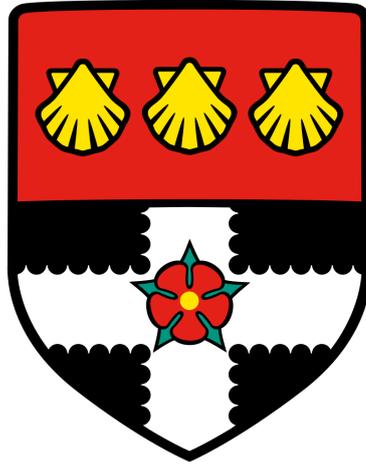# UNIVERSITY OF READING



# Weather Predictability:
# Some Theoretical Considerations

A thesis submitted for the degree of Doctor of Philosophy

**Tsz Yan LEUNG**

Centre for Doctoral Training in the Mathematics of Planet Earth

Department of Mathematics and Statistics

August 2020

(revised November 2020)

# Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Tsz Yan Leung

# Acknowledgements

First of all, I would like to thank Ted Shepherd, my primary supervisor, for his guidance and support. The breath of his knowledge has always sparked new inspirations. I am deeply grateful for his generous offer to bridge my funding gap, without which my studies would not have effected. Thanks for his encouragement towards conference attendance, where I have benefitted a lot through constructive discussions with other researchers.

I would also like to thank my co-supervisors Sebastian Reich and Martin Leutbecher for their contributions towards more specific areas of my work and for allowing me to spend time working at their respective institutions. I am pleased to have such a great and complementary team of experts as my supervisors.

Thanks to Richard Scott, from whom I had been provided with a turbulence code that served as a basis for my numerical experiments. These simulations would not have been possible without the high-performance computing resources at the ECMWF.

I wish to thank the staff and students of the MPE CDT and Ted's research group, who have been excellent mentors, companions and friends. I have much enjoyed being part of this community.

Finally, to my family and to Regine, thank you so much for your unconditional love and support.

*Deo gratias*, Who in His Divine Providence willed that my journey be shaped with the support of all these people and others who prayed for me. I owe you all my prayers.

# Weather Predictability:
# Some Theoretical Considerations

Tsz Yan Leung

## Abstract

The chaotic nature of atmospheric dynamics presents a central challenge to the accurate prediction of future weather. It is a well-known fact that the predictability of instantaneous weather is inherently limited to about two weeks, beyond which skilful prediction is impossible no matter how small the initial error is. This study seeks to advance the knowledge related to the limited predictability by addressing three theoretical topics.

The first topic concerns the mathematical origins of the predictability barrier. In a simplified context, what appears to be a contradiction between the finite-time limit and the regularity of the governing equations is reconciled through understanding the practical role of the slope of the energy spectrum in the latter.

The next topic explores the properties of error growth under the hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ energy spectrum that approximates the atmosphere. With the aid of simplified turbulence models, the synoptic-scale $k^{-3}$ range is found to substantially dampen the fast error growth characteristic of a $k^{-\frac{5}{3}}$ spectrum in the first decade of wavenumbers in the mesoscale range, so that the fast growth may only emerge when global numerical weather prediction models begin to resolve scales on the order of a few kilometres.

The final topic focusses on the relationship between metrics that quantify error growth and predictability. Two popular metrics, namely the Continuous Ranked Probability Score and the root-mean-square error, are found to be mathematically related under certain conditions. Simulated results show that the relationship approximately holds in idealised turbulent environments despite the required conditions not being fully met.

This study demonstrates that simple models can often be useful in identifying key mechanisms of error growth that lead to the limit of predictability. Future work involving simple models is encouraged to substantiate such understanding further.

# Contents

# List of Abbreviations

**2D** two-dimensional

**3D** three-dimensional

**CDF** cumulative distribution function

**CRPS** Continuous Ranked Probability Score

**DA** data assimilation

**ECMWF** European Centre for Medium-Range Weather Forecasts

**MSE** mean squared error

**NWP** numerical weather prediction

**ODE** ordinary differential equation

**PDE** partial differential equation

**PDF** probability density function

**RMSE** root-mean-square error

**SPPT** Stochastically Perturbed Parametrisation Tendencies

**SQG** surface quasi-geostrophic

# 1 Introduction

## 1.1 The early days of weather forecasting

The weather has been an inseparable part of human civilisation since time immemorial. Across the world, agricultural decisions are based on weather and climate. These then influence a region's economy through trading of agricultural yields. In this way, weather is pivotal to one's income and livelihood in agricultural economies. Before the advent of scientific understanding to weather phenomena, many cultures regarded weather events, and perhaps the success or failure to predict them, as divine interventions. Some polytheistic religions even had deities for specific weather phenomena. The religious attribution of weather, however, did not prevent humans from pursing their fundamental desire to predict the weather. Conventional wisdom gave rise to weather lores, mainly in the form of rhymes and short poems. Presumably based on anecdotal evidence, they described how certain observations would correlate with the local weather some time ahead, thereby giving rise to a pathway for predicting the weather. For example, the English saying *red sky at night, shepherd's delight; red sky in the morning, shepherd's warning* tells that a red sky at dusk (in the west) usually brings good weather on the next day, whereas a red sky at dawn (in the east) is an alert for inclement weather. As things turn out, this saying is supported by a typical feature of synoptic weather in the mid-latitudes where upper-level winds are predominantly westerly. Red sky around sunrise and sunset results from high atmospheric pressure through Rayleigh scattering. When a region of high pressure in the west is brought in by the westerly winds, it leads to fine weather on the next day. If the high pressure is to the east instead, unsettled weather in low-pressure regions may be on the way, since the high pressure has passed.

Other weather lores involve longer-term predictions. In medieval European cultures in particular, predictions on the sub-seasonal or seasonal scale were made based on weather elements or biological signs on given religious feast days. Some of these traditions were brought into the New World through colonial expansion and emigration (Groundhog Day being an example), despite that these long-term predictions, if of any value, would have been geographically local. Into the 21$^{st}$ century, St Swithun's Day (15$^{th}$ July) still remains as a notable example of folklore sub-seasonal weather prediction here in England, although its weather has no obvious relation with rainfall over the next 40 days (Sutton 1955).

Weather lores also exist in oriental cultures. Indeed, the red-sky saying above has a Chinese equivalent 朝霞不出門，晚霞行千里(*morning glow, stay at home; evening glow, walk a thousand miles*). However, the best-known oriental weather lore by far — the set of 24 solar terms — is more descriptive than predictive. These half-monthly solar terms tell a mixture of the arrival of seasons, typical weather phenomena, agricultural practices and biological signs, and is part of the traditional agricultural wisdom in east Asia.

Weather lores can be seen as a forerunner of analogue weather prediction, a class of forecasting methods based on exploiting historical patterns. An analogue forecaster makes statistical inferences on the future weather using purely historical knowledge. The invention of meteorological instruments in the 17$^{th}$ century gave rise to systematic and quantitative recording of weather observations. In this way, analogue forecasting gained momentum, which is still practised today in lesser-developed countries where it is too expensive to run atmospheric models. The discovery of the telegraph in the 19$^{th}$ century enabled near-real-time transfer of weather information across geographical locations, the seas included thanks to advances in marine navigation. Its impacts are profound. For weather prediction, it allowed the incorporation of spatial information into analogue forecasting, which had up to then been impossible. It also gave birth to synoptic meteorology, the understanding of how regional weather systems evolve. The growing knowledge in synoptic meteorology together with the abundance of observations in the Atlantic resulted in the success of perhaps the most important forecast ever issued by the United Kingdom's Meteorological Office, the forecast for D-day, 6$^{th}$ June, 1944 (Meteorological

Office n.d.). However, forecasts at that time were still made by "crude techniques of extrapolation, knowledge of local climatology and guesswork based on intuition" (Lynch 2008), and failed to fully capture the complex, non-linear advective processes of weather systems. As such, predictability was probably limited to no more than two or three days. Indeed, the forecasting team at the Meteorological Office was only confident enough to advise the Allied Forces of the D-day weather one and a half days ahead of the Normandy landings.

## 1.2    Equations of motion

The atmosphere is a fluid, whose motion can be understood by the fundamental principles of fluid dynamics. Fluid dynamics as a rigorous mathematical discipline dates back to the 18th century, when Leonhard Euler (1757) derived partial differential equations (PDEs) representing the conservation of mass and momentum in an inviscid hydrodynamic environment with the gravity $\mathbf{g}$ being the only external force:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{1.1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + \mathbf{g} \tag{1.2}$$

where $t$ is the time, $\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$ the differential operator for the three canonical directions $(x, y, z)$, $\mathbf{u}$ the three-dimensional (3D) velocity field, $\rho$ the density and $p$ the pressure. These equations are known as the Euler equations. Claude-Louis Navier later modified the momentum equation (Equation 1.2) to incorporate the effects of viscosity for Newtonian fluids:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + \mathbf{g} + \nu \nabla^2 \mathbf{u} \tag{1.3}$$

where $\nu$ is the kinematic viscosity coefficient. This equation, together with the continuity equation (Equation 1.1), are collectively referred to as the Navier-Stokes equations, which are also named after George Gabriel Stokes for his contribution in the dynamics of strongly viscous flows (Bistafa 2018).

A fundamental difference between atmospheric flows and idealised environments

3

where Equations 1.1 and 1.3 apply is the fact that Earth is rotating. The continuity equation does not change in a rotating frame of reference, but the momentum equation is further modified as

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \nu \nabla^2 \mathbf{u} \tag{1.4}$$

where $\boldsymbol{\Omega}$ is the angular velocity of Earth's rotation. The gravitational term $\mathbf{g}$ now incorporates a contribution from the centrifugal force associated with it (Vallis 2017).

Equations 1.1 and 1.4 together give four scalar equations but five unknown scalar variables (the three components of $\mathbf{u}$, $\rho$ and $p$), an under-determined problem. The ideal gas law

$$p = \rho R T, \tag{1.5}$$

where $R$ is the universal gas constant, diagnostically draws a link between the pressure and the density yet introduces a new variable, $T$, the absolute temperature (Vallis 2017). To close the system, therefore, it is necessary to invoke thermodynamic principles. The first two laws of thermodynamics together with the assumption that fluid motions are thermodynamically reversible give rise to a diagnostic quantity

$$\theta = T \left(\frac{p}{p_0}\right)^{-\frac{R}{c_p}} \tag{1.6}$$

known as the potential temperature, the temperature of a fluid parcel had it been brought adiabatically to a certain reference pressure $p_0$. Here, $c_p$ is the specific heat capacity of air at constant pressure. Finally, by specifying the rate of diabatic heating $\dot{Q}$ one obtains the equation

$$c_p \left(\frac{\partial \theta}{\partial t} + \mathbf{u} \cdot \nabla \theta\right) = \frac{\theta}{T}\dot{Q} \tag{1.7}$$

which completes the set of equations of motion (Vallis 2017).

In summary, Equations 1.1, 1.4, 1.5, 1.6 and 1.7 form a closed set of seven scalar equations for seven unknown scalar variables ($\mathbf{u}$ (three components), $\rho$, $p$, $T$ and $\theta$). This is a combination of hydrodynamic and thermodynamic principles, whose understanding flourished quickly in the 19th century.

## 1.3 The birth of numerical weather prediction

Revolutionary ideas for scientific and rigorous weather forecasting emerged at the turn of the 20th century. In 1901, the American scientist Cleveland Abbe argued that longer-range predictions could be made possible by pointing out that meteorological observations are broadly consistent with physical laws (Abbe 1901). He was soon echoed by the Norwegian Vilhelm Bjerknes, who framed weather forecasting more precisely as an initial-value problem: the future state of the atmosphere can be determined by its current state and the seven aforementioned equations of motion that govern it (Bjerknes 1904). Despite acknowledging that an analytic solution is out of question, he raised hope for longer-range predictability when numerical methods for solving these equations are developed and the network of meteorological observations become more extensive.

Lewis Fry Richardson, an Englishman working at the Meteorological Office, was inspired by Bjerknes' idea. He envisioned a forecast factory of human 'computers' (Figure 1.1) and made an attempt to calculate a forecast in retrospect. Although the forecast went badly wrong[1], the numerical methods employed in solving the problem were solid and laid a foundation for modern-day numerical weather prediction (NWP) (Richardson 1922, Lynch 2006). That being said, however well NWP methods are able to accurately predict the future weather, they would be useless without a machine that is capable of solving the initial-value problem fast enough for the forecast to be issued before the weather event materialises. The lack of computers thus put an effective halt to further progress in NWP until the mid-century, when the arrival of the Electronic Numerical Integrator and Computer facilitated the first computer forecasts in 1949. This was a significant accomplishment, not only because the forecasts were computed faster (albeit very slightly) than weather advanced, but also because synoptic-scale features were predicted with reasonable accuracy despite the simplest barotropic model being used (Charney et al. 1950). With further preparations, real-time NWP became operational in 1954 in Sweden and the year after in the United States of America (Persson 2005).

In the meantime, the understanding of atmospheric dynamics as a mathematical

---

[1]The error was attributed to an imbalance in the initial condition which resulted in spurious gravity wave oscillations (Lynch 2006).

Figure 1.1: Artist Stephen Conlin's depiction of Richardson's forecast factory. Courtesy of the Irish Times.

discipline continued to develop. Through the study of linearised barotropic and baroclinic flows, it became clear that the idealised atmosphere is a dynamically unstable system in the sense that small disturbances would amplify and contaminate predictions (Charney 1947, Eady 1949, Thompson 1953). This naturally provokes the question how far ahead weather could potentially be predicted. Thompson (1957) foresaw the existence of a point beyond which fast-growing error would rapidly diminish the benefits of doubling the density of weather stations. Later, Edward Lorenz (1963) famously demonstrated the chaoticity of atmospheric motions using a simple 3D homogeneous, autonomous and non-linear system of ordinary differential equations (ODEs)

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= 10(y - x) \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= x(28 - z) - y \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= xy - \frac{8}{3}z
\end{aligned}
$$

$$(1.8)$$

that represents cellular convection. Mathematically, finite-dimensional chaos can be characterised by the co-existence of the following three properties on the attractor of the dynamical system (Devaney 1989):

1. topological transitivity, which can in a loose sense be thought of as 'nearly every solution trajectory visits everywhere';

6

2. sensitive dependence of solutions on initial conditions;

3. density of periodic solutions.

Lorenz's discovery is profound in two aspects. First, on the theoretical side, it establishes that three-dimensionality of the dynamical system is not only necessary but also sufficient for chaotic behaviour (the earlier Poincaré-Bendixson theorem stipulates that chaos cannot happen with only two degrees of freedom). The existence of a strange attractor — an attractor with a fractal structure — suggests that the dynamics of this three-variable system is extremely complicated. Second, it shows that all periodic orbits are unstable: unless an initial condition sits exactly on a periodic orbit, its solution trajectory may never approach a given periodic orbit. This means it would be futile to predict the weather in the long range by identifying the periodic solutions of the atmosphere's governing equations.

## 1.4 The predictability horizon

While Lorenz's 1963 paper precludes predicting the weather by an essentially analogue method taking advantage of the system's periodic solutions, skilful weather prediction is still possible by dynamical methods. The divergence of solution trajectories resulting from the system's sensitive dependence on initial conditions implies that such predictions would become useless beyond a certain range of predictability. However, this range is generally expected to increase as the initial condition generating the prediction becomes more accurate. Since solutions of Lorenz's three-variable system are known to depend *continuously* on initial conditions despite being *sensitive* to them (Palmer et al. 2014), there is no upper bound to the range of predictability in the limit of small initial errors. In other words, the predictability horizon is infinite.

The concept of a predictability horizon can be motivated in a mathematically rigorous manner. In a perfect-model context of a deterministic autonomous dynamical system $\Phi : \mathbb{R} \times X \to X$ where $X$ is the state space, the dynamics of the error $\varepsilon = \Phi(t, x + \varepsilon_0) - \Phi(t, x)$ can be fully described by the lead time $t$, the initial error $\varepsilon_0$ and

7

the initial state of the system $x$. The lead time is the time elapsed since the introduction of the error. To quantify the accuracy of a given forecast, a measure of accuracy $S$ can be defined as a continuously increasing function of some norm $\|\cdot\|$ of the error, which makes it a function of $t$, $\varepsilon_0$ and $x$. Here we have assumed that the accuracy measure is negatively oriented, i.e. lower $S$ implies better accuracy. Now, averaging over the initial states on some (non-trivial) attractor $D$ of the system, we may define the overall measure of accuracy $\overline{S}(t, \varepsilon_0) := \int_{x \in D} S(t, \varepsilon_0, x)\, \mathrm{d}x$ of the forecast system. If we further assume that the error norm increases with $t$ in the average, which is a generic property of atmospheric flows up to a certain threshold where it asymptotes, then $\overline{S}(t, \varepsilon_0)$ monotonically increases in time.

Let's say that a prediction loses its skill when $\overline{S}$ exceeds some threshold value $\alpha$. The range of predictability $\tilde{T}(\varepsilon_0)$ for a given initial error $\varepsilon_0$ is the solution to $\overline{S}(t, \varepsilon_0) = \alpha$, whose uniqueness is guaranteed by the monotonicity assumption of $\overline{S}$. The predictability horizon is the limit of $\tilde{T}$ as $\|\varepsilon_0\| \to 0$ if it exists, or $\liminf_{\|\varepsilon_0\| \to 0} \tilde{T}(\varepsilon_0)$ more generally. It should be noted that the range of predictability and therefore its limit depend on the specification of $\overline{S}$ (or $S$) and $\alpha$. In deterministic weather forecasts, $S$ is often the energy of the full error or some scale-filtered error, in which case $\alpha$ is a multiplicative factor of the basic-state flow's climatological energy level close to two, where the error eventually saturates (Leung 2017). Yet, since $\tilde{T}(\varepsilon_0 = 0) = \infty$ by the very definition of deterministic systems, a well-posed system like Lorenz's 1963 system (Equations 1.8) would have indefinite predictability, i.e. $\liminf_{\|\varepsilon_0\| \to 0} \tilde{T}(\varepsilon_0) = \infty$, regardless of how $\overline{S}$ and $\alpha$ are defined.

In a seminal paper a few years later, this time using the vorticity form of the two-dimensional (2D) incompressible Euler equations[2]

$$\frac{\partial \theta}{\partial t} + J(\psi, \theta) = 0, \qquad \theta = \Delta \psi \tag{1.9}$$

where $\psi$ is the velocity streamfunction[3], $\theta$ is the vorticity, $\Delta$ is the 2D Laplacian operator[4] and $J(A, B) = \frac{\partial A}{\partial x}\frac{\partial B}{\partial y} - \frac{\partial A}{\partial y}\frac{\partial B}{\partial x}$, Lorenz argued that the predictability horizon for Earth's

---

[2]This is also known as the 2D barotropic vorticity equation or the 2D barotropic vorticity model.

[3]The velocity streamfunction $\psi$ is related to the velocity $(u, v)$ by $u = -\frac{\partial \psi}{\partial y}$ and $v = \frac{\partial \psi}{\partial x}$.

[4]$\Delta = \nabla \cdot \nabla$, where $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$.

atmosphere is inherently finite (Figure 1.2), that is, $\tilde{T}$ cannot be made arbitrarily large by reducing $\|\varepsilon_0\|$ to anything below a threshold unless $\varepsilon_0$ is exactly zero (Lorenz 1969). Mathematically, this is equivalent to $\liminf_{\|\varepsilon_0\| \to 0} \tilde{T}(\varepsilon_0) < \infty$, a more radical notion than the mere chaos demonstrated in the 1963 model (Palmer et al. 2014). A crucial assumption behind Lorenz's conclusion is that the wavenumber spectrum of the flow's energy $E(k)$ (where $k$ is the scalar wavenumber) scales as $k^{-\frac{5}{3}}$. A remark near the end of the paper speculates that whether the predictability horizon is finite depends on the energy spectrum's so-called 'spectral slope' $-p$, wherein $E(k) \propto k^{-p}$. It also hypothesises, by extrapolation, that predictability would become indefinite if the small scales had sufficiently little energy compared to the large scales in such a way that $p \geq 3$. This claim will be reviewed in Chapter 2 as Lorenz's argument on the predictability horizon is being discussed in detail.
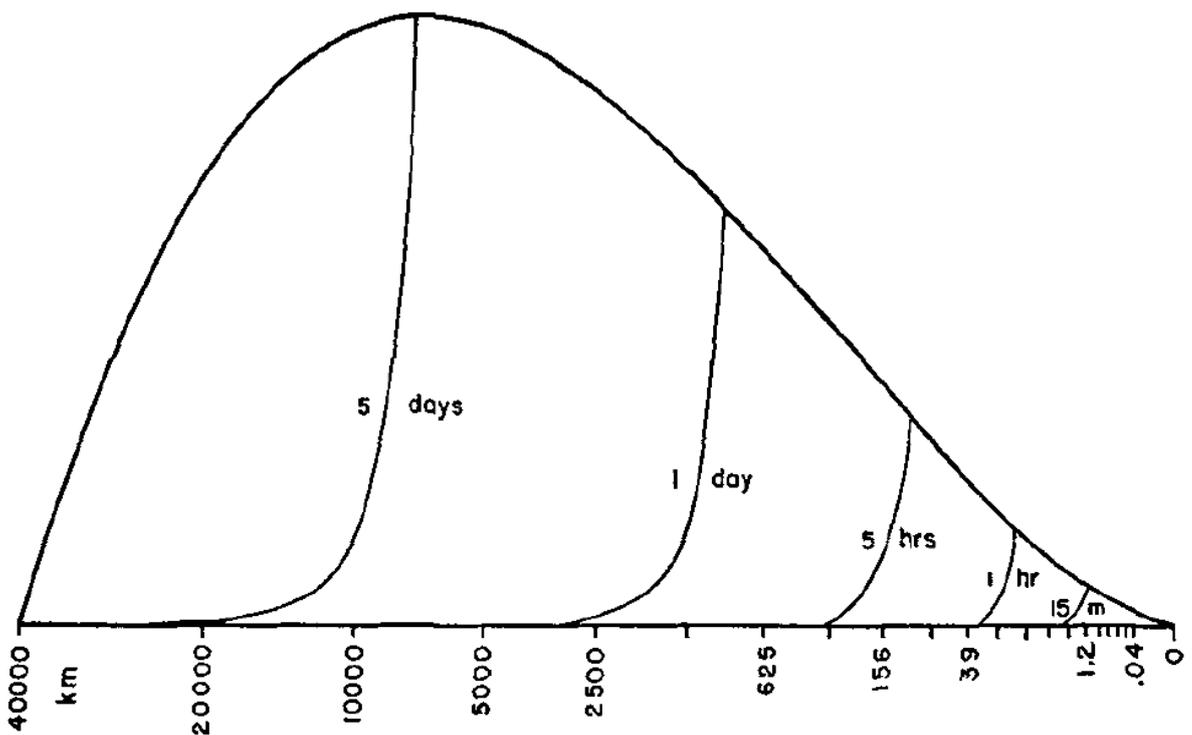


Figure 1.2: A depiction of faster-growing errors at smaller scales that results in a finite predictability horizon, taken from Lorenz (1969). The thick curve represents the basic-state energy spectrum, while each thin curve, together with the portion of the thick curve to the right of their intersection, represents the error energy spectrum at the specified forecast lead time.

Earth's atmospheric energy spectrum was not so well-known at Lorenz's time of writing. He seems to have based his choice of $p = \frac{5}{3}$, from which he derived that atmospheric predictability is inherently finite, on Kolmogorov's theory of 3D idealised

(incompressible, homogeneous and isotropic) turbulence (Kolmogorov 1941, Vallis 2017). The theory suggests that the $k^{-\frac{5}{3}}$ spectrum is generated through an inertial downscale cascade of energy. However, atmospheric flows are quasi-two-dimensional in scales down to $O(10)$ kilometres, and a theory of 2D turbulence had only emerged shortly before Lorenz wrote his 1969 paper. Apart from $p = \frac{5}{3}$, 2D idealised turbulence admits another spectral slope corresponding to $p = 3$ (Kraichnan 1967). Lorenz's conclusion would probably have changed had the latter spectral slope been chosen for the predictability analysis.

On the observational side, early indications using limited data showed that the large-scale atmospheric energy spectrum follows a power-law closer to $k^{-3}$ than $k^{-\frac{5}{3}}$ (Horn & Bryson 1963, Charney 1971). A global analysis had not been possible until satellite measurements became available a couple of decades later. Using a dataset derived from satellite data, Boer & Shepherd (1983) confirmed a $k^{-3}$ spectrum at the large scales. Later, with the aid of aircraft observations, Nastrom & Gage (1985) reported a transition from a $k^{-3}$ range to a $k^{-\frac{5}{3}}$ range in the mesoscale, at a wavelength of about 400 kilometres (Figure 1.3).

Yet, it should be kept in mind that atmospheric turbulence is far from being idealised. The classical picture of 3D and 2D turbulence (Kolmogorov 1941, Kraichnan 1967) assumes a spectrally localised forcing that generates an inertial cascade of some invariant quantity either upscale or downscale[5]. However, some sources of forcing have a continuous spectrum instead, orography being an example (van Niekerk et al. 2016). The presence of orography also makes the turbulence inhomogeneous. Moreover, the effects of Earth's rotation and the existence of coherent vortices add anisotropy to the turbulence.

A study by Maltrud & Vallis (1991) suggests that the hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ atmospheric spectrum can be simulated with 2D idealised turbulence by forcing separately at large and small scales. While there is a general consensus that the synoptic-scale $k^{-3}$ range, or indeed its logarithmically corrected version (Bowman 1996), is indeed a feature of 2D turbulence with a downscale enstrophy cascade, the physical origins of the $k^{-\frac{5}{3}}$ mesoscale range remain under debate. On one hand, it is thought to be a result of bal-

---

[5]The $k^{-3}$ range in 2D turbulence represents a downscale transfer of enstrophy, whereas the $k^{-\frac{5}{3}}$ range in 2D (resp. 3D) turbulence represents an upscale (resp. downscale) transfer of energy.
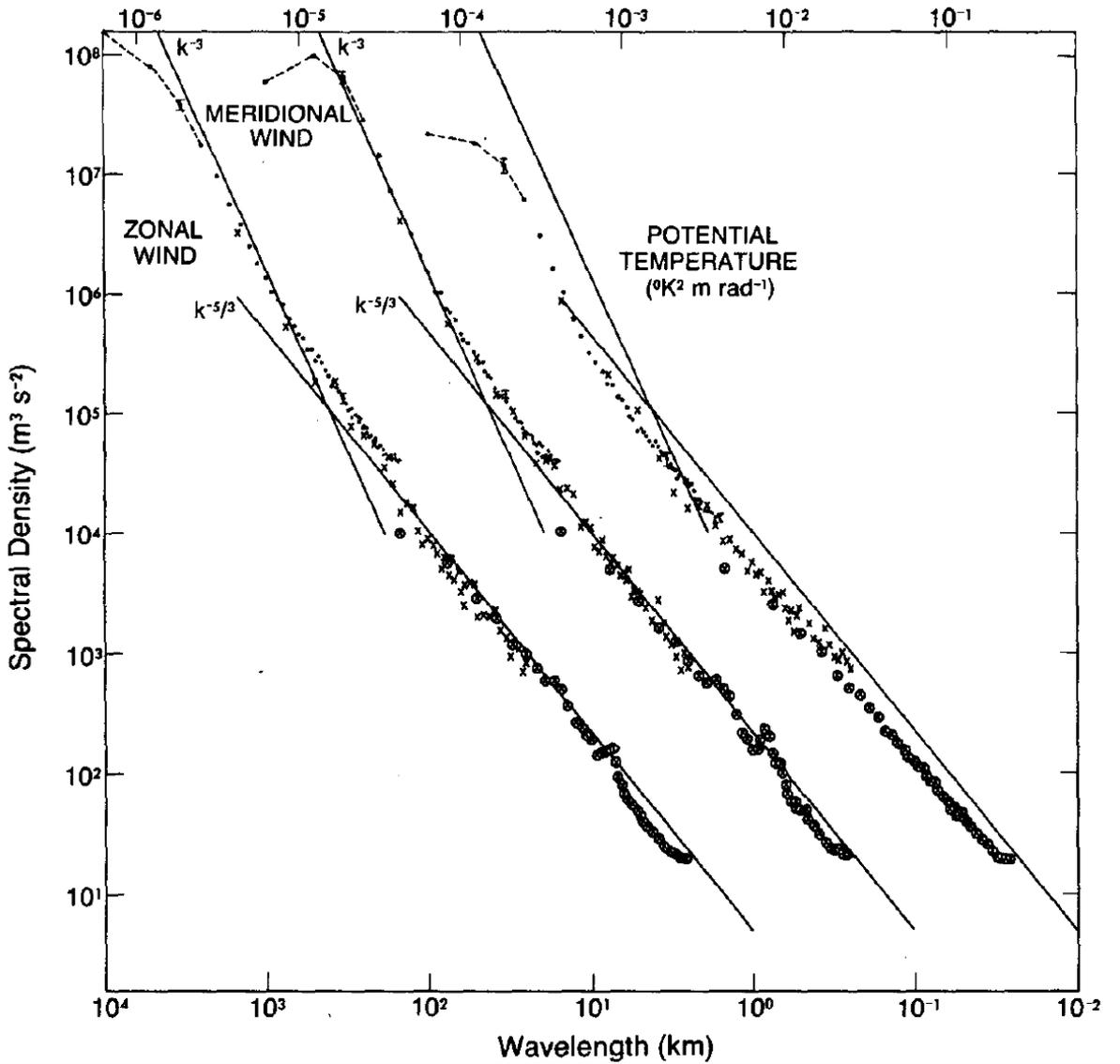
Figure 1.3: Power spectra of wind and potential temperature near the tropopause from aircraft data, taken from Nastrom & Gage (1985). The spectra for meridional wind and potential temperature are respectively shifted by one and two decades to the right. Lines showing the $k^{-3}$ and $k^{-\frac{5}{3}}$ scalings are inserted for reference.

anced[6] mesoscale motion, cascading energy upscale as in 2D idealised turbulence (Lilly 1989, Vallis et al. 1997). Lindborg (1999) however argued the opposite, that the mesoscale range cannot be explained by the inverse energy cascade in 2D turbulence theory. Rather, as it was later proposed, a $k^{-\frac{5}{3}}$ spectrum is generated through the downscale transfer of energy associated with unbalanced motion, but it is masked at larger scales by the $k^{-3}$ range for being too weak compared to the spectrum of balanced motion (Waite & Snyder 2009). More recently, Waite & Snyder (2013) discovered that moist processes such as convection and clouds play an important role in the shallowing of the spectrum from $k^{-3}$

---

[6]Atmospheric motion can be decomposed into a balanced part and an unbalanced part. Balanced motion is characterised by the conservation of potential vorticity, whereas the unbalanced part refers to the residual motion. Inertia-gravity waves are a primary contributor to unbalanced motion.

to $k^{-\frac{5}{3}}$, a result that was echoed by Sun & Zhang (2016).

Whatever the origins of the $k^{-\frac{5}{3}}$ range are, they may not be so relevant to the error growth properties and the finite predictability horizon associated to this shallower spectrum, for Rotunno & Snyder (2008) pointed out that predictability properties are primarily a function of the energy spectrum rather than model dynamics *per se*. They demonstrated this by generalising Lorenz (1969)'s model to the surface quasi-geostrophic (SQG) equations (Held et al. 1995), which can be equivalently expressed as

$$\frac{\partial \theta}{\partial t} + J(\psi, \theta) = 0, \qquad \theta = -(-\Delta)^{\frac{1}{2}} \psi, \tag{1.10}$$

and studied the sensitivity of the results to the dynamics and the energy spectrum independently. It is thought that the finite atmospheric predictability horizon is a result of the shallower mesoscale spectrum, independent of the spectral slope in the synoptic scale. Yet, it remains not so obvious as to how the presence of the $k^{-3}$ range in the synoptic scale could have an impact on error growth in the mesoscale and the predictability limit. This will be investigated in Chapter 3.

