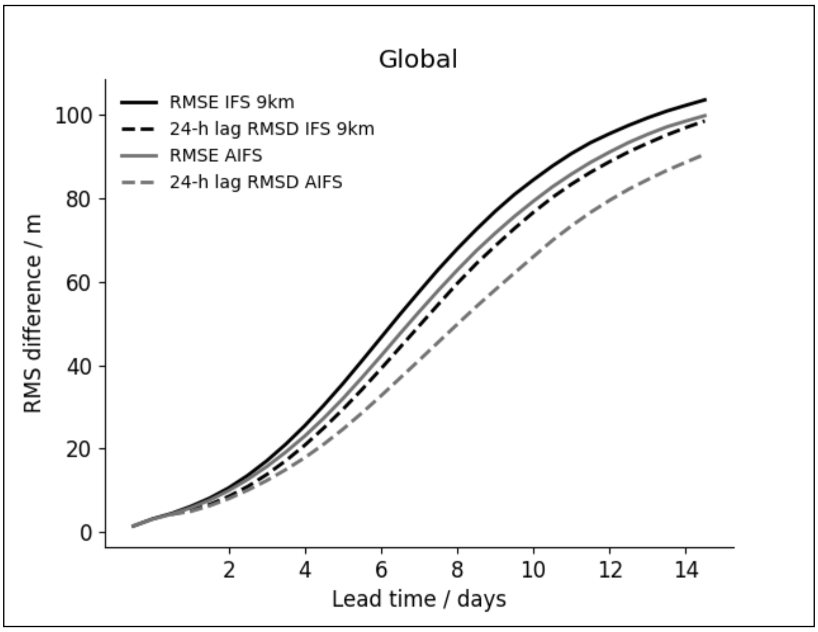

Figure 3 Global z500 RMS(E) for the period 1 April 2025 to 31 March 2026, shown for the 9-km IFS forecast (black) and AIFS forecast (grey). RMSE is shown by solid lines and the 24-hour lagged-forecast RMS difference by dashed lines.

In NWP, excessive numerical diffusion historically produced forecast fields that were too smooth. Over the past decade, however, forecast variability for upper-air variables such as z500 has become very close to that of the verifying analysis. The issue of smoothness has re-emerged with the advent of MLWP. Because these systems are currently trained to minimise the (R)MSE at specific lead times (e.g. 72 hours in the case of AIFS; Lang et al., 2024), the training procedure can favour different degrees of smoothing. In AIFS, the smoothing appears to be largely confined to sub-synoptic scales (wavenumbers greater than 20), which are less predictable (e.g. McTaggart-Cowan et al., 2026). This loss of forecast variability increases somewhat with lead time, so that AIFS acts, in effect, somewhat as a filter on the unpredictable scales. However, the relative reduction in the asymptotic error is much smaller than for an ensemble-mean forecast (not shown) and is probably insufficient to explain the full difference between IFS and AIFS.

One might also expect ML-based forecasts to reduce both systematic and random model errors. For several variables and regions, this is indeed the case. For z500 outside the tropics, however, the RMSE is dominated by the random component of the error. Applying a bias correction before calculating the RMSE for z500 produces only a small change (not shown).

A reduction in random model error in AIFS would act to reduce the term S in Eq. 3. However, if one considers the lagged-forecast difference, which contains no contribution from model error, the contrast between IFS and AIFS-single is even larger for the lagged-forecast curves (Figure 3, dashed lines). We can therefore conclude that the difference is (at least partly) associated with the intrinsic chaotic error growth (α) and/or with the forecast-variability term ($(F-C)$ in Eq. 3), which influences the asymptotic limit (E_∞), without excluding some contribution from S as well.

6. SO HOW MUCH FURTHER CAN THE FORECASTS BE IMPROVED?

In L82 it was stated, *"A study by Lorenz (1969a) indicated that even if the larger scales could be observed perfectly, the inevitable uncertainties in the smaller scales would after a day or so induce errors in the larger scales…".* This would put a limit on the predictability of the atmosphere. One way to explore this limit is to evaluate the divergence of two forecasts during model integration. In L82, this was done in a simple way by comparing consecutive operational forecasts. In 1980, forecasts were produced once per day, hence the 24-hour separation between initial times. In Figure 4, we compare the result from L82 with recent forecasts for the period 1 December 2023 to 10 March 2024 based on the 36-km ECMWF forecasts. The RMSE for the 9-km version, available out to 15 days, is also included, together with the RMS difference from the "accidental twin" experiment (see below). Finally, the figure includes the average RMSE of the 36-km forecast over days 40–45 as a measure of the asymptotic limit, together with 90% of that value.

