JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD043363

Intrinsic Predictability From the Troposphere to the Mesosphere/Lower Thermosphere (MLT)

H. Garny[1]

[1]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft und Raumfahrt (DLR), Oberpfaffenhofen, Germany

Key Points:

- Predictability timescales vary by atmospheric layer, with longest predictability in the stratosphere and shortest in the MLT
- MLT predictability is influenced by vertical wave propagation, reducing small-scale predictability and extending that of planetary scales
- Predictability horizons revealed by gravity-wave-permitting models are overstimated by a factor of 2 in coarse resolution models

Correspondence to:

H. Garny,
hella.garny@dlr.de

Citation:

Garny, H. (2025). Intrinsic predictability from the troposphere to the mesosphere/lower thermosphere (MLT). *Journal of Geophysical Research: Atmospheres*, *130*, e2025JD043363. https://doi.org/10.1029/2025JD043363

Received 15 JAN 2025
Accepted 31 MAY 2025

Abstract The atmosphere's flow becomes unpredictable beyond a certain time due to the inherent growth of small initial-state errors. While many research studies have focused on tropospheric predictability, predictability of the middle atmosphere remains less studied. This work contrasts the intrinsic predictability of different layers, with a focus on the mesosphere/lower thermosphere (MLT, 50–120 km altitude). Ensemble simulations with the UA-ICON model for an austral winter/spring season are conducted with a gravity-wave-permitting horizontal resolution of 20 km. Initially small perturbations grow fastest in the MLT, reaching 10% of saturation after 5–6 days, compared to 10 days in the troposphere and 2 weeks in the stratosphere. A saturation level of 50% is reached only after about 2 weeks in the MLT, similar to the troposphere. Saturation times are overestimated in a coarser resolution model (grid size 160 km) by up to a factor of two, highlighting the need for gravity wave-resolving models. Predictability in the MLT depends on horizontal scales. Motions on scales of hundreds of kilometers are predictable for less than 5 days, while larger scales (thousands of kilometers) remain predictable for up to 20 days. This scale-dependent progression of predictability cannot be explained by simple scaling for upscale error growth. Vertical wave propagation plays a significant role, with gravity waves transmitting perturbations upward at early lead times and planetary waves enhancing long-term predictability. In summary, the study shows that MLT predictability is scale-dependent and highlights the necessity of high-resolution models to capture fast-growing perturbations and assess intrinsic predictability limits accurately.

Plain Language Summary The atmosphere's movements are difficult to predict after a certain time because small errors in the initial conditions grow over time. While predictability of weather close to the surface is much studied, less is known about how predictable the atmosphere is at higher layers. This study explores how predictability differs in the atmospheric layers, with a focus on the mesosphere/lower thermosphere (MLT), located 50–120 km above Earth, using advanced numerical simulations that resolve smaller-scale atmospheric waves. The findings show that initially small errors grow fastest in the MLT, making it less predictable than lower atmospheric layers. Typical predictability horizons in the MLT are 5–6 days, compared to 10 days in the lower atmosphere. However, motion in the MLT on larger horizontal scales of few thousand kilometers can remain predictable for up to 3 weeks. Simulations with lower resolution models are shown to overestimate predictability by up to two times. Furthermore, it is shown that MLT predictability is strongly influenced by the vertical propagation of waves from the lower atmosphere. This research highlights the importance of high-resolution models to understand and predict the atmosphere at higher altitudes, in the MLT.

1. Introduction

Earth's atmosphere is unpredictable beyond a certain time limit. This inherent property of atmospheric flow was discovered by Edward Lorenz in the 1960s. In his work from 1969 (Lorenz, 1969), he laid the ground work for our theoretical understanding of how, in a flow with many scales of motion, small-scale perturbations amplify over time and spread from smaller to large scales ("upscale error growth"). Thus, even if the initial conditions of a forecast are well known, any tiny uncertainty (in scale and/or magnitude) is sufficient to lead to an unpredictable flow after a certain time. The resulting predictability limit is inherent to a flow with given properties, and can only be assessed within a self-consistent model of the flow of interest. Specifically, forecast simulations with a numerical model of the atmosphere are conducted that only differ slightly in their initial condition (sometimes referred to as "identical twin" or "error growth" experiments). In such simulations, the growth of the initial small difference (the "perturbation" or "error") can be measured, referred to as "error growth" (used here interchangeably with perturbation growth). Once the difference in the flow between the simulations saturates to a level equivalent to the kinetic energy of the background flow, the intrinsic predictability limit has been reached. This approach to quantify intrinsic predictability necessitates the implicit assumption that processes relevant for the

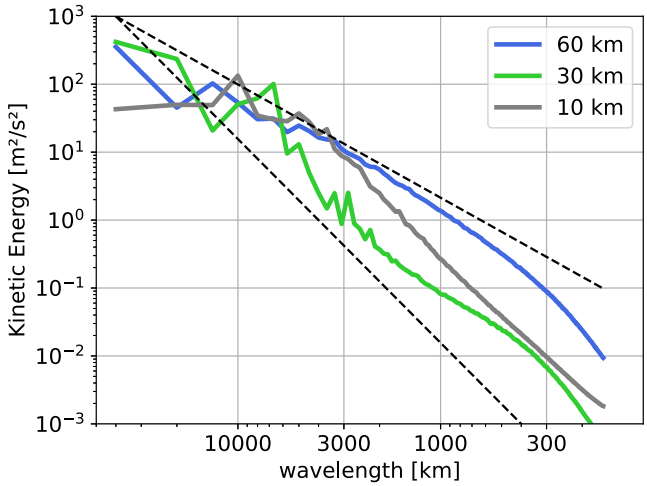

Figure 1. Kinetic energy power spectra at three different altitudes: in the troposphere (10 km, gray), the stratosphere (30 km, green), and the mesosphere (60 km, blue). Spectra are averaged over 25 August–30 September from the control UA-ICON simulation in R2B7 resolution (Base-HR). Dashed lines display the $k^{-5/3}$ and k^{-3} lines, respectively.

growth of the perturbation, and in general for the underlying flow of the region of interest, are well represented in the model (often referred to as "perfect model assumption"). Given that no model is perfect, this implies that estimations of the predictability limit are model-dependent. Thus, it is of interest to test the sensitivity of perturbation growth and the resulting predictability limit to underlying model assumptions, such as resolution and parameterizations.

For the tropospheric flow, early theoretical estimates of a predictability limit in the range of 2 weeks (Lorenz, 1969) have since been confirmed with modern numerical model simulations (e.g., Selz et al., 2022). Furthermore, many studies have assessed the processes relevant for growth of perturbations, and their representation in numerical models (see e.g. Baumgart et al., 2019). In particular, convection, and the associated latent heat release, have been found to play an important role for the initial growth of perturbations, and a stochastic representation of convective processes leads to faster growth that mimics higher resolution models (e.g., Baumgart et al., 2019; Selz & Craig, 2015).

Predictability of the middle atmosphere has so far gotten much less attention. Middle atmospheric error growth has been addressed to date only by a few studies (Liu et al., 2009; Pedatella et al., 2019; Smith et al., 2017), and they generally agree on rapid error growth in the mesosphere/lower thermosphere (MLT) compared to the stratosphere. These studies estimate a saturation of the perturbation in the MLT after roughly 10–30 days. However, all those studies have been conducted with coarse resolution models with a grid size on the order of 100 km. In the middle atmosphere, gravity waves with scales on the order from thousands to tens of kilometers have high amplitudes and are of first-order importance for vertical momentum transfer. In the coarse resolution models, gravity wave need to be parameterized, and those parameterizations rely on crude assumptions on wave sources, propagation, and dissipation. Smith et al. (2017) reported a high sensitivity of error growth in the MLT on the representation of gravity waves: when degrading aspects of the representation of gravity waves in the parameterizations further (by disabling variations in sources, or even applying pure wind damping), error growth was found to be reduced. With increasing computational power, it is now possible to explicitly resolve part of the gravity wave spectrum in global models. This allows to address the question how estimates of middle atmospheric predictability limits will be affected in models, which explicitly resolve gravity waves.

The layers of the atmosphere have distinct flow properties, which may lead to different predictability limits. In the troposphere, baroclinic instability and resulting wave motion on synoptic scales are crucially shaping the flow, and the flow on those synoptic scales can be well described as 2-dimensional turbulence, associated with a kinetic energy spectrum with a slope of k^{-3} (where k is the horizontal wave number; e.g., Koshyk and Hamilton (2001); see also Figure 1, further discussed in Section 2). For mesoscales, measurements show that the spectral slope turns to $k^{-5/3}$ (Nastrom & Gage, 1985). In the stratosphere, flow is dominated by planetary scales, while synoptic scales waves are filtered by the background winds, leading to lower kinetic energy than the troposphere for scales below about 7,000 km (see Figure 1). In the mesosphere (and lower thermosphere), high kinetic energy levels are found at mesoscales, with a spectral slope of close to $k^{-5/3}$ for scales below around 4,000 km (see blue line in Figure 1). The high energy at mesoscales is often associated with the high amplitudes of gravity waves at those altitudes, which grow exponentially while propagating up from the troposphere (Knobloch et al., 2023).

In early work by Lorenz (1969), the growth rate of initially small-scale perturbations was linked to the background kinetic energy spectrum. His work was based on an idealized flow system (namely the 2-dimensional vorticity equation), but the results were later confirmed by Rotunno and Snyder (2008) to hold also for other assumed flow properties. Their results showed that initial small-scale perturbations spread rapidly to larger scales in flows with a spectral slope more shallow than k^{-3}. Lilly (1972) introduced a simple scaling argument to describe this behavior, assuming that perturbations spread upscale via local scale interactions. Specifically, the assumption is that the time it takes the error to propagate from one scale k, on which the error is initially saturated, to the scale $2k$, which is initially unperturbed, is proportional to the eddy turnover time on scale k. The eddy

turnover time can be estimated by scaling arguments from the kinetic energy density $E(k)$ with $\tau_k \approx \left[k^3 E(k)\right]^{-1/2}$. Thus, the timescale for perturbations initially saturated at scale k_0 to spread to a larger scale k_1 is given by (see Vallis (2005), their Equation 8.80; hereafter referred to as "Lilly scaling"):

$$T = \int_{k_0}^{k_1} \tau_k d(\ln k) = \int_{k_0}^{k_1} \left[k^3 E(k)\right]^{-1/2} d(\ln k) \quad (1)$$

This scaling reveals the special case of a spectral slope of k^{-3}: with this slope, the eddy turnover timescale τ_k is constant with horizontal scale. Thus, the predictability time will increase linearly with ln k, and in the limit of an initial error at infinitely small scales, the predictability time becomes infinite. For any slopes that are shallower than k^{-3}, smaller scales have faster eddy turnover times, and thus faster error growth. This leads to a finite predictability limit for an infinitely small-scale initial perturbation.

Given the distinct energy spectra in the different layers of the atmosphere, what can we expect in terms of error growth and predictability based on those theoretical arguments? Given an initial perturbation at small scales, the Lilly scaling suggests that the shallow kinetic energy spectrum in the mesosphere should lead to fast upscale growth, and thus shortest predictability timescales, while the stratosphere with lowest kinetic energy at horizontal scales below about 5,000 km can be expected to display slowest upscale error growth. In the troposphere, the k^{-3} spectral slope for scales between about 300 and 3,000 km leads to the expectation of constant exponential growth rates for those scales. However, all those expectations are based on the assumption that error growth is dominated by local scale interactions. While this might be appropriate for the quasi-2-d turbulence at synoptic scales in the troposphere, it is not clear whether it holds in other dynamical regimes. At mesoscales in the troposphere, convection and associated latent heat release is key for spreading initially small-scale perturbations (Baumgart et al., 2019), and it has been shown that in an isolated midlatitude cyclone, error growth is not sensitive to the background kinetic energy spectrum (Lloveras et al., 2022). In the middle atmosphere, important contributions to the kinetic energy spectral budget stem from vertical fluxes of kinetic energy (Koshyk & Hamilton, 2001); thus, this process might lead to error growth in the middle atmosphere via a different pathway than local scale interactions.