Lorenz's hypothesis that the predictability horizon is finite if and only if the spectral slope $-p$ is shallower than $-3$ is also supported by dimensional analysis (Vallis 1985, Lilly 1990), a tool often used in turbulence studies. Perhaps for this reason the hypothesis has now become an accepted theory attributed to Lorenz himself. In the dimensional analysis formulation, the characteristic timescale $T(k)$ taken for an error to grow from wavenumber $2k$ to $k$ depends only on $k$ itself and the energy spectral density $E(k) \propto k^{-p}$, due to the turbulence's self-similarity. Since $k$ has the dimension of inverse length and $E(k)$ has the dimension of length cubed divided by time squared, the only way of combining them to form a characteristic timescale is $T(k) \propto (k^3 E(k))^{-0.5}$. Therefore, for an initial perturbation at wavenumber $k_p = k^n$ (where $n \in \mathbb{N}$), the range of predictability at wavenumber $k$ is

$$\tilde{T}_k(k_p) = \sum_{j=0}^{n-1} T(2^j k) \propto \sum_{j=0}^{n-1} (2^j k)^{\frac{p-3}{2}} = \begin{cases} k^{\frac{p-3}{2}} \left( \frac{1 - 2^{\frac{p-3}{2}n}}{1 - 2^{\frac{p-3}{2}}} \right) & p \neq 3 \\ n & p = 3 \end{cases}. \tag{1.11}$$

Hence, the predictability horizon $\lim_{k_p \to \infty} \tilde{T}_k(k_p) = \lim_{n \to \infty} \tilde{T}_k(k^n)$ is finite if and only if $p < 3$. For a spectral slope equal to or steeper than $-3$, $\tilde{T}_k(k_p)$ can indefinitely be extended by reducing the scale of the initial error (i.e. increasing $k_p$).

When predictability is inherently finite, the predictability horizon is a function of spatial scale. Taking the limit of Relation 1.11 for the case $p < 3$ gives

$$\lim_{k_p \to \infty} \tilde{T}_k(k_p) \propto k^{\frac{p-3}{2}}. \tag{1.12}$$

In particular, for a $k^{-\frac{5}{3}}$ spectrum, the predictability horizon scales as $k^{-\frac{2}{3}}$ with the wavenumber of interest. This is consistent with Lorenz (1969)'s observations and suggests that doubling the length scale increases the predictability horizon by a factor of roughly 1.6 (Table 1.1).

| Length scale | Predictability horizon | Successive ratio |
|---|---|---|
| $20000 - 40000$ km | 16.8 days | 1.66 |
| $10000 - 20000$ km | 10.1 days | 1.80 |
| $5000 - 10000$ km | 5.6 days | 1.75 |
| $2500 - 5000$ km | 3.2 days | 1.78 |
| $1250 - 2500$ km | 1.8 days | 1.64 |
| $625 - 1250$ km | 1.1 days | 1.68 |
| $313 - 625$ km | 15.7 hours | 1.65 |
| $156 - 313$ km | 9.5 hours | 1.64 |
| $78.1 - 156$ km | 5.8 hours | 1.56 |
| $39.1 - 78.1$ km | 3.6 hours | 1.64 |
| $19.5 - 39.1$ km | 2.2 hours | 1.69 |
| $9.77 - 19.5$ km | 1.3 hours | 1.53 |
| $4.88 - 9.77$ km | 51.1 minutes | 1.59 |
| $2.44 - 4.88$ km | 32.1 minutes | 1.58 |
| $1.22 - 2.44$ km | 20.3 minutes | 1.56 |
| $610$ m $- 1.22$ km | 13.0 minutes | 1.55 |
| $305 - 610$ m | 8.4 minutes | 1.47 |
| $153 - 305$ m | 5.7 minutes | 1.43 |
| $76.2 - 153$ m | 4.0 minutes | 1.29 |
| $38.1 - 76.2$ m | 3.1 minutes | 1.07 |
| $19.1 - 38.1$ m | 2.9 minutes | — |

Table 1.1: Predictability horizons at various length scales estimated by Lorenz (1969), and their successive ratios (the predictability horizon in the current row divided by that in the next row) correct to two decimal places.

The popular term 'butterfly effect' refers to this stronger notion of unpredictability. While earlier authors had used the same or similar metaphors to describe certain degrees of unpredictability of the atmospheric system (Franklin (1898) used grasshoppers

whereas Smagorinsky (1969) used butterflies), it was only until 1972 when the term 'butterfly effect' was coined after a scientific presentation given by Lorenz about his recent (1969) discovery of the finite predictability horizon. Some have instead attributed the butterfly effect to his three-variable chaotic system of 1963 (Equations 1.8), probably because its attractor coincidentally resembles the shape of a butterfly when viewed from a certain angle. However, this is just one of the many examples in which the true meaning of a phrase "has become distorted by the passage of time" (Palmer et al. 2014).

## 1.5   Probabilistic forecasts

The discoveries that atmospheric motions depend sensitively on initial conditions and a finite-time predictability horizon exists make it impossible to fully base weather forecasting on deterministic methods. Bjerknes' vision that weather can be deterministically predicted as an initial-value problem (Bjerknes 1904) would be ruined by the inevitable errors in estimating the atmosphere's initial state, let alone the incomplete understanding of the physical laws, the effects of boundaries, and numerical errors resulting from the discretisation of the equations and the limited precision of the machines that solve them. This is not to say that deterministic models are of no use, as operational forecasting centres are still using them today (Haiden et al. 2019), but an alternative avenue of prediction had to be sought. One way of such is to frame the forecasting problem as a probabilistic one. Conceptually, the initial conditions and the model parameters are treated as random variables in abstract probability spaces, and the evolution of weather is treated as a stochastic-dynamic process Epstein (1969). Yet it should be noted that the probabilistic framework does not suggest a random element in the evolution of weather; it merely accounts for our imperfect knowledge about it.

On a practical level, instead of running a single forecast using the forecaster's best estimate of the initial state and the best model, an ensemble of 'equally probable' scenarios is integrated in time using models with stochastically perturbed parameters. This first became operational at the United States' National Meteorological Center and the European Centre for Medium-Range Weather Forecasts (ECMWF) in 1992 (Toth &

Kalnay 1993, Molteni et al. 1996). The statistics of the results are reported as probabilities, which facilitate decision-making with a risk-based approach. Richardson (2000) showed that probabilistic predictions have generally better potential economic value than deterministic forecasts.

The accurate estimation of probability distributions of the initial conditions and the model parameters, or equivalently the accurate sampling of 'equally probable' scenarios, remains the central challenge in the development of probabilistic forecasting. With an idealised turbulence model, Leith (1974) argued that a Monte-Carlo sampling from a known Gaussian distribution around the best initial-condition estimate would produce a more skilful mean-square wind forecast than a deterministic prediction. Yet, the applicability of Leith's approach to the real atmosphere is uncertain, since the estimation of covariances in an inhomogeneous and anisotropic environment is a difficult task even when Gaussian distributions are assumed (Massart 2019). Even if the covariances can be reasonably well-modelled, Monte-Carlo sampling alone does not guarantee that the model's prognostic variables could be initialised in a dynamically consistent way. Such dynamical inconsistencies would induce unrealistic shocks in the short-range forecast that would compromise the forecast model's skill.

At the ECMWF, the ensemble of initial conditions is generated using a combination of singular vectors and ensemble data assimilation approaches. The singular-vector method (Leutbecher & Lang 2014) provides a dynamical estimate of the initial state's uncertainty. It identifies the most unstable modes of the atmospheric system on which random Gaussian perturbations are applied. The ensemble data assimilation algorithm (Isaksen et al. 2010) uses a Bayesian framework to compute, for each ensemble member, the optimal combination between the previous model integration for that member and the latest observations. Information about covariances between errors of model variables are implicitly incorporated into the optimisation algorithm, as are the covariances between observation errors. Its results are combined with the singular vectors to initialise ensemble forecasts.

Accounting for model uncertainty is a much more challenging task. The ideal is to build a stochastic equivalent of the dynamical model (Equations 1.1, 1.4–1.7), thus

allowing a stochastic representation of uncertainty in every dynamical and physical process. However, the work around the mathematics of such a complex model, let alone its discretisation, has only started in recent years (Holm 2015). While the NWP community awaits the development of such new science, model uncertainty is represented by empirical methods, primarily in a scheme known as SPPT — Stochastically Perturbed Parametrisation Tendencies (Palmer et al. 2009). Parametrisation of sub-grid-scale processes is unavoidable due to the closure problem inherent to turbulence (Vallis 2017). It empirically seeks an aggregated representation of motions beyond the model's resolution in terms of motions at larger scales. By perturbing the effects (the 'tendencies') these parametrisations have on the model's variables, the SPPT scheme improves the variability produced by the ensemble, especially in the tropics where convective activity is strong (Buizza et al. 1999, Palmer et al. 2009). The strength of the SPPT scheme is experimentally tuned so that the variance of the ensemble may match the mean squared error (MSE) of the ensemble mean in a broad sense. Recent studies have shown that introducing independent SPPT patterns among different parametrisation schemes further improves the ensemble's skill (Christensen et al. 2017). Apart from the SPPT scheme, other options of representing model uncertainty such as perturbing the parameters themselves (instead of their tendencies on model variables) are being explored (Ollinaho et al. 2017, Leutbecher et al. 2017).

We have seen in Section 1.4 that for deterministic forecasts, measures of accuracy are functions of error norm. In the context of probabilistic forecasts, they become functions of error distribution[7], i.e. distribution of $\varepsilon = \Phi(t, x + \varepsilon_0) - \Phi(t, x)$. A range of such measures, often known as 'scoring rules' or simply 'scores' in a probabilistic context, are commonly used in NWP operations to evaluate forecast performance. They can be categorised by type of state space $X$ of the forecast (Gneiting & Raftery 2007). For scalar variables ($X \subseteq \mathbb{R}$), the Continuous Ranked Probability Score (CRPS) is most commonly used. It is the integral of the squared difference between the cumulative distribution function (CDF) of $\Phi(t, x + \varepsilon_0)$ and the Heaviside step function[8] at $\Phi(t, x)$, or equivalently the integral of the squared difference between the CDF of $\varepsilon$ and the Heaviside step function

---

[7]In a probabilistic framework, the initial error $\varepsilon_0$ and therefore the error $\varepsilon$ are interpreted as random variables to account for the ensemble of initial conditions. For the sake of notational simplicity, the effects of model errors are not included so that $\Phi$ remains a deterministic function.

[8]The Heaviside step function $H_y(x)$ at a given threshold $y$ takes the value 1 if $x \geq y$ and 0 if $x < y$.

at zero. Scoring rules need not take into account the full forecast distribution like the CRPS does. For example, the quantile score compares the observed value with a certain quantile of the predictive distribution (Gneiting & Raftery 2007). Scores for scalar fields can be obtained through spatial integration of scores for scalar variables. For dichotomous forecasts ($X = \{0, 1\}$ depending on whether an event takes place), a popular choice is the Brier score (Wilks 2019), which for a single forecast is $(\mathbb{E}[\varepsilon])^2$, the squared difference between the forecast probability $\mathbb{E}[\Phi(t, x + \varepsilon_0)]$ and the outcome $\Phi(t, x) \in \{0, 1\}$. Gneiting & Raftery (2007) also discussed scores more generally for categorical variables ($X$ being a finite set). All scoring rules named above are strictly proper, an essential requirement for them to be put into practical use. Being strictly proper essentially means that the overall score $\overline{S}$ (i.e. the averaged score over many instances) is optimised if and only if the forecast and observed distributions agree. This will be discussed in more detail in Chapter 4, in which the relationship between forecast verification metrics will be investigated in light of concepts related to strictly proper scoring rules.

Probabilistic forecasts are shown to perform better than deterministic predictions in the medium-range. Haiden et al. (2015) quantified the additional skill in terms of the CRPS and saw a 30% improvement for five-day forecasts for the extra-tropics, up from 15%–20% ten years before their study. Another aspect in which ensembles outperform deterministic forecasts is forecast consistency, which measures the jumpiness of forecasts issued at successive times yet valid at the same time. Jumpy forecasts, even if relatively rare, can damage the credibility of the forecast system as a whole. Richardson et al. (2020) showed that ensembles greatly reduce these jumps, especially at longer lead times about two weeks ahead of the actual event.

## 1.6    Estimates of the predictability horizon

The first estimate of the predictability horizon was provided by Lorenz (1969) in the same paper as the revolutionary idea of inherently finite predictability was proposed. With the overall measure of accuracy $\overline{S}$ being the error energy and the threshold $\alpha$ for the loss of predictability being about 1.3 times the basic-state energy level, the global

predictability horizon was estimated as 16.8 days. Lorenz also came up with estimates of the predictability horizons at smaller scales which, as already noted in Section 1.4, nicely conform to the scaling estimate for a $k^{-\frac{5}{3}}$ spectrum (Relation 1.12, Table 1.1). In addition, the predictability horizon was found to be insensitive to the spatial scale of the initial error.

In probabilistic forecasts, the threshold $\alpha$ for the loss of predictability is customarily chosen as the $\overline{S}$ of the climatological distribution. Predictability is deemed to have been lost when the prediction ceases to be statistically distinguishable from a random prediction from the climatology. It is not clear from a theoretical perspective whether probabilistic predictions extend the predictability horizon. However, empirical estimates reaffirm the existence of a finite-time global predictability horizon for probabilistic forecasts on the order of two to three weeks (Buizza & Leutbecher 2015, Selz 2019, Zhang et al. 2019). This is echoed by a recent estimate of the deterministic predictability horizon using global convection-permitting model simulations (Judt 2018). Based on these estimates, it appears unlikely that the predictability horizon for probabilistic predictions is any different from that for deterministic predictions, although probabilistic predictions are more skilful at shorter lead times, as discussed earlier in Section 1.5.

Some of these studies estimated the predictability horizon with the aid of parametric error growth models. Essentially, error growth data were fitted to a simple parametric equation representing the generic pattern of error growth. The predictability horizon was then inferred using the parametric equation with the fitted parameters but with the initial error changed to a small value compared to the typical initial error of today's NWP models. One of such equations used is a scaled and translated hyperbolic tangent function

$$E(t) = A \tanh(at + b) + B, \tag{1.13}$$

where $E$ is some measure of error energy, $t$ is the lead time, and $A \geq 0, B \in \mathbb{R}, a \geq 0$ and $b \in \mathbb{R}$ are parameters to be fitted (Žagar et al. 2017). The measure of error energy can either be the total error energy or that at selected wavenumbers. It may or may not be normalised by the saturation energy level. The $E$ given in Equation 1.13 satisfies the

autonomous ODE

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{a}{A}(E_{\max} - E)(E - E_{\min}) \tag{1.14}$$

where $E_{\max} := A + B$ and $E_{\min} := B - A$ are respectively the supremum and infimum attainable values of $E$ over all $t \in \mathbb{R}$. Equation 1.14 can thus be considered as an evolution equation for the error, with an initial condition of $E(0) = A\tanh(b) + B$. From this equation, it can be seen that the parametric model is equivalent to the one proposed by Dalcher & Kalnay (1987)

$$\frac{\mathrm{d}E}{\mathrm{d}t} = (\alpha_1 E + \alpha_2)\left(1 - \frac{E}{E_{\max}}\right) \tag{1.15}$$

by noting that $\alpha_1 = \frac{a}{A}E_{\max}$ and $\alpha_2 = -\frac{a}{A}E_{\max}E_{\min}$ (Žagar et al. 2017). Loosely speaking, $\alpha_1$ represents the rate of exponential growth of the initial-condition error, $\alpha_2$ accounts for the short-term linear growth induced by model errors, and the factor $1 - \frac{E}{E_{\max}}$ parametrises the slowing down of error growth towards saturation. Yet, the growth of model errors is not necessarily linear (Vannitsem & Toth 2002, Leung 2017), and strictly speaking the effects of initial and model errors cannot be decoupled (Leung 2017, Žagar et al. 2017). Although the physical basis for Equation 1.15 has not been fully justified, it nevertheless often provides a nice fit to error growth data.

Although the fit using Equations 1.13–1.15 is popular among the meteorological community, it is not unique. Another parametric error growth model used in the literature is

$$E(t) = E_0 \exp\left(\beta_1\left(1 - e^{-\beta_2 t}\right)\right) \tag{1.16}$$

where $E_0, \beta_1, \beta_2 > 0$ (Selz & Craig 2015, Selz 2019), which is equivalent to the autonomous ODE

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta_2(\beta_1 - \log\frac{E}{E_0})E \tag{1.17}$$

with the initial condition $E(0) = E_0$. Here, the meaning of the parameters is less obvious than in Equation 1.15. Yet, both parametric models consistently suggest a predictability horizon of about 17 days (Selz 2019, Zhang et al. 2019) and the result agrees with other estimates of the predictability horizon (Buizza & Leutbecher 2015, Judt 2018). Therefore, it is safe to conclude that the estimate is robust, although its agreement with Lorenz

(1969)'s original estimate might be a coincidence, as Lorenz's assumption of a single-range $k^{-\frac{5}{3}}$ energy spectrum is now known to be incorrect (Boer & Shepherd 1983, Nastrom & Gage 1985, Judt 2018).

Judt (2020) recently discovered that the predictability horizon varies substantially among latitudinal zones due to their different dynamical characteristics. Notably, the tropical region remains predictable beyond 20 days, despite a faster error growth in the short term than the mid-latitudes and the polar regions. This is attributed to the presence of equatorial waves which modulate the weather over longer timescales, such as equatorial Kelvin, Rossby and mixed Rossby-gravity waves. Compared to baroclinic waves in the mid-latitudes, these waves are less prone to error growth and therefore provide the additional predictability.

The predictability horizon also varies in time. The recent study by Selz (2019) concluded that the predictability horizon changes with hemispheres and seasons. It was found to be longer in the Southern Hemisphere than in the Northern Hemisphere, and in winter than in summer. Moreover, retrospective analyses of 'forecast busts' — cases in which all major NWP models experienced a temporary drop of skill and failed to produce the right forecast — have shown that the predictability in these cases was intrinsically lower (Rodwell et al. 2013). This suggests that forecast busts are sometimes unavoidable, although it is not clear whether the flow-dependent predictability can itself be predicted so that potential cases of forecast bust can be warned in advance.

## 1.7   Practical predictability

Over the past decades, NWP performance and hence the (practical) range of predictability has steadily improved. Figure 1.4 shows such improvement at the ECMWF since the early 1980s. On average, about one day of predictive skill has been gained every decade — a five-day forecast today is as skilful as a four-day forecast ten years ago or a three-day forecast 20 years ago.

The practical predictability of the atmosphere is influenced by the overall mag-
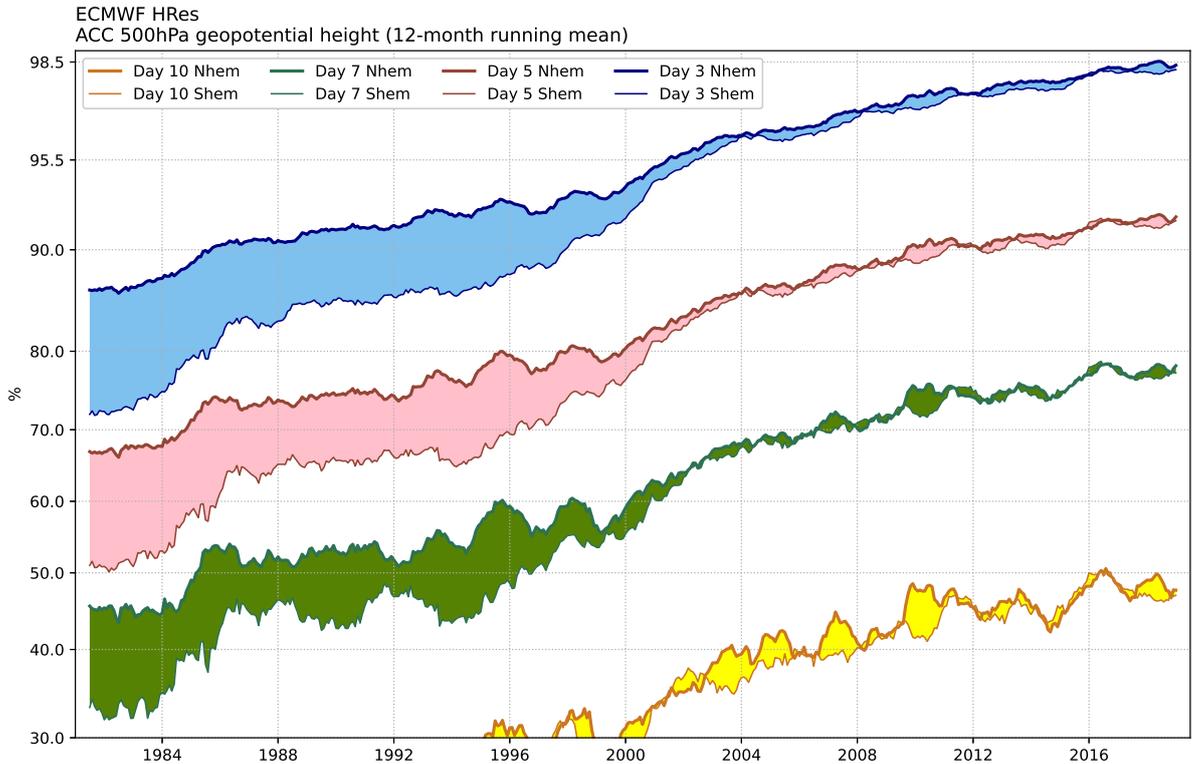
Figure 1.4: 12-month running-mean Anomaly Correlation Coefficients of the ECMWF's three-, five-, seven- and ten-day deterministic forecasts of the 500 hecto-pascal geopotential height in the Northern and Southern Hemispheres respectively, 1982–2019. Courtesy of the ECMWF.

nitude of initial error which can be reduced by extending the network of observations and improving the data assimilation (DA) system. DA is the process of finding the best estimate of the present atmospheric state to initialise a forecast, by optimally combining information from the previous forecast and the latest observations (Daley 1991, Kalnay 2002). A better observational network helps reducing the uncertainty and hence the error in the DA algorithm. This has much improved over the recent decades, thanks to the increasing availability of meteorological satellites (Rabier 2005). Satellite observations are particularly helpful in the Southern Hemisphere, where conventional observations (aircraft, ships, weather stations, etc.) are sparse. Together with an improving DA system, the inter-hemispheric predictability gap has much reduced over the years (Figure 1.4). Today, the vast majority of observations used in the DA algorithm are from satellites, without which the forecast quality would have been substantially worse (English et al. 2013).

Despite the gradual increase in forecast performance, there is still some room for further improvement before the predictability horizon is reached. Figure 1.4 suggests

that synoptic-scale features can now be skilfully predicted up to about nine days ahead[9]. Zhang et al. (2019) estimated that another three to five days of potential synoptic-scale predictability can be gained. Work on multiple fronts towards this goal are in constant progress: increase in model resolutions to allow explicit simulation of convection (Bauer et al. 2015), a unified ('coupled') DA system across components of the Earth system (such as the atmosphere, oceans and sea ice) to reduce initial errors (Lea et al. 2015), and more scalable computational infrastructure to support these calculations (Bauer et al. 2020), just to name a few.

## 1.8    Predicting beyond the predictability horizon

The existence of the finite predictability horizon seems to imply that there is no hope in getting information from weather forecasts beyond two or three weeks that is more useful than a random guess from the climatological distribution. However, this is not true. Behind the discussion of the finite predictability horizon sits a crucial assumption: the dynamical system being predicted governs the *instantaneous* pattern of fluid flow. Therefore, the two-week (or three-week) predictability horizon only applies to the prediction of instantaneous weather. For example, a dynamical forecast made on $1^{st}$ July for the weather at 14:27 on $31^{st}$ July may only be about as good as a climatological guess. Yet, it does not imply that skilful prediction of the *average* weather of the week from $28^{th}$ July to $3^{rd}$ August is impossible. This is because mathematically, the dynamical system that governs the averaged weather is no longer the one specified by Equations 1.1, 1.4–1.7. Instead, it is governed by the time-average of these equations, a separate dynamical system whose theoretical predictability properties have never been studied before in the same way as the instantaneous system. We do not know whether the predictability horizon for such a time-averaged system is finite. One thing we know, however, is that given the same initial error profile, the time-averaged system has a longer range of predictability than the instantaneous system, because time-averaging filters out high-frequency motions, which are often less predictable as they typically correspond to spatial structures with a short wavelength. Thus it is logical to deduce that even if the predictability horizon remains

---

[9]A 60% Anomaly Correlation Coefficient is commonly adopted by the NWP industry as the standard threshold of skill (Owens & Hewson 2018).

finite, it increases with the timescale with which the predictions are averaged (Wheeler et al. 2017).

This provides the basis of long-range forecasts beyond the predictability limit of instantaneous weather. A probabilistic approach enables the extraction of predictable signals in the extended range (Buizza & Leutbecher 2015). Conceptually, long-range predictions can be thought of as more of a problem of boundary forcings and less of an initial-value problem (Robertson et al. 2020). The source of predictability at these long timescales can be attributed to characteristic patterns of variability in the atmosphere (or more generally, the Earth system) at the respective timescale, such as the Madden-Julian Oscillation (Zhang 2005) and stratospheric signals for sub-seasonal forecasts, and the El Niño-Southern Oscillation (Wang & Fiedler 2006) for annual predictions. The role of the oceans becomes important at even longer timescales, as are anthropogenic forcings such as carbon dioxide emissions. The literature in the dynamics and modelling of these patterns is rich and research has been active, yet they are beyond the scope of this thesis.

## 1.9 Thesis outline

Despite the many developments in the prediction of weather over the past century, many fundamental problems remain unanswered. This thesis attempts to address a few of them:

- The finite predictability horizon suggests some form of discontinuity in the dynamical system governing the error, since $\tilde{T}(\varepsilon_0 = 0) = \infty$ but $\liminf_{\|\varepsilon_0\| \to 0} \tilde{T}(\varepsilon_0) < \infty$. Lorenz (1969) arrived at this conclusion using the incompressible 2D Euler equations, yet the closely related system of 2D Navier-Stokes equations is known to be well-posed (Robinson 2001) which implies indefinite predictability. This suggests a potential contradiction with Lorenz's theory that is now accepted as a 'canon of dynamical meteorology' (Rotunno & Snyder 2008). While the equations governing Earth's atmospheric flow together with the boundary conditions are too complex for a rigorous mathematical analysis of the dynamical system's regularity, could understanding the paradox in the context of 2D flows shed some light on the admissibility of a truly finite predictability barrier in Earth's atmospheric system, and therefore

the admissibility of its mathematical irregularity?

- Although the atmospheric energy spectrum is known to consist of a $k^{-3}$ range in the synoptic scale and a $k^{-\frac{5}{3}}$ range in the mesoscale (Nastrom & Gage 1985, Judt 2018), and the characteristic error dynamics in each of these ranges stand-alone has been extensively studied (Lorenz 1969, Lilly 1990, Rotunno & Snyder 2008), the interplay between these two ranges under the hybrid spectrum is a largely unexplored territory. Would error dynamics resemble mixed characteristics of the two ranges? Would this change as atmospheric models begin to resolve the mesoscale $k^{-\frac{5}{3}}$ range, and what would happen in the limit of infinite model resolution?

- The philosophy of forecast verification is rather different between deterministic and probabilistic predictions. Scores for deterministic forecasts quantify *differences* between the forecast and the verification (known as the 'analysis' in the NWP context), whereas metrics for probabilistic forecasts assess the *reliability* of ensembles and the *sharpness* of their distributions. While the loss of information and reliability in probabilistic prediction should be somehow related to the growth of the deterministic error of individual ensemble members, is this relationship quantifiable?

These questions will be discussed in Chapters 2, 3 and 4 respectively. Underlying these discussions will be Lorenz's model of 1969 and a simple idealised 2D barotropic turbulence model, which allow one to gain insights into the essential properties of multi-scale fluids in the context of these questions. With the aid of these simple models, it will be shown, respectively, that the finite predictability horizon may not be incompatible with the possible regularity of the atmospheric system, that the synoptic-scale $k^{-3}$ range in the hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum acts to slow down mesoscale error growth, and that certain verification metrics such as the CRPS and the root-mean-square error (RMSE) can be functionally related under certain conditions. Towards the end of the thesis (Chapter 5), some conclusions will be drawn and a few possible directions outlined for future research.

## 1.10  Publications

The work presented in this thesis has resulted in a couple of journal articles. Chapter 2 is based on a paper published in the *Journal of the Atmospheric Sciences* (Leung et al. 2019). Chapter 3 is based on a paper published in the same journal (Leung et al. 2020). Additionally, small parts of the present chapter are drawn from these two papers: the rigorous mathematical motivation of the predictability horizon in Section 1.4 is based on an appendix of Leung et al. (2019), whereas the discussion of parametric error growth models in Section 1.6 follows on from the corresponding material in Leung et al. (2020).

The candidate for this thesis was responsible in preparing the first draft of both papers in full, and was the first-drafter of all revisions in response to the reviewers' comments. The supervisors as co-authors were only involved in commenting and improving these drafts. Overall, the candidate is estimated to have contributed towards 90% of the work of each of these papers.

Apart from the two said papers, a third paper is being written up based on the material in Chapter 4 and is expected to be submitted soon to the *Quarterly Journal of the Royal Meteorological Society.*

# 2 Revisiting the inherent finite-time barrier

Lorenz's argument which led to his famous conclusion in 1969 that atmospheric predictability is inherently limited will be revisited in detail in this chapter. A counterargument based on analytic results of the 2D Navier-Stokes equations will be presented, which suggests that the predictability horizon cannot be finite. It turns out that both are right: the paradox will be reconciled through understanding the practical role of the spectral slope in a certain inequality among the analytic results, in the context of increasing model resolutions and decreasing spatial scales of the initial error.

## 2.1 Lorenz's model

The starting point of Lorenz (1969)'s derivation is the linearised error equation

$$\frac{\partial}{\partial t}(\Delta\varepsilon) + J(\psi, \Delta\varepsilon) + J(\varepsilon, \Delta\psi) = 0 \tag{2.1}$$

of Equation 1.9 around a basic-state $\psi(t, \mathbf{x})$, where $t$ is the time and $\mathbf{x} = (x, y) \in \mathbb{R}^2$ is the generic 2D position vector. $\varepsilon(t, \mathbf{x}) := \Psi(t, \mathbf{x}) - \psi(t, \mathbf{x})$ denotes a small departure from $\psi(t, \mathbf{x})$, where $\Psi(t, \mathbf{x})$ refers to another solution to Equation 1.9. In a nutshell, Lorenz reduced Equation 2.1 to a system of ODEs through a series of assumptions and by solving it argued inherently finite predictability for spectral slopes shallower than $-3$.