Comparing the RMSE and the lagged-forecast difference shows that the two curves for the recent period are much closer together than in L82, indicating that the current forecasting system is much closer to the "perfect-model". However, the lagged-forecast curve in L82 was not a good estimate of the upper limit of predictability because smoothing reduced the asymptotic level of the lagged-forecast difference.

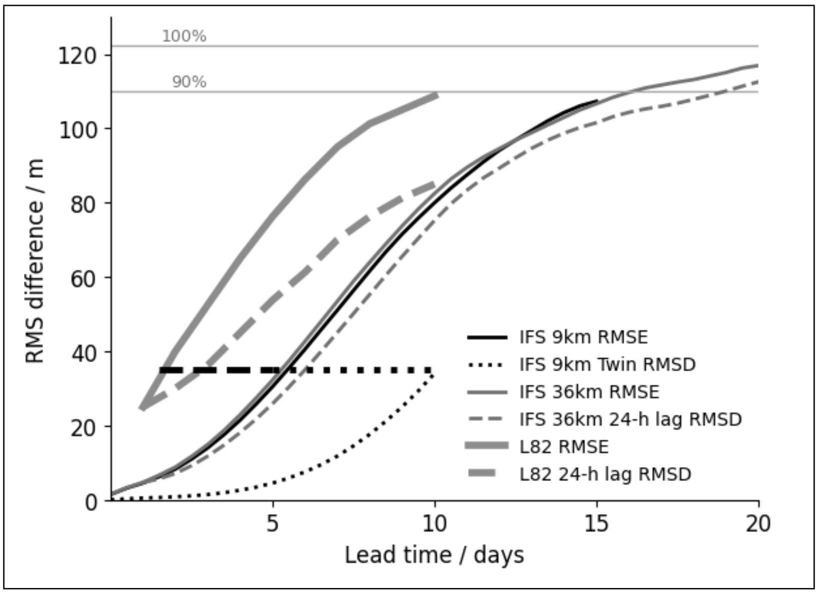

Figure 4 z500 RMS(E) for the period 1 December 2023 to 10 March 2024. Results are shown for 1980/81 (thick grey; based on Figure 1 of Lorenz, 1982), for the 2023/24 36-km forecast system (grey), and for the 2023/24 9-km forecast system (black). The lagged-forecast differences from Lorenz (1982) and from the 36-km forecast system are shown by dashed lines, and the RMS difference from the accidental 9-km twin experiment is shown by a black dotted line. The figure also includes the 100% and 90% levels of the mean RMSE over days 30–45. The black dashed line marks the 35 m error level between days 1.5 and 5, and the black dotted line marks the same error level between days 5 and 10.

An alternative approach to explore the potential predictability is to add perturbations to the initial conditions and rerun the forecast with the same model to examine how the difference between the two “twins” grows. This provides an opportunity to test different perturbation amplitudes, corresponding to different assumptions about the quality of future initial conditions. The method has the caveat that the perturbations must be constructed so that their growth is representative of the expected forecast-error growth. Twin experiments also assume a perfect model, that is, zero model error, and require that the model realistically simulates the intrinsic growth of errors. While this is generally believed to be a reasonable assumption on synoptic scales in the mid-latitudes, it has been more questionable at convective scales, as already noted in L82.

In recent years, several studies have attempted to estimate the limit of predictability using model twin experiments (e.g. Judt, 2018; Selz, 2019; Zhang et al., 2019). These studies used kilometre-scale global models, which are thought to capture error growth on convective scales more realistically, as discussed above. They all applied perturbations of very small amplitude, typically 1–10% of the current analysis error. Zhang et al. (2019), however, found little sensitivity of synoptic-scale error growth to explicitly resolving convection. That study concluded that the then-current predictability limit could be extended by about 5 days if initial-condition errors were reduced to extremely low levels.