The goal of this study is to estimate altitude-dependent predictability limits in numerical simulations with gravity-wave-permitting resolution, and contrast them with estimates of coarser resolution models. Furthermore, the study aims to assess in how far differences in error growth and predictability in different layers of the atmosphere can be explained by their distinct flow properties. To this end, numerical simulations are conducted to quantify the predictability limit inherent to the flow in the different atmospheric layers, utilizing a model that simulates the atmosphere from the troposphere up to the MLT region with a model lid at 150 km altitude (introduced in Section 2, along with the simulation setup and diagnostics used). The results on predictability limits and error growth for different layers in the atmosphere with the model in gravity-wave-permitting resolution (20 km horizontal grid spacing) are presented in Section 3. The sensitivity of the results with respect to horizontal resolution is analyzed by comparison to a coarse resolution setup (with 160 km horizontal grid spacing) and presented in Section 4. In Section 5, the relevant processes for error growth in the MLT are discussed before general implications are discussed, and the paper is concluded in Section 6.

2. Model, Experiments, and Diagnostics

2.1. The UA-ICON Model

In this study, the "Upper-Atmosphere" (UA)-ICOsahedral Nonhydrostatic (ICON) general circulation model is used (Borchert et al., 2019). UA-ICON is an upward extension of the German weather and climate model ICON (Zängl et al., 2015), which is based on a nonhydrostatic dynamical core with a triangular horizontal grid. UA-ICON incorporates an extended physics package with processes relevant for the MLT region, and an extension of the dynamical core from shallow to deep atmospheric approximations of the governing equations. For this study, UA-ICON based on ICON version 2.6.5 was used with the physics package developed for numerical weather prediction (NWP) combined with the extension for upper atmospheric physics including ion drag, molecular diffusion, and frictional heating as described by Borchert et al. (2019).

The UA-ICON model is used here in two resolutions. The "coarse" resolution setup closely follows Kunze et al. (2024), with a horizontal grid spacing of 160 km (R2B4) and 120 vertical layers extending to 150 km altitude, and the same choice of parameterizations. However, because the study by Kunze et al. (2024) was not available by the time the simulations were conducted, the tuning parameters of the gravity wave parameterizations are mostly set to their default value here. An exception is the parameter C^* controlling the saturation criterion of momentum fluxes in the nonorographic gravity wave parameterization by Orr et al. (2010), which is set to $C^* = 20$. The time step in the coarse resolution setup is set to 240 s for the dynamics, but reduced for the parameterizations (480 *s* for the upper atmosphere radiation calculation, 720 *s* for convection and UA physics, 1,200 *s* for the gravity wave parameterizations, and 3,600 *s* for main ecRAD radiation parameterization). Mean winds and temperature for the simulated period (not shown) are similar to those from the tuned model setup presented and evaluated by Kunze et al. (2024).

The "high-resolution" setup used here has a horizontal grid spacing of approximately 20 km (R2B7 resolution) and 250 vertical layers that extend from the surface to a 150 km altitude (similar to Charuvil Asokan et al., 2022). It uses the identical physics packages as the coarse resolution setup. While the higher resolutions allow to partially resolve gravity waves, it is still necessary to parameterize the unresolved part of the spectrum. The same gravity wave parameterizations as in the coarse resolution simulation are used, including same parameter settings. The time step in the high-resolution setup is 60 *s* for dynamics, 600 *s* for the upper atmosphere radiation calculation, 360 *s* for convection, 200 *s* for UA physics, 720 *s* for the gravity wave parameterizations, and 1,440 *s* for the radiation parameterization. Mean circulation and temperature averaged over the simulated period in the high-resolution setup differ somewhat from those in the coarse resolution, in particular, in the winter high latitudes above about 80 km and in the tropical middle atmosphere, but are an at least as realistic representation of the mean flow in the middle atmosphere (not shown).

In both model setups, a sponge layer that prevents wave reflection at the model top is active above 110 km altitude. Furthermore, both setups use identical external forcing with prescribed greenhouse gases and ozone for the radiation, and prescribed climatologically mean sea surface temperatures and sea ice cover.

The kinetic energy spectra from the high-resolution simulation at three different altitudes are shown in Figure 1. As discussed in Section 1, the energy spectra show distinct shapes in the atmospheric layers. Furthermore, the spectra can be used to estimate the effective resolution of the model, that is, estimate which horizontal scales are properly resolved. The kinetic energy in the stratosphere and the mesosphere strongly decline for scales below 300 km, indicating strong effects of numerical damping on motion on those scales. In the troposphere, a near-constant slope of near k^{-3} holds for scales also below 300 km. In ICON simulations with km-scale resolution, a flattening of the kinetic energy slope is found for scales below about 400 km (Stephan et al., 2022). Thus, it appears likely that in the simulation presented here, a physical flattening of the spectral slope at mesoscales is compensated by numerical dampening, leading to the apparent constant slope. In the mesosphere, the spectral slope is close to $k^{-5/3}$ for scales $\leq$3,000 km, but starts to deviate for scales $\leq$700 km. Simulations with a model that uses a similar horizontal resolution, but more advanced and physical diffusion parameterization, demonstrated that the spectrum in the MLT closely follows a $k^{-5/3}$ slope down to scales of 200 km (Becker et al., 2022). This strongly suggests that the deviation from the $k^{-5/3}$ slope in our model is due to numerical effects rather than a physical feature. Overall, the spectra indicate that scales above about 300 km are reasonably well resolved while smaller scales are subject to strong dampening. Thus for further analysis, only scales $\geq$300 km will be considered.

2.2. Experiments

The base simulations, from which perturbed experiments are spun off, are conducted for one austral winter/spring season, covering the period from mid-August to the end of October (Base-HR and Base-LR in Table 1). The base simulation is initialized on 13th of August 2015 with analysis of the ECMWF Integrated Forecast System (IFS) up to 80 km altitude. Above, an isothermal atmosphere is used as initial state, and wind velocities are zero initially. However, the atmospheric state at these levels spins up within a few days for both horizontal resolutions used in this study: For example, at 80 km, the total kinetic energy is close to the time-averaged background value of the simulations after 1 day, and has fully converged on all scales to the time-averaged background kinetic energy spectrum after 5 days. The simulations are analyzed here only after 25 August, and the unrealistic isothermal initial conditions are assumed to have no major influence on the analysis performed in this study.

Table 1
Table of Experiments

Experiment	Resolution	Perturbation	# Ensemble	Duration
Base-HR	R2B7	–	1	Aug 13–Oct 30
Base-LR	R2B4	–	1	Aug 13–Oct 30
Base-NoT-HR[a]	R2B7	–	1	Aug 13–Oct 30
P0001-HR	R2B7	$c = 0.001$	3	Aug 25–Sep 15 (1–30 Sep)
P0001-LR	R2B4	$c = 0.001$	3	Aug 25–Sep 30
P0001-NoT-HR[a]	R2B7	$c = 0.001$	1	Aug 25–Sep 30
P01-HR	R2B7	$c = 0.1$	1	Aug 25–Sep 15
P01-lowA-HR	R2B7	$c = 0.1$ below 30 km, $c = 0.001$ above 50 km	1	Aug 25–Sep 30
P01-upA-HR	R2B7	$c = 0.001$ below 30 km, $c = 0.1$ above 50 km	1	Aug 25–Sep 30

[a]With a constant solar zenith angle.

Perturbed simulations are conducted as spin-off simulations from those base simulations. They are initialized with the state of the base simulations on 25 August, both for the "high" resolution (R2B7) and the "coarse" resolution (R2B4). To study error growth, small perturbations are added to this initial state. Those are constructed with a simplified version of the random field perturbation method by Magnusson et al. (2009). In this method, the scaled difference between two randomly chosen days are used as perturbation, having the advantage that (a) the perturbations are physically balanced and (b) the same relative perturbation magnitude is applied in different regions/altitudes of the atmosphere. Magnusson et al. (2009) suggests to use dates within the same season as the initial day, but drawn from different years. Given that here such a database of multiyear simulations with UA-ICON is not available, the fields to construct the perturbation have to be taken from the base simulations themselves. To ensure that the fields are as uncorrelated as possible, days were chosen in August and subtracted from days in September, being at least 20 days apart. The perturbation that is added to a variable $X(t,x,y,z)$ is calculated as $c*(X(t_1,x,y,z) - X(t_2,x,y,z))$, where c is the scaling factor. The perturbed variables are the prognostic ICON variables' horizontal and vertical winds, temperature, density, and pressure. Perturbations were applied globally and at all altitudes. The scaling parameter c is set to 0.001 for the default perturbation experiments (Experiments P0001-HR and P0001-LR in Table 1). Three perturbed members are conducted and run until mid-September both for the high-resolution and coarse resolution model setup. One of the high resolution and all of the coarse resolution simulations were extended to the end of September.

The scaling parameter c was chosen so that the perturbations lie in the regime of the "intrinsic limit", that is, further reducing the perturbation does not change the predictability timescales. Selz et al. (2022) showed that the intrinsic limit is approached when reducing the current operational initial condition uncertainty by about two orders of magnitude. A scaling parameter c of 0.1 approximately corresponds to the operational initial condition uncertainty, as verified by comparison to operational ECMWF ensemble simulations. Thus, perturbations with $c = 0.001$ lie in the intrinsic limit in the troposphere, according to Selz et al. (2022). It was verified that this also holds for the middle atmosphere by testing for convergence of saturation timescales for simulations with $c = 0.1$, 0.07, 0.01, and 0.001.

The perturbation method chosen here adds perturbations on all spatial scales, not only at small scales as would be appropriate to study upscale error growth (e.g., as assumed by the Lilly scaling). Durran and Gingrich (2014) have shown that any small perturbations in large-scale fields will almost immediately lead to fast perturbation growth in small scales, effectively leading to upscale error growth behavior thereafter. Thus, the scales of the initial perturbations are not critically important for the resulting error growth and predictability estimates.

A set of sensitivity experiments were conducted on the role of vertical coupling for predictability in the middle atmosphere, with enhanced perturbations at specific altitudes regions only. An enhancement of the perturbations of two orders of magnitude was chosen (i.e.., $c = 0.1$) in order to be able to observe an impact away from the intrinsic limit. In one of the experiments, the enhanced perturbation is inserted in the lower atmosphere, below 30 km, with a linear transition to $c = 0.001$ between 30 and 50 km (P01-lowA-HR in Table 1). In the complementing experiment, enhanced perturbations with $c = 0.1$ are inserted only above 50 km, again with a linear

transition to $c = 0.001$ between 50 and 30 km (P01-upA-HR in Table 1). For comparison, an experiment with enhanced perturbation strength of $c = 0.1$ throughout the atmosphere is conducted (P01-HR in Table 1). Moreover, experiments without a diurnal cycle in the solar radiation were conducted in order to test the role of tidal waves for predictability. This includes a base simulation without tides (Base-NoT-HR), and a perturbed experiment with $c = 0.001$ (P0001-NoT-HR).

2.3. Diagnostics

The main diagnostic used in this study to quantify error growth is the evolution of total perturbation energy (e.g., Magnusson et al., 2009; Zhang et al., 2003), defined as

$$E_{tot}(t) = \frac{1}{2}\left[(u(t) - u_{pert}(t))^2 + (v(t) - v_{pert}(t))^2 + \frac{c_p}{T_{ref}}(T(t) - T_{pert}(t))^2\right] \quad (2)$$

where u, v and T are the horizontal wind components and temperature from the base simulation, respectively, and u_{pert}, v_{pert} and T_{pert} the same quantities, respectively, from the perturbed simulation at same time t. T_{ref} is a reference temperature, here set to 300 K (following Magnusson et al., 2009). Predictability timescales are then calculated as the time, when the perturbation energy reaches a certain fraction of the expected saturation level. A fraction of 10% is used to quantify the timescale for the fast initial error growth, and a fraction of 50% is used as a level where the flow has mostly decorrelated (following Selz et al., 2022).