### 2.1.1 An evolution equation for the error energy spectrum

The derivation of Lorenz (1969) will be reproduced in full, with a few modifications for mathematical consistency which will be discussed below. To begin, consider an ensemble $\mathcal{M}_0$ of basic-state streamfunction fields $\psi(t, \mathbf{x})$. The streamfunction can be characterised by its spatial Fourier transform

$$S(t, \mathbf{k}) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \psi(t, \mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{2.2}$$

whereupon $\psi(t, \mathbf{x})$ can be recovered by

$$\psi(t, \mathbf{x}) = \int_{\mathbb{R}^2} S(t, \mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \, \mathrm{d}\mathbf{k}. \tag{2.3}$$

Here $\mathbf{k} = (k_x, k_y) \in \mathbb{R}^2$ is the 2D wavevector. It follows from multiplying Equation 1.9 by $\psi(t, \mathbf{x})$ and integrating spatially that the total energy[1]

$$E_\psi(t) := \frac{1}{2} \int_{\mathbb{R}^2} |\nabla \psi(t, \mathbf{x})|^2 \, \mathrm{d}\mathbf{x} \tag{2.4}$$

is a time-invariant quantity, so that $E_\psi(t) = E_\psi$. In the ensemble mean, the total energy $E$ can be written as

$$
\begin{aligned}
E &:= \mathbb{E}_{\mathcal{M}_0}[E_\psi] \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \mathbb{E}_{\mathcal{M}_0}\left[|\nabla \psi(t, \mathbf{x})|^2\right] \, \mathrm{d}\mathbf{x} \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \mathrm{Var}_{\mathcal{M}_0}\left[\nabla \psi(t, \mathbf{x})\right] \, \mathrm{d}\mathbf{x} + \frac{1}{2} \int_{\mathbb{R}^2} |\mathbb{E}_{\mathcal{M}_0}\left[\nabla \psi(t, \mathbf{x})\right]|^2 \, \mathrm{d}\mathbf{x}.
\end{aligned}
\tag{2.5}
$$

By assuming that

1. the streamfunction averages to zero over the ensemble $\mathcal{M}_0$, i.e. $\mathbb{E}_{\mathcal{M}_0}[\psi(t, \mathbf{x})] \equiv 0$ and therefore $\mathbb{E}_{\mathcal{M}_0}[S(t, \mathbf{k})] \equiv 0$,

---

[1]Strictly speaking, this should be the total energy per unit mass. Yet, throughout the rest of the thesis, the designation 'per unit mass' is implied whenever an energy quantity is referred.

the second term on the right-hand-side of Equation 2.5 drops out. This reduces $E$ to

$$E = \frac{1}{2} \int_{\mathbb{R}^2} \text{Var}_{\mathcal{M}_0} \left[ \nabla \psi(t, \mathbf{x}) \right] \, \mathrm{d}\mathbf{x}. \tag{2.6}$$

$E$ can also be decomposed in terms of Fourier coefficients. Starting from the definition of $E_\psi$ in Equation 2.4 and using the fact that $i\mathbf{k}S(t, \mathbf{k})$ and $\nabla\psi(t, \mathbf{x})$ is a Fourier-transform pair,

$$
\begin{aligned}
E &= \mathbb{E}_{\mathcal{M}_0} \left[ \frac{1}{2} \int_{\mathbb{R}^2} |\nabla \psi(t, \mathbf{x})|^2 \, \mathrm{d}\mathbf{x} \right] \\
&= \mathbb{E}_{\mathcal{M}_0} \left[ \frac{1}{2} \int_{\mathbb{R}^2} \nabla \psi(t, \mathbf{x}) \cdot \left( \int_{\mathbb{R}^2} i\mathbf{k}S(t, \mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k} \right) \mathrm{d}\mathbf{x} \right] \\
&= \mathbb{E}_{\mathcal{M}_0} \left[ \frac{1}{2} \int_{\mathbb{R}^2} i\mathbf{k}S(t, \mathbf{k}) \cdot \left( \int_{\mathbb{R}^2} \nabla \psi(t, \mathbf{x}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{x} \right) \mathrm{d}\mathbf{k} \right] \\
&= \mathbb{E}_{\mathcal{M}_0} \left[ \frac{1}{2} \int_{\mathbb{R}^2} i\mathbf{k}S(t, \mathbf{k}) \cdot \left( \int_{\mathbb{R}^2} \nabla \psi(t, \mathbf{x}) \exp\left(-i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{x} \right)^* \mathrm{d}\mathbf{k} \right] \\
&= \mathbb{E}_{\mathcal{M}_0} \left[ \frac{1}{2} \int_{\mathbb{R}^2} i\mathbf{k}S(t, \mathbf{k}) \cdot 4\pi^2 \left[ i\mathbf{k}S(t, \mathbf{k}) \right]^* \mathrm{d}\mathbf{k} \right] \\
&= 2\pi^2 \int_{\mathbb{R}^2} |\mathbf{k}|^2 \mathbb{E}_{\mathcal{M}_0} \left[ S(t, \mathbf{k}) S(t, -\mathbf{k}) \right] \mathrm{d}\mathbf{k},
\end{aligned}
\tag{2.7}
$$

where an asterisk denotes complex conjugation.

Now, for each basic-state streamfunction $\psi(t, \mathbf{x})$, consider an ensemble $\mathcal{M}_\psi$ of error streamfunctions $\varepsilon(t, \mathbf{x})$. Instead of defining the mean error energy

$$F_\psi(t) = \mathbb{E}_{\mathcal{M}_\psi} \left[ \frac{1}{2} \int_{\mathbb{R}^2} |\nabla \varepsilon(t, \mathbf{x})|^2 \, \mathrm{d}\mathbf{x} \right] = \frac{1}{2} \int_{\mathbb{R}^2} \mathbb{E}_{\mathcal{M}_\psi} \left[ |\nabla \varepsilon(t, \mathbf{x})|^2 \right] \mathrm{d}\mathbf{x} \tag{2.8}$$

over $\mathcal{M}_\psi$, we shall work with the variance-like quantity

$$G_\psi(t) = \frac{1}{2} \int_{\mathbb{R}^2} \mathbb{E}_{\mathcal{M}_\psi} \left[ |\nabla \varepsilon'(t, \mathbf{x})|^2 \right] \mathrm{d}\mathbf{x} = \frac{1}{2} \int_{\mathbb{R}^2} \text{Var}_{\mathcal{M}_\psi} \left[ \nabla \varepsilon(t, \mathbf{x}) \right] \mathrm{d}\mathbf{x}, \tag{2.9}$$

where $\varepsilon'(t, \mathbf{x}) := \varepsilon(t, \mathbf{x}) - \mathbb{E}_{\mathcal{M}_\psi} \left[ \varepsilon(t, \mathbf{x}) \right]$ is the departure of $\varepsilon(t, \mathbf{x})$ from its mean over $\mathcal{M}_\psi$. With $\varepsilon'(t, \mathbf{x})$ as the primary variable for the error, we can define its spatial Fourier transform

$$e(t, \mathbf{k}) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \varepsilon'(t, \mathbf{x}) \exp\left(-i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{x} \tag{2.10}$$

and inverse transform

$$\varepsilon'(t, \mathbf{x}) = \int_{\mathbb{R}^2} e(t, \mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \mathrm{d}\mathbf{k}. \tag{2.11}$$

Proceeding in a similar manner to Equation 2.7, we may write

$$G_\psi(t) = 2\pi^2 \int_{\mathbb{R}^2} |\mathbf{k}|^2 \mathbb{E}_{\mathcal{M}_\psi}\left[e(t, \mathbf{k})e(t, -\mathbf{k})\right] \mathrm{d}\mathbf{k}. \tag{2.12}$$

Equation 2.12 can be averaged over $\mathcal{M}_0$ to obtain an expression for the statistical-mean error energy over the grand ensemble:

$$G(t) = 2\pi^2 \int_{\mathbb{R}^2} |\mathbf{k}|^2 \overline{e(t, \mathbf{k})e(t, -\mathbf{k})} \, \mathrm{d}\mathbf{k}, \tag{2.13}$$

where the overbar denotes an average over $\mathcal{M}_0 \times \mathcal{M}_\psi$. Since Equation 2.7 is constant over $\mathcal{M}_\psi$, it shall follow that

$$E = 2\pi^2 \int_{\mathbb{R}^2} |\mathbf{k}|^2 \overline{S(t, \mathbf{k})S(t, -\mathbf{k})} \, \mathrm{d}\mathbf{k}. \tag{2.14}$$

In the remainder of this section, $E$ and $G(t)$ shall simply be referred to as the *basic-state energy* and *error energy* respectively.

A comment should be made about the choice of $\varepsilon'(t, \mathbf{x})$ in place of $\varepsilon(t, \mathbf{x})$ as the primary prognostic variable for the error. Note that the $G_\psi(t)$ in Equation 2.9 is equivalent to

$$\frac{1}{2} \int_{\mathbb{R}^2} \mathrm{Var}_{\mathcal{M}_\psi}\left[\nabla\Psi(t, \mathbf{x})\right] \mathrm{d}\mathbf{x} \tag{2.15}$$

as far as the ensemble $\mathcal{M}_\psi$ is concerned. In the limit $t \to \infty$ when the statistical properties of $\Psi(t, \mathbf{x})$ in $\mathcal{M}_\psi$ become indistinguishable from those of $\psi(t, \mathbf{x})$ in $\mathcal{M}_0$, we have

$$G(t) = \mathbb{E}_{\mathcal{M}_0}[G_\psi(t)] = \frac{1}{2} \int_{\mathbb{R}^2} \mathrm{Var}_{\mathcal{M}_0}\left[\nabla\psi(t, \mathbf{x})\right] \mathrm{d}\mathbf{x} = E \tag{2.16}$$

in view of Equation 2.6. Had $F_\psi(t)$ (Equation 2.8) been used to define the error energy

instead, the asymptotic $(t \to \infty)$ statistical-mean error energy would have been

$$
\begin{aligned}
\mathbb{E}_{\mathcal{M}_0} & [F_\psi(t)] \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \overline{|\nabla(\Psi(t,\mathbf{x}) - \psi(t,\mathbf{x}))|^2} \, \mathrm{d}\mathbf{x} \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \overline{|\nabla(\Psi(t,\mathbf{x}) - \mathbb{E}_{\mathcal{M}_\psi}[\Psi(t,\mathbf{x})] + \mathbb{E}_{\mathcal{M}_0}[\psi(t,\mathbf{x})] - \psi(t,\mathbf{x}))|^2} \, \mathrm{d}\mathbf{x} \\
&= \frac{1}{2} \int_{\mathbb{R}^2} \left( \overline{|\nabla(\Psi(t,\mathbf{x}) - \mathbb{E}_{\mathcal{M}_\psi}[\Psi(t,\mathbf{x})])|^2} + \overline{|\nabla(\psi(t,\mathbf{x}) - \mathbb{E}_{\mathcal{M}_0}[\psi(t,\mathbf{x})])|^2} \right) \, \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^2} \mathrm{Var}_{\mathcal{M}_0}[\nabla\psi(t,\mathbf{x})] \, \mathrm{d}\mathbf{x} \\
&= 2E
\end{aligned}
\tag{2.17}
$$

under the same condition of statistical indistinguishability. Therefore, using $\varepsilon'(t,\mathbf{x})$ allows the error energy to saturate at the basic-state energy level rather than twice of it.

Now, substituting Equations 2.3 and 2.11 into Equation 2.1 and rearranging, we have

$$
\begin{aligned}
\frac{\partial}{\partial t} \int_{\mathbb{R}^2} -|\mathbf{k}|^2 e(t,\mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k} = \\
- J\left( \int_{\mathbb{R}^2} S(t,\mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k}, \int_{\mathbb{R}^2} -|\mathbf{k}|^2 e(t,\mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k} \right) \\
- J\left( \int_{\mathbb{R}^2} e(t,\mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k}, \int_{\mathbb{R}^2} -|\mathbf{k}|^2 S(t,\mathbf{k}) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right) \, \mathrm{d}\mathbf{k} \right).
\end{aligned}
\tag{2.18}
$$

By the orthogonality of the Fourier basis, Equation 2.18 implies

$$
\begin{aligned}
-|\mathbf{k}|^2 \frac{\mathrm{d}}{\mathrm{d}t} e(t,\mathbf{k}) = & -\int_{\mathbb{R}^2} i(k_x - l_x) S(t, \mathbf{k} - \mathbf{l}) \left(-|\mathbf{l}|^2 i l_y e(t,\mathbf{l})\right) \, \mathrm{d}\mathbf{l} \\
& + \int_{\mathbb{R}^2} i(k_y - l_y) S(t, \mathbf{k} - \mathbf{l}) \left(-|\mathbf{l}|^2 i l_x e(t,\mathbf{l})\right) \, \mathrm{d}\mathbf{l} \\
& - \int_{\mathbb{R}^2} i l_x e(t,\mathbf{l}) \left(-|\mathbf{k} - \mathbf{l}|^2 i(k_y - l_y) S(t, \mathbf{k} - \mathbf{l})\right) \, \mathrm{d}\mathbf{l} \\
& + \int_{\mathbb{R}^2} i l_y e(t,\mathbf{l}) \left(-|\mathbf{k} - \mathbf{l}|^2 i(k_x - l_x) S(t, \mathbf{k} - \mathbf{l})\right) \, \mathrm{d}\mathbf{l}
\end{aligned}
\tag{2.19}
$$

for all $\mathbf{k}$, where $\mathbf{l} = (l_x, l_y)$ is the dummy integration variable. Defining

$$
A(\mathbf{k}, \mathbf{l}) := -\frac{|\mathbf{k} - \mathbf{l}|^2 - |\mathbf{l}|^2}{|\mathbf{k}|^2} \left(k_x l_y - k_y l_x\right),
\tag{2.20}
$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}e(t,\mathbf{k}) = \int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})\,\mathrm{d}\mathbf{l}, \tag{2.21}$$

an evolution equation for the error's Fourier coefficients.

We would like to build an evolution equation for the error energy $G(t)$ or its Fourier decomposition. As an intermediate step, we can use Equation 2.21 to write an equation for $\frac{\mathrm{d}}{\mathrm{d}t}(e(t,\mathbf{k})e(t,-\mathbf{k}))$. Since it is clear from the definition of $A(\mathbf{k},\mathbf{l})$ that $A(-\mathbf{k},-\mathbf{l}) = A(\mathbf{k},\mathbf{l})$, we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(e(t,\mathbf{k})e(t,-\mathbf{k})\right)\\
&= e(t,-\mathbf{k})\frac{\mathrm{d}}{\mathrm{d}t}e(t,\mathbf{k}) + e(t,\mathbf{k})\frac{\mathrm{d}}{\mathrm{d}t}e(t,-\mathbf{k})\\
&= e(t,-\mathbf{k})\int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})\,\mathrm{d}\mathbf{l} + e(t,\mathbf{k})\int_{\mathbb{R}^2} A(-\mathbf{k},-\mathbf{l})S(t,\mathbf{l}-\mathbf{k})e(t,-\mathbf{l})\,\mathrm{d}(-\mathbf{l})\\
&= e(t,-\mathbf{k})\int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})\,\mathrm{d}\mathbf{l} + e(t,\mathbf{k})\int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})S(t,\mathbf{l}-\mathbf{k})e(t,-\mathbf{l})\,\mathrm{d}\mathbf{l}\\
&= \int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})\left(S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})e(t,-\mathbf{k}) + S(t,\mathbf{l}-\mathbf{k})e(t,-\mathbf{l})e(t,\mathbf{k})\right)\,\mathrm{d}\mathbf{l}.
\end{aligned} \tag{2.22}$$

Hence, in the statistical mean,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{e(t,\mathbf{k})e(t,-\mathbf{k})}\right)\\
&= \int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{l})\left(\overline{S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})e(t,-\mathbf{k})} + \overline{S(t,\mathbf{l}-\mathbf{k})e(t,-\mathbf{l})e(t,\mathbf{k})}\right)\,\mathrm{d}\mathbf{l}.
\end{aligned} \tag{2.23}$$

In Equation 2.23 we have a new, cubic quantity $\overline{S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})e(t,-\mathbf{k})}$, for which we would like to find an expression. Using Equation 2.21 again and proceeding similarly, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{S(t,\mathbf{k}-\mathbf{l})e(t,\mathbf{l})e(t,-\mathbf{k})}\right) &= \overline{e(t,\mathbf{l})e(t,\mathbf{k})\frac{\mathrm{d}}{\mathrm{d}t}S(t,\mathbf{k}-\mathbf{l})}\\
&+ \int_{\mathbb{R}^2} A(\mathbf{l},\mathbf{m})\overline{S(t,\mathbf{k}-\mathbf{l})S(t,\mathbf{l}-\mathbf{m})e(t,\mathbf{m})e(t,-\mathbf{k})}\,\mathrm{d}\mathbf{m}\\
&+ \int_{\mathbb{R}^2} A(\mathbf{k},\mathbf{m})\overline{S(t,\mathbf{k}-\mathbf{l})S(t,\mathbf{m}-\mathbf{k})e(t,\mathbf{l})e(t,-\mathbf{m})}\,\mathrm{d}\mathbf{m},
\end{aligned} \tag{2.24}$$

where $\mathbf{m}$ is yet another dummy variable. Now a quartic quantity enters a cubic equation. It should be clear at this point that we would never be able to close the system should

we continue in this way by writing an equation for the quartic quantity. It would only give rise to higher-order expressions. This is the classic closure problem of turbulence (Orszag 1970, Vallis 2017). To proceed in a meaningful way, it is necessary to represent higher-order moments in terms of lower-order moments at some point. Such a functional relation is known as a closure scheme or closure approximation. Here, Lorenz has chosen to close the system using a form of quasi-normal approximation (Lorenz 1969, Orszag 1970) which expresses quartic statistics in terms of quadratic statistics. More precisely, he made the following assumptions:

2. $\frac{\mathrm{d}}{\mathrm{d}t} S(t, \mathbf{k} - \mathbf{l})$ is a quadratic quantity in $S$ only (it cannot be a function of $e$, since the basic-state flow cannot depend on the error);

3. quadratic quantities of $S$ and $e$ are independent;

4. the ensemble for which the statistics is taken is homogeneous.

A few implications are drawn from these assumptions. First of all, Assumptions 2 and 3 suggest that the term $\overline{e(t, \mathbf{l})e(t, \mathbf{k})\frac{\mathrm{d}}{\mathrm{d}t} S(t, \mathbf{k} - \mathbf{l})}$ in Equation 2.24 is equal to the quantity $\overline{e(t, \mathbf{l})e(t, \mathbf{k})\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{S(t, \mathbf{k} - \mathbf{l})}\right)}$, which vanishes after applying Assumption 1. Using Assumption 3 (the closure assumption), the quartic expression in the second term of Equation 2.24 can be expressed as $\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{m})}\ \overline{e(t, \mathbf{m})e(t, -\mathbf{k})}$. The homogeneity Assumption 4 demands the invariance of statistical quantities to spatial translations. This implies that

$$\overline{\psi(t, \mathbf{x})^2} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \overline{S(t, \mathbf{k}')S(t, \mathbf{l}')} \exp\left(i(\mathbf{k}' + \mathbf{l}') \cdot \mathbf{x}\right) \mathrm{d}\mathbf{k}'\, \mathrm{d}\mathbf{l}' \tag{2.25}$$

cannot depend on $\mathbf{x}$, which forces $\overline{S(t, \mathbf{k}')S(t, \mathbf{l}')} = 0$ whenever $\mathbf{k}' + \mathbf{l}' \neq \mathbf{0}$. Hence $\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{m})} = 0$ for all $\mathbf{k} \neq \mathbf{m}$. Similarly, $\overline{e(t, \mathbf{m})e(t, -\mathbf{k})} = 0$ whenever the same condition is satisfied. This allows us to write

$$\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{m})e(t, \mathbf{m})e(t, -\mathbf{k})} = \delta(\mathbf{k} - \mathbf{m})\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{m})}\ \overline{e(t, \mathbf{m})e(t, -\mathbf{k})}$$
$$\tag{2.26}$$

where $\delta(\cdot)$ is the Dirac delta function at the zero vector. The quartic expression in the

third term of Equation 2.24 can likewise be simplified as

$$\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{m} - \mathbf{k})e(t, \mathbf{l})e(t, -\mathbf{m})} = \delta(\mathbf{l} - \mathbf{m})\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{m} - \mathbf{k})}\ \overline{e(t, \mathbf{l})e(t, -\mathbf{m})}.$$

(2.27)

Now, applying the quasi-normal approximation (Equations 2.26 and 2.27), Equation 2.24 can be re-written as

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}&\left(\overline{S(t, \mathbf{k} - \mathbf{l})e(t, \mathbf{l})e(t, -\mathbf{k})}\right)\\
&= \overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{k})}\left(A(\mathbf{k}, \mathbf{l})\overline{e(t, \mathbf{l})e(t, -\mathbf{l})} + A(\mathbf{l}, \mathbf{k})\overline{e(t, \mathbf{k})e(t, -\mathbf{k})}\right).
\end{aligned}
$$

(2.28)

This can be combined with Equation 2.23 to form a closed system of second-order ODEs

$$
\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}&\left(\overline{e(t, \mathbf{k})e(t, -\mathbf{k})}\right)\\
&= 2\int_{\mathbb{R}^2}\overline{S(t, \mathbf{k} - \mathbf{l})S(t, \mathbf{l} - \mathbf{k})}\left(A(\mathbf{k}, \mathbf{l})^2\overline{e(t, \mathbf{l})e(t, -\mathbf{l})} + A(\mathbf{k}, \mathbf{l})A(\mathbf{l}, \mathbf{k})\overline{e(t, \mathbf{k})e(t, -\mathbf{k})}\right)\mathrm{d}\mathbf{l}.
\end{aligned}
$$

(2.29)

We are now in a position to write an equation for the evolution of the 2D error energy spectrum $Z_{2D}(t, \mathbf{k})$, whose integral over $\mathbb{R}^2$ in spectral space gives the $G(t)$ in Equation 2.13. As such,

$$Z_{2D}(t, \mathbf{k}) = 2\pi^2|\mathbf{k}|^2\overline{e(t, \mathbf{k})e(t, -\mathbf{k})}.$$

(2.30)

Similarly, with reference to Equation 2.14, the basic state's 2D energy spectrum $X_{2D}(t, \mathbf{k})$ is

$$X_{2D}(t, \mathbf{k}) = 2\pi^2|\mathbf{k}|^2\overline{S(t, \mathbf{k})S(t, -\mathbf{k})}.$$

(2.31)

The basic-state spectrum is stationary when the turbulence is spun up to a steady state, so that $X_{2D}(t, \mathbf{k}) = X_{2D}(\mathbf{k})$. Substituting Equations 2.30 and 2.31 into Equation 2.29 returns

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}Z_{2D}(t, \mathbf{k}) = \int_{\mathbb{R}^2}\frac{1}{\pi^2}\frac{X_{2D}(\mathbf{k} - \mathbf{l})}{|\mathbf{k} - \mathbf{l}|^2}\left(A(\mathbf{k}, \mathbf{l})^2\frac{|\mathbf{k}|^2}{|\mathbf{l}|^2}Z_{2D}(t, \mathbf{l}) + A(\mathbf{k}, \mathbf{l})A(\mathbf{l}, \mathbf{k})Z_{2D}(t, \mathbf{k})\right)\mathrm{d}\mathbf{l}.$$

(2.32)

This corresponds to Equation 22 of Lorenz's original work in 1969. Our derivation up

to this point is different from Lorenz's primarily in one aspect. While we decomposed the $\psi$ and $\varepsilon'$ fields using Fourier transforms, Lorenz used Fourier series instead, and only at this equation did he take the limit to obtain an equation for continuous wavevectors. As we shall see later, the equation will be discretised for numerical computations after some further manipulations. Therefore, Lorenz's approach of starting with a discrete system, taking the continuous limit then discretising again seems to be mathematically not as neat as starting the derivation with a continuous framework. In particular, the above Assumption 4 could not be correctly applied to yield the Dirac delta functions in Equations 2.26 and 2.27 without considering the continuous model, since the Dirac delta would have to be preceded by a factor of the inverse square of the discretisation interval in spectral space. The extra factor is not obvious, and even Lorenz himself failed to mention it in his work.

Another difference is in the choice of definition of the Fourier basis. The definition that we have adopted in Equations 2.2, 2.3, 2.10 and 2.11 assumes that waves corresponding to wavevectors of unit magnitude have a period of $2\pi$. While the corresponding definition in Lorenz's paper would have suggested the same (his Equations 8 and 9), his subsequent analysis implicitly presumed a unit period for unit-magnitude wavevectors. This can be reflected in, for example, his Equations 10 and 11 which define expressions for the energy quantities. We have decided to adhere to the former definition as it is physically more intuitive: a zonal wave of wavenumber 1 in the atmosphere has only a crest and a trough along a given latitudinal circle and therefore has a wavelength of $2\pi$ radians. For this reason, our Equations 2.7, 2.12–2.14 and 2.30–2.32 differ from Lorenz's corresponding expressions by a factor of $4\pi^2$.

Now, Equation 2.32 is concerned with 2D energy spectra. To simplify the analysis, we make another assumption:

5. the turbulence is isotropic, so that statistical quantities depend on $\mathbf{k}$ only through its magnitude $k := |\mathbf{k}|$ but not its direction.

The isotropy assumption is standard in turbulence analysis. It allows us to define energy spectra in terms of the scalar wavenumber, thus highlighting the dependence of energy on spatial scale regardless of direction.

The scalar energy spectral density $\frac{X(k)}{k}$ and $\frac{Z(t,k)}{k}$, whereupon[2]

$$
\begin{aligned}
E &= \int_{\mathbb{R}} X(k)\, \mathrm{d}(\log k) = \int_0^\infty \frac{X(k)}{k}\, \mathrm{d}k, \\
G(t) &= \int_{\mathbb{R}} Z(t,k)\, \mathrm{d}(\log k) = \int_0^\infty \frac{Z(t,k)}{k}\, \mathrm{d}k,
\end{aligned}
\tag{2.33}
$$

can be obtained by integrating the 2D spectra (Equations 2.30 and 2.31) over circles of constant $k$. Thanks to the isotropy Assumption 5, this is as simple as multiplying $X_{2D}(\mathbf{k})$ and $Z_{2D}(t, \mathbf{k})$ by $2\pi k$. Hence

$$
\begin{aligned}
X(k) &= 2\pi k^2 X_{2D}(\mathbf{k}) = 4\pi^3 k^4 \overline{S(t,k)S(t,-k)}, \\
Z(t,k) &= 2\pi k^2 Z_{2D}(t, \mathbf{k}) = 4\pi^3 k^4 \overline{e(t,k)e(t,-k)}.
\end{aligned}
\tag{2.34}
$$

We would like to write Equation 2.32 in terms of scalar wavenumbers. For a fixed $\mathbf{k}$, each vector $\mathbf{l}$ can almost be characterised by the scalar values $l := |\mathbf{l}|$ and $m := |\mathbf{k} - \mathbf{l}|$, the only issue being that $l$ and $m$ also give rise to another vector which is the reflection of $\mathbf{l}$ along $\mathbf{k}$. But this does not prevent us from transforming the integration element $\mathrm{d}\mathbf{l}$ into $\mathrm{d}l\, \mathrm{d}m$, if we account for the fact that each $(l, m)$ pair corresponds to two possible vectors $\mathbf{l}$.

To find the Jacobian determinant of the transformation $(l_x, l_y) \to (l, m)$, it is easier to work it out via the Jacobian determinant $J$ of the inverse transformation $(l, m) \to (l_x, l_y)$ than to compute it directly. With $l = \sqrt{l_x^2 + l_y^2}$ and $m = \sqrt{(k_x - l_x)^2 + (k_y - l_y)^2}$, we have

$$
J = \left| \det \begin{pmatrix} \frac{\partial l}{\partial l_x} & \frac{\partial l}{\partial l_y} \\ \frac{\partial m}{\partial l_x} & \frac{\partial m}{\partial l_x} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \frac{l_x}{l} & \frac{l_y}{l} \\ \frac{l_x - k_x}{m} & \frac{l_y - k_y}{m} \end{pmatrix} \right| = \frac{|k_x l_y - k_y l_x|}{lm}.
\tag{2.35}
$$

The numerator $|k_x l_y - k_y l_x|$ is the magnitude of the 3D vector product $(k_x, k_y, 0) \times (l_x, l_y, 0)$, which is also the the area of the parallelepiped spanned by these two vectors. Hence,

$$
J = \frac{2\alpha(k, l, m)}{lm},
\tag{2.36}
$$

where $\alpha(k, l, m)$ is the area of the triangle with side lengths $k$, $l$ and $m$. The Jacobian

---

[2] In this section, the notation used by Lorenz in respect of integration limits is retained. For integrals involving the element $\mathrm{d}(f(x))$ where $f(\cdot)$ is a function, the integration limits are expressed in $f(x)$-coordinates instead of $x$-coordinates.

determinant of the forward transformation $(l_x, l_y) \to (l, m)$ is therefore $J^{-1} = \frac{lm}{2\alpha(k,l,m)}$.
Now, taking into account that each pair of $(l, m)$ corresponds to two pairs of $(l_x, l_y)$, we
have

$$d\mathbf{l} = dl_x \, dl_y = \frac{lm}{\alpha(k, l, m)} \, dl \, dm = \frac{l^2 m^2}{\alpha(k, l, m)} \, d(\log l) \, d(\log m). \tag{2.37}$$

The integration limits are determined by the triangle inequality. For fixed $k$
and $l$, it is only possible that $m$ falls in the range $(|k - l|, k + l)$. If the $m$-integration
is performed before the $l$-integration, then the proper integration limits are $\log m \in$
$(\log|k - l|, \log(k + l))$ and $\log l \in \mathbb{R}$. In this case, Equation 2.32 becomes

$$\frac{d^2}{dt^2} Z(t, k) = 2\pi k^2 \int_{\mathbb{R}} \int_{\log|k-l|}^{\log(k+l)} \frac{1}{\pi^2} \frac{X(m)}{2\pi m^4} \left( A(\mathbf{k}, \mathbf{l})^2 \frac{k^2}{l^2} \frac{Z(t, l)}{2\pi l^2} \right.$$
$$\left. + A(\mathbf{k}, \mathbf{l}) A(\mathbf{l}, \mathbf{k}) \frac{Z(t, k)}{2\pi k^2} \right) \frac{l^2 m^2}{\alpha(k, l, m)} \, d(\log m) \, d(\log l), \tag{2.38}$$

where we have also used Equations 2.34. Now, Equation 2.20 together with the reference
above on the area of the parallelepiped implies that

$$|A(\mathbf{k}, \mathbf{l})| = 2 \frac{m^2 - l^2}{k^2} \alpha(k, l, m). \tag{2.39}$$

Finally, substituting Equation 2.39 into Equation 2.38 gives

$$\frac{d^2}{dt^2} Z(t, k) = \int_{\mathbb{R}} \left( C_1(k, l) Z(t, l) - C_2(k, l) Z(t, k) \right) d(\log l), \tag{2.40}$$

where

$$C_j(k, l) = \int_{\mathbb{R}} B_j(k, l, m) X(m) \, d(\log m), \qquad j = 1, 2 \tag{2.41}$$

and

$$B_1(k, l, m) = \begin{cases} \frac{2}{\pi^3} \frac{(m^2 - l^2)^2}{l^2 m^2} \alpha(k, l, m) & m \in (|k - l|, k + l) \\ 0 & \text{otherwise} \end{cases},$$

$$B_2(k, l, m) = \begin{cases} \frac{2}{\pi^3} \frac{(m^2 - l^2)(m^2 - k^2)}{k^2 m^2} \alpha(k, l, m) & m \in (|k - l|, k + l) \\ 0 & \text{otherwise} \end{cases}. \tag{2.42}$$

We have extended the range of integration in the $m$-integral to allow $\log m$ to run along

the whole real line, by defining $B_1(k, l, m)$ and $B_2(k, l, m)$ to take the value zero whenever the lengths $k$, $l$ and $m$ cannot form a triangle.

Equations 2.40–2.42 define the evolution of the error energy spectrum under some fairly general assumptions, most notably homogeneity, isotropy and quasi-normal closure of the turbulence. They require that the basic-state energy spectrum be prescribed. Since the equations are too complex to be analytically solvable, it will be necessary to introduce numerical approximations so as to simplify the system and solve it. These will be discussed in the next sub-section.

## 2.1.2   Discretisation and reduction to a system of ODEs

By performing Fourier transforms on $\psi(t, \mathbf{x})$ and $\varepsilon'(t, \mathbf{x})$ it has been implicitly assumed that the transformed functions $S(t, \cdot)$ and $e(t, \cdot)$ belong to the Schwartz space of rapidly decaying functions (as $|\mathbf{k}| \to \infty$). It follows that $X(\cdot)$ and $Z(t, \cdot)$ are also Schwartz functions. Together with Equations 2.34 which imply $X(k), Z(t, k) \to 0$ as $k \to 0$, we may choose some $N_0 \in \mathbb{R}^+$, $\rho > 1$ and $n \in \mathbb{N}$ such that $X(k)$ and $Z(t, k)$ are negligibly small whenever $k \notin (N_0, N_n]$. Here, $N_0$ and $N_n := \rho^n N_0$ are respectively the minimum and maximum resolved wavenumbers, $\rho$ is the resolution factor, and $n$ is the number of resolved scales.