At ECMWF, two almost identical operational forecasts were run from late June 2023 to November 2024 using a 9-km model grid spacing. Although the two forecasts had identical initial conditions and model formulation, the surface orography differed by less than one metre, and some other static surface specifications also differed slightly, so the runs were not bit-identical. This “accidental twin experiment” therefore provided a pair of forecasts whose differences grew from minimal perturbations. The RMS difference between the two forecasts is included in Figure 4.

The results in Figure 4 show that the current 5-day forecast error (35 m) is similar to the 1.5-day error in 1980/81, implying an improvement of about 3.5 days over the past 44 years. In the accidental twin experiment, the same level of difference is reached at around day 10, as marked by the black dashed and dotted lines in Figure 4. For this dataset, 90% of the 25–30 day error level is reached just after day 15 for the 36-km version of the current model (the final point of the 15-day forecast for the 9-km version lies just below the 90% line). If the accidental twin experiment were to continue growing beyond day 10 in a similar manner to the lagged forecast difference from the operational forecast, this would suggest an upper bound on predictability of approximately 20–22 days for the metric used here. However, the choice of threshold, here 90%, for defining the point at which no useful skill remains introduces an unavoidable degree of arbitrariness into estimates of the predictability limit.

7. FITTING THE LORENZ MODEL TO OTHER PARAMETERS

In the conclusion of L82, it is also stated:

> *First, we have examined only 500-mb height data. It seems probable that strongly coupled elements will have similar ranges of predictability, so that our conclusion may well also apply to middle-latitude tropospheric temperature and wind prediction, but they may be quite unrealistic for such elements as tropical cloudiness and rainfall.*

To illustrate this point, Figure 5 shows the RMSE over Europe for both z500 and 10-metre wind speed, verified against observations (SYNOP for wind and radiosondes for z500) and averaged over the period 1 April 2025 to 31 March 2026. It is immediately apparent that the 12-hour forecast error (the first point) for wind speed is much larger, relative to the asymptotic limit, than for z500. This is partly related to sub-grid-scale variability (representativeness error), which is captured by the observations but not by

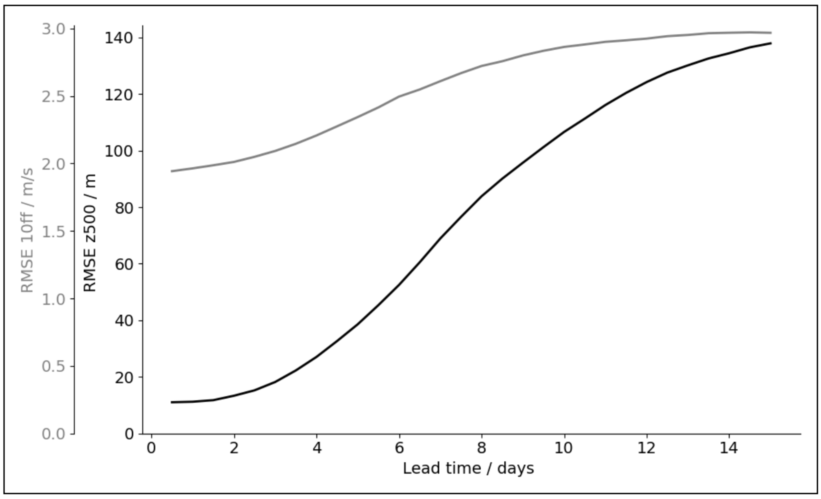

Figure 5 RMSE of 10-metre wind speed and z500 over Europe, verified against observations, for the period 1 April 2025 to 31 March 2026.

the model output, as well as to rapidly developing model biases (see the discussion of different error types in the Introduction section). Another possible contributor is larger analysis error in near-surface winds. Nevertheless, during the forecast the loss of the remaining predictability is broadly similar to that seen for z500, as anticipated in L82. The maximum error growth, however, appears to occur at somewhat shorter lead times for 10-metre wind speed than for z500. One possible explanation is that different variables are most active on different horizontal scales and therefore exhibit different characteristic time scales of error growth, since predictability is lost more rapidly on shorter time scales (Lorenz, 1969a; Žagar et al., 2017).