To determine the exact saturation limit, one would need to calculate the difference energy at the respective time of the year between the atmospheric states in different years. This would require multiyear simulations, which are not available here. Therefore, here the saturation level is assumed to equal the perturbation energy calculated from the difference fields between the 2 days as used to construct the initial perturbation. Thus, for a scaling factor of 0.001, the 10% saturation level corresponds to the time when the perturbation energy has grown five orders of magnitude (because the initial perturbation is 10^{-6} times the perturbation energy of the difference field).

Next to the perturbation energy for the full fields, a scale decomposition is performed to quantify error growth for specific horizontal scales. For this purpose, a standard spherical harmonics decomposition is used (following e.g. Koshyk & Hamilton, 2001) and applied to the ICON output fields interpolated on a Gaussian grid (corresponding roughly to the original model resolution). For the scale decomposition, only the kinetic energy component is taken into account, but no difference in behavior is found whether the potential energy is considered or not. Predictability timescales for individual scales (or ranges of scales) are calculated as described above for the total perturbation energy. However, since the time series of perturbation kinetic energy for individual scales can be noisier, a fit function for error growth was applied to obtain smoother results (e.g., Dalcher & Kalnay, 1987):

$$E(t) = E_{sat} - (E_{sat} + S/\alpha)/(1 + \mu) \quad (3)$$

with $\mu = (E_0 + S/\alpha)/(E_{sat} - E_0) * \exp((\alpha + S/E_{sat})\,t)$, where E_0 is the initial perturbation energy, E_{sat} is the expected saturation energy, and S and α are the fit parameters. The predictability timescale is then derived by interpolation of the fitted line to the chosen saturation level.

3. Error Growth in Different Atmospheric Layers

The growth of global perturbation energy in the perturbed UA-ICON simulations at different altitudes is shown in Figure 2 (based on difference between Base-HR and P0001-HR simulations). The first observation is that error growth clearly behaves similar within each layer, but is distinct between different layers: error growth curves for layers within the troposphere (gray lines), within the stratosphere (green), and within the MLT (blue) lie close together, but each set of curves displays distinct behavior. In all layers, perturbation energy grows rapidly by 2–3 orders of magnitude within the first day. Thereafter, the growth rates in the troposphere and the stratosphere strongly decrease. In the troposphere, a nearly constant exponential growth rate is found up until about day 12, after which growth of the perturbation energy further decreases to a smaller rate. In the stratosphere, the perturbation energy grows slowest compared to the other layers, while growth is fastest in the MLT. Growth rates in the MLT are particularly high in the first 4–5 days, but with growth rates attenuating steadily.

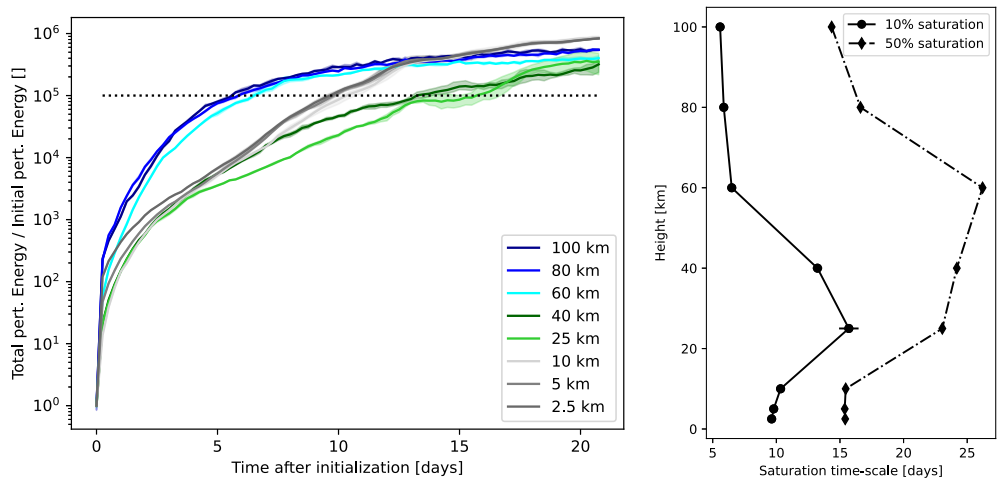

Figure 2. Left: total perturbation energy at different altitude levels (see legend). The energy is normalized with the initial perturbation energy at each level. Shading shows the range covered by the three ensemble members; lines are ensemble means. Right: saturation timescale at the different altitudes for saturation thresholds of 10% (i.e., time when perturbation energy has grown by 5 orders of magnitude; solid line) and 50% (dash-dotted line).

The different behavior of perturbation energy growth in the vertical layers of the atmosphere is characterized by the saturation timescales for two saturation levels in Figure 2 (right). Saturation to 10% of the expected saturation value (i.e., time it takes to grow the initial perturbation by 5 orders of magnitude, as indicated by dashed line) measures timescales for the fast initial error growth, while saturation to 50% also measures the subsequent slower growth phase (see also Section 2.3 for details). In the troposphere, a saturation level of 10% is reached after about 10 days, in the stratosphere after about 2 weeks, while in the MLT, perturbation energy has grown by 5 orders of magnitude already after 4–6 days. However, the saturation timescale for a saturation level of 50% is similar in the troposphere and at the higher levels of 80 and 100 km altitudes (both about 15 days). At the 60 km level, saturation timescales to a 50% level are about 3 weeks, as in the stratosphere. Thus, while initially the perturbations grow most rapidly in the MLT, thereafter a second slower phase of error growth is found there.

To further examine the behavior of error growth in the different layers, the perturbation energy is decomposed by horizontal scale. Figure 3 shows the spectra of perturbation energy at altitudes of 10 km (troposphere) and 60 km (mesosphere) for different simulation times. At both altitudes, the perturbation energy grows quickest at the smallest resolved scales, which dominate the perturbation spectrum after 1 day into the simulation, consistent with Durran and Gingrich (2014) (see also Section 2). The progression of the perturbation energy spectra at the

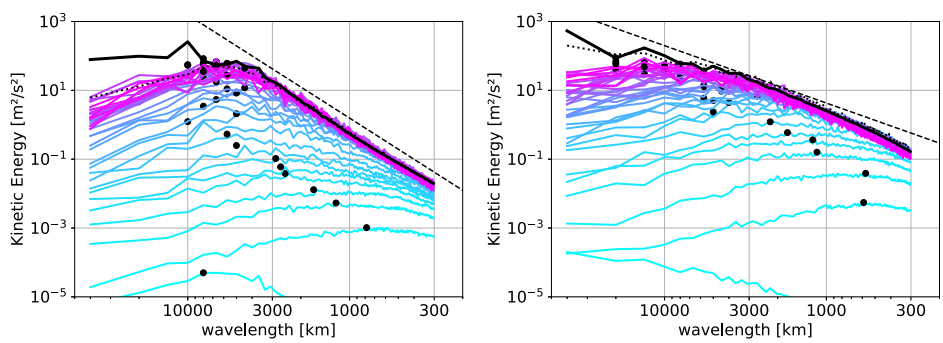

Figure 3. Kinetic energy spectra of perturbation energy at 10 km (left) and 60 km (right), averaged over the 3 ensemble members. Black solid lines are background spectra (scaled by factor 2) and the lowest light blue line is the initial perturbation. Perturbation energy spectra are displayed every 24 hr, with time progressing from light blue to magenta, up to 37 days after initialization. The black dots mark the maximum of the energy spectrum at each time. The black dotted line is the energy spectrum of the difference between the 2 days that was used for initialization (i.e., 10^6 times the initial perturbation). The black dashed lines are the k^{-3} and $k^{-5/3}$ lines in the left and the right panel, respectively.

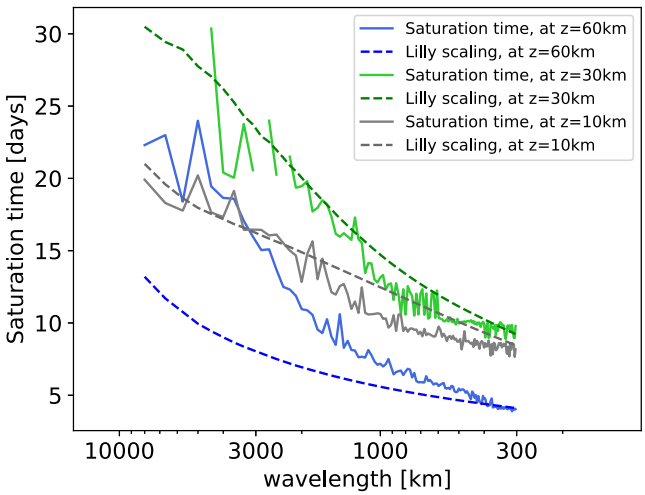

Figure 4. Saturation timescale as a function of horizontal scale estimated from the modeled perturbation energy spectra (i.e., time when perturbation energy reaches 50% of background energy) in solid lines at altitudes of 10 km (gray), 30 km (green), and 60 km (blue), and estimates by the Lilly scaling (i.e., integral over turnover time from the background kinetic energy spectra, see Equation 1) in dashed lines for three different altitudes (see legend). The Lilly scaling timescales are set to the modeled timescale for the smallest depicted wavelength of 300 km. Model saturation time is not displayed if it was not reached until the end of the simulation.

two altitudes shows distinct behavior: At 10 km, after the initial fast growth of smallest scales in the first 1–2 days, the maximum in perturbation energy (marked by black dots) progresses steadily from scales of about 1,000 km to about 8,000 km, and the nearly constant growth rate can be seen by roughly equidistant lines of perturbation energy. After about 7 days, the maximum perturbation energy remains at scales of around 8,000 km, but the amplitude still grows steadily for all scales until saturation is reached. At 60 km, the maximum in perturbation energy also progresses from smaller to larger scales, but with a less constant rate, and with the maximum closer to the saturation (i.e., background) line. Perturbation energy overall progresses with an initially strong, and a steadily decreasing growth rate (i.e., decreasing distance between lines). This behavior is overall qualitatively in agreement with the theoretical expectations as laid out by Rotunno and Snyder (2008) for error growth in a k^{-3} versus $k^{-5/3}$ regime (compare to their Figure 1).

The saturation of perturbation kinetic energy as a function of horizontal scale is further quantified for the different altitudes in Figure 4. Here, the saturation timescale is calculated as the time, when the perturbation kinetic energy of each specific wave number reaches 50% of the expected background saturation value. At all altitudes, saturation progresses monotonically from smaller to larger scales (note that perturbation energy for individual scales is rather noisy for the limited number of experiments conducted here, resulting in noise in the saturation times). Saturation for motion on horizontal scales smaller than 3,000 km is reached fastest in the mesosphere and slowest in the stratosphere. However, saturation times in the mesosphere increase strongly with scale, and for scales larger than 3,000 km, saturation times in the mesosphere are as long or even longer than those in the troposphere.

As outlined in Section 1, based on early theoretical work, Lilly (1972) suggested a scaling that relates error growth timescales to the background kinetic energy spectrum (Equation 1). To test whether this scaling can explain the different timescales of error growth in the atmospheric layers, the saturation timescales according to the Lilly scaling based on the background kinetic energy spectra in the different layers are added in Figure 4. For the smallest displayed scale, the timescales of the Lilly scaling are set to those from the model simulations. Interestingly, this simple scaling for the timescale of error growth from 300 km upscale is qualitatively consistent with the model simulations in the troposphere (10 km, gray lines) and the stratosphere (30 km, green lines). However, this is not the case in the mesosphere (60 km, blue lines). At this altitude, the kinetic energy spectrum with its high energy at smaller scales and a shallow slope (see Figure 1) suggests rapid upscale error growth, but according to the model simulations, the error takes substantially longer to propagate from scales of 300 km to about 8,000 km (see Figure 4 blue lines; this result also holds for 80 km, not shown). Thus, while the Lilly scaling argument is generally consistent with the slower error growth in the stratosphere than the troposphere, and the rapid initial error growth in the mesosphere, it appears to be unsuited to explain the progression of upscale error growth in the MLT. One reason for this discrepancy could be that the underlying assumption of local scale interactions as a dominant process for error growth might not be valid in the MLT region. This will further be discussed in Section 5.