Under this approximation, we may write $E = \sum_{K=1}^n X_K$ and $G(t) = \sum_{K=1}^n Z_K(t)$ such that

$$X_K = \int_{a_{K-1}}^{a_K} X(k) \, \mathrm{d}(\log k),$$

$$Z_K(t) = \int_{a_{K-1}}^{a_K} Z(t, k) \, \mathrm{d}(\log k), \tag{2.43}$$

where $a_K := \log N_K$ and $N_K := \rho^K N_0$ for $K = 0, 1, \ldots, n$. Note that we have introduced a new index $K$ representing the logarithm of scalar wavenumbers $k$. We may now integrate both sides of Equation 2.40 with respect to $\log k$ over the interval $(a_{K-1}, a_K]$ to obtain

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} Z_K(t) = \int_{a_{K-1}}^{a_K} \int_{\mathbb{R}} (C_1(k, l) Z(t, l) - C_2(k, l) Z(t, k)) \, \mathrm{d}(\log l) \, \mathrm{d}(\log k). \tag{2.44}$$

Applying the negligibility approximation on the right-hand-side gives

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}Z_K(t) = \int_{a_{K-1}}^{a_K} \sum_{L=1}^{n} \left( \int_{a_{L-1}}^{a_L} (C_1(k,l)Z(t,l) - C_2(k,l)Z(t,k)) \, \mathrm{d}(\log l) \right) \mathrm{d}(\log k). \quad (2.45)$$

We now introduce another approximation, that $X(k)$ and $Z(t,k)$ are step functions in $\log k$. Owing to the constraint in Equations 2.43, it is necessary that they take their mean values $(\log \rho)^{-1} X_K$ and $(\log \rho)^{-1} Z_K(t)$ respectively on the interval $\log k \in (a_{K-1}, a_K]$. Using this, Equation 2.45 becomes

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}Z_K(t) = \sum_{L=1}^{n} \left( C_{(1)KL} Z_L(t) - C_{(2)KL} Z_K(t) \right) \quad (2.46)$$

where

$$C_{(j)KL} = (\log \rho)^{-1} \int_{a_{K-1}}^{a_K} \int_{a_{L-1}}^{a_L} C_j(k,l) \, \mathrm{d}(\log l) \, \mathrm{d}(\log k), \qquad j = 1, 2. \quad (2.47)$$

Substituting Equation 2.41 and applying the two approximations again, Equation 2.47 can be re-written as

$$C_{(j)KL} = (\log \rho)^{-2} \int_{a_{K-1}}^{a_K} \int_{a_{L-1}}^{a_L} \sum_{M=1}^{n} \left( \int_{a_{M-1}}^{a_M} B_j(k,l,m) X_M \, \mathrm{d}(\log m) \right) \mathrm{d}(\log l) \, \mathrm{d}(\log k),$$

$$j = 1, 2.$$

$$(2.48)$$

A third approximation specifies that for fixed $k$ and $l$, $B_j(k,l,m)$ takes the constant value $B_j(k,l,N_M)$ over each interval $\log m \in (a_{M-1}, a_M]$. Unlike the previous approximation wherein $X(k)$ and $Z(t,k)$ take their mean values over the appropriate interval, $B_j(k,l,m)$ takes the value at an end-point of the interval. This is presumably to save some computational effort when the function is integrated later. With this approximation, Equation 2.48 simplifies to

$$C_{(j)KL} = (\log \rho)^{-1} \int_{a_{K-1}}^{a_K} \int_{a_{L-1}}^{a_L} \sum_{M=1}^{n} (B_j(k,l,N_M) X_M) \, \mathrm{d}(\log l) \, \mathrm{d}(\log k), \qquad j = 1, 2.$$

$$(2.49)$$

To further simplify $C_{(j)KL}$, we note from Equations 2.42 that $B_j(k,l,m) = m^2 B_j(\frac{k}{m}, \frac{l}{m}, 1)$

for any $m \in \mathbb{R}^+$. It follows that

$$
\begin{aligned}
C_{(j)KL} \\
&= (\log \rho)^{-1} \int_{a_{K-1}}^{a_K} \int_{a_{L-1}}^{a_L} \sum_{M=1}^{n} \left( N_M^2 B_j \left( \frac{k}{N_M}, \frac{l}{N_M}, 1 \right) X_M \right) \mathrm{d}(\log l) \, \mathrm{d}(\log k) \\
&= (\log \rho)^{-1} \sum_{M=1}^{n} \left( \int_{(K-M-1)\log \rho}^{(K-M)\log \rho} \int_{(L-M-1)\log \rho}^{(L-M)\log \rho} N_M^2 B_j(k', l', 1) X_M \right) \mathrm{d}(\log l') \, \mathrm{d}(\log k') \\
&= \sum_{M=1}^{n} B_{(j)K-M,L-M} N_M^2 X_M, \qquad j = 1, 2
\end{aligned}
$$

$$(2.50)$$

where

$$
B_{(j)K,L} = (\log \rho)^{-1} \int_{(K-1)\log \rho}^{K \log \rho} \int_{(L-1)\log \rho}^{L \log \rho} B_j(k', l', 1) \, \mathrm{d}(\log l') \, \mathrm{d}(\log k'), \qquad j = 1, 2. \quad (2.51)
$$

Now, putting Equations 2.46 and 2.50 together, we end up with

$$
\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2} Z_K(t) &= \sum_{L=1}^{n} \sum_{M=1}^{n} \left( B_{(1)K-M,L-M} N_M^2 X_M Z_L(t) - B_{(2)K-M,L-M} N_M^2 X_M Z_K(t) \right) \\
&= \sum_{L=1}^{n} \sum_{M=1}^{n} \left( B_{(1)K-M,L-M} - \delta_{KL} \sum_{L'=1}^{n} B_{(2)K-M,L'-M} \right) N_M^2 X_M Z_L(t) \qquad (2.52) \\
&= \sum_{L=1}^{n} C_{KL} Z_L(t)
\end{aligned}
$$

where

$$
C_{KL} = \sum_{M=1}^{n} \left( B_{(1)K-M,L-M} - \delta_{KL} \sum_{L'=1}^{n} B_{(2)K-M,L'-M} \right) N_M^2 X_M \qquad (2.53)
$$

and where $B_{(j)K,L}, j = 1, 2$ are as specified in Equation 2.51. This is different from Lorenz's original work, in which Equation 2.53 (his Equations 41 and 42 combined) would have been

$$
C_{KL} = \sum_{M=1}^{n} \left( B_{(1)K-M,L-M} - \delta_{KL} \sum_{L'=-\infty}^{\infty} B_{(2)K-M,L'} \right) N_M^2 X_M. \qquad (2.54)
$$

The change was not discussed in Lorenz's paper, nor was it justified in the more recent work of Rotunno & Snyder (2008) which revisited Lorenz's model with great detail. After some investigation, we speculate that Lorenz tweaked the definition of $C_{KL}$ in order that the error may grow as physically expected upon integration of the model (Equation 2.52). Had Lorenz's results been reproduced using the expression in Equation 2.53 instead, the error would have been stuck in an unrealistic oscillatory regime. In order to proceed

meaningfully, we therefore have to accept Lorenz's modification, albeit reluctantly. The physical and technical reasons for such a qualitative discrepancy are beyond the scope of this thesis.

To allow for the computation of $C_{KL}$ — which now includes an infinite sum — on a computer, we further approximated Equation 2.54 by

$$C_{KL} = \sum_{M=1}^{n} \left( B_{(1)K-M,L-M} - \delta_{KL} \sum_{L'=-(n-1)}^{n-1} B_{(2)K-M,L'} \right) N_M^2 X_M. \qquad (2.55)$$

We believe that Lorenz (1969) and Rotunno & Snyder (2008) handled the infinite sum in a similar manner although neither of them discussed this explicitly. We do not expect the results of the model's integration to depend sensitively on the choice of the approximation.

Now, we may write Equation 2.52 as a linear, homogeneous and autonomous system of second-order ODEs

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \mathbf{Z}(t) = \mathbf{C}\mathbf{Z}(t), \text{ where } \mathbf{Z}(t) = \begin{pmatrix} Z_1(t) \\ \vdots \\ Z_n(t) \end{pmatrix} \text{ and } \mathbf{C} = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{pmatrix}. \qquad (2.56)$$

This is equivalent to the first-order system

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \mathbf{Z}(t) \\ \mathbf{W}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Z}(t) \\ \mathbf{W}(t) \end{pmatrix}, \qquad (2.57)$$

where $\mathbf{W}(t)$ is the first derivative of $\mathbf{Z}(t)$, and $\mathbf{I}$ and $\mathbf{0}$ are respectively the $n \times n$ identity and zero matrices[3].

It remains to specify $N_0$, $\rho$, $n$ and $X_M$ (or $X_K$, since $M$ is a dummy variable for the summation). Lorenz (1969) suggested that the basic-state energy spectrum follow a $-\frac{5}{3}$ power-law, i.e. $\frac{X(k)}{k} \sim k^{-\frac{5}{3}}$. It follows that $X(k) \sim k^{-\frac{2}{3}}$, or $X_K \sim \rho^{-\frac{2}{3}K}$. A correction to the power-law is added so that $X(k)$ may indeed be small enough to be negligible when

---

[3]Where there is no ambiguity, the notations $\mathbf{I}$ and $\mathbf{0}$ may later refer to the identity and zero matrices (resp.) of other sizes, and $\mathbf{0}$ may also refer to the zero vertical vector of an appropriate size.

$k \leq N_0$, as demanded by the earlier approximation. For this, Lorenz prescribed

$$X_K = c \left( \rho^{-\frac{2}{3}K} - \rho^{-K} \right) \tag{2.58}$$

where $c$ is chosen to normalise the total basic-state energy to unit value, i.e. $E = \sum_{K=1}^{n} X_K = 1$. Additionally, he set $N_0 = 1$, $\rho = 2$ and $n = 21$. In this thesis, $N_0$ and $\rho$ will be fixed throughout, but $n$ and $X_K$ will remain variable.

### 2.1.3 Solving the ODEs

Solving the system of ODEs (Equation 2.56 or 2.57) first requires that the coefficients of $\mathbf{C}$ be computed using Equations 2.51 and 2.55. It involves integrating $B_j(k', l', 1), j = 1, 2$ which, as Rotunno & Snyder (2008) pointed out, are nearly singular in certain regions of the $(k', l')$ plane. Therefore, the evaluation of the integrals has to be handled with care. Rather than simply integrating over the box $((K-1)\log\rho, K\log\rho] \times ((L-1)\log\rho, L\log\rho]$ as Equation 2.51 suggests, the geometry of the region in which $B_1(k', l', 1)$ and $B_2(k', l', 1)$ are non-zero (Figure 2.1) was taken into account so that resources were not wasted in integrating the zero regions of these functions. In addition, the integrals were evaluated using two different ways on Python (with `scipy.integrate.nquad`) to make sure that the results were accurate:

1. with logarithmic coordinates as stated in Equation 2.51;

2. applying a coordinate change to canonical coordinates: $\mathrm{d}(\log l)\,\mathrm{d}(\log k) \to \frac{1}{kl}\,\mathrm{d}l\,\mathrm{d}k$.

The first method returned a warning message about the bad behaviour of the integrand whereas the second method did not. Moreover, the second method was about 10 times more efficient in terms of the computation's wall-clock time. As such, the second method was preferred over the first.

The coefficients $C_{KL}$ were then computed according to Equation 2.55, with Lorenz's choices of $n$ and the basic-state energy spectrum. Since the entries of $\mathbf{C}$ computed by these two methods were found to differ by no more than 0.0025%, we are confident that the computations are accurate. Table 2.1 displays the first eight rows and columns
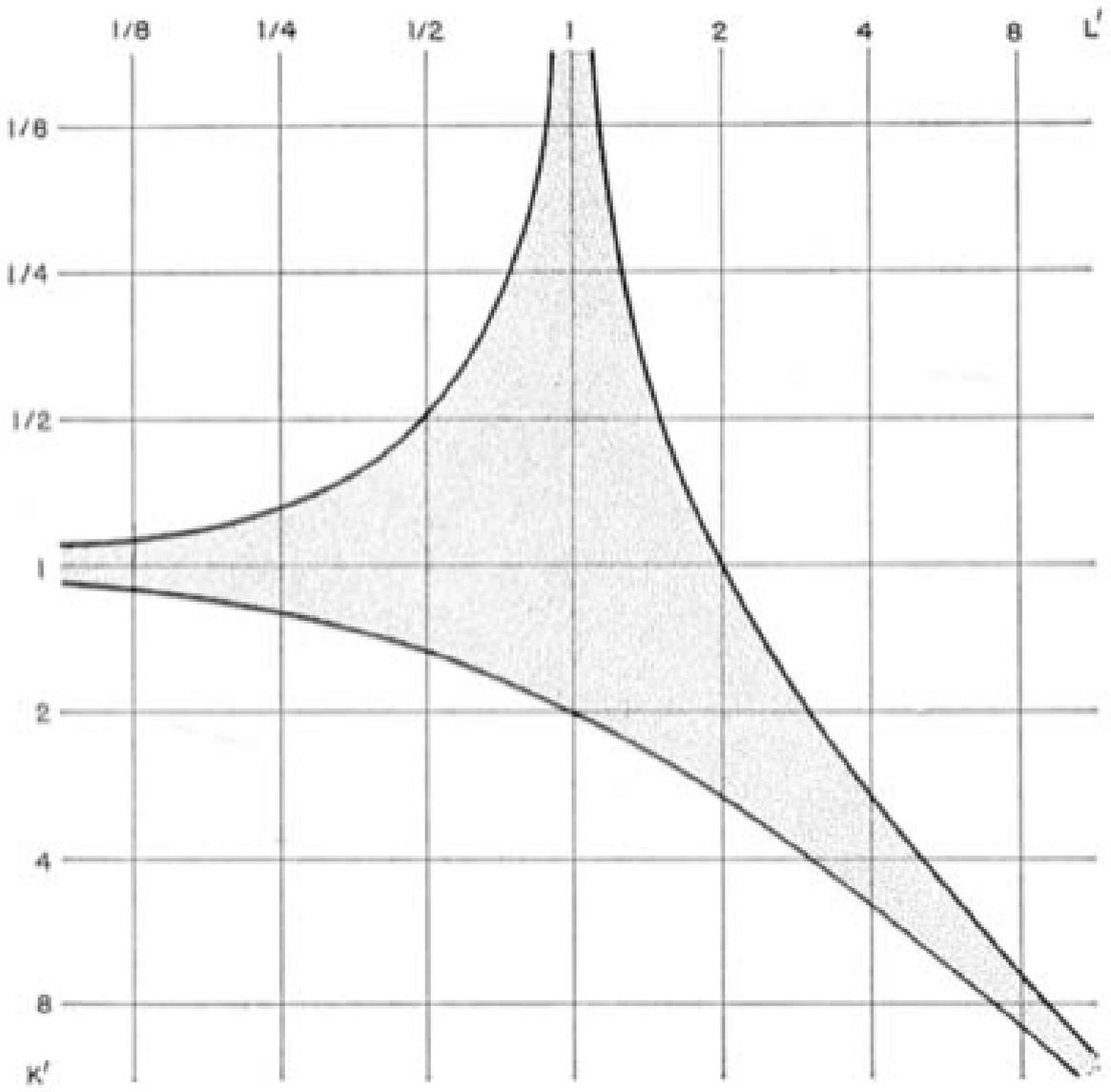
Figure 2.1: The region of the $(k', l')$ plane in which $B_1(k', l', 1)$ and $B_2(k', l', 1)$ are non-zero. The figure is taken from Lorenz (1969). Note that the axes, which Lorenz named as $(K', L')$ instead, are logarithmic.

of the matrix $\mathbf{C}$, which agree with the values computed by Rotunno & Snyder (2008) up to the factor-of-$4\pi^2$ difference that had been carried over from the very beginning of the derivation. As Rotunno & Snyder (2008) remarked, the small discrepancy between their entries and Lorenz's remained unknown, since they were unable to access the numerical code which Lorenz used for his own computations. A similar pattern goes beyond the $8^{\text{th}}$ row and column: the entries on the diagonal are negative, and are of a similar magnitude but opposite sign as the entries to their left (smaller $L$); whereas entries to their right are positive yet much smaller except the ones immediately next to them which are typically the largest of the row. Asymptotically, the large positive entries at small $L$ increase by a factor of about 2.5 every row.

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.005 | 0.006 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.073 | 0.010 | 0.046 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 |
| 3 | 0.339 | 0.259 | −0.031 | 0.221 | 0.017 | 0.003 | 0.001 | 0.000 |
| 4 | 1.14 | 1.05 | 0.838 | −0.320 | 0.865 | 0.049 | 0.009 | 0.002 |
| 5 | 3.37 | 3.30 | 3.04 | 2.57 | −1.56 | 2.98 | 0.134 | 0.025 |
| 6 | 9.43 | 9.38 | 9.19 | 8.45 | 7.55 | −6.00 | 9.50 | 0.359 |
| 7 | 25.6 | 25.5 | 25.4 | 24.9 | 22.9 | 21.6 | −20.4 | 28.6 |
| 8 | 68.0 | 68.0 | 68.0 | 67.6 | 66.2 | 60.8 | 60.1 | −64.0 |

Table 2.1: The first eight rows and columns of the matrix $\mathbf{C}$ for the $k^{-\frac{5}{3}}$ spectrum according to Equation 2.58. The entries are correct to 3 significant figures or 3 decimal places as appropriate.

The matrix $\mathbf{C}$ carries all the information about scale-interactions, which is in itself an important and much studied topic of theoretical turbulence (see, for example, Burgess & Shepherd (2013)). According to Equation 2.56, the rate of change of the error growth rate (the 'acceleration') at a particular scale $K$ is the dot product of row $K$ of $\mathbf{C}$ with the vector $\mathbf{Z}$ of error energies. Hence, the coefficient $C_{KL}$ represents scale $L$'s instantaneous effect on accelerating error growth at scale $K$. As such, the primarily negative entries along the diagonal indicate the slowing down of error growth at a scale itself as a result of energy transfer to other scales, whereas the large positive entries in the strictly lower-triangular part of the matrix represent the fast downscale spread of errors (Rotunno & Snyder 2008). They contrast the strictly upper-triangular part, where the only entries substantially different from zero are found in the super-diagonal. The latter feature suggests that upscale error growth is dominantly controlled by local interactions. That said, the role of the very weak non-local interactions in upscale error growth (the small numbers at the top-right of the matrix) cannot be ignored. This is because once a small-scale error triggers a tiny error at the large scales, the latter error in turn spreads to the small scales quickly (as a result of the large entries in the matrix's bottom-left), thus creating a positive feedback effect. In fact, this feedback mechanism is crucial to the error growth of the system as a whole, with the size of the bottom-left entries often serving as a proxy for the scale-dependency of the growth rate.

To integrate Equation 2.57, it is necessary that the initial conditions for $\mathbf{Z}$ and $\mathbf{W}$ be specified. For Lorenz's first experiment, the initial error was concentrated at the second smallest scale. He set $Z_{n-1}(0) = 2^{-16}E$ and $Z_K(0) = 0$ for $K = 1, \ldots n - 2$ and $K = n$. An initial condition for $\mathbf{W}$ was not explicitly provided, but it would be natural to

assume that the error spectrum initially has zero tendency, i.e. $\mathbf{W}(0)$ is the zero vector. With such initial conditions, Lorenz integrated Equation 2.57 using a simple second-order Runge-Kutta scheme. However, we note that the equation admits an analytic solution, since it is a linear system with constant coefficients. We have thus decided to solve it explicitly by writing out the general solution in terms of the eigenvalues and eigenvectors of

$$\begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$$

and projecting the vector $(\mathbf{Z}(0), \mathbf{W}(0))^{\mathrm{T}}$ of initial conditions onto such an eigenspace to determine the constants of the general solution. Such an exact approach is a good and easy alternative to the numerical schemes used by Lorenz (1969) and Rotunno & Snyder (2008) and its extension by Durran & Gingrich (2014).

Although Lorenz's error growth model results from a linearised equation (Equation 2.1) and is therefore linear, the linear approximation breaks down as the error becomes large. Indeed, the error energy saturates at the basic-state energy level (Equation 2.16), which would not have happened had the turbulence been linear. As such, a treatment of non-linear effects has to be included in the error growth model. In Lorenz (1969), the non-linear effects were incorporated into the solution procedure by removing the corresponding components of $\mathbf{Z}(t)$, $\mathbf{W}(t)$ and $\mathbf{C}$ when the relative error energy $\frac{Z_K(t)}{X_K}$ reached a certain threshold ($\approx 0.815$), but here we require $Z_K(t) = X_K$ for such components to be removed. Although the effects of the saturated scale on error growth at the unsaturated scales were no longer represented in the prognostic equation thereafter, they were nevertheless retained through an inhomogeneous forcing term added to the right-hand-side of Equation 2.57. This treatment was non-linear because its implementation was dependent on the solution itself. Time-integration with the resulting inhomogeneous, lower-dimensional system was carried on, until all scales became saturated.

We recognise that Durran & Gingrich (2014) has suggested another treatment of the turbulence's non-linearity by modifying the governing equation (Equation 2.57) itself. Their modification allows error growth to gradually slow down towards saturation and is no doubt a better description of it than the scheme above. However, we wish to stick to Lorenz's original proposal in this thesis as it would enable us to easier work with the

model's analytic properties.

Considering all the above, we have solved Lorenz's model using the following procedure:

1. Set a time-step $h$; in this case, $h = 10^{-3}$.

2. Initialise the run by setting time $t = 0$. Also initialise $t_0 = 0$. ($t_0$ is the time when the previous saturation occurs.)

3. Project the initial condition $(\mathbf{Z}(0), \mathbf{W}(0))^{\mathrm{T}}$ onto the eigenspace of the block matrix

$$
\begin{pmatrix}
\mathbf{0} & \mathbf{I} \\
\mathbf{C} & \mathbf{0}
\end{pmatrix}
$$

to determine the constants of the general solution.

4. Compute the solution at time $t - t_0$ and check if any of the scales $K$ saturates by time $t$.

5. If none of the scales saturates, reset $t = t + h$ and repeat step 4.

6. If, for any $K$, $Z_K(t) > X_K$ by time $t$, then the clock is reset to the previous time-step $t - h$, and $h$ is refined to $10^{-6}$.

7. Repeat steps 4 and 5 with the new value of $h$ until $Z_K(t) > X_K$. The saturation time of scale $K$ is determined as if $Z_K(t)$ increased linearly between times $t - h$ and $t$.

8. Reset $h = 0.001$ and set $t_0$ to be the current time $t$. Remove the row and column of the matrix $\mathbf{C}$ corresponding to the saturated scale $K$ and the corresponding entries of $\mathbf{Z}(t_0)$, $\mathbf{W}(t_0)$ and all the $\mathbf{F}_j$ terms (more on the $\mathbf{F}_j$ terms below). The reduced-size system

$$
\frac{\mathrm{d}^2}{\mathrm{d}t^2} \mathbf{Z}(t) = \mathbf{C}\mathbf{Z}(t) + \sum_j \mathbf{F}_j, \tag{2.59}
$$

where the summation is performed over all saturated scales, has a new inhomogeneous term $\mathbf{F}_K$ which accounts for the contribution of scale $K$'s saturated energy

to error growth at the remaining scales. ($\mathbf{F}_K$ equals to $X_K$ multiplied by the $K^{\text{th}}$ column of $\mathbf{C}$ restricted to the rows corresponding to the remaining scales.) As the new system is equivalent to the reduced-size first-order system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\mathbf{Z}(t)\\\mathbf{W}(t)\end{pmatrix}=\begin{pmatrix}\mathbf{0}&\mathbf{I}\\\mathbf{C}&\mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{Z}(t)\\\mathbf{W}(t)\end{pmatrix}+\begin{pmatrix}\mathbf{0}\\\sum_j\mathbf{F}_j\end{pmatrix}\qquad(2.60)$$

and $\sum_j\mathbf{F}_j$ is a constant vector, its solution $(\mathbf{Z}(t),\mathbf{W}(t))^{\mathrm{T}}$ can be expressed as the sum of a particular solution $(\mathbf{Z}_p,\mathbf{W}_p)^{\mathrm{T}}$ and a solution $(\mathbf{Z}_h(t-t_0),\mathbf{Z}_h(t-t_0))^{\mathrm{T}}$ of the homogeneous system in the variable $t-t_0$. A particular solution to the differential equation can be obtained by solving

$$-\begin{pmatrix}\mathbf{0}&\mathbf{I}\\\mathbf{C}&\mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_p\\\mathbf{W}_p\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\\sum_j\mathbf{F}_j\end{pmatrix}.\qquad(2.61)$$

Re-calculate the eigenvalues and eigenvectors of the new

$$\begin{pmatrix}\mathbf{0}&\mathbf{I}\\\mathbf{C}&\mathbf{0}\end{pmatrix}$$

and project $(\mathbf{Z}(t_0),\mathbf{W}(t_0))^{\mathrm{T}}-(\mathbf{Z}_p,\mathbf{W}_p)^{\mathrm{T}}$ onto such an eigenspace to determine the constants of $(\mathbf{Z}_h(t-t_0),\mathbf{Z}_h(t-t_0))^{\mathrm{T}}$ and thus the full solution.

9. Repeat steps $4-8$ until all scales saturate.

The distribution of eigenvalues of the block matrix

$$\begin{pmatrix}\mathbf{0}&\mathbf{I}\\\mathbf{C}&\mathbf{0}\end{pmatrix}$$

deserves a mention. In the original matrix, i.e. before any rows and columns are removed from the matrix by the above procedure, the $2n=42$ eigenvalues consist of 5 pairs of real numbers and 16 pairs of complex conjugates. Each pair of real eigenvalues are of the same magnitude but opposite-signed. The remaining eigenvalues can be treated as purely imaginary, as their real parts have magnitudes smaller than $10^{-12}$. As such, we

may consider that all the eigenvalues fall on either the real or the imaginary axis. The magnitudes of these eigenvalues span rather evenly (on a logarithmic scale) across a few orders of magnitude, from about $5.5 \times 10^3$ down to about $8.6 \times 10^{-2}$. The largest positive real eigenvalue which determines the overall rate of error growth has a magnitude of about $1.9 \times 10^3$. This gradually decreases as the size of the matrix is reduced throughout the solution procedure, by about an order of magnitude for every five scales removed from the prognostic system. Yet the basic structure of the eigenvalues' distribution remains the same, namely that they exist in opposite-signed pairs of the same magnitude on either axis, and that the ratio between the number of real and imaginary eigenvalues is approximately $1:3$. Having said that, the only remaining pair of eigenvalues just before the saturation of the error energy at the largest scale is real-valued, or else the largest scale would never be able to saturate.

The mathematics of the dynamical system (Equation 2.57) in no way precludes the physically unrealistic occurrence of negative error energies. Indeed, it is a known shortcoming of the quasi-normal approximation of turbulence closure (Orszag 1970). As Lorenz (1969) remarked himself, the solution trajectory corresponding to his initial condition is found to oscillate back and forth a few times before it gains a sufficient projection on the real, positive eigenvalues to grow monotonically. This is especially the case when the initial condition has a substantial projection on eigenvectors whose corresponding eigenvalues have an imaginary component. Given that about three quarters of the eigenvalues are imaginary, it is hardly surprising that such oscillatory behaviour happens. Qualitatively, the oscillations represent a time-delay in error growth, which in the real world may correspond to an initial geostrophic adjustment process of unbalanced mesoscale errors (Bierdel et al. 2017, 2018).

Figure 2.2 shows the results of solving Lorenz's model for the $k^{-\frac{5}{3}}$ basic-state spectrum (Equation 2.58) and the small-scale initial condition specified above. The error energy spectrum spreads upscale in a more or less self-similar fashion and peaks at the smallest scale yet to be saturated (Figure 2.2(a)). The only exception to this is near the initial time, when the error is still concentrated at the very small scales and is transitioning from the oscillatory regime to the monotonic growth regime. The saturation times $t_K$ (Figure 2.2(b)) largely agree with the values which Lorenz (1969) himself reported (see

also Table 1.1), and scale nicely as $\rho^{-\beta K}$ with a $\beta > 0$, thus implying that the error at small scales saturates much faster than at large scales. By fitting the $t_K \propto \rho^{-\beta K}$ scaling to the data, it was found that $\beta = 0.6933$.
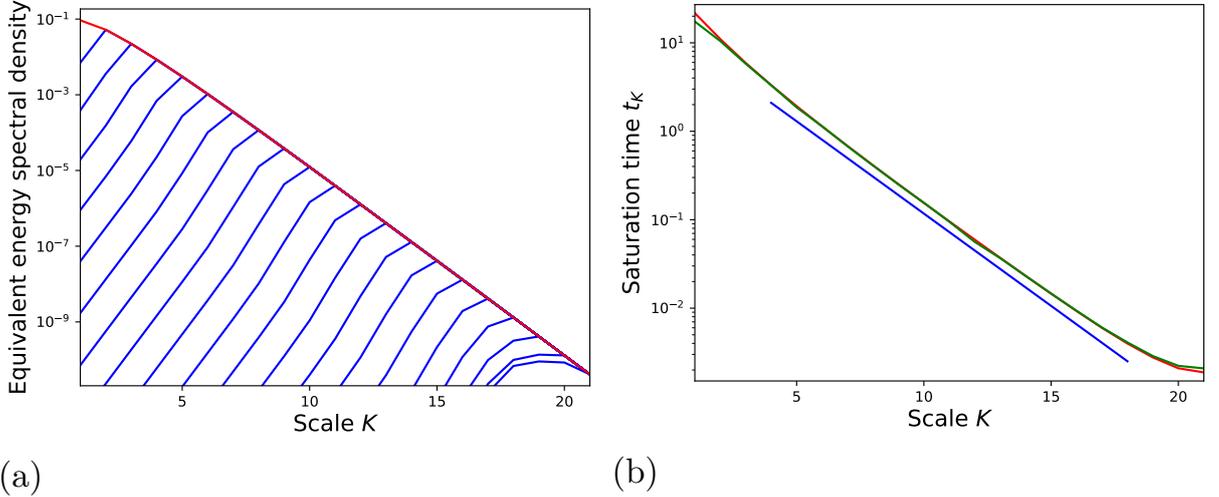


(a)                          (b)

Figure 2.2: (a) Evolution of the error energy spectrum (blue, from bottom-right to top-left) in Lorenz (1969)'s model, under the basic-state energy spectrum of slope $-\frac{5}{3}$ shown in the red curve. The equivalent energy spectral densities $\rho^{-K}Z_K(t)$ and $\rho^{-K}X_K$ are depicted in the vertical axis. These are functions that smoothly distribute $Z_K(t)$ and $X_K$ (resp.), which would have been densities in $k$ had $K$ been a continuous variable. The error energy spectrum is plotted whenever it saturates at some scale, rather than at equal time-intervals. (b) Saturation times $t_K$ of the error energy as a function of scale $K$ (red), compared to the values reported by Lorenz but appropriately non-dimensionalised (green). The blue curve shows a line-of-fit $t_K \propto \rho^{-\beta K}$ for the red curve, where $\beta = 0.6933$.

### 2.1.4   An argument for the finite predictability horizon

The parameter $\beta$ is an important indicator for determining the predictability horizon. In Lorenz (1969)'s original formulation, the fitting was performed on the successive differences $t_K - t_{K+1}$ of the saturation times, rather than the saturation times $t_K$ themselves. He suggested that $t_K - t_{K+1}$ would scale as $\rho^{-\beta K}$ so that, given an initial error at an infinitesimally small scale, the predictability horizon would be finite if and only if the telescoping series

$$t_K = \sum_{j=K}^{\infty}(t_j - t_{j+1}) \propto \sum_{j=K}^{\infty} \rho^{-\beta j} \tag{2.62}$$

is summable. This happens if and only if $\beta > 0$, in which case the predictability horizon $t_K$ may be estimated as

$$t_K \propto \sum_{j=K}^{\infty} \rho^{-\beta j} = \frac{\rho^{-\beta K}}{1 - \rho^{-\beta}} \propto \rho^{-\beta K}, \tag{2.63}$$

hence our choice of fitting $t_K$ to the scaling $\rho^{-\beta K}$.