In the medium range, variability in the mid-latitudes is dominated by Rossby-wave dynamics, which strongly imprint on z500. In the tropics, variability is more complex: strong convection in the ITCZ can produce rapid error growth, while the large-scale flow is also modulated by much slower modes such as the MJO and ENSO. The implications for predictability were explored by Judt (2020). On synoptic scales, tropical variability is also influenced by atmospheric tropical wave modes. Žagar et al. (2017) applied the L82 framework for different types of tropical waves using a normal-mode basis that takes both wind and geopotential into account. Using the ensemble spread as a proxy for the forecast error, they found that unbalanced (gravity-wave) modes exhibited a faster loss of predictability than balanced modes such as equatorial Rossby waves.

8. DISCUSSION

The landmark paper by Lorenz (1982) provided one of the first published assessments of global forecast skill from an operational NWP model and, at the same time, introduced a conceptual framework for understanding forecast errors and their possible reduction.

L82 examined forecast skill averaged over 100 days during boreal winter. While long-term improvements in forecast systems are important to document, it is also reasonable to ask how relevant an average predictability limit is in practice. Although the RMSE for a season or a year provides a measure of mean forecast performance, forecast skill fluctuates substantially from day to day. Occasional episodes of particularly rapid error growth are commonly referred to as forecast busts or drop-outs. The error growth associated with such events has been the subject of extensive research (e.g. Grams, Magnusson, & Madonna, 2018; Hauser et al., 2026; Magnusson, 2017; Rodwell et al., 2013).

One way to address the day-to-day variability of forecast skill is to run an ensemble forecast system in order to estimate the evolving forecast uncertainty. To obtain a reliable estimate of that uncertainty, however, all major sources of forecast error must be represented (Rodwell et al., 2016). Here too, the L82 framework remains informative, as explored by Bengtsson, Magnusson, & Källén (2008). In this context, an ensemble system must simulate realistic initial-condition errors (Magnusson, Nycander, & Källén, 2009), include a representation of non-systematic model error (Leutbecher et al., 2017), and employ a model formulation that yields realistic intrinsic error growth and asymptotic error limits (Bengtsson et al., 2008). In the past, it was often difficult to obtain sufficient ensemble spread relative to the ensemble-mean error. In many cases, this was due to insufficient variability in the forecast model, which led to too low an asymptotic limit of the ensemble spread. This behaviour is captured much better today, at least for upper-air variables such as z500 (see, for example, Figure 9 in Haiden et al., 2025). For surface variables, however, sub-grid variability, observation errors, and systematic forecast errors must be taken into account before ensemble spread can be meaningfully compared with ensemble-mean error (Saetra et al., 2004). Understanding the different components of forecast error is therefore essential when evaluating an ensemble prediction system.

Coming back to the first question of "How far into the future can weather be predicted? and following "how can the forecasting system be improved to narrow the gap between practical and potential predictability?", the accidental twin experiment presented here may serve as an upper estimate of predictability. However, it assumes that it realistically captures the intrinsic growth of errors, but probably more importantly, it is also not realistic to expect that such minimal initial-condition errors can be achieved in practice. This is partly due to limitations in observational coverage and to observation errors that constrain mesoscale data assimilation (Majumdar et al., 2021). Moreover, the rapid growth of convective-scale errors and their upscale transfer raise the question of whether reducing initial errors on convective scales will be worth the required effort. On the other hand, the results for 10-metre wind error suggest that, for near-surface variables, there remains considerable room for improvement before this limit is approached. Whether this is most effectively achieved through model development, post-processing, or ML-based forecasting remains to be seen. While efforts to reduce these errors continue, the associated uncertainties must also be estimated and communicated appropriately.

Recent results from MLWP have begun to challenge current estimates of error growth and to suggest a substantially longer potential predictability horizon (Vonich & Hakim, 2025). With regard to chaotic error growth, Selz and Craig (2023), followed by Selz and Craig (2025), discussed slower difference growth on small, convective scales in MLWP models. However, the results presented here also indicate reduced growth at larger error magnitudes. This raises several questions: (1) What is the relationship between lead-time-dependent smoothing at specific scales, which affects E_∞, and the intrinsic chaotic error growth parameter α? (2) Does smoothing at convective scales reduce α for synoptic scales? (3) How does α diagnosed from the difference between two forecasts (e.g lagged forecast), relate to α diagnosed from the difference between a forecast and the truth (i.e forecast error)?