4. Resolution Sensitivity of Error Growth

Previous work has assessed error growth in the MLT in models with typical climate model resolution of around 100 km horizontal grid spacing (Liu et al., 2009; Pedatella et al., 2019; Smith et al., 2017). Here, predictability limits and the evolution of perturbation energy is compared between the simulation with about 20 km grid spacing (R2B7, "high resolution" in the following) and a coarser resolution simulation with about 160 km grid spacing (R2B4, "coarse resolution in the following"), as shown in Figure 5. The general behavior of the growth of perturbation energy is similar to the higher resolution simulation (compare to Figure 2), but in particular, at altitudes in the MLT, the perturbation energy growth rates are constant already after about 5 days. The perturbation energy evolution at 60 km falls in between those at higher altitudes in the MLT and stratospheric levels at

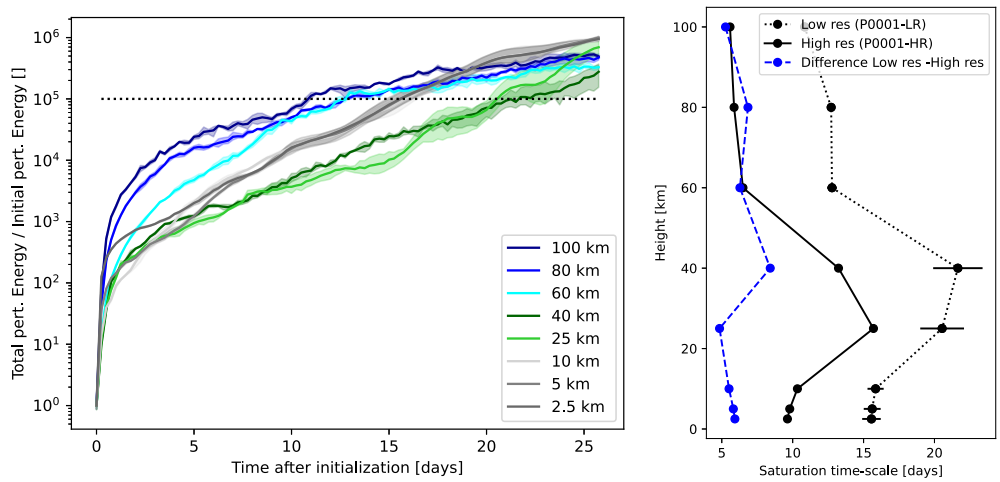

Figure 5. Left: as Figure 2, but for the lower resolution simulation with R2B4 (P0001-LR). Right: saturation timescales to a saturation level of 10% for the coarse resolution simulation (dotted), the high resolution (solid, repeated from Figure 2), and their difference (blue dashed).

this resolution. Generally, slower growth is found at all altitudes, translating into a longer saturation timescale than the higher resolution experiments: at all altitudes, the time to grow the initial perturbation by 5 orders of magnitude is about 4–7 days longer in the coarse than in the high-resolution simulation. Interestingly, this offset is about the same at all altitudes, but in relative terms, the effect is strongest in the MLT (100% longer at coarse than at high resolution).

Given the lack of small scales, which lead to the fast error growth in the initial days, the slower growth of perturbation energy in the coarse resolution simulation might be expected. The evolution of perturbation energy growth separated into three ranges of horizontal scales is shown in Figure 6 for altitudes of 10 and 60 km. At both altitudes, the initial weaker growth of perturbation energy in the first days in the coarse resolution simulation is clearly attributable to the underresolved motion at scales less than 2,000 km (compare yellow dashed and solid lines in Figure 6). Note that 2,000 km is chosen here as upper limit for the "small scales" range because this is the smallest scale properly resolved in the coarse resolution simulation, with smaller scales being subject to strong numerical diffusion. Thus, there is less energy at those scales in the coarse resolution model, and therewith, error growth is also limited. More interestingly, not only the perturbation energy of those underresolved scales grows more slowly in the coarse resolution simulation. Both the intermediate as well as the planetary-scale perturbation energies in the high-resolution simulation exceed the respective energies in the low-resolution simulation after about 3–5 days after initialization. Prior to that, the initial error growth at those scales is stronger in the coarse

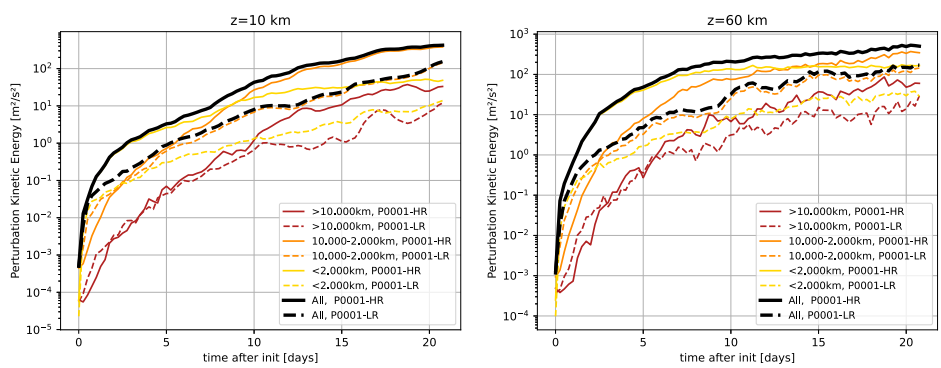

Figure 6. Evolution of perturbation kinetic energy at 10 km (left) and 60 km (right) for the higher resolution simulation (P0001-HR, solid lines) and the coarser resolution simulations (P0001-LR, dashed lines). Perturbation kinetic energy is shown for planetary scales (larger than 10.000 km, dark red), for intermediate scales (between 10.000 and 2.000 km, orange), and for smaller scales (less than 2,000 km, yellow). Black lines are perturbation energies of all scales.

resolution simulation, reasons for which are not immediately clear. For the planetary scales, there is higher variability in the perturbation energy, limiting the robustness of the difference between the coarse and high resolution simulations. More ensemble member and/or initial dates would be required for a robust quantification of this difference. The intermediate-scale perturbation energies do, however, clearly grow faster in the high-resolution simulation, showing that the lack of error growth on smaller scales also limits the error growth at well-resolved scales.

Overall, the comparison of error growth in simulations with different horizontal resolutions shows that error growth is slower at coarser resolution, and thus, the predictability horizon is vastly overestimated in the coarser resolution model. For the model setup used here, this overestimation is up to 100% in the MLT, but likely depends on details of the model such as the gravity wave parameterization. The slower growth in the coarse resolution model is not only due to the lack of small scales that lead to initial fast error growth, but also at larger scales, which are well resolved in the coarse resolution model; perturbation energy growth is inhibited when smaller scales are not well resolved.

5. Role of Vertical Coupling for Error Growth in the MLT

The results presented in Section 3 revealed a two-stage behavior of error growth in the MLT. In an initial phase, errors grow rapidly for motion on horizontal scales of less than a few thousand kilometers, making them unpredictable after a few days. Thereafter, slower error growth is found, associated with motion on larger scales that remains predictable for lead times of 10–20 days.

At the end of Section 3, it was discussed that the scaling arguments based on the background kinetic energy spectrum under the assumption of local scale interactions ("Lilly scaling") can qualitatively explain upscale error growth timescales in the troposphere and the stratosphere. Thus, the scaling appears to be applicable despite its simplified and idealized assumptions (see Sections 1 and 2). In the MLT, the Lilly scaling is inconsistent with the model results; in particular, it cannot explain the longer saturation times for scales exceeding a few thousand kilometers. This discrepancy might indicate that local scale interactions are not the dominant mechanism driving error growth in the MLT. Flow in the MLT is strongly influenced by vertically upward propagating momentum fluxes, making it likely that they have a role to play for error growth. Indications for the influence of vertical wave propagation on MLT error growth are presented in the following. First, the focus is on the phase of rapid error growth within the first few days, and second, on the later stage beyond 10-day lead time, when smaller scales have decorrelated but larger scales remain predictable.

5.1. Early Growth Phase

In the troposphere, the initial rapid error growth is typically associated with the prediction of convection, which happens on timescales of hours and quickly decorrelates (Baumgart et al., 2019; Selz & Craig, 2015). Convection can generate gravity waves, and any convectively generated small-scale perturbations in the troposphere might propagate upward to the middle atmosphere via gravity waves. Is such upward propagation of perturbation energy relevant for the early fast error growth phase in the mesosphere, or are error growth processes within the MLT (e.g., by local scale interactions) more important?

To illustrate the early growth phase, Figure 7 (left) presents snapshots of the perturbation in meridional wind after 6-hr lead time, for two randomly chosen ensemble members. The wind fields have been decomposed into their divergent and rotational components, and the divergent component is shown here to emphasize perturbations associated with convection and with propagating gravity waves. At 10 km altitude in the troposphere, the largest perturbations are observed in the summer hemisphere, likely associated with convective activity. At 30 and 60 km, largest perturbations are found just above the regions of strong perturbations in the troposphere, that is, perturbations are vertically coherent. The height-longitude cross-section of vertical wind differences at the location of strongest tropospheric perturbations (in the bottom panel of Figure 7, left) provides further evidence that the perturbations at early lead times in the middle atmosphere are vertically coherent features, likely associated with propagating gravity waves.

Given those vertically coherent perturbations at early lead times, it would be expected that the perturbation energy grows at similar rates in the different layers. This is indeed the case when considering only the divergent component of the perturbation kinetic energy, as shown in Figure 8. The growth of the divergent component scaled by the initial perturbation in each layer (dotted lines) is nearly identical in the stratosphere and the mesosphere during the first

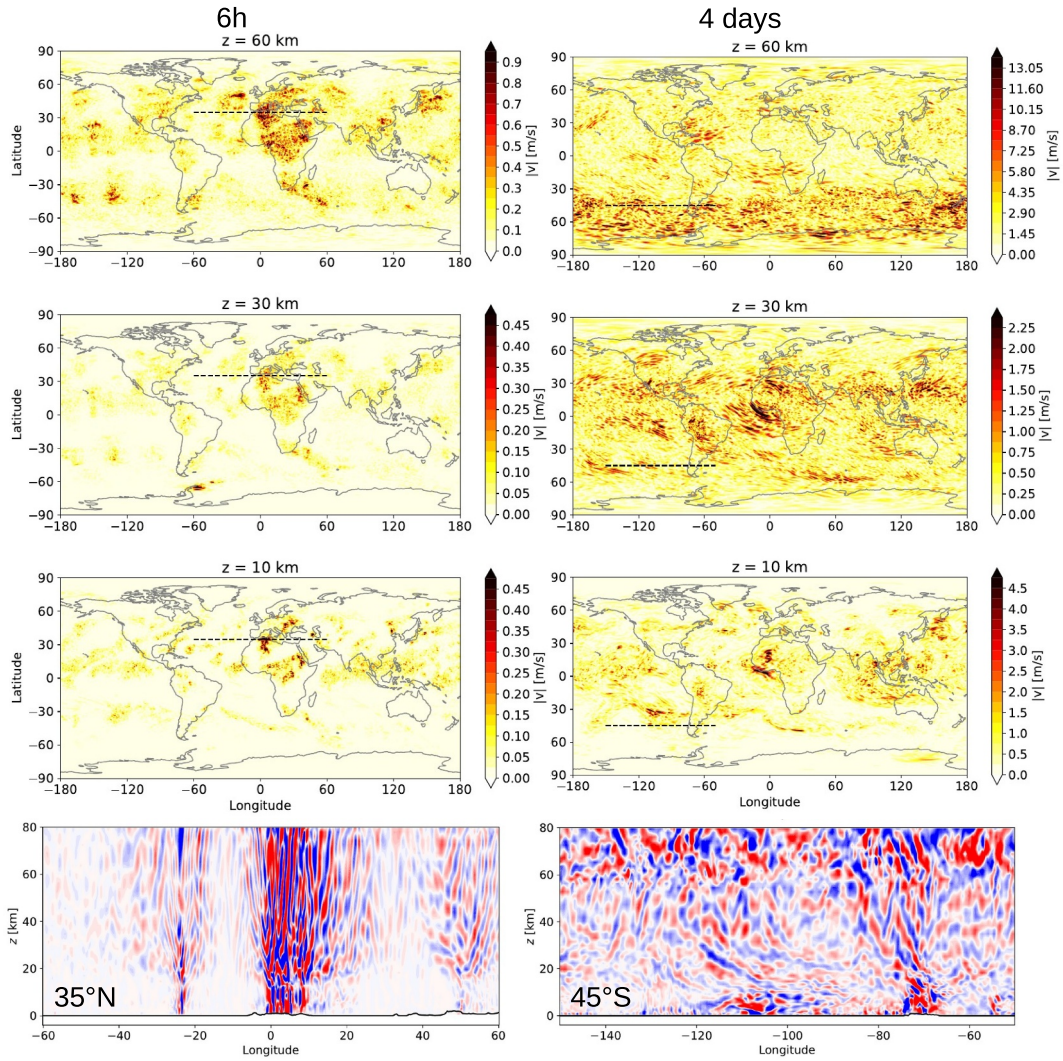

Figure 7. Snapshots of perturbations after lead times of 6 hr (left panels) and 4 days (right panels) between two members of the P0001-HR ensemble. Upper three rows show global maps of amplitudes of the perturbations in the divergent component of meridional velocity at altitudes of 60 km (top), 30 km (second from top), and 10 km (third from top). Bottom row shows perturbations in scaled vertical velocity as a function of altitude for a selected longitude range and latitude (as indicated by black dashed lines in upper panels). Vertical velocities are scaled by their zonally and temporally averaged vertical profile at the respective latitude (displayed range: from −150% to 150% in steps of 10%).