Lorenz observed that $\beta = \frac{2}{3}$ for a $k^{-\frac{5}{3}}$ spectrum, which is slightly different from the value of $\beta$ that we have obtained by a standard least-squares fitting algorithm. According to the argument above, this would imply a finite predictability horizon for the atmosphere, since the atmospheric energy spectrum was presumed to be $k^{-\frac{5}{3}}$. Additionally, he found that $\beta$ depends on the spectral slope, with $\beta = \frac{1}{3}$ for a hypothetical $k^{-\frac{7}{3}}$ spectrum. By linear extrapolation, Lorenz thus hypothesised that $\beta$ would become zero and predictability would become indefinite if the spectral slope were steepened to $-3$.

Lorenz's result is supported by arguments on dimensional grounds (Vallis 1985, Lilly 1990). Relation 1.12 in Section 1.4 suggests that the predictability horizon scales with the wavenumber as $k^{\frac{p-3}{2}} \approx \rho^{\frac{p-3}{2}K}$, where $-p$ is the spectral slope. Hence $\beta = \frac{3-p}{2}$, implying that $\beta = \frac{2}{3}, \frac{1}{3}$ and $0$ respectively for spectral slopes $-\frac{5}{3}, -\frac{7}{3}$ and $-3$.

## 2.1.5 A $k^{-3}$ basic-state energy spectrum

It would be of interest to study the theoretical error growth and predictability properties under a $-3$ spectral slope, as it has been shown to appear in the atmospheric energy spectrum in the synoptic scale (Boer & Shepherd 1983, Nastrom & Gage 1985), and also because such properties are thought to be very different from those under a $k^{-\frac{5}{3}}$ spectrum (Lorenz 1969). This can be done by modifying the inputs of Lorenz's model. Rotunno & Snyder (2008) have already solved for the growth of the error energy spectrum in this context. To assess the range of predictability in Lorenz's framework, we extend their calculations to investigate the dependence of $t_K$ on $K$. This will enable us to check whether indefinite predictability is indeed achieved with a spectral slope of $-3$, as Lorenz hypothesised.

The input spectrum is updated according to the proposal of Rotunno & Snyder (2008) to reflect a $k^{-3}$ spectrum. This is done by multiplying Equation 2.58 by a factor of $\rho^{-\frac{4}{3}(K-3)}$ when $K > 3$, and retaining the original values otherwise so that the spectrum at the largest scales remains unaffected. This implies

$$X_K = c\left(\rho^{-\frac{2}{3}K} - \rho^{-K}\right) \times \min\{1, \rho^{-\frac{4}{3}(K-3)}\}, \qquad (2.64)$$

where $c$ is as in Equation 2.58 (as a result, the total energy in this case is not normalised to unit value). In addition, the initial condition for $Z_{n-1}$ is updated as $2^{-40}$, since the original value would now exceed the basic-state energy level $X_{n-1}$. Apart from these changes, everything in the set-up remains the same as before.

The re-computed matrix $\mathbf{C}$, whose first 8 rows and columns are displayed in Table 2.2, is distinctively different from that for the $k^{-\frac{5}{3}}$ spectrum (Table 2.1). In any given row, the largest entries in magnitude are those on the diagonal, super-diagonal and sub-diagonal. This indicates that the primary mechanism for error growth is local triad interactions, both upscale and downscale. The relatively small entries in the bottom-left of the matrix suggest that the quick spreading of large-scale errors into the small scales is absent. Since it is the primary mechanism of the fast error growth under a $k^{-\frac{5}{3}}$ spectrum, one may expect that error growth under a $k^{-3}$ spectrum is much slower.

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.005 | 0.006 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.073 | 0.010 | 0.046 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.339 | 0.259 | −0.031 | 0.202 | 0.006 | 0.000 | 0.000 | 0.000 |
| 4 | 0.608 | 0.538 | 0.412 | −0.453 | 0.692 | 0.007 | 0.001 | 0.000 |
| 5 | 0.733 | 0.709 | 0.625 | 0.744 | −1.79 | 1.81 | 0.007 | 0.001 |
| 6 | 0.824 | 0.816 | 0.789 | 0.694 | 1.64 | −4.89 | 4.09 | 0.008 |
| 7 | 0.894 | 0.891 | 0.882 | 0.852 | 0.749 | 3.67 | −11.4 | 8.67 |
| 8 | 0.948 | 0.947 | 0.944 | 0.935 | 0.903 | 0.792 | 7.96 | −24.8 |

Table 2.2: The first eight rows and columns of the matrix $\mathbf{C}$ for the $k^{-3}$ spectrum according to Equation 2.64. The entries are correct to 3 significant figures or 3 decimal places as appropriate.

The eigenvalues of this $\mathbf{C}$ range from approximately $1.1 \times 10^{-1}$ to $5.0 \times 10^2$ and are therefore less spread out than those for the matrix corresponding to the $k^{-\frac{5}{3}}$ spectrum. Like the latter matrix, the real eigenvalues come in pairs of opposite-signed numbers of the same magnitude. However, there are only two such pairs out of the 21,

and the larger pair has a magnitude of only about $9.4 \times 10^{-1}$. This may also explain the slow error growth under the $k^{-3}$ spectrum. On the other hand, the many remaining pairs of purely imaginary eigenvalues suggest that the error trajectory is likely to spend a long time in an oscillatory regime while struggling to project itself sufficiently on the eigenspace corresponding to the real and positive eigenvalues. Indeed, the trajectory for the initial condition specified above has spent almost the same amount of time to saturate the smallest scale $K = 21$ as the time it takes to grow from $K = 21$ to $K = 3$ (Figure 2.3(b)).
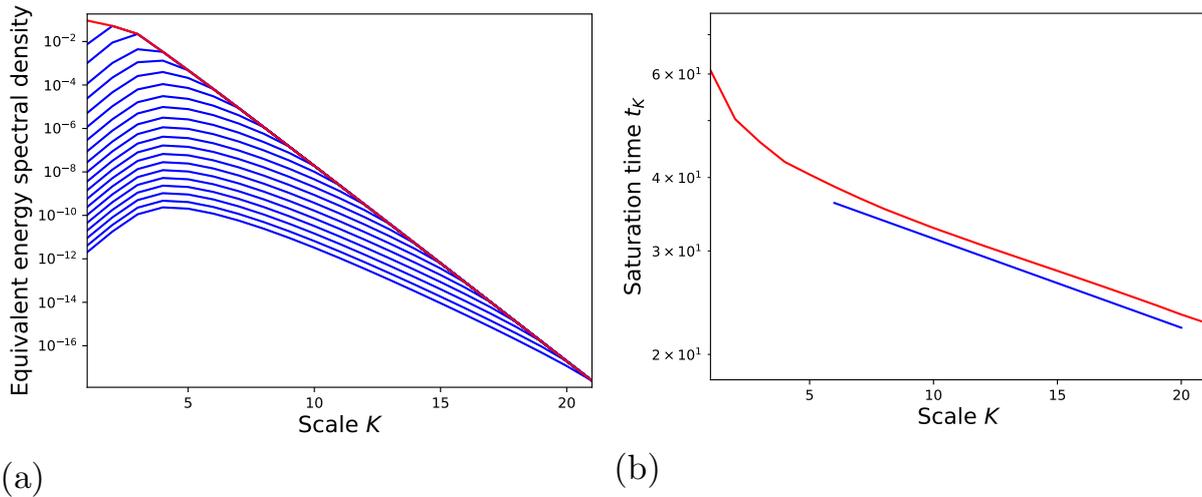


(a)             (b)

Figure 2.3: As in Figure 2.2, but for a $k^{-3}$ spectrum according to Equation 2.64 and a smaller initial error magnitude of $2^{-40}$ at $K = 20$. There is no green curve in (b), as Lorenz did not run the model for a $k^{-3}$ spectrum himself. The line-of-fit shown by the blue curve in (b) corresponds to $\beta = 0.0504$.

Rather than growing upscale, the error energy spectrum seems to grow up-magnitude at a more or less uniform rate across all scales (Figure 2.3(a)). This presents a major qualitative difference from error growth under a $k^{-\frac{5}{3}}$ spectrum. The $\beta$ for which $t_K \propto \rho^{-\beta K}$ is found to be small but positive (about 0.05), so that the predictability horizon represented by the sum in Equation 2.62 remains finite. Although this is contrary to Lorenz's prediction as well as the dimensional theory, we acknowledge that this $\beta$ is only marginally away from the critical value of zero. Having said that, one should bear in mind that Equation 2.63 is invalid when $\beta = 0$. Instead, since $t_K - t_{K+1}$ is constant in $K$, the range of predictability (given a certain initial error) should scale as $\beta_1 - \beta_2 K$, which would have been a curve of constantly changing slope had it been plotted with a logarithmic $t_K$ axis as in Figure 2.3(b). Hence, the straight segment of the red curve in Figure 2.3(b) reaffirms our deduction that Lorenz's model fails to predict the indefinite

predictability associated with the $k^{-3}$ spectrum.

The results in Figure 2.3 can be compared with the corresponding results (Figure 2.4) from an identical-twin perturbation experiment with idealised 2D turbulence, whose set-up will be discussed in Section 3.1. Figure 2.4(a) shows the growth of the error energy spectrum. Like in Lorenz's model (Figure 2.3(a)), the error spectrum grows quite uniformly across the scales. However, its flatness presents a contrast with the sharp large-scale peak predicted by Lorenz's model. The range of predictability $\tilde{T}$ scales nicely as $k^{-\beta}$ with $\beta = 0.2365 > 0$ (Figure 2.4(b)), again reinforcing (this time with greater confidence as $\beta$ is further away from zero) that indefinite predictability may have not been reached with a $k^{-3}$ spectrum. This could be an indication that the existing theory on dimensional grounds might be too simplistic to describe the non-linear dependence between the predictability diagnostic $\beta$ and the slope $-p$ of the basic-state energy spectrum. On top of this, the unrealistic shape of the error energy spectrum that results from Lorenz's model in the $k^{-3}$ case may further complicate matters.



(a)                                          (b)
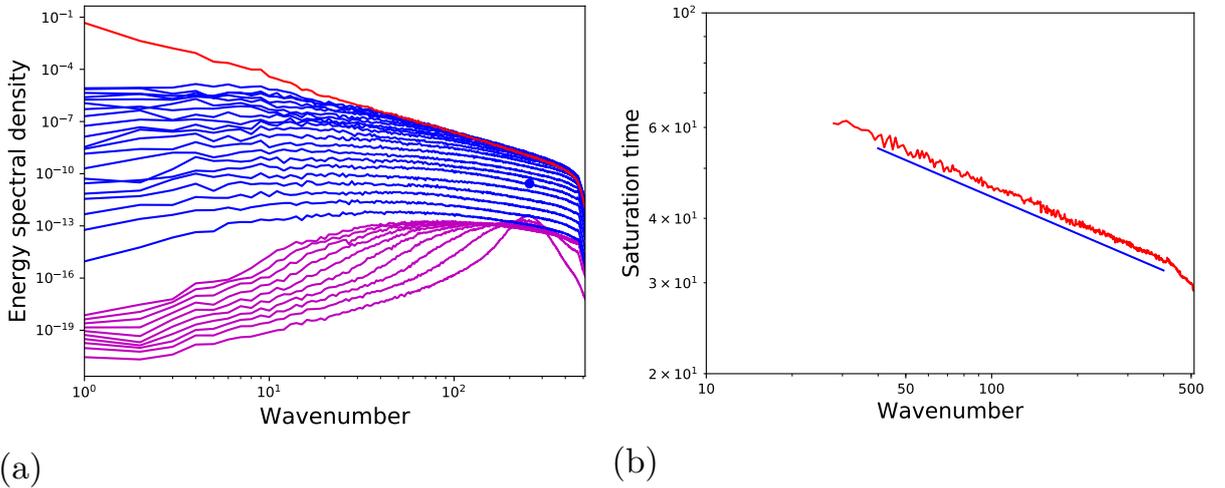
Figure 2.4: As in Figure 2.3, but for directly simulated 2D turbulence averaged over four independent realisations. The blue dot in (a) shows the initial perturbation. The magenta curves show the error spectrum at non-dimensional times $t = 0.3, 0.6, \ldots, 3$, and the blue curves at $t = 6, 9, \ldots, 66$. The basic-state energy spectrum (scaled by a factor of two, see Equation 2.17) at the final time of the simulation ($t = 66$) is shown in red. In (b), the blue line-of-fit corresponds to $\beta = 0.2365$.

## 2.2 Aspects from PDE theory: the incompressible 2D Navier-Stokes equations

A very different approach to understanding predictability is the analytic theory of PDEs. The Navier-Stokes equations (Equations 1.1 and 1.3), which are essentially fluid-dynamical representations of the mass-conservation principle and Newton's Second Law of Motion, are always useful as a pedagogical first step towards understanding and modelling motions of real fluid flows in the atmosphere. As such, their analytical properties have been extensively studied. Building on these analytic results, we now consider their implications for predictability.

Here, we restrict our attention to the 2D incompressible case but with a general forcing term $\mathbf{f}(t, \mathbf{x}) \in \mathbb{R}^2$. By an appropriate rescaling of the variables, the constant density can assume a unit value without loss of generality. In this case, the Navier-Stokes problem can be mathematically reduced to

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \mathbf{f} + \nu \nabla^2 \mathbf{u}$$
$$\nabla \cdot \mathbf{u} = 0$$

(2.65)

where $\nu > 0$ is a constant, and the fields $\mathbf{u}(t, \mathbf{x}) \in \mathbb{R}^2$ and $p(t, \mathbf{x}) \in \mathbb{R}$ are to be solved for $t > 0$ in a square domain $\Omega := [0, 2\pi) \times [0, 2\pi)$ subject to an initial condition $\mathbf{u}_0(\mathbf{x}) := \mathbf{u}(0, \mathbf{x})$ and doubly periodic boundary conditions. We shall further assume that both $\mathbf{u}_0(\mathbf{x})$ and $\mathbf{f}(t, \mathbf{x})$ integrate spatially to zero. In the remainder of this chapter, we shall drop the reference 'incompressible' and simply refer to these equations as the '2D Navier-Stokes equations'.

Unlike its 3D counterpart whose regularity problem remains open, the above initial-boundary-value problem for the 2D Navier-Stokes equations is known to be well-posed: a solution exists, is unique, and depends continuously on initial conditions. This style of rigorous analysis can be traced back to the early work of Leray (1933, 1934). An overview of such analytic results is given in Robinson (2013), which are discussed in detail in a book by the same author (Robinson 2001). The proof of the solution's uniqueness

and continuous dependence on initial conditions will be reproduced below, as it is relevant to the study of predictability.

### 2.2.1 Weak formulation

We shall focus on the weak formulation of the initial-boundary-value problem described by Equations 2.65, which admits a more general class of solutions than the strong formulation. It casts the 2D Navier-Stokes equations in the form of an ODE in an appropriate function space. To begin, we note that the solution $\mathbf{u}(t, \mathbf{x})$ always has a spatial mean of zero. This can be shown by spatially integrating the first of Equations 2.65:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \mathbf{u}\,\mathrm{d}\mathbf{x} + \int_\Omega \mathbf{u} \cdot \nabla \mathbf{u}\,\mathrm{d}\mathbf{x} = - \int_\Omega \nabla p\,\mathrm{d}\mathbf{x} + \int_\Omega \mathbf{f}\,\mathrm{d}\mathbf{x} + \int_\Omega \nu \nabla^2 \mathbf{u}\,\mathrm{d}\mathbf{x}. \qquad (2.66)$$

The first term on the right-hand-side is zero, since each vector component[4] is an integral of a derivative of a periodic function. The same applies to the third term, because the divergence included in the $\nabla^2$ can be interpreted as a sum of derivatives. The second term on the right-hand-side is also zero due to the zero-mean assumption of $\mathbf{f}$. For each component of $\mathbf{u}$, the second term on the left-hand-side equals

$$\int_\Omega \mathbf{u} \cdot \nabla u_i\,\mathrm{d}\mathbf{x} = \int_\Omega \nabla \cdot (u_i \mathbf{u})\,\mathrm{d}\mathbf{x} - \int_\Omega u_i \nabla \cdot \mathbf{u}\,\mathrm{d}\mathbf{x} = \int_{\partial\Omega} u_i \mathbf{u} \cdot \mathrm{d}\mathbf{S} - \int_\Omega u_i \nabla \cdot \mathbf{u}\,\mathrm{d}\mathbf{x} = 0 \quad (2.67)$$

($\partial\Omega$ denotes the boundary of $\Omega$), thanks to the periodic boundary conditions and the non-divergent condition $\nabla \cdot \mathbf{u} = 0$ of Equations 2.65. Equation 2.66 thus suggests a constant spatial mean of $\mathbf{u}$, whose value is given by the initial condition which has been assumed to be zero.

We thus define a function space

$$\mathbb{V} := \left\{ \mathbf{u}(t, \cdot) \in \left[C_p^\infty(\Omega)\right]^2 : \nabla \cdot \mathbf{u} = 0, \int_\Omega \mathbf{u}\,\mathrm{d}\mathbf{x} = \mathbf{0} \right\}, \qquad (2.68)$$

where $\left[C_p^\infty(\Omega)\right]^2$ is the space of two-component smooth and doubly periodic functions on $\Omega$. Let $V$ be the completion of $\mathbb{V}$ in the $\mathbb{H}^1(\Omega)$ norm, where $\mathbb{H}^1(\Omega)$ is the space of two-

---

[4]Throughout this section, components of vectors are denoted by subscripts, such as $u_i$, $v_j$ and $x_i$ for the components of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{x}$ respectively.

component, divergence-free, zero-mean and continuously differentiable functions whose $L^2$ norm and whose first derivative's $L^2$ norm are finite. In the weak formulation of the 2D Navier-Stokes equations, $V$ is the function space to which the test function $\mathbf{v}$ belongs, so that solutions belong to its dual space $V^*$, the space of linear functionals on $V$.

Now, taking the inner product of the evolution equation (the first of Equations 2.65) with $\mathbf{v}$ and applying Green's first identity, we have

$$\int_\Omega \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{v}\,\mathrm{d}\mathbf{x} + \int_\Omega (\mathbf{u}\cdot\nabla\mathbf{u})\cdot\mathbf{v}\,\mathrm{d}\mathbf{x} = -\int_\Omega \nabla p \cdot \mathbf{v}\,\mathrm{d}\mathbf{x} + \int_\Omega \mathbf{f}\cdot\mathbf{v}\,\mathrm{d}\mathbf{x} - \int_\Omega \nu\nabla\mathbf{u}\cdot\nabla\mathbf{v}\,\mathrm{d}\mathbf{x}, \quad (2.69)$$

where $\nabla\mathbf{u}\cdot\nabla\mathbf{v} = \sum_{i=1}^2 \nabla u_i \cdot \nabla v_i$. Arguing in the same way as Equation 2.67, the first term on the right-hand-side vanishes, and thus the dependence on pressure is dropped. This can be expected, as the pressure is a diagnostic quantity in the Navier-Stokes equations — there is no equation for the time-evolution of the pressure. Defining a bi-linear form

$$a(\mathbf{u},\mathbf{v}) := \int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{v}\,\mathrm{d}\mathbf{x} \quad (2.70)$$

and a tri-linear form

$$b(\mathbf{u},\mathbf{v},\mathbf{w}) := \int_\Omega (\mathbf{u}\cdot\nabla\mathbf{v})\cdot\mathbf{w}\,\mathrm{d}\mathbf{x} = \sum_{i,j=1}^2 \int_\Omega u_i\frac{\partial v_j}{\partial x_i}w_j\,\mathrm{d}\mathbf{x}, \quad (2.71)$$

and writing $(\cdot,\cdot)$ for the inner product on $V$, Equation 2.69 reduces to

$$\left(\frac{\partial \mathbf{u}}{\partial t}, \mathbf{v}\right) + \nu a(\mathbf{u},\mathbf{v}) + b(\mathbf{u},\mathbf{u},\mathbf{v}) = (\mathbf{f},\mathbf{v}). \quad (2.72)$$

To require Equation 2.72 to hold for all $\mathbf{v}\in V$ is equivalent to satisfying the ODE[5]

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} + \nu A\mathbf{u} + B(\mathbf{u},\mathbf{u}) = \mathbf{f} \quad (2.73)$$

in $V^*$, where $A: V \to V^*$ is the linear operator defined by

$$(A\mathbf{u},\mathbf{v}) = a(\mathbf{u},\mathbf{v}) \quad (2.74)$$

---

[5]The use of the total derivative $\frac{\mathrm{d}}{\mathrm{d}t}$ in Equation 2.73 in place of the partial derivative $\frac{\partial}{\partial t}$ is to emphasise that the spatial dependence of $\mathbf{u}$ and $\mathbf{f}$ have already been treated through the introduction of the space $V^*$ so that only the temporal dependence remains (see Robinson (2001), pp. 194–195 and 240–241).

and $B : V \times V \rightarrow V^*$ is the bi-linear operator defined by

$$(B(\mathbf{u}, \mathbf{v}), \mathbf{w}) = b(\mathbf{u}, \mathbf{v}, \mathbf{w}). \tag{2.75}$$

Equation 2.73 in $V^*$ defines the weak formulation of the 2D Navier-Stokes equations' initial-boundary-value problem described above.

## 2.2.2 Uniqueness of solutions and their continuous dependence on initial conditions

Using Equation 2.73, we shall prove the uniqueness of weak solutions and their continuous dependence on $\mathbf{u}_0$. Since any strong solution is also a weak solution, it shall automatically follow from the proof that strong solutions are unique and depend continuously on initial conditions.

We first demonstrate an elementary property of the tri-linear form $b(\mathbf{u}, \mathbf{v}, \mathbf{w})$, that it is anti-symmetric with respect to the second and third arguments. Using the product rule and the divergence theorem,

$$
\begin{aligned}
b(\mathbf{u}, \mathbf{v}, \mathbf{w}) &= \int_\Omega \sum_{j=1}^2 \left( (\mathbf{u} \cdot \nabla) v_j \right) w_j \, \mathrm{d}\mathbf{x} \\
&= \int_\Omega \sum_{j=1}^2 \left( \nabla \cdot (v_j \mathbf{u}) \right) w_j \, \mathrm{d}\mathbf{x} - \int_\Omega \sum_{j=1}^2 v_j (\nabla \cdot \mathbf{u}) w_j \, \mathrm{d}\mathbf{x} \\
&= \int_\Omega \sum_{j=1}^2 \nabla \cdot (v_j w_j \mathbf{u}) \, \mathrm{d}\mathbf{x} - \int_\Omega \sum_{j=1}^2 v_j \mathbf{u} \cdot \nabla w_j \, \mathrm{d}\mathbf{x} - \int_\Omega \sum_{j=1}^2 v_j (\nabla \cdot \mathbf{u}) w_j \, \mathrm{d}\mathbf{x} \\
&= \int_{\partial\Omega} \sum_{j=1}^2 v_j w_j \mathbf{u} \cdot \mathrm{d}\mathbf{S} - \int_\Omega \sum_{j=1}^2 \left( (\mathbf{u} \cdot \nabla) w_j \right) v_j \, \mathrm{d}\mathbf{x} - \int_\Omega (\mathbf{v} \cdot \mathbf{w})(\nabla \cdot \mathbf{u}) \, \mathrm{d}\mathbf{x} \\
&= -b(\mathbf{u}, \mathbf{w}, \mathbf{v})
\end{aligned}
\tag{2.76}
$$

due to the periodic boundary conditions and the non-divergent condition $\nabla \cdot \mathbf{u} = 0$.

We now proceed with the main proof. For two solutions $\mathbf{u}$ and $\mathbf{v}$ of Equation 2.73, not necessarily corresponding to the same initial condition, the error velocity field

$\mathbf{w} \coloneqq \mathbf{u} - \mathbf{v}$ satisfies

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} + \nu A\mathbf{w} + B(\mathbf{w}, \mathbf{u}) + B(\mathbf{v}, \mathbf{w}) = \mathbf{0}. \tag{2.77}$$

Taking the inner product of this error equation with $\mathbf{w}$, we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{w}\|^2 + \nu A\|\nabla\mathbf{w}\|^2 = -b(\mathbf{w}, \mathbf{u}, \mathbf{w}) - b(\mathbf{v}, \mathbf{w}, \mathbf{w}), \tag{2.78}$$

where $\|\cdot\|$ is the $L^2$ norm. The second term on the right-hand-side drops out following Equation 2.76. The remaining terms thus satisfy the inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{w}\|^2 + \nu A\|\nabla\mathbf{w}\|^2 \le |b(\mathbf{w}, \mathbf{u}, \mathbf{w})|, \tag{2.79}$$

where $|\cdot|$ simply denotes the magnitude. The term on the right-hand-side can be bounded above by $M_0\|\mathbf{w}\|\|\nabla\mathbf{w}\|\|\nabla\mathbf{u}\|$, following an application of Ladyzhenskaya's and Hölder's inequalities (Robinson 2001), whose scope is beyond this thesis. The bound here, in which $M_0$ here is a positive constant, is specific to the 2D version of the tri-linear form $b(\cdot, \cdot, \cdot)$. Using the elementary inequality $|2xy| \le x^2 + y^2$, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{w}\|^2 + \nu A\|\nabla\mathbf{w}\|^2 \le \frac{\nu}{2}\|\nabla\mathbf{w}\|^2 + \frac{M_0^2}{2\nu}\|\mathbf{w}\|^2\|\nabla\mathbf{u}\|^2 \tag{2.80}$$

which upon rearrangement gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{w}\|^2 + \nu A\|\nabla\mathbf{w}\|^2 \le \frac{M_0^2}{\nu}\|\mathbf{w}\|^2\|\nabla\mathbf{u}\|^2. \tag{2.81}$$

We can now invoke Grönwall's inequality on

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{w}\|^2 \le \frac{M_0^2}{\nu}\|\mathbf{w}\|^2\|\nabla\mathbf{u}\|^2, \tag{2.82}$$

which Inequality 2.81 implies, and arrive at

$$\|\mathbf{w}(t, \mathbf{x})\|^2 \le \exp\left(\int_0^t \frac{M_0^2}{\nu}\|\nabla\mathbf{u}(s)\|^2\,\mathrm{d}s\right)\|\mathbf{w}(0, \mathbf{x})\|^2. \tag{2.83}$$

Solutions thus depend Lipschitz-continuously on initial conditions, since

$$\|\mathbf{u}(t,\mathbf{x}) - \mathbf{v}(t,\mathbf{x})\| \leq L(T)\|\mathbf{u}(0,\mathbf{x}) - \mathbf{v}(0,\mathbf{x})\|, \qquad t \in [0,T] \tag{2.84}$$

where

$$L(T) := \sup_{t \in [0,T]} \sqrt{\exp\left(\int_0^t \frac{M_0^2}{\nu}\|\nabla\mathbf{u}(s)\|^2\,\mathrm{d}s\right)}. \tag{2.85}$$

In particular, uniqueness follows by setting $\mathbf{u}(0,\mathbf{x}) = \mathbf{v}(0,\mathbf{x})$, or equivalently $\mathbf{w}(0,\mathbf{x}) \equiv \mathbf{0}$.

### 2.2.3  Implications for predictability

As an immediate corollary to the Lipschitz-continuity of the solutions' dependence on initial conditions (Inequality 2.84), predictability is indefinite for the 2D Navier-Stokes equations (Palmer et al. 2014). Indeed, if a prediction is defined to lose its skill when $\|\mathbf{w}(t,\mathbf{x})\| \geq \varepsilon$, then for any given time $T \in \mathbb{R}^+$, the prediction is skilful for at least up to $T$ when the initial error $\|\mathbf{w}(0,\mathbf{x})\|$ can be made sufficiently small, that is, smaller than $\frac{1}{L(T)}\varepsilon$. This argument is independent of the spectral slope of the basic-state energy spectrum, unlike Lorenz's argument in Section 2.1.

## 2.3  Reconciling the paradox

At first glance, the indefinite predictability implied by the well-posedness theorem of the 2D Navier-Stokes equations seems to contradict Lorenz's result for any spectral slope shallower than $-3$. However, we have not discussed the role of the spectral slope in $L(T)$ which, as we will see in the following, reconciles the conflict.

Central to our argument are the Inequality 2.84 and Equation 2.85 above, and an understanding of the distinction between the real atmosphere and its numerical model. For simplicity, suppose the real fluid system has only one inertial range of slope $-p$ in its energy spectrum, which is stationary. In this case, $|\hat{\mathbf{u}}(t,k)|^2 = A_0 k^{-p}$ between its large-scale cutoff wavenumber $k_1$ and small-scale cutoff wavenumber $k_2$, with $A_0$ being a constant. Here we have changed the notation: Fourier coefficients (denoted by the hat

symbol) are now defined in terms of scalar wavenumbers $k$. It follows that

$$\|\nabla \mathbf{u}_s\|^2 = \int_0^\infty k^2 |\hat{\mathbf{u}}_s|^2 \, \mathrm{d}k = \int_0^{k_1} k^2 |\hat{\mathbf{u}}_s|^2 \, \mathrm{d}k + A_0 \int_{k_1}^{k_2} k^{2-p} \, \mathrm{d}k + \int_{k_2}^\infty k^2 |\hat{\mathbf{u}}_s|^2 \, \mathrm{d}k, \quad (2.86)$$

where the subscript $s$ distinguishes the system itself from a model for the system which we will denote with the subscript $m$. The three terms on the right-hand-side of Equation 2.86 represent contributions from the large scale, the inertial range and the viscous range respectively. Compared to the first two terms, the term representing the viscous range is assumed to be small. In particular, the integrand is assumed to decay rapidly enough so that $\|\nabla \mathbf{u}_s\|^2$ remains finite (this is in fact part of the definition of the function space $V^*$ to which $\mathbf{u}_s$ belongs).

Now, suppose the model truncates at wavenumber $k_t \ll k_2$ and numerical dissipation kicks in at some wavenumber $k_0$ close to $k_t$ yet substantially greater than $k_1$. We therefore have $k_0 \ll k_2$, which is a key assumption of the present analysis. For the model,

$$\|\nabla \mathbf{u}_m\|^2 = \int_0^{k_t} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k = \int_0^{k_1} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k + A_0 \int_{k_1}^{k_0} k^{2-p} \, \mathrm{d}k + \int_{k_0}^{k_t} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k. \quad (2.87)$$

Again, we may neglect the contribution from the viscous range, so that

$$\|\nabla \mathbf{u}_m\|^2 \approx \int_0^{k_1} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k + A_0 \int_{k_1}^{k_0} k^{2-p} \, \mathrm{d}k. \quad (2.88)$$

Because $k_0, k_t \ll k_2$, the second integral in Relation 2.88 with $p < 3$ *appears* to diverge as the model resolution $k_t$ and thus $k_0$ increases. Combining this with Equation 2.85, $L(T)$ — until $k_2$ is reached — grows exponentially with $k_0$. This makes the Lipschitz-continuous bound in Inequality 2.84 almost useless. To keep the error $\|\mathbf{u}(t, \mathbf{x}) - \mathbf{v}(t, \mathbf{x})\|$ under control, the initial error $\|\mathbf{u}(0, \mathbf{x}) - \mathbf{v}(0, \mathbf{x})\|$ would have to decrease exponentially, but decreasing the scale of the initial error without changing its magnitude relative to the basic-state energy spectral density (as presumed in Lorenz's thought experiment) would only give a polynomial decrease. Therefore in practice, since the initial error cannot be made to decrease fast enough as more scales are resolved (let alone the large-scale error be constrained to zero (Durran & Gingrich 2014)), the corollary of indefinite predictability discussed in Sub-section 2.2.3 fails to hold. As such, there is no contradiction between

Lorenz's argument of inherently finite predictability and the well-posedness theorem of the 2D Navier-Stokes equations.

This concept, known as 'asymptotic ill-posedness', was put forward by Palmer et al. (2014) as they argued that whether the system of 3D Navier-Stokes equations is analytically well-posed is practically irrelevant to the theory of inherently finite predictability. We have now extended the discussion to the 2D system and given a mathematical basis to the concept in our context.