Answering these questions will require further research. In particular, addressing the first question may require extending the growth model to individual wavenumbers, or to spectral bands, while allowing E_∞ to depend explicitly on lead time. The second question leads back to the problem of scale interactions and upscale error growth discussed already by Lorenz (1969a) and more recently by Rotunno and Snyder (2008) and Sun and Zhang (2020). One possible hypothesis is that smoothing a strong but intrinsically unpredictable convective outflow could reduce errors in the Rossby-wave forcing, following the mechanism described by Zhang et al. (2007), and thereby reduce downstream error growth along the Rossby-wave guide. This idea, however, remains to be tested. Addressing the third question may require genuinely new developments in the study of forecast-error growth.

Taken together, these considerations show that forecast-error growth and the limit of atmospheric predictability remain active and important areas of research, nearly half a century after Edward Lorenz's seminal study in *Tellus*.

NOTE

1 In L82 the RMS difference marker is on the lead-time for the older forecast, i.e for Day 2 – Day 1 forecasts is on Day 2. If a forecaster was to make a forecast based on a "lagged forecast principles", the skill should be measure at the lead time of the latter forecast in the pair, not to make use of forecast information from the future. In this paper we keep the definition from L82 for consistency.

ACKNOWLEDGEMENTS

I would like to thank the editor, Bjorn Stevens, and my ECMWF colleagues Adrian Simmons and Thomas Haiden for providing valuable insights during the preparation of this paper. I am also grateful for the comments of two reviewers, which helped to improve the paper.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Linus Magnusson orcid.org/0000-0003-4707-2231
European Centre for Medium-Range Weather Forecast (ECMWF), Department of Meteorology, Stockholm University, Sweden

REFERENCES

Bauer, P., Thorpe, A., & **Brunet, G.** (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*, 47–55. https://doi.org/10.1038/nature14956

Baumgart, M., Riemer, M., Wirth, V., Teubler, F., & **Lang, S. T. K.** (2018). Potential-vorticity dynamics of forecast errors: a quantitative case study. *Monthly Weather Review*, *146*, 1405–1425. https://doi.org/10.1175/MWR-D-17-0196.1

Bengtsson, L. K., Magnusson, L., & **Källén, E.** (2008). Independent estimations of the asymptotic variability in an ensemble forecast system. *Monthly Weather Review*, *136*, 4105–4112. https://doi.org/10.1175/2008MWR2526.1

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & **Tian, Q.** (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*, 533–538. https://doi.org/10.1038/s41586-023-06185-3

Boffetta, G., Giuliani, P., Paladin, G., & **Vulpiani, A.** (1998). An extension of the Lyapunov analysis for the predictability problem. *Journal of the Atmospheric Sciences*, *55*, 3409–3416. https://doi.org/10.1175/1520-0469(1998)055<3409:AEOTLA>2.0.CO;2

Burridge, D. M., & **Källén, E.** (Eds.) (1984). *Problems and prospects in long and medium range weather forecasting*. Springer. https://doi.org/10.1007/978-3-642-82132-5

Charney, J. G., Fleagle, R. G., Riehl, H., Lally, V. E., & **Wark, D. Q.** (1966). The feasibility of a global observation and analysis experiment. *BAMS*, 47, 200–220.

Dalcher, A., & **Kalnay, E.** (1987). Error growth and predictability in operational ECMWF forecasts. *Tellus A*, *39*, 474–491. https://doi.org/10.3402/tellusa.v39i5.11774

Grams, C. M., Magnusson, L., & **Madonna, E.** (2018). An atmospheric-dynamics perspective on the amplification and propagation of forecast error in numerical weather prediction models: a case study. *Quarterly Journal of the Royal Meteorological Society*, *144*, 2577–2591. https://doi.org/10.1002/qj.3353

Haiden, T., Janousek, M., Vitart, F., Prates, F., Maier-Gerber, M., Li, C. W. Y., & **Chevallier, M.** (2025). *Evaluation of ECMWF forecasts*. ECMWF Technical Memorandum 931.