2 days, and the relative growth in the troposphere slightly exceeds that of the middle atmospheric layers. The stronger growth in the troposphere is likely associated with nonpropagating wave modes and divergent motions linked to convection, while only the propagating modes contribute to growth in the middle atmosphere.

The overall fastest growth in total perturbation kinetic energy in the mesosphere compared to the other layers thus stems from the relative fastest increase in rotational perturbation kinetic energy (Figure 8, see dashed lines). Perturbations in the rotational component of the meridional velocity peak in the winter extratropics at 60 km altitude (not shown), with no corresponding anomalies at lower altitudes. This suggests that these perturbations develop within this layer rather than propagating upward, and upscale growth by local scale interactions are a plausible mechanism for the rapid growth of rotational perturbation kinetic energy in the mesosphere.

After about 3 days, the relative increase in the divergent component of perturbation energy in the mesosphere exceeds that of the troposphere (Figure 8). On the right hand panel of Figure 7, snapshots of the difference in the divergent component of the meridional wind amplitudes after 4 days are shown. In the troposphere, largest differences are still found in the summer subtropics and tropics. In the mesosphere, on the other hand, perturbation

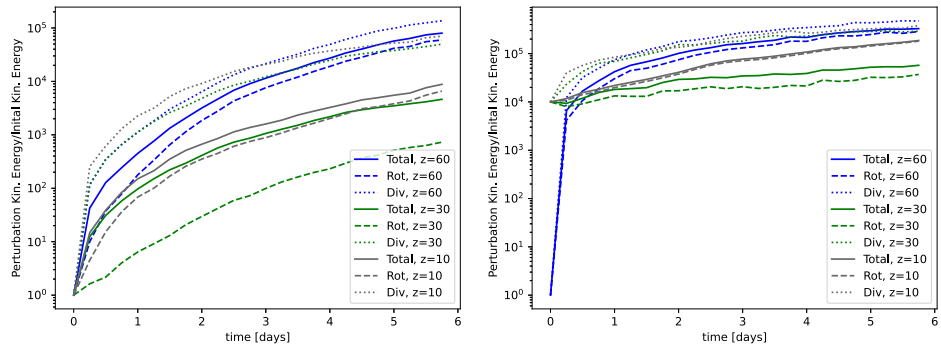

Figure 8. Global mean perturbation kinetic energy as function of lead time for the troposphere (gray), the stratosphere (green), and the mesosphere (blue). Solid lines are from total velocity fields, while dashed lines are the rotational component of velocity, and dotted lines the divergent component. All perturbation energies are normalized by their initial perturbation energy. Left: experiment with small perturbations (P0001-HR). Right: experiment with larger initial perturbations in the lower atmosphere (P01-lowA-HR). Note that the energies at 10 and 30 km are scaled by a factor of 10^4 to account for the different initial perturbation scaling in this experiment.

amplitudes are largest in the winter extratropics, the region with strongest gravity wave activity in this layer. In the cross-section of differences in vertical velocity at 45°S (lowest panel), some signatures of propagating waves are apparent, but perturbations are largest at 60 km and above, and the irregular structures at those altitudes indicate wave breaking.

The role of wave breaking is further highlighted in Figure 9, which shows the vertical profile of amplitudes in vertical velocity at 45°S. The mean amplitudes averaged zonally and over time (the black line in the left panel) reveal a local maximum in the upper troposphere, likely associated with convection, and a minimum in the lower

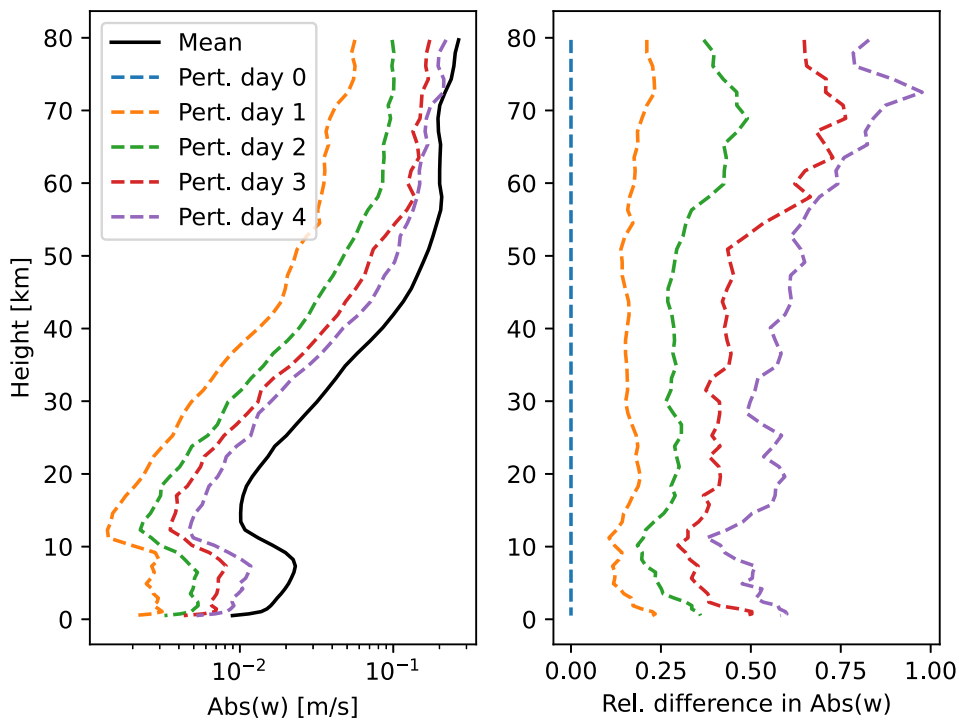

Figure 9. Left: amplitudes in vertical velocity as function of height at 45°S, the black line shows average over 25–30 August, and colored dashed lines show amplitudes of perturbation in vertical velocity for lead times of 0–4 days (see legend). Right: amplitudes of vertical velocity perturbations as in left panel, but relative to the average background values (i.e., black profile from the left panel). Colors are different lead times as in the left panel.

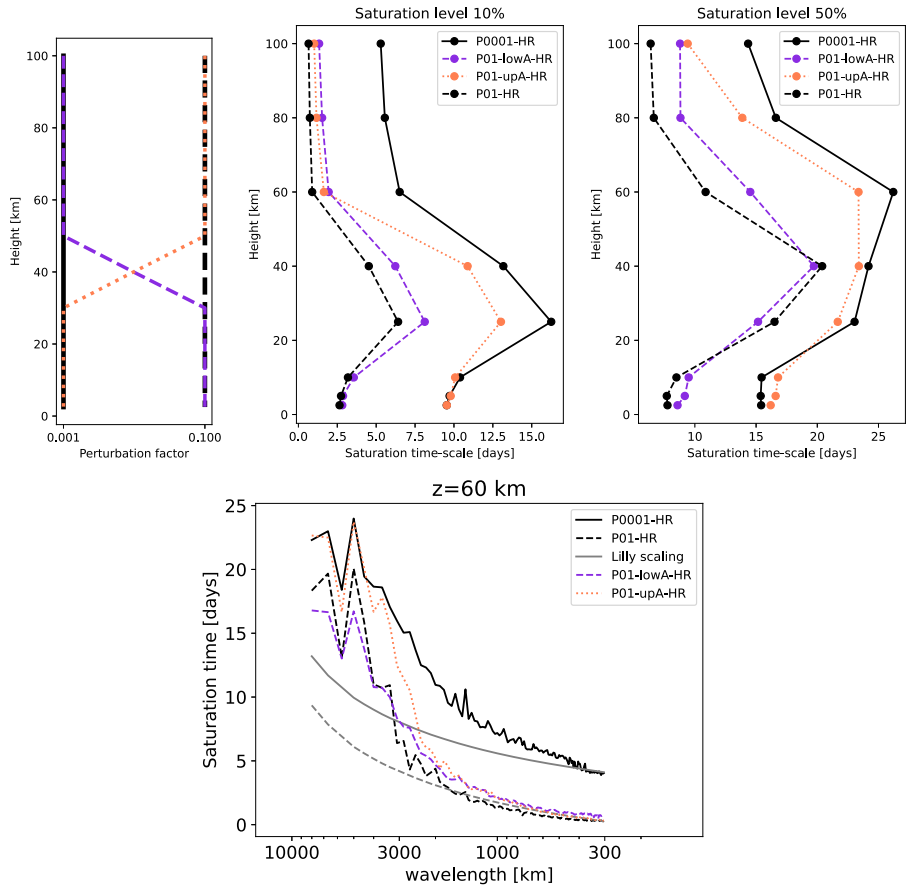

Figure 10. Top left: initial condition perturbation factor for experiments with small initial perturbation (P0001-HR), for larger initial perturbation (P01-HR; dotted black lines), and for sensitivity simulations with enhanced initial condition perturbation in the lower and upper atmospheric domains (P01-lowA-HR and P01-upA-HR, see legend on right panels). Top middle and right: saturation timescales (as in Figure 2) for the different experiments (see legend) to a saturation level of 10% (middle) and 50% (right). Bottom: as in Figure 4, but only at 60 km, and comparing the different experiments as in upper panels (see legend). The gray and gray dashed lines are the Lilly scaling timescales, shifted to the saturation timescale at 300 km for the small perturbation run P0001-HR (solid), and the large perturbation run P01-HR (dashed).

stratosphere. Between about 20 and 50 km altitude, amplitudes increase exponentially with height, indicating linear wave propagation. However, this exponential increase halts above 50 km, with vertical velocity amplitudes stagnating beyond 55 km. Thus, the profile of vertical wind amplitudes indicates a region of wave dissipation above about 55-km altitude.

In addition to the background profile, Figure 9 displays the amplitudes of vertical velocity perturbations for lead times ranging from 0 to 4 days, shown both in absolute terms (left panel) and relative to the background profile (right panel). Within the first day, perturbation amplitudes grow to roughly 15% of the background amplitudes throughout the atmosphere. After 2–3 days, the relative perturbation amplitudes grow considerably faster in the mesosphere (above 50 km altitude) compared to lower levels. This accelerated growth occurs in the region of wave dissipation, as indicated by the background profile of vertical velocity amplitudes. In regions of wave breaking, wave structures are expected to decorrelate on short timescales. Thus, the rapid perturbation growth observed in the mesosphere compared to lower layers maybe attributed to the amplifying effects of wave breaking processes in these regions.