When $p > 3$, the second integral on the right-hand-side of Relation 2.88 does not appear to diverge as $k_0 \rightarrow k_2$. This means one may indeed approximate $\|\nabla \mathbf{u}_s\|^2$ by the $\|\nabla \mathbf{u}_m\|^2$ in Relation 2.88 with a sufficiently large value of $k_0$. So would the $L(T)$ in Equation 2.85 be approximated without regard to the model resolution, making it possible for $\|\mathbf{u}(t, \mathbf{x}) - \mathbf{v}(t, \mathbf{x})\| < \varepsilon$ by making $\|\mathbf{u}(0, \mathbf{x}) - \mathbf{v}(0, \mathbf{x})\|$ small enough in scale and thus achieving indefinite predictability.

So far our argument for the cases $p < 3$ and $p > 3$ are in harmony with Lorenz's result. For the borderline case $p = 3$, the analysis here suggests a possibly finite predictability horizon in practice. This is because Relation 2.88 gives

$$\|\nabla \mathbf{u}_m\|^2 \approx \int_0^{k_1} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k + A_0 \int_{k_1}^{k_0} k^{-1} \, \mathrm{d}k = \int_0^{k_1} k^2 |\hat{\mathbf{u}}_m|^2 \, \mathrm{d}k + A_0 \log \frac{k_0}{k_1}, \qquad (2.89)$$

which appears to diverge as $k_0 \rightarrow k_2$. This disagrees with the established theory by Lorenz (1969) and on dimensional grounds. Even with the logarithmic correction

$$|\hat{\mathbf{u}}(t, k)|^2 \sim k^{-3} \left[ \log \left( \frac{k}{k_r} \right) \right]^{-\frac{1}{3}} \qquad (k_r > 0 \text{ constant}), \qquad (2.90)$$

or more generally

$$|\hat{\mathbf{u}}(t, k)|^2 \sim k^{-3} \left[ A_1 \log \left( \frac{k}{k_r} \right) + A_2 \right]^{-\frac{1}{3}} \qquad (A_1, \, A_2, \, k_r > 0 \text{ constants}), \qquad (2.91)$$

to the $-3$ spectrum (Bowman 1996), an easy calculation along the previous lines still suggests that the predictability horizon can be finite in practice. As such, we are unable to explain the disagreement and leave the problem open.

The above analysis has assumed that the fluid system exhibits only one inertial range in its energy spectrum. This can easily be extended to systems with multiple inertial ranges, since only the range immediately preceding the viscous range pertains to the argument concerning the large-$k_0$ behaviour. As such, the relevant $p$ for Earth's atmosphere is $\frac{5}{3}$ (Nastrom & Gage 1985). Since the typical $k_t$ for current-day atmospheric models is smaller than $k_2$ by "at least seven or eight orders of magnitude" (Palmer et al. 2014), the crucial assumption ($k_t \ll k_2$) is satisfied for our analysis to be applicable to the real atmosphere. We therefore conclude that the inherently finite atmospheric predictability as understood by the meteorological community is not inconsistent with the regularity of the 2D Navier-Stokes equations.

## 2.4 Summary

Half a century since Lorenz's pioneering discovery of the finite predictability horizon for the atmosphere, it has now become an "accepted part of the canon of dynamical meteorology" (Rotunno & Snyder 2008). His argument is based on solving a simple system of ODEs representing error growth in the dynamics of the 2D barotropic vorticity equation, and on exploring patterns in the error's saturation times.

In this chapter, we have re-assessed the details behind Lorenz's conclusion (Section 2.1). His model of ODEs has been re-derived in a mathematically more consistent manner, and solved analytically rather than numerically. The behaviour of error growth produced by the model in relation to its mathematical properties has been discussed. Furthermore, we have extended his calculations to the $k^{-3}$ energy spectrum. While the $k^{-3}$ spectrum is thought to be associated with indefinite predictability, this has however been refuted within the context of Lorenz's model.

We have also discussed how the finite predictability horizon could be compatible with an analytic theorem for the 2D Navier-Stokes equations, which implies indefinite predictability regardless of the slope of the energy spectrum (Sections 2.2 and 2.3). The apparent contradiction could be reconciled by noting how quickly the initial error has to be brought down with increasing resolution to guarantee indefinite predictability, and

by understanding how this is practically impossible when the spectrum is shallower than $k^{-3}$. Our analysis gives rise to an inconsistent result with the mainstream understanding of indefinite predictability in the borderline case of a $k^{-3}$ spectrum, yet it nevertheless agrees with our own computations of Lorenz's model.

Until recently, energy spectra in global NWP models had only resolved the synoptic-scale $k^{-3}$ range. As model resolutions extend into the mesoscale $k^{-\frac{5}{3}}$ range (European Centre for Medium-Range Weather Forecasts 2016), the strong constraints on the range of predictability envisaged by Lorenz is expected to become visible (Judt 2018). However, the limits on predictability arising from initial errors on the large scales will also limit predictability in practice (Durran & Gingrich 2014), and the interplay between the two spectral ranges could be an interesting area to explore. These will be discussed in the next chapter.

# 3 Impact of the mesoscale range on error growth and the predictability limit

Despite the conflicting opinions on the predictability horizon associated with the $k^{-3}$ range when the problem is presented in different perspectives, the previous chapter has unequivocally established that a $k^{-\frac{5}{3}}$ energy spectrum imposes a finite-time predictability barrier for all practical purposes, as errors develop more rapidly on smaller scales than on larger scales. The $k^{-\frac{5}{3}}$ range, which appears only in the mesoscale part of the atmospheric energy spectrum (Nastrom & Gage 1985), is now being increasingly resolved by global NWP models (European Centre for Medium-Range Weather Forecasts 2016). It is therefore becoming more relevant to ask how errors in the mesoscale range would grow under the influence of the synoptic-scale $k^{-3}$ range of the atmospheric energy spectrum, and how it would be different from the classic pattern of error growth for a stand-alone $k^{-\frac{5}{3}}$ spectrum. In this chapter, the error growth behaviour under such a hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum reminiscent of Earth's atmosphere will be examined in Lorenz's model and in a series of identical-twin perturbation experiments using an idealised 2D barotropic turbulence model at a range of resolutions. Implications on the predictability limit will also be discussed.

## 3.1 Numerical experiments using an idealised 2D turbulence model

The numerical experiments are performed on a forced-dissipative version of the 2D barotropic vorticity model (Equation 1.9), namely

$$\frac{\partial \theta}{\partial t} + J(\psi, \theta) = f + d, \qquad \theta = \Delta \psi, \tag{3.1}$$

where $f$ is the forcing and $d$ the dissipation. The model is run pseudo-spectrally in a doubly periodic domain $D$ at various resolutions $k_t \in \{256, 512, 1024, 2048\}$ (where $k_t$ is the truncation wavenumber), with the $f$ and $d$ prescribed in spectral space.

In the absence of $f$ and $d$, Equation 3.1 possesses two invariant quadratic integral quantities: the energy $E = \int_D -\frac{1}{2}\psi\theta \, d\mathbf{x}$ and the enstrophy $Z = \int_D \frac{1}{2}\theta^2 \, d\mathbf{x}$. This can be seen by multiplying the equation by $-\psi$ and by $\theta$ respectively, followed by integrating over $D$, upon which the Jacobian term $J(\cdot, \cdot)$ vanishes since the boundary conditions are doubly periodic. $E$ and $Z$ are defined in such a way that they are positive-definite. This should be obvious for $Z$, whereas the result for $E$ can be obtained by applying Green's first identity.

The main reason for which this particular 2D model is chosen for the numerical experiments is that its dynamics supports both $k^{-3}$ and $k^{-\frac{5}{3}}$ energy spectra (Lindborg 1999, Vallis 2017). In addition, it is closely related to classical 2D turbulence models which serve as a starting point for understanding large-scale atmospheric dynamics. If $d$ were chosen to be $\nu\Delta\theta$ where $\nu > 0$, then the incompressible 2D Navier-Stokes equations (Equation 2.65) would be recovered in the vorticity form. Equation 3.1 could also represent 2D motions in a rotating frame if the rate of rotation were constant, since the term $2\mathbf{\Omega} \times \mathbf{u}$ representing rotation (cf. Equation 1.4) would drop out in the 2D vorticity formulation.

### 3.1.1 Construction of the hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ energy spectrum

The addition of the $f$ and $d$ terms in Equation 3.1 is necessary for numerically sustaining a quasi-stationary turbulence and therefore an inertial range in the energy spectrum. Without these terms, the turbulence could quickly decay by numerical dissipation or blow up by having unrealistically too much energy at the small scales. The shape of the spectrum is determined by the way in which the turbulence is forced. Under the classical picture of 2D turbulence (Kraichnan 1967), a large-scale forcing generates a $k^{-3}$ inertial range through a downscale enstrophy cascade, whereas a small-scale forcing generates a $k^{-\frac{5}{3}}$ range through an upscale energy cascade. One way therefore to construct a hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum is to force the system at both large and small scales. Inspired by Maltrud & Vallis (1991), forcing is injected in limited bands of wavenumbers $k \in [10, 14]$ and $k \in [\frac{5}{8}k_t, \frac{165}{256}k_t]$. Physically, the former represents synoptic-scale baroclinic forcing, and the latter represents mesoscale forcing. The mesoscale forcing is made to depend on the resolution $k_t$ of the model so that the extent to which the $k^{-\frac{5}{3}}$ range is resolved may vary across model resolutions.

Compared to the choice of wavenumbers at which the forcing is injected, the specific description of the forcing at a given wavenumber is less crucial. Maltrud & Vallis (1991) provided a form of stochastic forcing in its time-discrete representation

$$f_{n_t} = \hat{A}\sqrt{1 - R^2}e^{i\phi} + Rf_{n_t-1} \tag{3.2}$$

where $n_t > 0$ is the number of time-steps elapsed, $\hat{A} > 0$, $R \in (0, 1)$, and $\phi$ is a uniform random number in $[0, 2\pi)$ independent of $f_1, \ldots, f_{n_t-1}$. This form of forcing is to be followed in the remainder of this thesis[1]. Yet, it would be useful to construct the time-continuous form of Equation 3.2 in order to obtain a physical interpretation of the

---

[1]The simulation that generated Figure 2.4 is forced by a white-noise process at $k \in [18, 22]$ only, so as to obtain a $k^{-3}$ spectrum.

parameters. For this we compute the mean

$$
\begin{aligned}
\mu &= \mathbb{E}\left[f_{n_t}\right] \\
&= \mathbb{E}\left[\hat{A}\sqrt{1-R^2}e^{i\phi} + Rf_{n_t-1}\right] \\
&= 0 + R\,\mathbb{E}\left[f_{n_t-1}\right] \\
&= 0 + R\mu
\end{aligned}
\tag{3.3}
$$

which implies

$$
\mu = \frac{0}{1-R} = 0,
\tag{3.4}
$$

and the variance

$$
\begin{aligned}
V &= \mathbb{E}\left[f_{n_t}f_{n_t}^*\right] \\
&= \mathbb{E}\left[\left(\hat{A}\sqrt{1-R^2}e^{i\phi} + Rf_{n_t-1}\right)\left(\hat{A}\sqrt{1-R^2}e^{-i\phi} + Rf_{n_t-1}^*\right)\right] \\
&= \hat{A}^2(1-R^2) + R\hat{A}\sqrt{1-R^2}\,\mathbb{E}\left[e^{-i\phi}f_{n_t-1} + e^{i\phi}f_{n_t-1}^*\right] + R^2\,\mathbb{E}\left[f_{n_t-1}f_{n_t-1}^*\right] \\
&= \hat{A}^2(1-R^2) + 0 + R^2 V
\end{aligned}
\tag{3.5}
$$

which implies
$$
V = \frac{\hat{A}^2(1-R^2)}{1-R^2} = \hat{A}^2.
\tag{3.6}
$$

$\hat{A}$ is therefore the standard deviation of the stochastic process. The auto-covariance of the stochastic process is

$$
\begin{aligned}
\lambda(\tau) &= \mathbb{E}\left[f_{n_t}f_{n_t+\tau}^*\right] - \mu^2 = \mathbb{E}\left[f_{n_t}f_{n_t+\tau}^*\right] \\
&= \mathbb{E}\left[f_{n_t}\hat{A}\sqrt{1-R^2}e^{-i\phi}\right] + \mathbb{E}\left[f_{n_t}Rf_{n_t+\tau-1}^*\right] \\
&= 0 + R\lambda(\tau-1) \\
&= \cdots \\
&= R^\tau\lambda(0) = R^\tau V,
\end{aligned}
\tag{3.7}
$$

so that the auto-correlation is $V^{-1}\lambda(\tau) = R^\tau$. Hence, the $e$-folding de-correlation time $t_f$ is one such that

$$
R^{\frac{t_f}{\Delta t}} = \frac{1}{e}
\tag{3.8}
$$

where $\Delta t$ is the time-step. We therefore obtain

$$R = \exp\left(-\frac{\Delta t}{t_f}\right), \tag{3.9}$$

with which Equation 3.2 can be expressed as

$$f_{n_t} = \hat{A}\sqrt{\frac{2}{t_f}} e^{i\phi} \sqrt{\Delta t} + \left(1 - \frac{\Delta t}{t_f}\right) f_{n_t - 1} \tag{3.10}$$

in the limit of small $\Delta t$. In view of this, Equation 3.2 is therefore equivalent to the complex-valued stochastic process

$$\mathrm{d}f = -\frac{1}{t_f} f \,\mathrm{d}t + \hat{A}\sqrt{\frac{2}{t_f}} \,\mathrm{d}\tilde{W}, \tag{3.11}$$

which is an Ornstein-Uhlenbeck process except that the noise $\tilde{W}$ is a uniform random number on the unit circle in the complex plane. This will be applied independently on each 2D wavevector in the wavebands specified above.

To mimic real-world models which do not compromise the quality of large-scale predictions as the model resolution progressively increases, the fully resolved part of the energy spectra must agree among runs of different $k_t$. This is achieved by controlling the forcing parameters in Equation 3.11. Unfortunately, this has to be done *ad experimentum*, since no known formulae to our knowledge relate the parameters with the shape of the spectrum. For convenience, $t_f$ is fixed at 0.5. The following choices of $\hat{A}$ are found to be appropriate following a series of fine-tuning tests: $\hat{A} = 0.004$ for the large-scale forcing for all $k_t$; and $\hat{A} = 0.005, 0.006, 0.007, 0.008$ for the small-scale forcing for the $k_t = 256$, 512, 1024, 2048 experiments respectively.

Dissipation is introduced to remove energy cascaded into the largest scales and enstrophy cascaded into the smallest scales. A standard dissipation operator is in the form $d = (-1)^{s+1} \nu \Delta^s \theta$, where $\nu > 0$ and $s \in \mathbb{R}$. Like the forcing, the dissipation is introduced scale-selectively[2]. At the largest scales $k \in [1, 3]$, it is in the form of a linear drag $d = -0.0029\,\theta$. At the smallest scales $k \geq \frac{25}{32} k_t$, $d = -0.083\,\Delta^8\theta$, which is a hyper-

---

[2]Except for the simulation corresponding to Figure 2.4, in which $d$ is applied to all scales and is proportional to $-\Delta^6\theta$.

viscosity. At most wavenumbers, therefore, forcing and dissipation are both absent, thus enabling clean energy and enstrophy cascades along the inertial ranges.

Figure 3.1 shows the basic-state energy spectra of the turbulence spun up using the above choices of forcing and dissipation parameters. Such choices make the transition between the $k^{-3}$ and $k^{-\frac{5}{3}}$ ranges happen on the order of $k = 100$, in agreement with the atmospheric energy spectrum observed by Nastrom & Gage (1985) where the spectral break sits at a length scale of about 400 kilometres. The spectra shown in Figure 3.1 are averages over five independent realisations, and are scaled by $k^{\frac{5}{3}}$ so that a perfect $k^{-\frac{5}{3}}$ range would appear as a horizontal line in the figure. It is apparent that the transition to a $k^{-\frac{5}{3}}$ spectrum is gradual, and is not even achieved in the highest-resolution run ($k_t = 2048$), although it is getting very close.
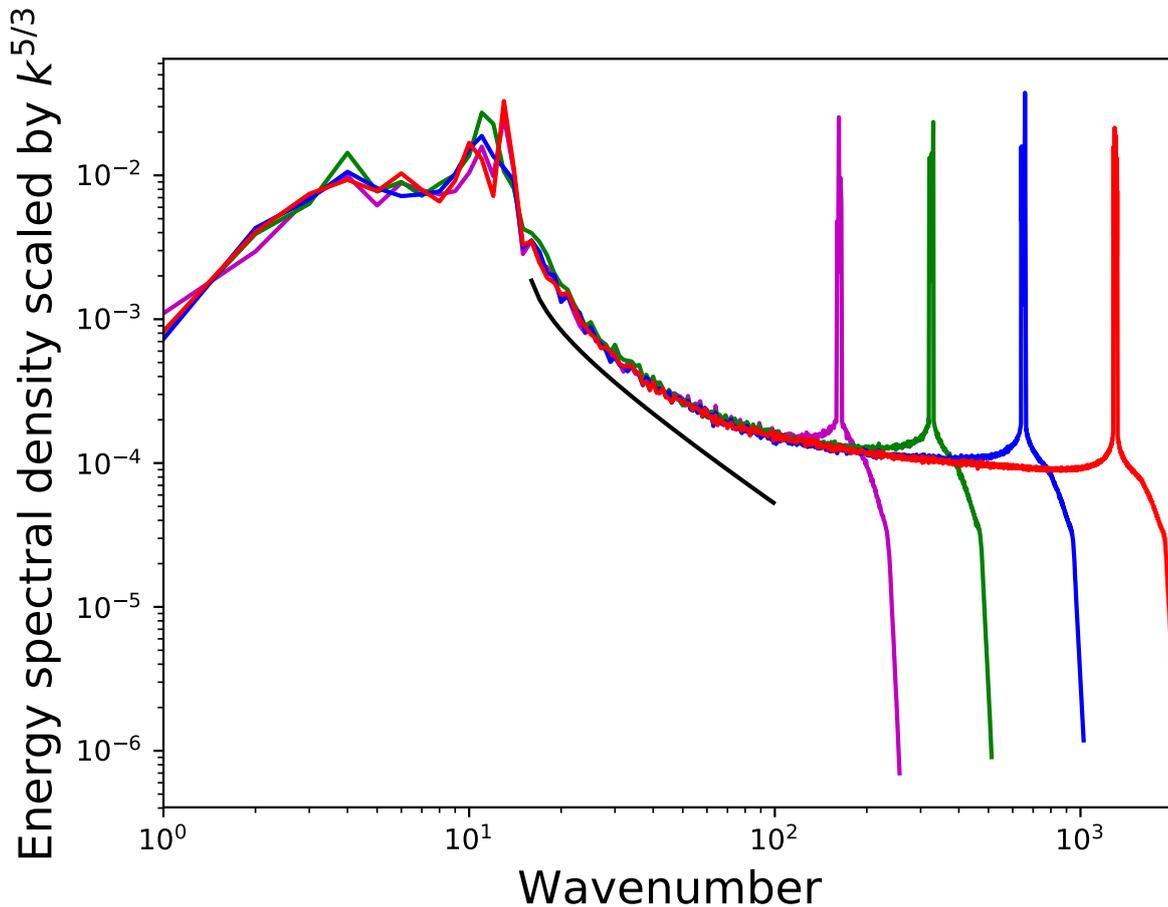


Figure 3.1: Basic-state energy spectra, scaled by a factor of $k^{\frac{5}{3}}$, for model resolutions $k_t = 256$ (magenta), 512 (green), 1024 (blue) and 2048 (red). The black curve shows a logarithmically corrected $k^{-3}$ reference spectrum $E(k) \sim k^{-3} \left[ \log \left( \frac{k}{15} \right) \right]^{-\frac{1}{3}}$ (cf. Relation 2.90), again scaled by a factor of $k^{\frac{5}{3}}$. The spectra are averaged over five independent realisations that differ in the random seed. The prominent peaks are associated with the mesoscale forcing, while the steep drop-off at the smallest scales is associated with the hyper-viscosity.

### 3.1.2 Experimental design

The turbulence corresponding to the energy spectra in Figure 3.1 serves as a starting point for the perturbation experiments described as follows. These experiments are in the form of identical twins — pairs of runs that differ only in the initial condition. In particular, the realisation of the stochastic forcing (Equation 3.11) is kept fixed within each pair throughout the model's integration, so that any error developed throughout the integration can be solely attributed to the error in the initial condition.

The initial perturbations are introduced at a single wavenumber $k_p$ at a relative magnitude of 1%, using the procedure established by Leung (2017) and Leung et al. (2019). For each wavevector $\mathbf{k}$ whose modulus equals $k_p$ when rounded to the nearest integer, a random phase shift $S_{\boldsymbol{\delta}} := \exp(-i\mathbf{k} \cdot \boldsymbol{\delta})$ is applied on a pre-determined part $\gamma \in [0, 1]$ of the spectral coefficients $\hat{\psi}(0, \mathbf{k})$ and thus $\hat{\theta}(0, \mathbf{k})$. Here, $\boldsymbol{\delta}$ is a random vector drawn from a bi-variate uniform distribution in $[0, 2\pi) \times [0, 2\pi)$, and $\hat{\psi}(t, \cdot)$ and $\hat{\theta}(t, \cdot)$ indicate the spatial Fourier coefficients of $\psi(t, \mathbf{x})$ and $\theta(t, \mathbf{x})$ respectively. On average, the 2D error energy spectral density $E_e(0, \mathbf{k})$ for that wavevector at the initial time is

$$
\begin{aligned}
&\mathbb{E}\left[E_e(0, \mathbf{k})\right] \\
&= \mathbb{E}\left[-\frac{1}{2}\left((\gamma S_{\boldsymbol{\delta}} + (1-\gamma))^* \hat{\psi}(0, \mathbf{k})^* - \hat{\psi}(0, \mathbf{k})^*\right)\left((\gamma S_{\boldsymbol{\delta}} + (1-\gamma))\hat{\theta}(0, \mathbf{k}) - \hat{\theta}(0, \mathbf{k})\right)\right] \\
&= \mathbb{E}\left[-\frac{1}{2}\gamma^2 |S_{\boldsymbol{\delta}} - 1|^2 \hat{\psi}(0, \mathbf{k})^* \hat{\theta}(0, \mathbf{k})\right] \\
&= \mathbb{E}\left[-\frac{1}{2}\left(2\gamma^2(1 - \cos \mathbf{k} \cdot \boldsymbol{\delta})\right)\hat{\psi}(0, \mathbf{k})^* \hat{\theta}(0, \mathbf{k})\right] \\
&= 2\gamma^2 E(0, \mathbf{k}),
\end{aligned}
\tag{3.12}
$$

where $E(t, \mathbf{k})$ is the 2D energy spectral density of the basic-state flow. Therefore, $\gamma$ is set to be $\frac{1}{\sqrt{2}}$ for a 1% relative error.

Two sets of perturbation experiments are performed. The first set explores the dependence of error growth properties on the scale $k_p$ of the initial error. There the model resolution is fixed to be the highest possible, i.e. $k_t = 2048$, and perturbations are introduced at $k_p = 128$, 256, 512 and 1024. The second set explores the sensitivity of error growth to the model resolution by making $k_t$ variable. Model resolutions of

$k_t = 256, 512, 1024$ and $2048$ are considered. $k_p$ is fixed relative to $k_t$ at $\frac{k_p}{k_t} = 0.5$ so that the initial error is confined to a small scale yet unaffected by the forcing and dissipation. The combination $(k_t, k_p) = (2048, 1024)$ is therefore included in both sets. To reduce noise in the results, the experiment for each combination $(k_t, k_p)$ is repeated using the five independently generated initial profiles specific to $k_t$. The results are then averaged over these realisations[3] before being reported in Figures 3.2–3.7.

### 3.1.3   Numerics of the model

The model is run in pseudo-spectral space, with $\hat{\theta}(t, \mathbf{k}) = -|\mathbf{k}|^2 \hat{\psi}(t, \mathbf{k})$ being the prognostic variable. Numerically, $\mathbf{k}$ runs inclusively from $-(k_t-1)$ to $+k_t$ in both components, taking integer values. The scalar wavenumber $k$ is determined by rounding $|\mathbf{k}|$ to the nearest integer. Due to the square nature of the $\mathbf{k}$ grid, dynamics at scales beyond $k = k_t$ cannot be fully resolved by the model. Without loss of generality, $\psi(t, \mathbf{x})$ is assumed to have a spatial mean of zero, so that both $\hat{\theta}(t, \mathbf{0})$ and $\hat{\theta}(t, \mathbf{0})$ are fixed at zero.

The Jacobian term $J(\cdot, \cdot)$ in Equation 3.1 is computed in physical space via a pair of discrete Fourier transforms with a spectral de-aliasing filter introduced by Hou & Li (2007). The dissipation term $d$ is treated exactly by means of an integrating factor. We have adopted the original formula for the forcing term $f$ provided by Maltrud & Vallis (1991) (Equation 3.2), with the parameter $R$ computed exactly using Equation 3.9. The model is integrated in time up to $t = 150$ using a standard fourth-order Runge-Kutta method with an adaptive time-stepping scheme to ensure numerical stability. Since the realisation of the stochastic forcing has to be uniform within each identical-twin pair, any changes in the time-step $\Delta t$ as a result of the adaptive time-stepping scheme are applied simultaneously to the control and perturbed runs.

---

[3]The result in Figure 2.4 is an average of only four independent realisations generated using the forcing and dissipation terms described in the earlier footnotes of this chapter.

### 3.1.4 Results

**Error growth and its dependence on perturbation scale**

Figure 3.2 shows the evolution of the error energy spectra for the various perturbation scales $k_p$ in the highest-resolution ($k_t = 2048$) model, in which a substantial part of the basic-state energy spectrum follows the $k^{-\frac{5}{3}}$ power-law reasonably well (Figure 3.1). The error spectra grow up-magnitude more or less uniformly across scales. As the mesoscale saturates, the error growth slows down, as indicated by the more closely packed spectra at later times. These observations are broadly consistent with the findings of Boffetta & Musacchio (2001), who simulated error growth in the inverse-cascade regime of 2D turbulence (i.e. a $k^{-\frac{5}{3}}$ basic-state spectrum). They also agree with Judt (2018)'s study using a global convection-permitting NWP model.

Figure 3.2 also suggests that the dependence of error growth behaviour on perturbation scale $k_p$ is minimal, as manifested by the largely similar shapes of the error spectra across the panels. This is in good agreement with Durran & Gingrich (2014), who pointed out that the qualitative pattern of error growth is not so sensitive to the scale of the initial error. Apart from these features shown in the figure, it is found that decreasing the perturbation scale (increasing $k_p$) introduces a time-lag in saturating a given synoptic scale. Yet, this lag decreases as the synoptic-scale wavenumber decreases and becomes negligible at the largest scales.

**Dependence on model resolution**

The results for the second set of experiments, in which the model resolution $k_t$ is variable, are shown in Figure 3.3. There is a qualitative difference between the error energy spectra of the low-resolution runs, where the $k^{-\frac{5}{3}}$ range is barely resolved (Figure 3.3(a,b)), and those of the high-resolution runs where the $k^{-\frac{5}{3}}$ range is well-resolved (Figure 3.3(c,d)). Without a resolved mesoscale range, the error spectra peak at the synoptic scale (about $k = 10$) throughout the growth process, following a short initial adjustment. This is consistent with previous studies (Rotunno & Snyder 2008, Durran & Gingrich 2014). In
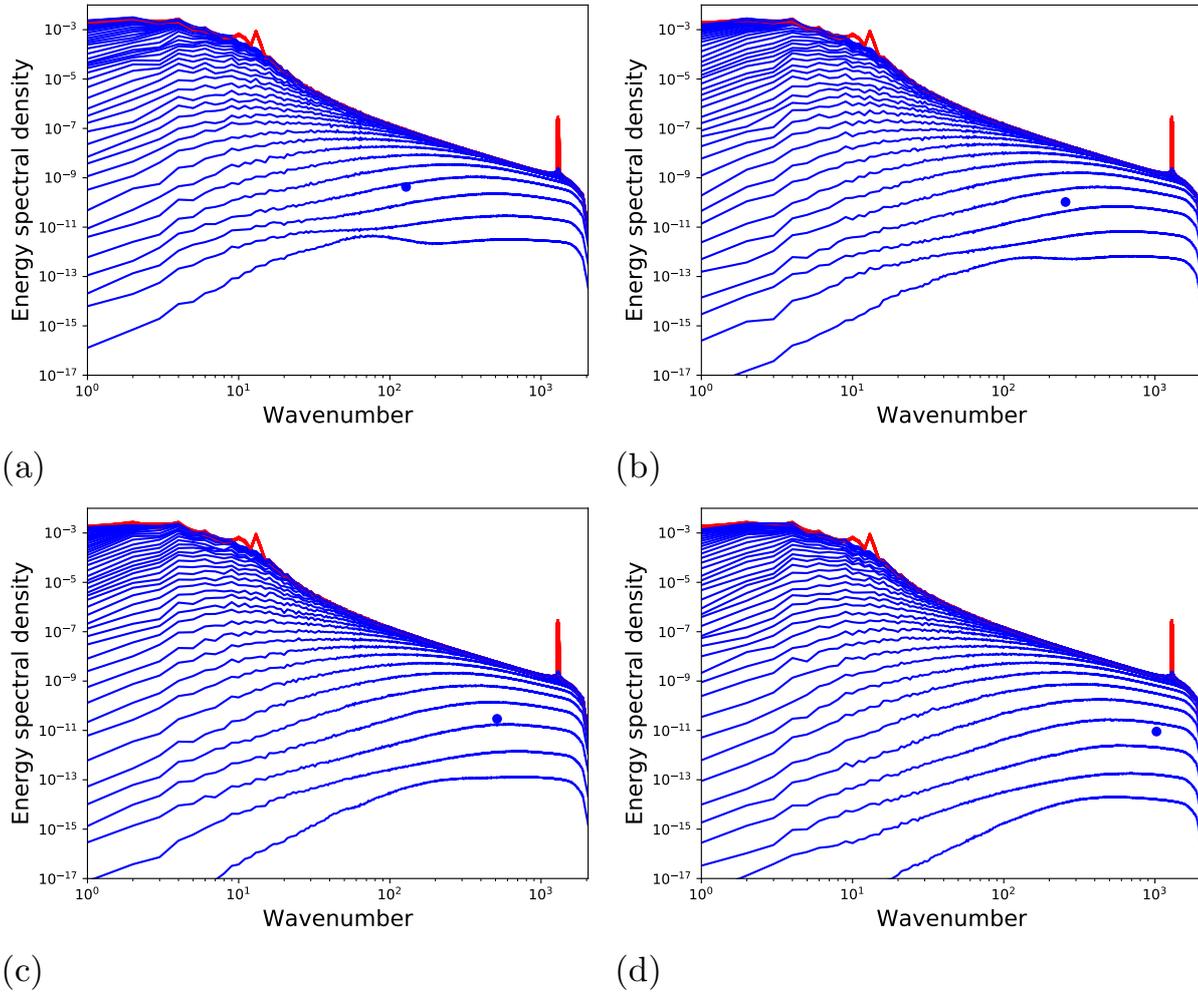
Figure 3.2: Evolution of error energy spectra (blue, from bottom to top within each panel) for identical-twin experiments with $k_t = 2048$ and $k_p =$ (a) 128, (b) 256, (c) 512 and (d) 1024. The error spectra are plotted at equal time-intervals. The blue dots indicate the scale ($k_p$) and magnitude of the initial perturbations, and the red curves indicate the basic-state energy spectra (scaled by a factor of two, see Equation 2.17) at the same times as the error spectra are plotted.

the presence of a mesoscale range, however, they initially peak at nearly the smallest resolved scale, i.e. towards the end of the $k^{-\frac{5}{3}}$ range, again echoing earlier studies (Lorenz 1969, Rotunno & Snyder 2008, Durran & Gingrich 2014). After the mesoscale error saturates, a separate, broad peak in the synoptic scale begins to emerge in the error spectra, resembling the error growth paradigm under a $k^{-3}$ range (cf. Figure 2.4). The same has been reported by Judt (2018) in the context of a high-resolution global NWP model.