Hauser, S., Cavallo, S. M., Magnusson, L., Martin, J. E., & **Parsons, D. B.** (2026). Exceptionally poor and good medium-range forecasts of the large-scale circulation over Europe in ERA5 reforecasts. *Quarterly Journal of the Royal Meteorological Society*, *152*, e70117. https://doi.org/10.1002/qj.70117

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M. *et al.* (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Hoskins, B. J., & **Karoly, D. J.** (1981). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *Journal of the Atmospheric Sciences*, *38*, 1179–1196. https://doi.org/10.1175/1520-0469(1981)038%3C1179:TSLROA%3E2.0.CO;2

Judt, F. (2018). Insights into atmospheric predictability through global convection-permitting model simulations. *Journal of the Atmospheric Sciences*, *75*, 1477–1497. https://doi.org/10.1175/JAS-D-17-0343.1

Judt, F. (2020). Atmospheric predictability of the tropics, middle latitudes, and polar regions explored through global storm-resolving simulations. *Journal of the Atmospheric Sciences*, *77*, 257–276. https://doi.org/10.1175/JAS-D-19-0116.1

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & **Battaglia, P.** (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*, 1416–1421. https://doi.org/10.1126/science.adi2336

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Ben Bouallègue, Z., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., & **Rabier, F.** (2024). AIFS: ECMWF's data-driven forecasting system. *arXiv*. https://doi.org/10.48550/arXiv.2406.01465

Leith, C. E. (1978). Objective methods for weather prediction. *Annual Review of Fluid Mechanics*, *10*, 107–128. https://doi.org/10.1146/annurev.fl.10.010178.000543

Leutbecher, M. *et al.* (2017). Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, *143*, 2315–2339. https://doi.org/10.1002/qj.3094

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141. https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2

Lorenz, E. N. (1969a). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*, 289–307. https://doi.org/10.3402/tellusa.v21i3.10086

Lorenz, E. N. (1969b). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, *26*, 636–646. https://doi.org/10.1175/1520-0469(1969)026%3C0636:APARBN%3E2.0.CO;2

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*, 505–513. https://doi.org/10.1111/j.2153-3490.1982.tb01839.x

Magnusson, L. (2017). Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society*, *143*, 2129–2142. https://doi.org/10.1002/qj.3072

Magnusson, L., & **Källén, E.** (2013). Factors influencing skill improvements in the ECMWF forecasting system. *Monthly Weather Review*, *141*, 3142–3153. https://doi.org/10.1175/MWR-D-12-00318.1

Magnusson, L., Nycander, J., & **Källén, E.** (2009). Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus A*, *61*, 194–209. https://doi.org/10.1111/j.1600-0870.2008.00385.x

Majumdar, S. J. *et al.* (2021) Multiscale forecasting of high-impact weather: current status and future challenges. *Bulletin of the American Meteorological Society*, *102*, E635–E659. https://doi.org/10.1175/BAMS-D-20-0111.1

McTaggart-Cowan, R. *et al.* (2026) WP-MIP: An artificial intelligence, hybrid and physically based model intercomparison project for weather prediction. *arXiv*. https://doi.org/10.5194/ems2026-598

Moldovan, G., Pinnington, E., Prieto Nemesio, A., Lang, S., Ben Bouallègue, Z., Dramsch, J., Alexe, M., Santa Cruz, M., Hahner, S., Cook, H., Theissen, H., Clare, M., O'Brien, C., Polster, J., Magnusson, L., Mertes, G., Pinault, F., Raoult, B., de Rosnay, P., Forbes, R., & **Chantry, M.** (2026). AIFS 1.1.0: an update to ECMWF's machine-learned weather forecast model AIFS. *EGUsphere*, accepted. https://doi.org/10.5194/egusphere-2025-4716

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, *116*, 2417–2424. https://doi.org/10.1175/1520-0493(1988)116%3C2417:SSBOTM%3E2.0.CO;2

Nicolis, C., Vannitsem, S., & **Royer, J.-F.** (1995). Short-range predictability of the atmosphere: mechanisms for superexponential error growth. *Quarterly Journal of the Royal Meteorological Society*, *121*, 705–722. https://doi.org/10.1002/qj.49712152312

Palmer, T. N. (1993). Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, *74*, 49–65. https://doi.org/10.1175/1520-0477(1993)074%3C0049:ERAPAT%3E2.0.CO;2

Peña, M., & **Kalnay, E.** (2004). Separating fast and slow modes in coupled chaotic systems. *Nonlinear Processes in Geophysics*, *11*, 319–327. https://doi.org/10.5194/npg-11-319-2004

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., & **Simmons, A.** (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, *126*, 1143–1170. https://doi.org/10.1002/qj.49712656415

Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., Richardson, D. S., & **Yamaguchi, M.** (2016). Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *142*, 443–454. https://doi.org/10.1002/qj.2663

Rodwell, M. J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L., Källén, E., Klocke, D., Lopez, P., McNally, T., Persson, A., Prates, F., & **Wedi, N.** (2013). Characteristics of occasional poor medium-range weather forecasts for Europe. *Bulletin of the American Meteorological Society*, *94*, 1393–1405. https://doi.org/10.1175/BAMS-D-12-00099.1

Rotunno, R., & **Snyder, C.** (2008). A generalization of Lorenz's model for the predictability of flows with many scales of motion. *Journal of the Atmospheric Sciences*, *65*, 1063–1076. https://doi.org/10.1175/2007JAS2449.1

Saetra, Ø., Hersbach, H., Bidlot, J., & **Richardson, D. S.** (2004). Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, *132*, 1487–1501. https://doi.org/10.1175/1520-0493(2004)132<1487:EOOEOT>2.0.CO;2

Selz, T. (2019). Estimating the intrinsic limit of predictability using a stochastic convection scheme. *Journal of the Atmospheric Sciences*, *76*, 757–765. https://doi.org/10.1175/JAS-D-17-0373.1

Selz, T., & **Craig, G. C.** (2023). Can artificial intelligence-based weather prediction models simulate the butterfly effect? *Geophysical Research Letters*, *50*, e2023GL105747. https://doi.org/10.1029/2023GL105747

Selz, T., & **Craig, G. C.** (2025). Can AI-based weather prediction models simulate the butterfly effect? The role of architecture and implementation. *ESS Open Archive*. https://doi.org/10.22541/essoar.176556004.46250262/v1

Simmons, A. J., & **Hollingsworth, A.** (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *128*, 647–677. https://doi.org/10.1256/003590002321042135

Simmons, A.J., **Mureau, R.**, & **Petroliagis, T.** (1995), Error growth and estimates of predictability from the ECMWF forecasting system. *Q.J.R. Meteorol. Soc.*, 121, 1739–1771. https://doi.org/10.1002/qj.49712152711

Stan, C., **Straus, D. M.**, **Frederiksen, J. S.**, **Lin, H.**, **Maloney, E. D.**, & **Schumacher, C.** (2017). Review of tropical–extratropical teleconnections on intraseasonal time scales. *Reviews of Geophysics*, *55*, 902–937. https://doi.org/10.1002/2016RG000538

Stroe, R., & **Royer, J.-F.** (1993). Comparison of different error growth formulas and predictability estimation in numerical extended-range forecasts. *Annals of Geophysics*, *11*, 296–316.

Sun, Y. Q., & **Zhang, F.** (2020). A new theoretical framework for understanding multiscale atmospheric predictability. *Journal of the Atmospheric Sciences*, *77*, 2297–2309. https://doi.org/10.1175/JAS-D-19-0271.1

Vonich, P. T., & **Hakim, G. J.** (2025). Testing the limit of atmospheric predictability with a machine-learning weather model. *arXiv*, arXiv:2504. https://doi.org/10.1175/AIES-D-26-0009.1

Žagar, N., **Horvat, M.**, **Zaplotnik, Ž.**, & **Magnusson, L.** (2017). Scale-dependent estimates of the growth of forecast uncertainties in a global prediction system. *Tellus A*, *69*, 1287492. https://doi.org/10.1080/16000870.2017.1287492

Zhang, F., **Bei, N.**, **Rotunno, R.**, **Snyder, C.**, & **Epifanio, C. C.** (2007). Mesoscale predictability of moist baroclinic waves: convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, *64*, 3579–3594. https://doi.org/10.1175/JAS4028.1

Zhang, F., **Sun, Y. Q.**, **Magnusson, L.**, **Buizza, R.**, **Lin, S.-J.**, **Chen, J.-H.**, & **Emanuel, K.** (2019). What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, *76*, 1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

TO CITE THIS ARTICLE:

Magnusson, L. (2026). Perspective Paper on Lorenz (1982): “Atmospheric Predictability Experiments with a Large Numerical Model”. *Tellus*, *79*(1), 204–215. https://doi.org/10.16993/tellus.4141

Submitted: 20 January 2026 **Accepted:** 04 July 2026 **Published:** 17 August 2026