To further confirm the important role of vertical propagation of perturbations for early lead times, a sensitivity experiment with enhanced initial perturbation strength in the lower atmosphere (below 30 km) is conducted (see description in Section 2, P01-lowA-HR). The perturbation of the initial conditions below an altitude of 30 km is enhanced by a factor of 100 in this experiment (see Figure 10 top left). The relative growth of perturbation energy

for this sensitivity experiment is shown in Figure 8 (right). Perturbation energy in the mesosphere is initially 4 orders of magnitude smaller than in the lower layers, but it rapidly increases within the first hours of the simulation, and even exceeds the relative perturbation energies in the troposphere and the stratosphere after 12 hr. The divergent component of the perturbation energy in the mesosphere rises more strongly initially, and approaches values similar to the troposphere after 2 days. Thus, perturbation energy propagates upward on timescales of few hours, confirming that the early stage of error growth in the mesosphere is strongly affected by the lower atmosphere. This is further confirmed by comparing saturation timescales in this sensitivity simulation with large perturbations in the lower atmosphere only to a simulation with large initial perturbations throughout the atmosphere (P01-HR; see Figure 10, compare purple to back dashed lines). The perturbation energy in the sensitivity experiment with large initial perturbations only in the lower atmosphere (P01-lowA-HR) saturate only few days after the experiment with large perturbations everywhere, both for a saturation level of 10% and 50%. For motion on horizontal scales from 300 to around 3,000 km, the delay in saturation is less than a day (see Figure 10, bottom). Saturation times for larger (planetary) scales are, however, not delayed for the sensitivity experiment P01-lowA-HR compared to P01-HR. Saturation of those larger scales at later lead times will be discussed in the next subsection.

Overall, results presented here indicate that the early phase of error growth in the MLT is determined by a combination of upward propagation of perturbation energy from the troposphere, and further accelerated error growth by wave breaking at those altitudes. The former manifests in comparable relative growth of the divergent component of perturbation energy in all atmospheric layers in the first days, and is confirmed by fast upward propagation of the larger perturbations from the lower atmosphere in the P01-lowA-HR experiment. The latter is likely associated with the more rapid increase in the rotational (and at lead times beyond 3 days also divergent) component of perturbation energy in the mesosphere compared to lower layers, likely associated with scale interactions occurring within the MLT layer.

5.2. Lead Times Beyond 10 Days

Flow on horizontal scales of few thousands of km has been found to be predictable for about 10–20 days in the mesosphere (see Section 3). If upscale error growth via local scale interactions would be acting, perturbation energies on those scales would be expected to saturate after less than 10 days. In the following, which processes might lead to those long lead times for larger scales in the MLT will be discussed. First, it will be tested in how far atmospheric solar tides contribute to predictability, and second the role of planetary-scale waves and their vertical propagation will be discussed.

5.2.1. Role of Tides

The role of tides in leading to predictable components in the large-scale flow are tested by sensitivity simulations in which tides are suppressed. To do so, the diurnal cycle of the solar zenith angle is switched off, that is, is set to a constant value for all times of the day and all longitudes (Experiment Base-NoT-HR in Table 1). Note that this leads to overall strongly enhanced heating in the lower thermosphere, with consequences for background winds (enhanced westerlies, not shown). Also winds in the mesosphere and the stratosphere are altered on average over the August–September period. The background kinetic energy spectra in the MLT in the simulation without tides shows the intended decrease in kinetic energy at planetary scales, but also kinetic energy of smaller scales is somewhat altered compared to the reference simulation. The different background kinetic energy spectra result in slightly different saturation timescales calculated by the Lilly scaling (compare solid and dashed gray lines in Figure 11), but the differences are smaller than differences between layers. Thus, overall, it is assumed here that despite differences in the background flow, comparison of error growth between the reference simulation and the simulation without tides can give insight into the role of tides for error growth.

As presented in Figure 11, the progression of perturbation energy saturation from small to large scales at 60 km altitude in the simulation without tides is remarkably similar to the reference simulation. At 80 km, predictability timescales for larger scales of several thousands of km are reduced in the simulation without tides, from over 30 days to about 22 days for scales of 7,000 km. However, saturation timescales are still far longer than the prediction by the Lilly scaling when excluding tides. Thus overall, the sensitivity simulation suggests that tides contribute to predictability of large horizontal scales at altitudes of 80 km and above, but are not the main contributor to the long saturation times of those scales in the MLT.

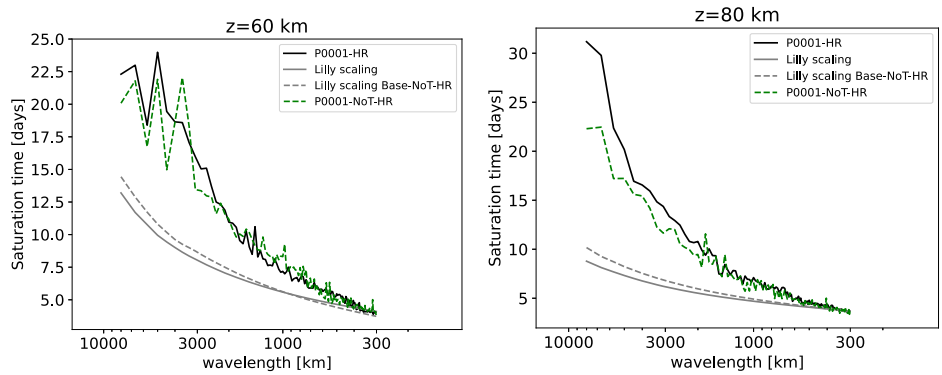

Figure 11. Comparison of saturation timescales (as in Figure 4) for the reference experiments with tides (black, based on difference of Base-HR and P0001-HR) and without tides (green dashed, based on difference of Base-NoT-HR and P0001-NoT-HR). Saturation timescales estimated from the base simulation kinetic energy spectra with the Lilly scaling are shown in gray (Base-HR) and gray dashed (Base-NoT-HR).

5.2.2. Planetary Waves

Total perturbation energy in the MLT has reached about 30% of its expected saturation after 10 days, but thereafter, perturbation energy rises much slower, and this is in particular due to motion on horizontal scales of few thousands of km (see Figure 6). Figure 12 (left) shows the saturation time of perturbation kinetic energy, here decomposed by the zonal wave number. Long saturation times of 15–20 days are found for planetary waves (zonal wave number 1–3). The wave number decomposition along individual latitude circles further allows to separate the Northern and Southern Hemisphere (dashed and dotted curve in Figure 12). The long saturation timescales of planetary waves stem particularly from the Southern Hemisphere, that is, winter. The right-hand panel of Figure 12 shows the time evolution of perturbation kinetic energy for planetary waves only, here displayed with linear scaling of the perturbation energy axis to highlight the evolution at later lead times. Steady rise of perturbation energy up to about 40% of the background value is found within the first 10–15 days; thereafter, the perturbation energy appears to stagnate and rise only slowly up until day 25, with more rapid rise to roughly 100% of the background value between day 25 and 30. Note that this evolution likely is case-dependent, being a function of the specific planetary wave activity (as further discussed in Section 6). In the following, the focus is on better understanding the role of planetary waves for predictability in the particular case simulated here.

Figure 13 compares a snapshot of planetary wave structures at 30°S in two ensemble members after 16-day lead time, that is, at a time when the perturbation kinetic energy of planetary waves stagnates (see Figure 12, right). At this time and latitude, a wave number-2 structure is found with strong amplitudes extending from 40 to about 70 km altitude. This planetary wave structure is remarkably similar in the two members, that is, it is predictable

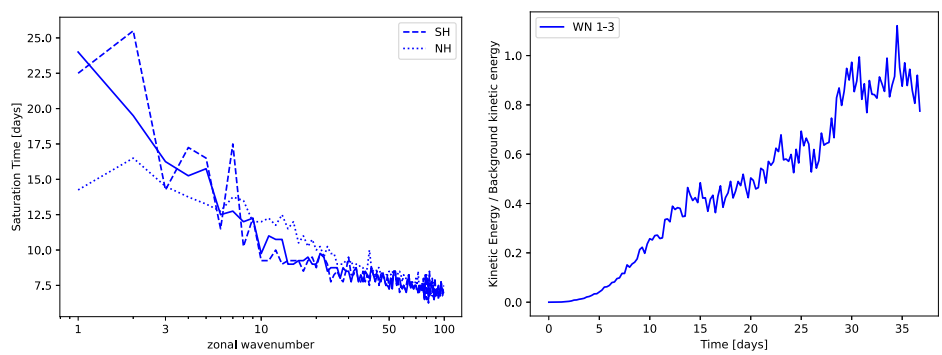

Figure 12. Left: saturation timescales at an altitude of 60 km for scales of motion decomposed by the zonal wave number. The solid line is the saturation time of perturbation kinetic energy of the global average over all latitudes, while the dashed and dotted lines show the Southern and Northern Hemisphere only, respectively. Right: time evolution of the globally averaged perturbation kinetic energy for planetary wave numbers from 1 to 3 only.

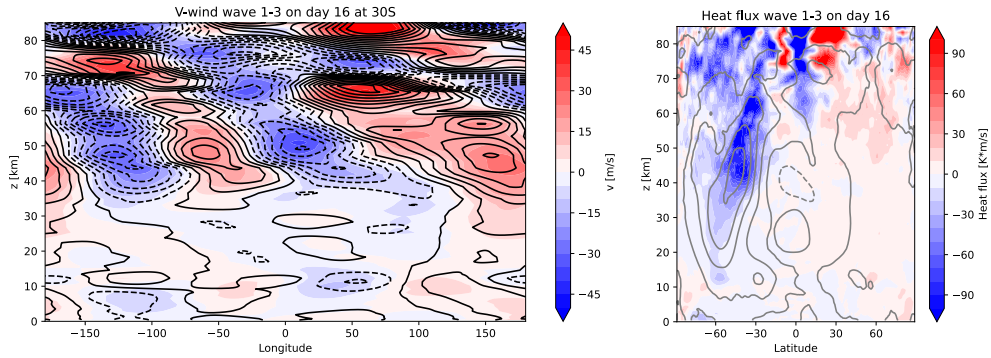

Figure 13. Left: snapshot of meridional velocity at 30°S after 16 days into the simulation for the control simulation (colors contours) and perturbed members (black contours). The velocity field has been filtered for zonal wave numbers from 1 to 3. Right: zonal mean heat flux of zonal wave numbers 1–3 on the same day (colored contours) and zonal mean zonal wind (gray contours, 20 m/s contour interval) from the control simulation.

even after 16-days lead time. The heat flux as measure of upward wave propagation for this day is shown in the right-hand panel of Figure 13 together with the zonal mean winds, indicating that the planetary waves propagate upward along the polar vortex. The enhanced wave amplitudes and heat fluxes above the vortex core at 40 km might either stem from lateral propagation along the vortex, or from local wave amplification or generation within the stratosphere (e.g., by barotropic/baroclinic instability as suggested by Sato et al., 2018). In either case, the planetary waves are stemming from the stratosphere with its long timescales.

To examine the lead time-dependent evolution of the planetary waves, in Figure 14, the evolution of heat fluxes of planetary waves averaged over 30–60°S as function of altitude and lead time is compared between two ensemble member. Initially, the fluxes are closely aligned, and both simulations capture an enhanced upward heat flux at days 10–20. Thereafter, heat flux anomalies begin to diverge more strongly between the two ensemble members. The lower panel of Figure 14 further displays the time series of heat fluxes at 60 km for all available ensemble members. This further confirms that planetary wave activity is very well predictable for about 10 days; between day 10 and 20, heat flux anomalies begin to diverge between ensemble members, but all simulations capture the enhanced (negative) heat flux around day 12–20.