Error energy spectra under a hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum thus show a stage-dependent peak and an up-magnitude growth at almost all stages. The analysis of the error growth behaviour may be done more quantitatively by fitting the error growth to a parametric model and extracting information from the fitted parameters, as will be
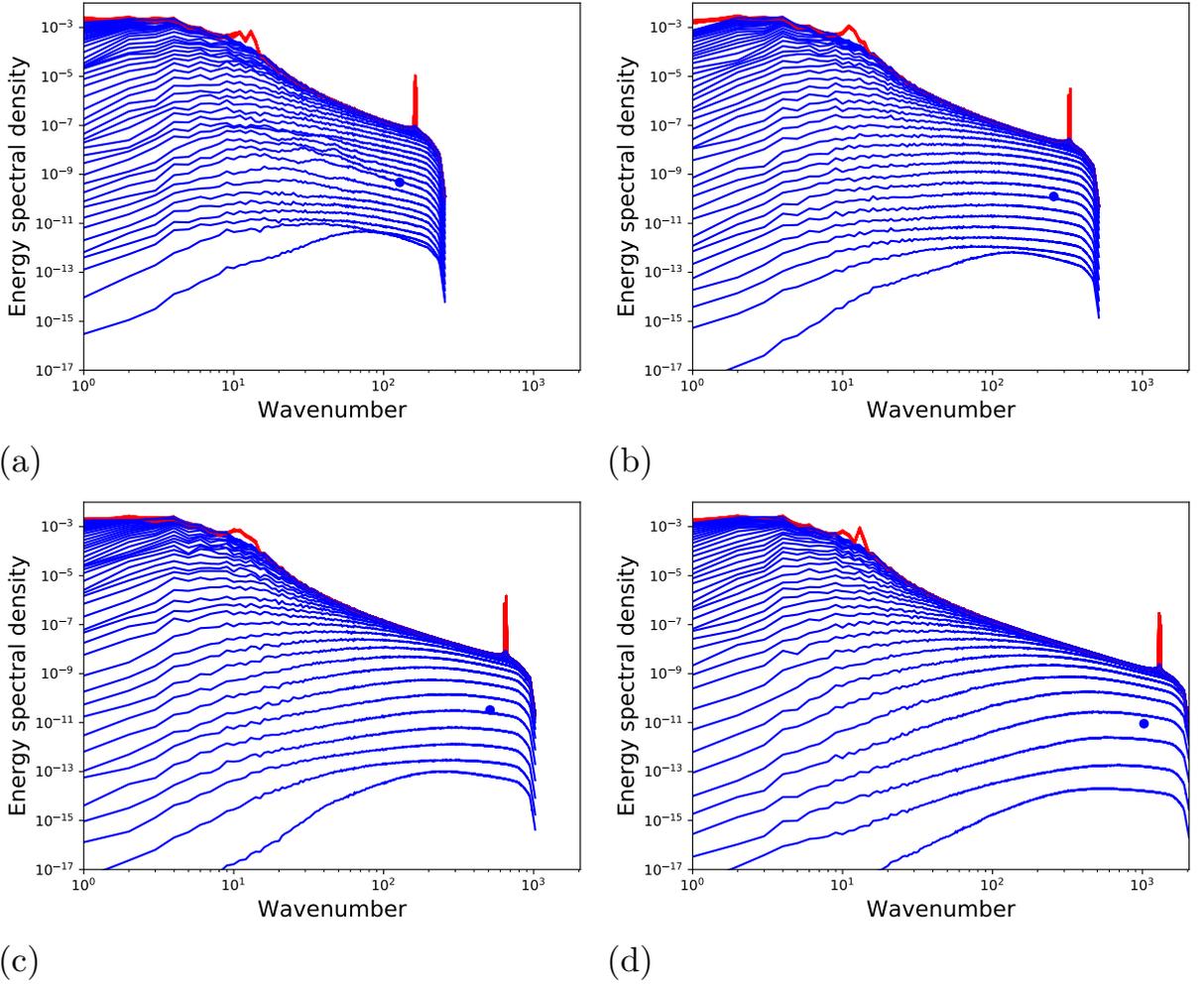
Figure 3.3: As in Figure 3.2, but for $k_t =$ (a) 256, (b) 512, (c) 1024 and (d) 2048, and $k_p = \frac{1}{2}k_t$. Note that (d) is identical to Figure 3.2(d).

discussed in the next section.

## 3.2 Assessing the error growth rate

### 3.2.1 Fitting the error growth to a parametric model

Parametric error growth models offer a simple description of the time-evolution of the error which is often useful for diagnostic analyses of error dynamics. A parametric model is used in this section to systematically investigate the dependence of the rate of error growth on spatial scale, based on the results of the numerical simulations in the previous section. The parametric model of Žagar et al. (2017) ('the Žagar model'), which has already been discussed in Section 1.6, is chosen for this purpose. Here, the relative error

energy spectral density $E_r(t, k)$, defined by

$$E_r(t, k) = \frac{\int_{|\mathbf{k}|=k} E_e(t, \mathbf{k}) \, \mathrm{d}\mathbf{k}}{2 \int_{|\mathbf{k}|=k} E(t, \mathbf{k}) \, \mathrm{d}\mathbf{k}}, \tag{3.13}$$

is fitted to a hyperbolic tangent function

$$E_r(t, k) = A \tanh(at + b) + B \tag{3.14}$$

individually for each wavenumber. The relative error energy is the ratio between the error energy and twice the energy of the basic-state flow, at which the error eventually saturates (see Equation 2.17). The parameters $A$, $B$, $a$ and $b$ are real-valued functions of the scalar wavenumber $k$, yet without loss of generality $A$ and $a$ are assumed to be non-negative. The fitting is carried out using Python's `scipy.optimize` package, in which a least-squares minimisation is performed by the Levenberg-Marquardt algorithm to compute the set of parameters that best approximates the evolution of the error. Since the algorithm returns only a local minimum but not necessarily the global minimum, it requires an appropriate initial guess.

The definition of the relative error in Equation 3.13 implies that $E_r(t, k) \to 1$ as $t \to \infty$. Moreover, one may expect $E_r(t, k) \to 0$ as $t \to -\infty$ as the initial error is small. As such, $A$ and $B$ are expected to be both 0.5 regardless of $k$. While setting $A = B = 0.5$ can reduce the dimensionality of the parameter space and hence make the fitting easier, this may compromise the quality of the fit as the error at especially the largest scales may have not yet fully saturated by the end of the numerical experiment. For this reason, no restriction on parameter values is imposed in the fitting algorithm.

To illustrate that the hyperbolic tangent function is capable in describing error growth, Figure 3.4 shows the evolution of the relative error energy at a specific wavenumber and its best fit according to Equation 3.14. The fit typically smoothens the error's fluctuations around the saturation level. Away from the saturation level, the fitting function matches the error almost perfectly.

The contour plot in Figure 3.5(a) is obtained by repeating the fitting procedure independently for all wavenumbers $k$ up to $k_t$. The corresponding plot for the raw, unfitted
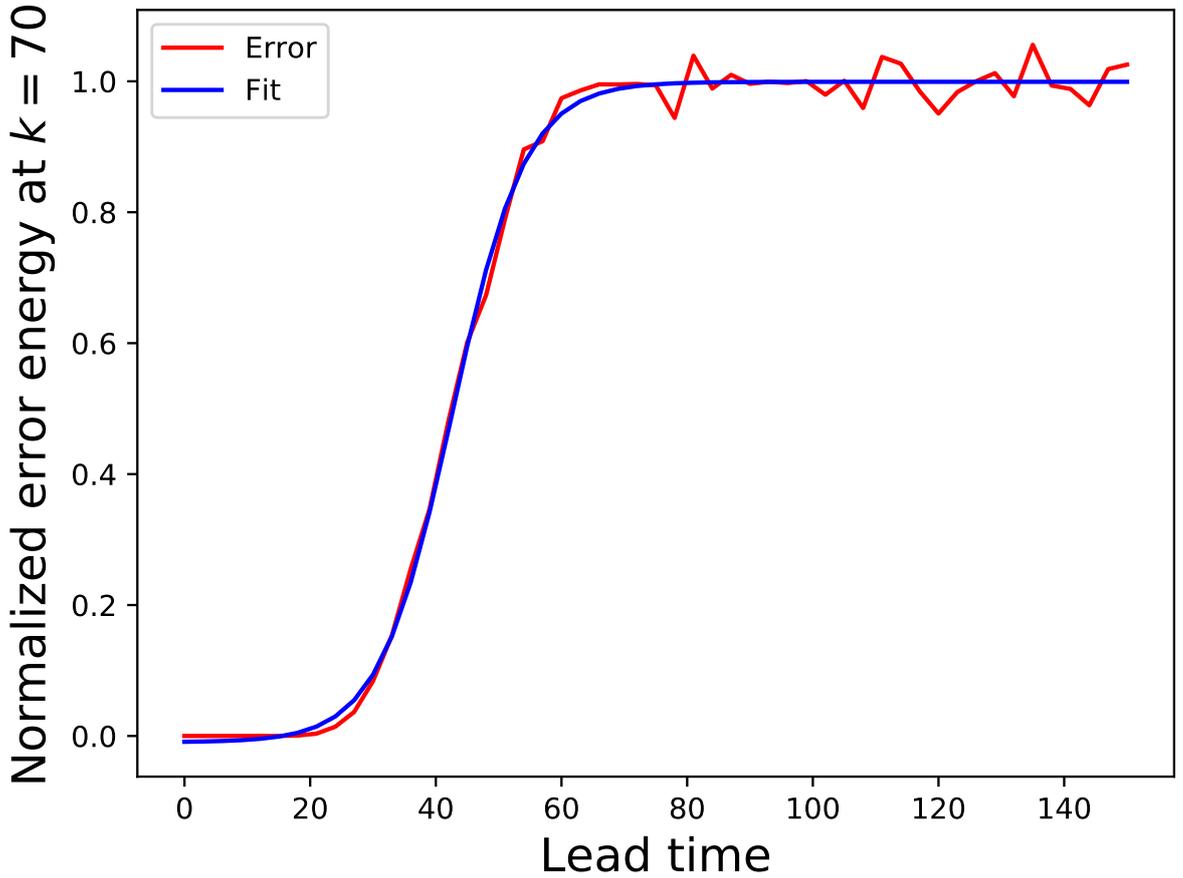
Figure 3.4: Growth of the relative error energy $E_r$ at $k = 70$ in the $(k_t, k_p) = (2048, 1024)$ simulation (red). The blue curve shows the best fit of the red curve to the Žagar model according to Equation 3.14.

error is shown in Figure 3.5(b). It is evident that the fitting removes noise and provides a cleaner signal to the error growth pattern.

### 3.2.2 Inferring predictability from the parameters

Parameter $a$ of Equation 3.14 carries a mathematical interpretation. It controls the width of the hyperbolic tangent curve. By studying its dependence on $k$, $k_t$ and $k_p$, the predictability of the system can be inferred. To see this, let $E_1$ and $E_2$ be two arbitrary error energy levels with $E_1 < E_2$, and $t_1$ and $t_2$ be the times when these levels are attained. If we write $F_i = \frac{E_i - B}{A}$, $i = 1, 2$, then Equation 3.14 implies $at_i + b = \tanh^{-1}(F_i)$, so that

$$t_2 - t_1 = \frac{1}{a} \left( \tanh^{-1}(F_2) - \tanh^{-1}(F_1) \right). \tag{3.15}$$
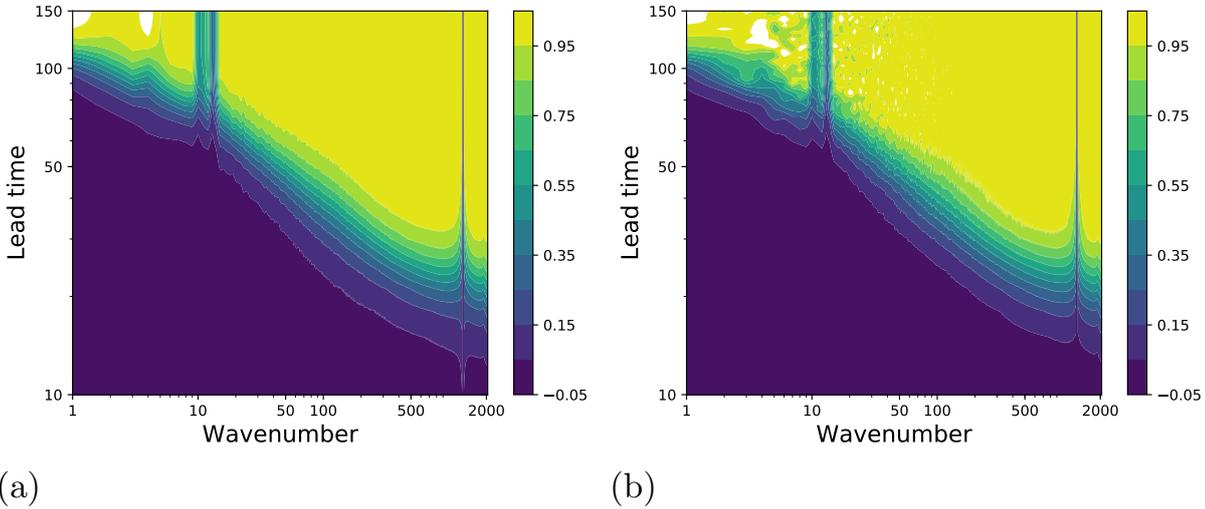
Figure 3.5: Growth of the (a) fitted and (b) raw relative errors as functions of the wavenumber, for the same simulations as in Figure 3.4. The colours and contours indicate the relative error energy $E_r$.

Since the hyperbolic tangent function is monotonically increasing, $\tanh^{-1}(F_2) - \tanh^{-1}(F_1)$ is always positive, meaning that a smaller $a$ always gives a larger (longer) $t_2 - t_1$. As $a$ becomes larger, the curve narrows and thus suggests a more rapid error growth.

For the first set of experiments in which $k_t = 2048$ and $k_p$ is variable, Figure 3.6 shows that $a$ increases with $k$ until the effects of the small-scale forcing become important[4]. Hence, by the above argument, the error grows faster as the spatial scale decreases. This is particularly apparent in the $k^{-\frac{5}{3}}$ mesoscale range, where the slope $\frac{\mathrm{d}a}{\mathrm{d}(\log k)}$ increases. This is a hallmark of inherently finite predictability, and reinforces the agreement with Judt (2018)'s earlier study using a more sophisticated NWP model.

It is interesting to see that $a$ increases more rapidly in the mesoscale when $k_p$ is smaller. In other words, error growth in the mesoscale is faster when the perturbation is applied at a larger scale. This may be attributable to the fast transfer of larger-scale errors into the smaller scales (Durran & Gingrich 2014).

Figure 3.7 shows $a(k)$ for the second set of experiments, in which $\frac{k_p}{k_t}$ is fixed at $\frac{1}{2}$. It is quite remarkable that the values of $a$ for the different resolutions are broadly consistent (as long as they lie outside the forcing ranges), meaning that the error growth at a given scale is not substantially altered by pushing the model to a higher resolution.

---

[4]The basic-state energy levels at the forcing ranges are substantially higher than the energy levels at their neighbouring wavenumbers (Figure 3.1), and it takes much longer for the error to saturate these scales if that ever happens (Figures 3.2 and 3.3). As such, the parameter $a$ is not representative of the error growth rate at these wavenumbers, and should therefore be discarded.
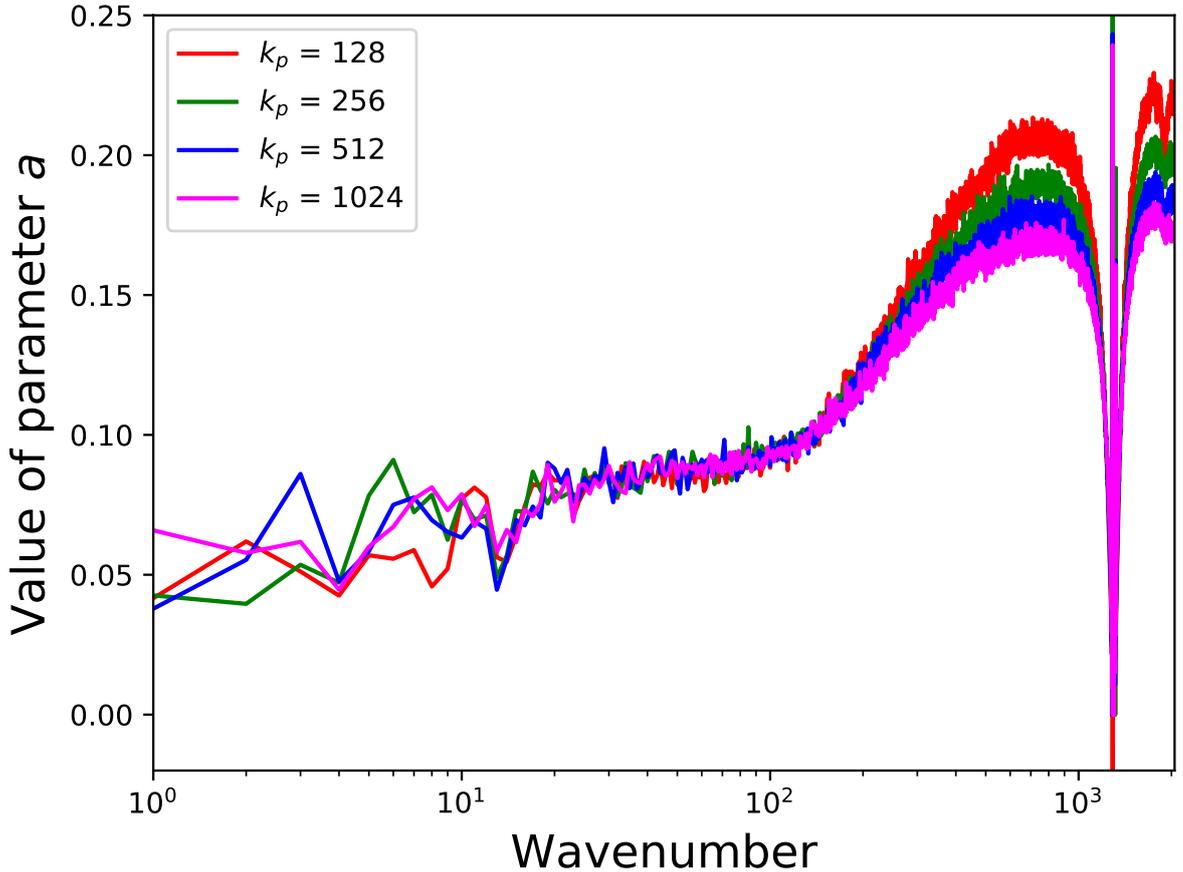
Figure 3.6: Parameter $a$ of the Žagar model, fitted to the relative error energy at individual wavenumbers according to Equation 3.14, as a function of wavenumber, for perturbation experiments of various $k_p$ for the highest-resolution model $k_t = 2048$. Note that the vertical axis is linear and not logarithmic.

Having said that, the distinctively changing slope $\frac{\mathrm{d}a}{\mathrm{d}(\log k)}$ for the highest-resolution run $k_t = 2048$ (the same magenta curve as in Figure 3.6) is not seen when $k_t$ is smaller.

The heuristic dimensional argument for homogeneous and isotropic turbulence (Lilly 1990) implies that the parameter $a$ should scale as $(k^3 E(k))^{\frac{1}{2}}$, since it carries the physical dimension of inverse time. Accordingly, $a$ should be constant in $k$ if the energy spectrum is $k^{-3}$, and should scale as $k^{\frac{2}{3}}$ if $E(k) \sim k^{-\frac{5}{3}}$. However, Figure 3.7 suggests that $a$ scales with $k$ logarithmically in the large scales. Into the small scales of the highest-resolution runs, a polynomial scaling seems to emerge, but in any case it falls well short of $k^{\frac{2}{3}}$ which demands a more-than-fourfold increase in $a$ for every decade of wavenumbers. Hence, the observed behaviour of $a$ remains in an intermediate, non-asymptotic regime, as might be expected under a hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ energy spectrum.
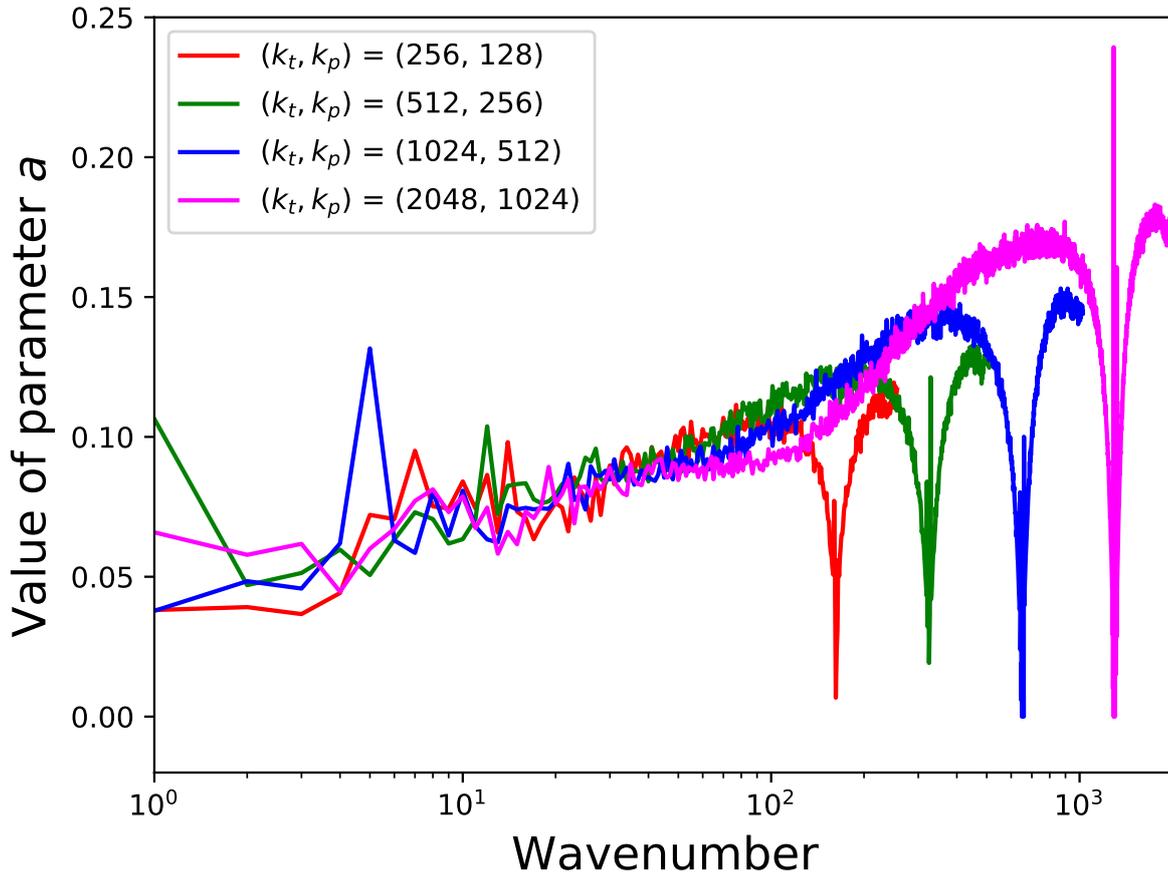
Figure 3.7: As in Figure 3.6, but for combinations $(k_t, k_p)$ in which $k_p = \frac{1}{2}k_t$.

## 3.3 Exploring the asymptotic behaviour using Lorenz's model

It is of interest to investigate the characteristics of error growth under the hybrid spectrum in the infinite-resolution limit. To achieve this, a much higher-resolution model is needed to reasonably serve as a proxy for the infinite-resolution case. The primitive model of Lorenz (1969) (Equation 2.56) is a good candidate for this purpose. Being computationally inexpensive, it enables the running of ultra-high-resolution simulations.

### 3.3.1 Reproducing the results of the numerical experiments

We first demonstrate that Lorenz's model is able to capture the essential aspects of error growth observed in the numerical experiments of Sections 3.1 and 3.2. Specifically, we show this for the set of experiments in which $\frac{k_p}{k_t}$ is fixed (cf. Figure 3.3). Since Lorenz's

80

model allows the basic-state energy spectrum to be independently specified, we recycle such basic-state spectra recorded at the end of the identical-twin simulations in Section 3.1. This is then used to compute the matrix $\mathbf{C}$ and hence run the model. For each $(k_t, k_p)$ pair, a single basic-state spectrum is formed by averaging over the 5 independent realisations. Next, the spikes induced by the forcing are removed, with the energy spectral densities at the forced wavenumbers replaced by interpolation of the densities at the neighbouring wavenumbers outside the forced ranges[5]. The resulting spectrum is then discretised into the scales $K$. The number $n$ of scales in the model depends on $k_t$ such that $n = \log_\rho \frac{k_t}{N_0} = \log_2 k_t = 8, 9, 10$ and 11 respectively. A small but necessary tweak is introduced: in Lorenz's original formulation, scale $K$ refers to wavenumbers in the half-open interval $(\rho^{K-1} N_0, \rho^K N_0]$, whereas here it refers to $[\rho^{K-1} N_0, \rho^K N_0)$ instead — the inclusion of end-points is reversed. In this way, contributions from wavenumber $N_0 = 1$ is not discarded by Lorenz's model. The contribution at wavenumber $N_n = \rho^n N_0 = k_t$ is now dropped, but its impact should be negligible as it carries little energy.

The model, with $\mathbf{C}$ computed from the discretised spectrum, is solved for one-half of the initial error drawn from the respective numerical experiment. The factor of one-half follows from Lorenz's definition of error energy which allows the error energy to saturate at the basic-state energy level rather than twice of it[6]. The initial condition for $\frac{d\mathbf{Z}}{dt}$ is set to be zero for all $K$, in the same way as it is for the rest of this thesis.

Figure 3.8 shows the parameter $a$ of the Žagar model, which is now a function of $K$. Compared to the growth rates for the numerical experiments (Figure 3.7), the single most distinctive feature — that $a$ generally increases as $k$ or $K$ increases, albeit much slower than the heuristic scaling would suggest — is captured in Lorenz's model. In other words, Lorenz's model is able to reproduce the moderate quickening of error growth in the mesoscale, though not to the same extent as in the numerical experiments themselves (the values of $a$ in the mesoscale range in Figure 3.8 are generally smaller than in Figure 3.7 by a factor of two). Lorenz's model also captures the suppression of error growth at intermediate scales in the higher-resolution simulations, as seen in Figure 3.7.

It should be noted that the possibly unrealistic emergence of transient negative

---

[5]The interpolation is linear in log-log space in order to respect the power-law nature of the spectrum.
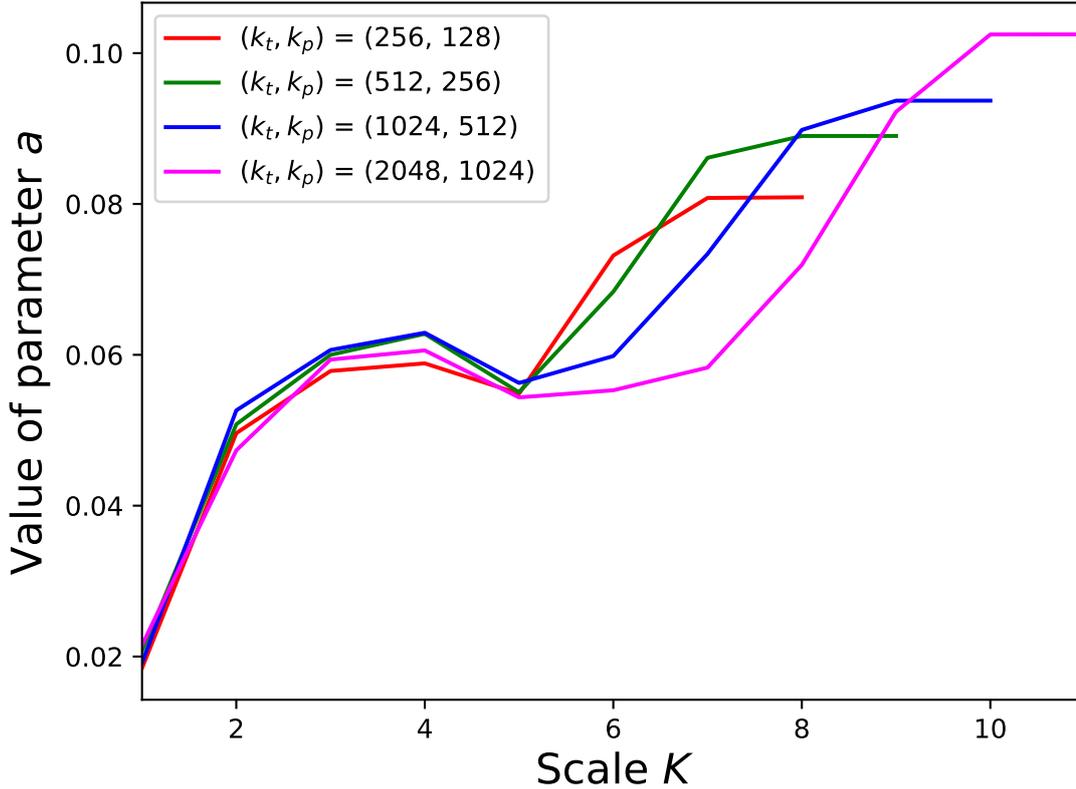[6]See the discussion around Equations 2.8–2.17.

Figure 3.8: As in Figure 3.7, but for Lorenz (1969)'s model.

error energy values discussed in Section 2.1.3 does not affect our concerned parameter $a$ of the Žagar model. This is because the erratic oscillatory behaviour qualitatively represents nothing more than a time-delay in error growth, which is already represented in the parameter $b$.

### 3.3.2 Error growth in the infinite-resolution limit

Having demonstrated the ability of Lorenz's model to reproduce the basic features of error growth, we turn our focus to an ultra-high-resolution case, $n = 21$. Physically, it corresponds to a minimum wavelength of about 19 metres on Earth, well beyond the resolution of today's NWP models.

The discretised basic-state spectrum used for the $n = 11$ simulation above is extended to $K = 21$, assuming a pure $k^{-\frac{5}{3}}$ range at these smaller scales. In other words, for all integers $K \in [11, 21)$,

$$\frac{X_{K+1}}{X_K} = \rho^{-\frac{2}{3}} = 2^{-\frac{2}{3}}. \tag{3.16}$$

As explained in Section 2.1.2, the scaling $\rho^{-\frac{2}{3}K} \approx k^{-\frac{2}{3}} = k^{-\frac{5}{3}+1}$ is proportional to the energy integrated over a unit logarithm of wavenumbers when the energy spectral density scales as $k^{-\frac{5}{3}}$.

Figure 3.9(a) illustrates the growth of a small-scale error under this hybrid basic-state spectrum extended to $K = 21$. The error spectrum exhibits a fairly sharp peak at all lead times, in contrast with the lower-resolution case (e.g. Figure 3.3(d)) in which the peak is much broader. Figure 3.9(b) shows the same but for a single $k^{-\frac{5}{3}}$ range, defined by Equation 2.58 yet normalised to such a level that the magnitude of the mesoscale part of the spectrum agrees with that in Figure 3.9(a). This enables a direct comparison between the two figures for examining the effects of an additional $k^{-3}$ range in the synoptic scale (it should be noted that in this way the hybrid spectrum is more energetic in absolute terms). There is a very close agreement between the nature of the mesoscale error growth in Figure 3.9(a) and in Figure 3.9(b). Moreover, the large positive entries in the bottom-left of the matrix $\mathbf{C}$ for the hybrid spectrum (Table 3.1) increase by a factor of about 2.5 every row, the same type of behaviour as the matrix for the single-range $k^{-\frac{5}{3}}$ spectrum (Table 2.1). It seems plausible, then, to suggest that the error under the hybrid spectrum asymptotically behaves as the error under a single $k^{-\frac{5}{3}}$ range, and that the presence of the $k^{-3}$ range does not affect the fast error growth at the smallest scales. This comparison also suggests that the resolution $n = 21$ is sufficient to be considered a proxy for the infinite-resolution limit.