The analysis of the planetary wave heat fluxes suggest that planetary waves propagating upward from the stratosphere along the polar vortex lead to the long saturation times for large scales in the mesosphere. To further confirm this, sensitivity experiments with enhanced perturbation strength only in the lower or the upper atmosphere are consulted. Comparison of the sensitivity experiment with enhanced perturbations only above 50 km (P01-upA-HR) to the experiment with enhanced perturbations throughout the atmosphere (P01-HR) can reveal whether the higher predictability of the lower atmosphere leads to gain in predictability above. Figure 10 shows that for the sensitivity experiment with enhanced perturbations above 50 km (P01-upA-HR, orange dotted line), perturbation kinetic energy in the MLT saturates quickly for small scales (lower panel). Saturation of global perturbation energy to the 10% level is reached at similar times than in the experiment with enhanced initial perturbations throughout the atmosphere (upper middle panel, compare orange dotted and black dashed lines). However, in the P01-upA-HR experiment, it takes substantially longer to reach 50% saturation (upper right panel) compared to the experiment P01-HR. This stems from horizontal scales larger than few thousand km, which have saturation times in P01-upA-HR similar to the experiment with small initial perturbations everywhere (see lower panel). Thus, the longer predictability of the lower atmosphere in P01-upA-HR leads to extended predictability of the large-scale flow at 60 km, despite the high initial condition uncertainty at this level. This further is consistent with the hypothesis that upward propagating planetary waves from the stratosphere lead to long saturation timescales of large scales in the mesosphere.

The evolution of heat fluxes in the sensitivity experiments is also consistent with this idea: As seen in the lower panel of Figure 14, the experiment with high initial condition uncertainty at upper levels (P01-upA-HR, orange line) shows stronger deviations from heat fluxes of the other ensemble members during the first about 10 days; however, it does capture the negative heat flux event during day 10–20 as well as the other members. In contrast,

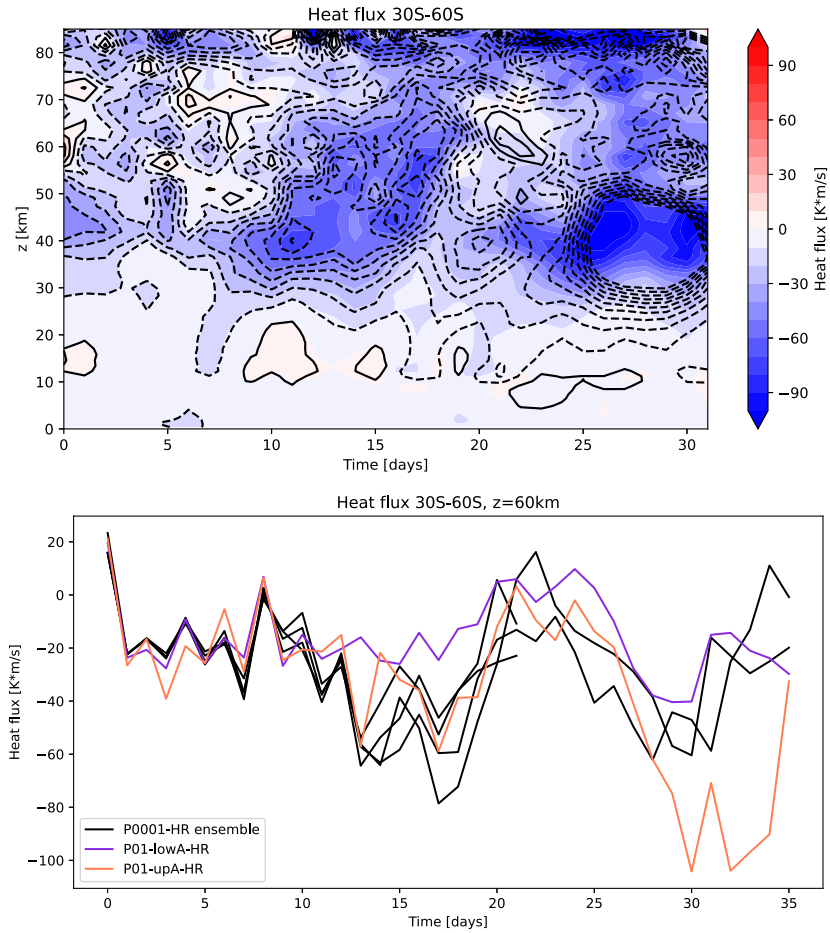

Figure 14. Top: Heat flux for planetary waves (wave number 1 to 3) averaged over 30–60°S as a function of altitude and lead time for the control simulation (colored contours) and a perturbed member (black contours). Bottom: as above, but time-series of averaged heat flux at 60 km altitude for the control and three perturbed simulation (black lines, P0001-HR) and the sensitivity experiments with enhanced initial condition perturbation in the lower atmosphere (P01-lowA-HR) and upper atmosphere (P01-upA-HR).

the experiment with enhanced initial condition uncertainty in the lower atmosphere (P01-lowA-HR, purple line) is aligned with the other ensemble member in the first days, consistent with low initial condition uncertainty at this level. However, it is the only simulation which does not capture the strong negative heat fluxes during day 10–20, again consistent with the higher initial condition uncertainty at lower levels.

Overall, the results presented in this section suggest that in the particular case analyzed here, upward propagation of planetary waves from the winter stratosphere, with their long timescales, lead to the long saturation timescales of planetary waves in the mesosphere.

6. General Discussion and Conclusions

This study aims to advance our basic understanding of intrinsic predictability in the different layers of our atmosphere, with a focus on the MLT. Overall, the results obtained with the gravity-wave-permitting model show that error growth in the different layers behaves as expected from the distinct kinetic energy spectra: predictability timescales are longest in the stratosphere, in agreement with the dominance of planetary scales as well as lowest kinetic energy at smaller scales. In the MLT, high kinetic energy at small scales and a flat (about $k^{-5/3}$) spectral slope lead to fast error growth at small scales, and generally shortest predictability timescales.

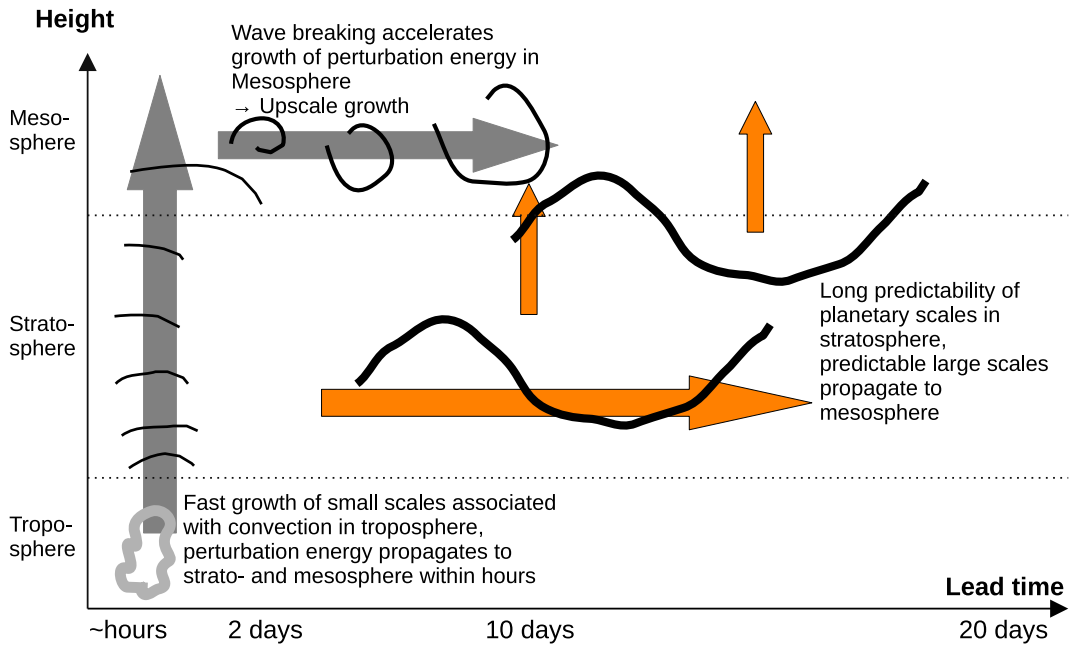

Figure 15. Schematic of error growth stages in the mesosphere. Gray arrows denote propagation of perturbation energy, orange arrows of predictable flow components, leading to predictable large scales in the mesosphere for lead times up to 10–20 days.

It was moreover demonstrated that predictability timescales are overestimated in a model with coarse resolution on the order of 100 km grid size, in agreement with previous results for the troposphere by Selz and Craig (2015). For the model setups with 20 versus 160 km horizontal resolution, time to reach 10% saturation of perturbation energy was overestimated by around 5 days throughout the atmosphere for the coarser resolution. This corresponds to a factor of two in the MLT. However, this overestimation in the coarse resolution model likely depends on the gravity wave parameterization, which can influence error growth (Smith et al., 2017). Previous studies using coarse resolution models provide estimates of saturation times in the MLT of roughly 20–30 days as intrinsic limit (Liu et al., 2009; Smith et al., 2017), and of 10 days in a setup aiming more at practical predictability limit (Pedatella et al., 2019). Those longer timescales than the present study might partly be due to differences in diagnostics, but it is likely that the coarse resolution models in these studies contribute to the longer predictability timescales. The grid spacing of 20 km used here is known to still not resolve important parts of the gravity wave spectrum (Polichtchouk et al., 2023). This opens the question how much estimates of the intrinsic predictability limit might further be reduced for even higher resolution. For the troposphere, results from Selz and Craig (2023) suggest a reduction of the 10% saturation times at 300 hPa of roughly 3 days when transitioning from the 20 km grid spacing to 2.5 km (see their Figure 1). If this translates to higher altitudes (as suggested by the constant offset in saturation times shown in Figure 5), this would further reduce the predictability of smaller scales in the MLT substantially. A potential avenue to mimic the fast error growth at small, unresolved scales might be the inclusion of a stochastic component in the gravity wave parameterization, following similar successful approaches in convection parameterizations for tropospheric weather predictions (Craig et al., 2021).

It was tested in how far the simple scaling of upscale error growth by Lilly (1972) might explain the differences in saturation times at the different altitudes. This scaling is in qualitative agreement with the progression of saturation times by horizontal scale in the troposphere and the stratosphere, but not in the MLT. Rather, at those altitudes the delay of saturation time from smaller scales of hundreds of km to scales of thousands of km is longer than what the Lilly scaling suggests. While this deviation might be caused by the quite simplified nature of the Lilly scaling, it could also indicate that processes other than the assumed local scale interactions contribute to error growth in the MLT region. More detailed investigation of the behavior of perturbations in the mesosphere combined with sensitivity experiments revealed that vertical wave coupling plays an important role for predictability at those altitudes. The stages of mesospheric error growth as hypothesized based on the analysis in this study are summarized in Figure 15. At early lead times, perturbations in the troposphere (e.g., associated with

convection), spread to higher altitudes within hours via gravity wave propagation. Thus, early growth of small scales in the mesosphere is controlled by tropospheric error growth processes, manifesting in a similar relative increase in the divergent component of perturbation kinetic energy. After a few days, growth in perturbation energy in the mesosphere exceeds that at lower layers, and the analysis suggest that this is associated with wave breaking at those altitudes, which accelerates error growth. This process can likely be understood as being dominated by upscale growth via local scale interactions. For later lead times, the results presented here suggest that upward propagation of planetary wave from the stratosphere introduces longer predictability horizons for those large scales. Thus, overall, the deviation of the saturation times for increasing horizontal scales from the Lilly scaling in the mesosphere might arise from two opposing effects of vertical wave coupling: the fast upward propagation of perturbation energy at small scales reduces their saturation times, while large scales become more predictable (later saturation times) because of the long timescales of upward propagating planetary waves from the stratosphere. Upscale error growth within the mesosphere by local scale interactions likely still is a dominant process to cause the fast overall error growth, but it is difficult to disentangle this process from the vertical propagation of perturbation energy. Interestingly, in the simulations with enhanced initial condition uncertainty in the mesosphere, saturation times follow the Lilly scaling more closely for scales below about 3,000 km (see Figure 10). This might indicate that local scale interactions are more dominant once perturbations within this layer are of a certain strength, so that upward propagation of small-scale perturbations is less important. The same behavior is found in the sensitivity experiment with large initial condition uncertainty in the lower atmosphere only (P01-lowA-HR); this is likely because the larger perturbations in the lower atmosphere propagate upward within the first hours, making the experiment indistinguishable from the experiment with large initial perturbations in the mesosphere thereafter. To better quantify the role of vertical propagation of perturbation energy versus upscale growth by local scale interactions, future work might consider budget analysis of the kinetic energy spectrum along the lines of Koshyk and Hamilton (2001) and Augier and Lindborg (2013), reformulated for perturbation energy.

It should be noted that the results presented here are deduced from a limited set of model simulations, with only one initial date in one season, and a limited number of ensemble members. Thus, state dependency of intrinsic predictability is not reflected in the results of this study. This is likely in particular relevant for the predictability timescales of larger horizontal scales. In the particular case analyzed here, the long timescales of planetary waves in the mesosphere are tied to a strong wave-2 event in the winter stratosphere. During episodes with less prominent planetary wave activity or upward propagation, saturation timescales might be reduced, as indicated by comparison of the winter to the summer hemisphere. For smaller scales waves, the global simulations integrate over many wave events; therefore, their saturation timescales are likely less case-dependent. Overall, the differences in the behavior of global perturbation energy growth at different altitudes is far larger than the spread between ensembles (see e.g. spread in Figure 2), and while there likely is state dependency of predictability estimates, it is unlikely that it would exceed the different behavior of the atmospheric layers, or the resolution dependency.

Next to the insights into fundamental processes of predictability in the vertically coupled atmosphere, results of this study might also be of practical value. Currently, whole atmosphere prediction models are being developed with the aim to exploit the predictable components of (lower) atmospheric forcing on the state of the ionosphere and the thermosphere, potentially relevant for space weather predictions. This study contributes to this endeavor in providing updated estimates of the intrinsic predictability of atmospheric motion in the MLT. With the higher resolution model simulations conducted here, it is revealed that previous estimates of the predictability horizon in the MLT might have been overestimated, at least for horizontal scales below about a few thousand kilometers; those scales are found to be predictable for only about 5 days according to the present study, and this predictability timescale might further be reduced when transitioning to fully gravity wave-resolving models. Larger scales are found to be predictable for around 20 days, thanks to stratospheric control. Next to leading to predictability of such individual planetary wave events, the stratosphere might facilitate probabilistic predictability through wave filtering during anomalous states such as sudden warmings. Given the otherwise short predictability horizon in the MLT, this is likely of key importance for predictions in the MLT and above (as suggested by e.g., Pedatella et al., 2018).

Data Availability Statement

The ICON model version 2.6.5 was obtained on a personal noncommercial research license distributed by MPI-M. ICON as open source model can be obtained at https://icon-model.org/. Figures were generated with python-matplotlib. For the spherical harmonics transformation, the python package pyshtools was used (Wieczorek & Meschede, 2018), obtained from https://shtools.github.io/SHTOOLS/ (last access 15 January 2025). Data of the UA-ICON simulations are made publicly available on a Zenodo data archive at https://doi.org/10.5281/zenodo.14646394 (Garny, 2025).

Acknowledgments
I thank Sebastian Borchert and Markus Kunze for support and advice in setting up the UA-ICON simulations. Furthermore, I am very grateful to Tobias Selz, Andreas Dörnbrack, Jonas Späth, and Sonja Gisinger for valuable discussions and for comments on an earlier version of the manuscript. I thank the three reviewers for valuable comments that significantly improved the paper. Open Access funding enabled and organized by Projekt DEAL.

References

Augier, P., & Lindborg, E. (2013). A new formulation of the spectral energy budget of the atmosphere, with application to two high-resolution general circulation models. *Journal of the Atmospheric Sciences*, *70*(7), 2293–2308. https://doi.org/10.1175/JAS-D-12-0281.1

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., & Riemer, M. (2019). Quantitative view on the processes governing the upscale error growth up to the planetary scale using a stochastic convection scheme. *Monthly Weather Review*, *147*(5), 1713–1731. https://doi.org/10.1175/MWR-D-18-0292.1

Becker, E., Vadas, S. L., Bossert, K., Harvey, V. L., Zülicke, C., & Hoffmann, L. (2022). A high-resolution whole-atmosphere model with resolved gravity waves and specified large-scale dynamics in the troposphere and stratosphere. *Journal of Geophysical Research: Atmospheres*, *127*(2), e2021JD035018. https://doi.org/10.1029/2021JD035018

Borchert, S., Zhou, G., Baldauf, M., Schmidt, H., Zängl, G., & Reinert, D. (2019). The upper-atmosphere extension of the icon general circulation model (version: Ua-icon-1.0). *Geoscientific Model Development*, *12*(8), 3541–3569. https://doi.org/10.5194/gmd-12-3541-2019

Charuvil Asokan, H., Chau, J. L., Larsen, M. F., Conte, J. F., Marino, R., Vierinen, J., et al. (2022). Validation of multistatic meteor radar analysis using modeled mesospheric dynamics: An assessment of the reliability of gradients and vertical velocities. *Journal of Geophysical Research: Atmospheres*, *127*(5), e2021JD036039. https://doi.org/10.1029/2021JD036039

Craig, G. C., Fink, A. H., Hoose, C., Janjić, T., Knippertz, P., Laurian, A., et al. (2021). Waves to weather: Exploring the limits of predictability of weather. *Bulletin of the American Meteorological Society*, *102*(22), E2151–E2164. 12.11.34; LK 01. https://doi.org/10.1175/BAMS-D-20-0035.1

Dalcher, A., & Kalnay, E. (1987). Error growth and predictability in operational ECMWF forecasts. *Tellus*, *39A*(5), 474–491. https://doi.org/10.1111/j.1600-0870.1987.tb00322.x

Durran, D. R., & Gingrich, M. (2014). Atmospheric predictability: Why butterflies are not of practical importance. *Journal of the Atmospheric Sciences*, *71*(7), 2476–2488. https://doi.org/10.1175/JAS-D-14-0007.1

Garny, H. (2025). Uaicon simulations on intrinsic predictability from the troposphere to the Mesosphere/Lower Thermosphere (MLT) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.14646394

Knobloch, S., Kaifler, B., Dörnbrack, A., & Rapp, M. (2023). Horizontalwavenumber spectra across the middleatmosphere from airborne lidarobservations during the 2019 southernhemispheric ssw. *Geophysical Research Letters*, *50*(14), e2023GL104357. https://doi.org/10.1029/2023GL104357

Koshyk, J. N., & Hamilton, K. (2001). The horizontal kinetic energy spectrum and spectral budget simulated by a high-resolution troposphere-stratosphere-mesosphere GCM. *Journal of the Atmospheric Sciences*, *58*(4), 329–348. https://doi.org/10.1175/1520-0469(2001)058<0329:THKESA>2.0.CO;2

Kunze, M., Zülicke, C., Siddiqui, T. A., Stephan, C. C., Yamazaki, Y., Stolle, C., et al. (2024). Ua-icon with NWP physics package (version: UA-icon-2.1): Mean state and variability of the middle atmosphere. *Geoscientific Model Development Discussions (GMDD)*, *2024*, 1–41. https://doi.org/10.5194/gmd-2024-191

Lilly, D. K. (1972). Numerical simulation studies of two-dimensional turbulence: Ii. Stability and predictability studies. *Geophysical & Astrophysical Fluid Dynamics*, *4*, 1–28. https://doi.org/10.1080/03091927208236087

Liu, J. J., Jones, D. B. A., Worden, J. R., Noone, D., Parrington, M., & Kar, J. (2009). Analysis of the summertime buildup of tropospheric ozone abundances over the middle East and North Africa as observed by the tropospheric emission spectrometer instrument. *Journal of Geophysical Research*, *114*(D5), d05304. https://doi.org/10.1029/2008jd010993

Lloveras, D. J., Tierney, L. H., & Durran, D. R. (2022). Mesoscale predictability in moist midlatitude cyclones is not sensitive to the slope of the background kinetic energy spectrum. *Journal of the Atmospheric Sciences*, *79*(1), 119–139. https://doi.org/10.1175/JAS-D-21-0147.1

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*(3), 289–307. https://doi.org/10.1111/j.2153-3490.1969.tb00444.x

Magnusson, L., Nycander, J., & Källen, E. (2009). Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus A: Dynamic Meteorology and Oceanography*, *61*(2), 194–209. https://doi.org/10.1111/j.1600-0870.2008.00385.x

Nastrom, G. D., & Gage, K. S. (1985). A climatology of aircraft wavenumber spectra observed by commercial aircraft. *Journal of the Atmospheric Sciences*, *42*(9), 950–960. https://doi.org/10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2

Orr, A., Bechtold, P., Scinocca, J., Ern, M., & Janiskova, M. (2010). Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, *23*(22), 5905–5926. https://doi.org/10.1175/2010JCLI3490.1

Pedatella, N. M., Chau, J. L., Schmidt, H., Goncharenko, L. P., Stolle, C., Hocke, K., et al. (2018). How sudden stratospheric warming affects the whole atmosphere. *Eos, Earth and Space Science News*, *99*, 6. https://doi.org/10.1029/2018EO092441

Pedatella, N. M., Liu, H.-L., Marsh, D. R., Raeder, K., & Anderson, J. L. (2019). Error growth in the mesosphere and lower thermosphere based on hindcast experiments in a whole atmosphere model. *Space Weather*, *17*(10), 1442–1460. https://doi.org/10.1029/2019SW002221

Polichtchouk, I., van Niekerk, A., & Wedi, N. (2023). Resolved gravity waves in the extratropical stratosphere: Effect of horizontal resolution increase from o(10) to o(1) km. *Journal of the Atmospheric Sciences*, *80*(2), 473–486. https://doi.org/10.1175/JAS-D-22-0138.1

Rotunno, R., & Snyder, C. (2008). A generalization of lorenzs model for the predictability of flows with many scales of motion. *Journal of the Atmospheric Sciences*, *65*(3), 1063–1076. https://doi.org/10.1175/2007JAS2449.1

Sato, K., Yasui, R., & Miyoshi, Y. (2018). The momentum budget in the stratosphere, mesosphere, and lower thermosphere. Part I: Contributions of different wave types and in situ generation of Rossby waves. *Journal of the Atmospheric Sciences*, *75*(10), 3613–3633. https://doi.org/10.1175/JAS-D-17-0336.1

Selz, T., & Craig, G. C. (2015). Simulation of upscale error growth with a stochastic convection scheme. *Geophysical Research Letters*, *42*(8), 3056–3062. https://doi.org/10.1002/2015GL063525

Selz, T., & Craig, G. C. (2023). Can artificial intelligence-based weather prediction models simulate the butterfly effect? *Geophysical Research Letters*, *50*(20), e2023GL105747. https://doi.org/10.1029/2023GL105747

Selz, T., Riemer, M., & Craig, G. C. (2022). The transition from practical to intrinsic predictability of midlatitude weather. *Journal of the Atmospheric Sciences*, *79*(8), 2013–2030. https://doi.org/10.1175/JAS-D-21-0271.1

Smith, A. K., Pedatella, N. M., Marsh, D. R., & Matsuo, T. (2017). On the dynamical control of the mesosphere-lower thermosphere by the lower and middle atmosphere. *Journal of the Atmospheric Sciences*, *74*(3), 933–947. https://doi.org/10.1175/JAS-D-16-0226.1

Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., et al. (2022). Atmospheric energy spectra in global kilometre-scale models. *Tellus A: Dynamic Meteorology and Oceanography*, *74*(2022), 280–299. https://doi.org/10.16993/tellusa.26

Vallis, G. K. (2005). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.

Wieczorek, M. A., & Meschede, M. (2018). Shtools—Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems*, *19*(8), 2574–2592. https://doi.org/10.1029/2018GC007529

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The icon (icosahedral non-hydrostatic) modelling framework of DWD and MPI-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579. https://doi.org/10.1002/qj.2378

Zhang, F., Snyder, C., & Rotunno, R. (2003). Effects of moist convection on mesoscale predictability. *Journal of the Atmospheric Sciences*, *60*(9), 1173–1185. https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2