This can be expressed in more quantitative terms by considering the parameter $a$ of the Žagar model (Figure 3.10). For $n = 21$ (black solid curve), $a$ grows exponentially beyond $K = 11$. This growth is very similar in simulations at intermediate resolutions, confirming that our results have converged in this respect. Indeed, the growth is even faster than the theoretically expected scaling of $k^{\frac{2}{3}} \approx \rho^{\frac{2}{3}K}$ for a $k^{-\frac{5}{3}}$ spectrum. The implication here is that it is necessary to fully resolve $K = 11$ (representing the range of length scales between 19.5 and 39.1 kilometres on Earth) for the model to pick up the fast error growth pertaining to the $k^{-\frac{5}{3}}$ range, despite it being more than a decade of wavenumbers beyond the spectral break between the $k^{-3}$ and $k^{-\frac{5}{3}}$ ranges (Figure 3.1). Moreover, the results suggest that the synoptic-scale $k^{-3}$ range acts to slow down error growth in the first decade of the mesoscale. This is also supported by $a(K)$'s approximate
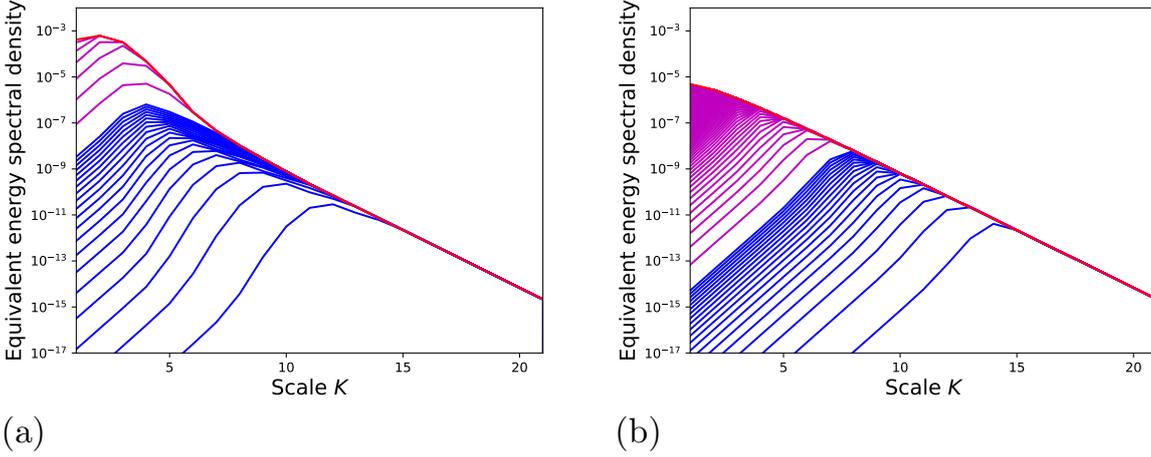
Figure 3.9: (a) Evolution of the error energy spectrum (blue and magenta, from bottom-right to top-left) in the Lorenz (1969) model under the basic-state energy spectrum (red) recovered from the $(k_t, k_p) = (2048, 1024)$ simulations in Section 3.1 (with modifications, details of which are given in the text) and extended to $K = 21$ via Equation 3.16, and an initial condition $Z_n(0) = 5 \times 10^{-7} \times \sum_{L=1}^{n} X_L$ and $Z_K(0) = 0$ for all other $K$. (b) As in (a), but for a single-range $k^{-\frac{5}{3}}$ basic-state energy spectrum according to Equation 2.58 yet normalised to such a level that the magnitude of the mesoscale part of the spectrum coincides with (a). The error spectra are plotted in blue at equal time-intervals of $\Delta t = 3$ up to $t = 60$, and in magenta at intervals of $\Delta t = 30$ thereafter. The vertical axes depict the equivalent energy spectral densities $\rho^{-K} Z_K(t)$ and $\rho^{-K} X_K$, as in Figures 2.2(a) and 2.3(a).

proportionality to $\rho^{\frac{2}{3}K}$ for all $K$ in the single-range $k^{-\frac{5}{3}}$ spectrum, which is not shown here.

### 3.3.3 The predictability horizon: a renewed estimate

By solving Lorenz's model for the hybrid spectrum, we can now update his estimate of the predictability horizon which presumed a $k^{-\frac{5}{3}}$ basic-state spectrum. Table 3.2 lists the error saturation time for each $K$, dimensionalised using Lorenz's estimate of the root-mean-square wind speed in the upper troposphere (17.1824 metres per second). The difference with Lorenz (1969)'s list is substantial. While the difference at the smallest scales may be attributed to the unrealistic oscillatory behaviour discussed earlier, which is intrinsic to Lorenz's model, the sensitivity of the saturation times to small changes in the shape of the error energy spectrum can also contribute to the difference, especially at the intermediate scales. This can be seen in Figure 3.9. In the case of the hybrid spectrum before the full saturation of the mesoscale, the typical error energy spectrum peaks at several scales larger than the scale most recently saturated, and forms a small angle with the basic-state energy

|    | 1       | 2       | 3       |
|----|---------|---------|---------|
| 1  | 0.003   | 0.008   | 0.003   |
| 2  | 0.122   | 0.030   | 0.062   |
| 3  | 0.693   | 0.529   | 0.032   |
| 4  | 1.27    | 1.12    | 0.797   |
| 5  | 1.25    | 1.20    | 1.04    |
| 6  | 0.833   | 0.822   | 0.784   |
| 7  | 0.642   | 0.640   | 0.634   |
| 8  | 1.01    | 1.01    | 1.01    |
| 9  | 2.05    | 2.05    | 2.05    |
| 10 | 4.51    | 4.51    | 4.51    |
| 11 | 10.3    | 10.3    | 10.3    |
| 12 | 25.3    | 25.3    | 25.3    |
| 13 | 63.7    | 63.7    | 63.7    |
| 14 | 161     | 161     | 161     |
| 15 | 405     | 405     | 405     |
| 16 | 1020    | 1020    | 1020    |
| 17 | 2570    | 2570    | 2570    |
| 18 | 6470    | 6470    | 6470    |
| 19 | 16 300  | 16 300  | 16 300  |
| 20 | 41 100  | 41 100  | 41 100  |
| 21 | 104 000 | 104 000 | 104 000 |

Table 3.1: The first three columns of the matrix $\mathbf{C}$ for the same hybrid energy spectrum as in Figure 3.9(a). The entries are correct to 3 significant figures or 3 decimal places as appropriate.

spectrum. Had the spectral distribution of the error been slightly sharper so that the error spectrum peaked at the scale most recently saturated (as in Figure 3.9(b)), the saturation times would have been made earlier. Hence, the saturation times could be sensitive to the shape of the error spectrum and should therefore not be taken too literally. Having said that, the predictability limit for the planetary scale is estimated to be about 15 to 20 days, in line with recent estimates using more sophisticated models (Buizza & Leutbecher 2015, Judt 2018, Zhang et al. 2019). The extra predictability (relative to Lorenz's prediction) at the large scales is likely a consequence of the additional $k^{-3}$ range in the synoptic scale. A more careful analysis is needed to confirm this, but a very recent result by Sun & Zhang (2020) based on Lorenz's model also supports extended predictability for a hybrid spectrum compared to $k^{-\frac{5}{3}}$. Indeed, our estimated saturation times in Table 3.2 are in good agreement with theirs, who constructed an evolution equation for the error by combining Equation 2.56 for matrices $\mathbf{C}$ corresponding to barotropic dynamics for a $k^{-3}$ spectrum and SQG dynamics[7] for a $k^{-\frac{5}{3}}$ spectrum, and solved the equation.

---

[7]An extension of Lorenz's model to SQG dynamics was discussed by Rotunno & Snyder (2008).
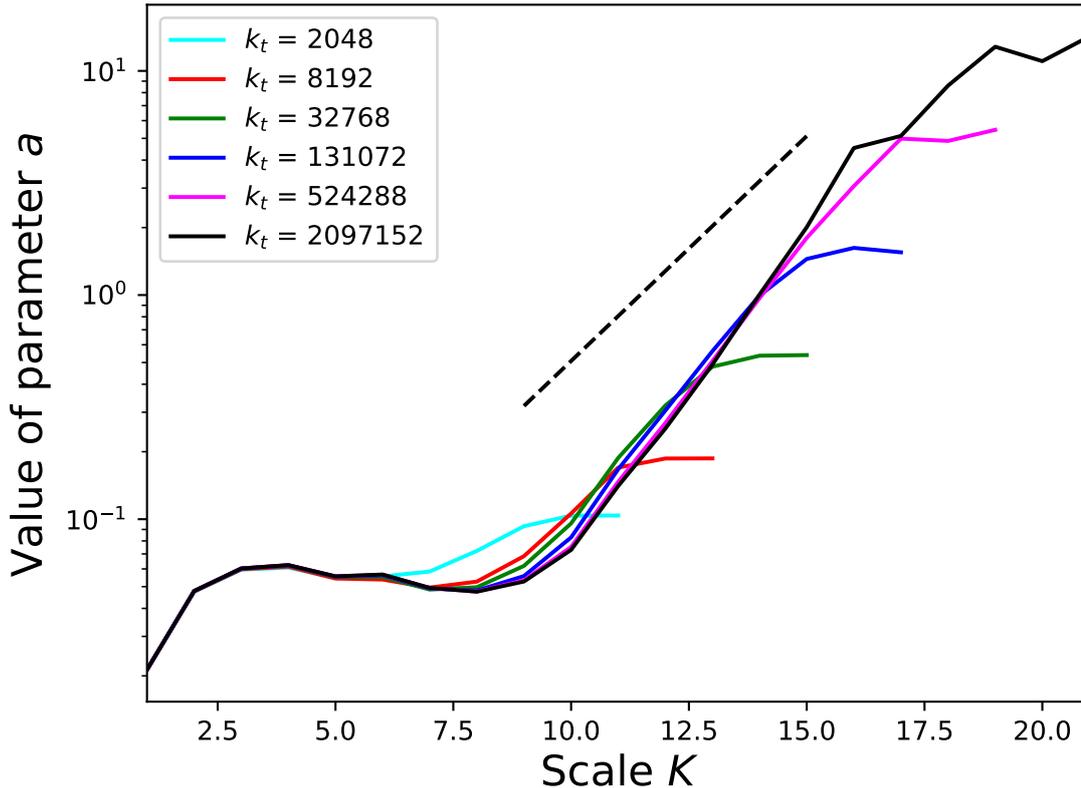
Figure 3.10: As in Figure 3.8, but for resolutions $n = 11$ (cyan), 13 (red), 15 (green), 17 (blue), 19 (magenta) and 21 (black), and an initial condition $Z_n(0) = 5 \times 10^{-7} \times \sum_{L=1}^{n} X_L$ and $Z_K(0) = 0$ for all other $K$. The vertical axis is logarithmic and the dashed line indicates an appropriately normalised $\rho^{\frac{2}{3}K}$ scaling.

## 3.4 Other initial error profiles

In Section 3.3, we focussed on cases where the initial error is concentrated at the smallest available scale, thereby approximating an infinitesimally small-scale error. This is analogous to Lorenz (1969)'s well-known Experiment A. Initial error spectra in realistic weather forecasts are, however, very different. To explore the sensitivity of error growth behaviour to the initial error spectrum, Lorenz performed the lesser-known Experiments B and C. In his Experiment B, the initial error was confined to the largest-available scale, whereas Experiment C was initialised with a fixed fraction of the control energy spectrum across all scales. He concluded that the predictability horizon at the planetary scale is barely dependent on the initial error spectrum. Durran & Gingrich (2014) expanded on Lorenz's results to show that, despite the insensitivity of the predictability horizon, the error spectra in Experiments B and C grow somewhat differently from Experiment A

| $K$ | Length scales | Predictability horizon | | |
|---|---|---|---|---|
| | | Our estimate | Sun and Zhang's estimate | Lorenz's estimate |
| 1 | $20000 - 40000$ km | 20.1 days | | 16.8 days |
| 2 | $10000 - 20000$ km | 15.8 days | | 10.1 days |
| 3 | $5000 - 10000$ km | 12.6 days | 19.2 days | 5.6 days |
| 4 | $2500 - 5000$ km | 10.3 days | 13.8 days | 3.2 days |
| 5 | $1250 - 2500$ km | 8.74 days | 10.4 days | 1.8 days |
| 6 | $625 - 1250$ km | 6.46 days | 7.9 days | 1.1 days |
| 7 | $313 - 625$ km | 5.31 days | 5.5 days | 15.7 hours |
| 8 | $156 - 313$ km | 4.30 days | 4.0 days | 9.5 hours |
| 9 | $78.1 - 156$ km | 3.53 days | 2.8 days | 5.8 hours |
| 10 | $39.1 - 78.1$ km | 2.52 days | 1.9 days | 3.6 hours |
| 11 | $19.5 - 39.1$ km | 1.24 days | 1.2 days | 2.2 hours |
| 12 | $9.77 - 19.5$ km | 20.4 hours | | 1.3 hours |
| 13 | $4.88 - 9.77$ km | 10.8 hours | | 51.1 minutes |
| 14 | $2.44 - 4.88$ km | 7.19 hours | | 32.1 minutes |
| 15 | $1.22 - 2.44$ km | 4.89 hours | | 20.3 minutes |
| 16 | $610$ m $- 1.22$ km | 2.62 hours | | 13.0 minutes |
| 17 | $305 - 610$ m | 1.88 hours | | 8.4 minutes |
| 18 | $153 - 305$ m | 1.35 hours | | 5.7 minutes |
| 19 | $76.2 - 153$ m | 58.0 minutes | | 4.0 minutes |
| 20 | $38.1 - 76.2$ m | 47.0 minutes | | 3.1 minutes |
| 21 | $19.1 - 38.1$ m | 41.1 minutes | | 2.9 minutes |

Table 3.2: Dimensionalised error saturation times $t_K$ for various length scales $K$, computed using Lorenz's model for the same hybrid energy spectrum and initial error as in Figure 3.9(a). The corresponding values for a hybrid spectrum reported by Sun & Zhang (2020) and for a $k^{-\frac{5}{3}}$ spectrum by Lorenz (1969) are shown for reference. (N.B. Lorenz's values are also shown in Table 1.1, and in Figure 2.2(b) in a non-dimensionalised form.)

(their Figures 2(a) and 3). They also demonstrated that additional small-scale 'butterflies' are practically irrelevant to the error growth pattern when the initial error spectrum has a non-negligible contribution from the large scales.

Here, Durran & Gingrich (2014)'s experiments are repeated for the hybrid basic-state spectrum with $n = 21$ scales. The growth of the error spectrum is shown in Figure 3.11. In Figure 3.11(a), the initial error is confined to the largest scale, whereas in Figure 3.11(b) the initial error is distributed across all scales in a uniform manner relative to the basic-state spectrum. The error spectra have similar shapes beyond the initial time, and both figures conform nicely to Durran & Gingrich (2014)'s results.

The Žagar error-growth parameter $a(K)$ for both alternative initial conditions is seen to follow the same general pattern as the case in which the initial error is at the smallest scale (Figure 3.12). In particular, the exponential growth of $a$ from $K = 11$
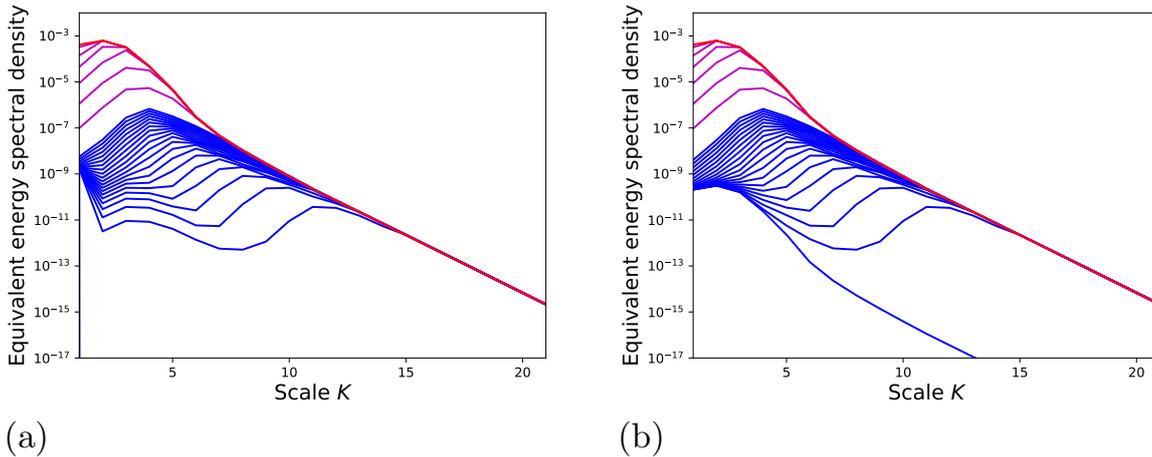
Figure 3.11: As in Figure 3.9(a), but for the following initial conditions for $\mathbf{Z}$: (a) $Z_1 = 5 \times 10^{-7} \times \sum_{L=1}^{n} X_L$ and $Z_K = 0$ for all other $K$; (b) $Z_K = 5 \times 10^{-7} \times X_K$ for all $K$.

and the sluggish variation at smaller $K$ still hold. Indeed, differences in $a(K)$ across the three cases are practically invisible for all $K \leq 14$. Beyond $K = 14$, the curves for the large-scale and proportional initial errors remain nearly identical to each other but are distinct from the curve for the small-scale initial error by a small margin. The overall excellent agreement across the three initial error profiles therefore extends Durran & Gingrich (2014)'s conclusion — that "the loss of predictability generated by initial errors of small but fixed absolute magnitude is essentially independent of their spatial scale" — to the hybrid spectrum. Yet the comparison also shows that the inferences obtained from our version of Lorenz's Experiment A (on the quickening of error growth beyond $K = 11$ and therefore the need to fully resolve this scale) are robust to initial error distributions.

## 3.5   Summary

Building on Judt (2018)'s study which shows that model-world errors in a convection-permitting global NWP model demonstrate mixed characteristics of error growth under a hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum, we examined in this chapter the sensitivity of error growth properties to the model resolution or, in other words, to the extent to which the $k^{-\frac{5}{3}}$ mesoscale range is explicitly resolved. This was done in a 2D barotropic vorticity model. The use of simple models for casting light on error growth and predictability properties in the real world is justified as long as the canonical hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ energy spectrum (Nastrom & Gage 1985) is well-modelled, since these properties are largely determined
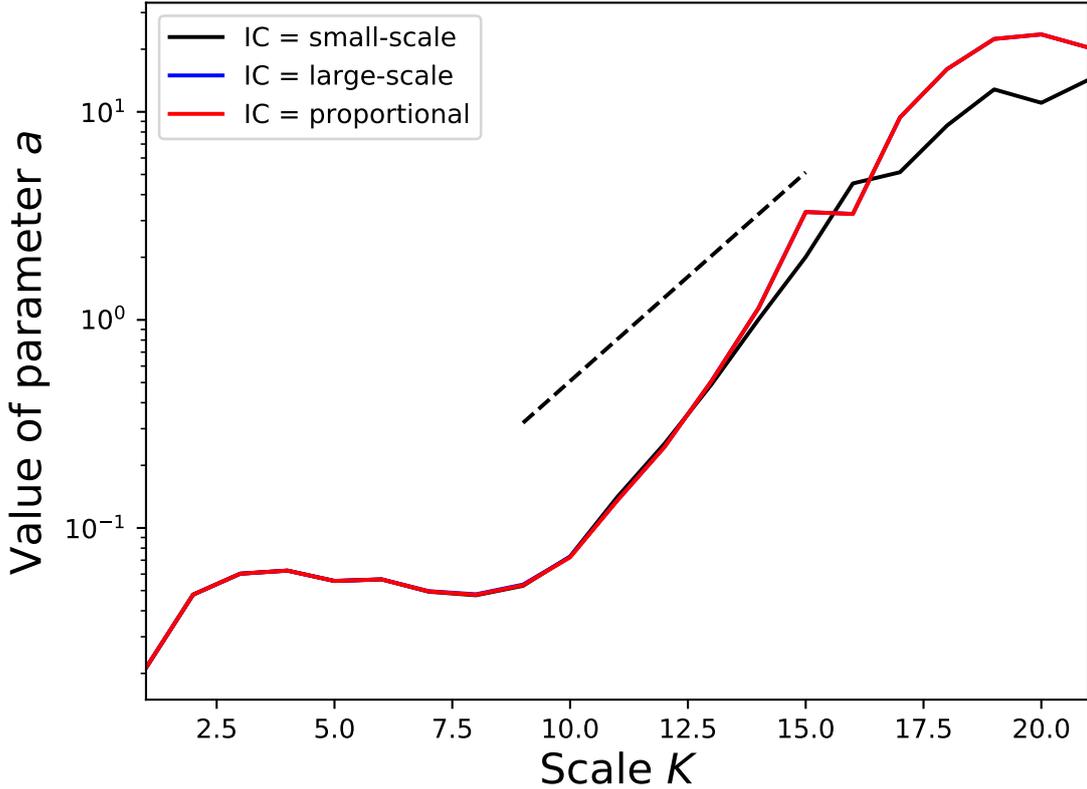
Figure 3.12: As in Figure 3.10, but for various initial conditions for the error with the $n = 21$ resolution only. The black curve is identical to the black curve in Figure 3.10. The red and blue curves, which are essentially indistinguishable from each other, correspond to the cases in which the initial condition of the same magnitude is moved to $K = 1$ (red) and redistributed as a uniform fraction of the background spectrum (blue), respectively.

by the shape of the spectrum (Rotunno & Snyder 2008).

Results from identical-twin perturbation experiments with the 2D barotropic vorticity model at a range of resolutions (Section 3.1) show that a stage-dependent peak in the error energy spectrum begins to emerge as the model resolution increases from $k_t = 256$ (where there is essentially no room for the $k^{-\frac{5}{3}}$ range) to $k_t = 2048$ (where the mesoscale range is substantially resolved). Under the hybrid spectrum, the error spectrum initially peaks at the small scales until the $k^{-\frac{5}{3}}$ range becomes saturated, then a synoptic-scale peak characteristic of error growth under a $k^{-3}$ spectrum starts to appear. These observations echo Judt (2018)'s findings, and confirm that the 2D barotropic vorticity equation can mimic the essential aspects of this process.

The dependence of error growth rate on spatial scale is used to quantitatively characterise the predictability of the system. A measure of such rate is the parameter $a$ of

the parametric error growth model of Žagar et al. (2017) (Section 3.2). By fitting the error energy data obtained from the perturbation experiments to this parametric model, it is shown that the error indeed grows faster as the spatial scale decreases, thereby providing a hint of limited predictability. This is particularly evident in the $k^{-\frac{5}{3}}$ range. However, the increase in the growth rate as the spatial scale decreases falls well short of the theoretical estimate, thus indicating that the error behaviour has not reached the asymptotic regime pertaining to this mesoscale range.

The model of Lorenz (1969), which is also based on the 2D barotropic vorticity equation, is used to investigate the asymptotic behaviour (Section 3.3). At a modest computational cost, Lorenz's model successfully captures the important characteristics of error growth, thus enabling ultra-high-resolution simulations for estimating growth patterns in the continuum. It is found that under the hybrid spectrum, the fast upscale cascade of error energy characteristic of limited predictability becomes unambiguously visible only beyond $k = 2048 = 2^{11}$ (19.5 kilometres), more than a decade of wavenumbers beyond the spectral break between the synoptic-scale and mesoscale ranges. Until then, the synoptic-scale range suppresses mesoscale error growth.

Applying these results to NWP would mean that models have to fully resolve the dynamics at the scale of the typical grid resolution of today's global ensembles ($\sim 20$ kilometres) in order for the fast mesoscale uncertainty growth to be accurately captured within the model. Based on Skamarock (2004), this would suggest a grid resolution seven times finer than typical of today, i.e. on the order of a few kilometres, after accounting for the need for a dissipation range. Pushing NWP models to such a resolution can be anticipated to provide a more realistic description of small-scale error growth and thus of the uncertainty in the forecast, even when the initial errors are not confined to the smallest scales (Section 3.4). Yet, we recognise that developing stochastic parametrisations for processes on the $O(1)$-kilometre scale (e.g. cloud processes) may also achieve the same purpose. It should also be noted that realistic initial error profiles have typically far greater amplitudes than those considered in this chapter, whose focus is on predictability properties in the limiting case.

Judt (2020) suggests that the canonical hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ spectrum, which has

been assumed here throughout, is restricted to the mid-latitude upper troposphere only. The applicability of these results to other parts of the atmosphere, or indeed to the atmosphere as a whole, remains a topic of further research.

# 4 Forecast verification: relating deterministic and probabilistic metrics

Lorenz's model is often said to be the prototype model for 'deterministic' predictability. Indeed, we have compared its predictions to results related to the error energy obtained in the deterministic, identical-twin numerical experiments in the last chapter. However, as we have seen earlier in Chapter 2, the so-called error energy in Lorenz's model is actually a measure of variance taken over an ensemble of error velocity fields (Equations 2.9–2.13). While the use of the ensemble variance, or equivalently the mean squared departure from the ensemble mean, allows us to view predictability from a probabilistic perspective, it is nevertheless only one of the many ways of assessing probabilistic predictions. In this chapter, we will present another verification metric, the CRPS, which is often used by the NWP community and which condenses information associated with the full forecast distribution (relative to the observed value) into a single number. Its relation with the ensemble variance will be derived, as an example to illustrate that verification metrics can be related under certain conditions. Finally, the CRPS-variance relation will be tested on ensemble simulations of idealised 2D turbulence, and deviations from the derived relation will be discussed.

## 4.1 Evaluating forecasts

### 4.1.1 Proper Scoring rules

Gneiting et al. (2007) defined the maxim of probabilistic forecasting as "maximizing the sharpness of the predictive distributions subject to calibration". In other words, probabilistic forecasting has two goals. The basic goal is to make reliable predictions: the weather event should materialise according to the same distribution as it is forecast. On top of this, forecast systems should strive to predict distributions whose variances are small, so that they may contain more information useful for their users. In the context of operational NWP, the practical need of monitoring forecast performance and making decisions about model upgrades gives rise to a range of scoring rules, functions which condense information contained in forecast distributions into scalar values. It is imperative that these scoring rules respect the goals of probabilistic forecasting: not only should they be optimised whenever the correct distribution is predicted, they should also reward the prediction whose distribution is sharper when two reliable predictions are concerned. This sub-section provides a few preliminaries related to the former of these two properties, which was discussed in depth by Gneiting & Raftery (2007). Scores which fulfil this property are known as proper scores.

Mathematically, let $P$ denote the predictive distribution of a scalar random variable $U$ which materialises at value $u$. A scoring rule $S(P, u)$ is a function of the predictive distribution and the verification value. If the latter follows some distribution $Q$, in which case it is also interpreted as a random variable, then the average score over many predictions with distribution $P$ can be denoted by $S(P, Q) := \mathbb{E}_Q[S(P, u)]$, with the second argument of the function $S(\cdot, \cdot)$ now being a distribution instead of a scalar value[1]. Assuming without loss of generality that scores are negative-oriented so that forecasts with lower scores are better, a score $S$ is said to be proper over a class $\mathcal{C}$ of distributions if

$$S(Q, Q) \leq S(P, Q) \qquad \forall P, Q \in \mathcal{C}, \tag{4.1}$$

---

[1] Without ambiguity, $S(\cdot, \cdot)$ can mean either the score for an individual prediction or the expected score over many predictions, depending on the second argument being a scalar variable or a distribution.

i.e. for any verifying distribution in $\mathcal{C}$, the minimal score is attained when the predictive distribution (which is also assumed to be in $\mathcal{C}$) agrees with it.

Note that the definition of proper scores does not require that the score associated with the rightly predicted distribution be its unique minimum. Scores which fulfil this extra requirement — that Condition 4.1 holds only when $P = Q$ (up to a set of measure zero) — are known to be strictly proper.

Two scoring rules popular among the NWP community will be discussed in the remainder of this section in light of their properness: the RMSE, which is closely related to the ensemble variance used in Lorenz's model, and the CRPS. This will become helpful for deriving a CRPS-RMSE relationship in the next section.

### 4.1.2 Root-mean-square error

As its name suggests, the RMSE is the square-root of the MSE. The latter is defined as

$$MSE(P, u) := \mathbb{E}_P \left[ (U - u)^2 \right] \tag{4.2}$$

for an outcome $u \in \mathbb{R}$ and a distribution $P$ of its forecast $U$. Mathematically speaking, this is the MSE of $u$ as an estimator of the ensemble mean, although it may seem counter-intuitive in a forecasting context. Nevertheless, the well-known bias-variance decomposition of MSE applies:

$$
\begin{aligned}
MSE(P, u) &= \mathbb{E}_P \left[ (\mathbb{E}_P \left[ U \right] - u + U - \mathbb{E}_P \left[ U \right])^2 \right] \\
&= (\mathbb{E}_P \left[ U \right] - u)^2 + \mathbb{E}_P \left[ (U - \mathbb{E}_P \left[ U \right])^2 \right] \\
&= (\mu_P - u)^2 + \sigma_P^2,
\end{aligned} \tag{4.3}
$$

where $\mu_P$ and $\sigma_P$ are respectively the mean and the standard deviation of $P$. Assuming that the verifying distribution $Q$ for $u$ has mean $\mu_Q$ and standard deviation $\sigma_Q$, the expected score $MSE(P, Q)$ is

$$
\begin{aligned}
MSE(P, Q) &= \mathbb{E}_Q \left[ (\mu_P - u)^2 + \sigma_P^2 \right] \\
&= \sigma_Q^2 + (\mu_P - \mu_Q)^2 + \sigma_P^2.
\end{aligned} \tag{4.4}
$$

The last equality can be established by observing that $\mathbb{E}_Q\left[(\mu_P - u)^2\right]$ is the MSE of $\mu_P$ as an estimator of $u$, whence the same bias-variance decomposition applies. From Equation 4.4, it follows that

$$
\begin{aligned}
RMSE(P,Q) &= \sqrt{MSE(P,Q)} \\
&= \sqrt{\sigma_Q^2 + (\mu_P - \mu_Q)^2 + \sigma_P^2} \\
&= \sigma_Q\sqrt{1 + b^2 + r^2},
\end{aligned}
\tag{4.5}
$$

where

$$
b := \frac{\mu_P - \mu_Q}{\sigma_Q}
\tag{4.6}
$$

is the relative bias and

$$
r := \frac{\sigma_P}{\sigma_Q}
\tag{4.7}
$$

is the ratio of standard deviations, or simply the spread ratio.

Note that we have not defined $RMSE(P, u)$. Should it be defined by taking the square-root of Equation 4.2, then the $RMSE(P, Q)$ defined in Equation 4.5 would generally not be equal to $\mathbb{E}_Q\left[RMSE(P, u)\right]$. Hence, strictly speaking, the RMSE does not fit into the framework of scoring rules described in the last sub-section. It is simply a convenient proxy for the scoring rule $MSE(P, u)$, since it has the same physical dimensions as the variable $u$ of interest. Given that the RMSE relates with the MSE bijectively and monotonically, we may nevertheless apply the concepts of proper scoring rules to the RMSE, bearing in mind that in this sense the two quantities are synonymous.

The RMSE is not a proper score over any non-trivial class of distributions, since $RMSE(Q, Q) = \sqrt{2}\sigma_Q$ is not the global minimum over $\mu_P, \mu_Q \in \mathbb{R}$ and $\sigma_P, \sigma_Q \in \mathbb{R}^+$. One could have achieved a better score by, for example, making an unbiased ($\mu_P = \mu_Q$) yet under-spread ($\sigma_P < \sigma_Q$) prediction. Figure 4.1, which shows $\frac{1}{\sigma_Q}RMSE(P,Q)$ as a function of $b$ and $r$, illustrates this point. A possible modification which makes the score proper while maintaining the property that it depends only on the first two moments of the predictive distribution is

$$
S(P, u) = \left(\frac{\mu_P - u}{\sigma_P}\right)^2 + \log\left(\sigma_P\right)^2,
\tag{4.8}
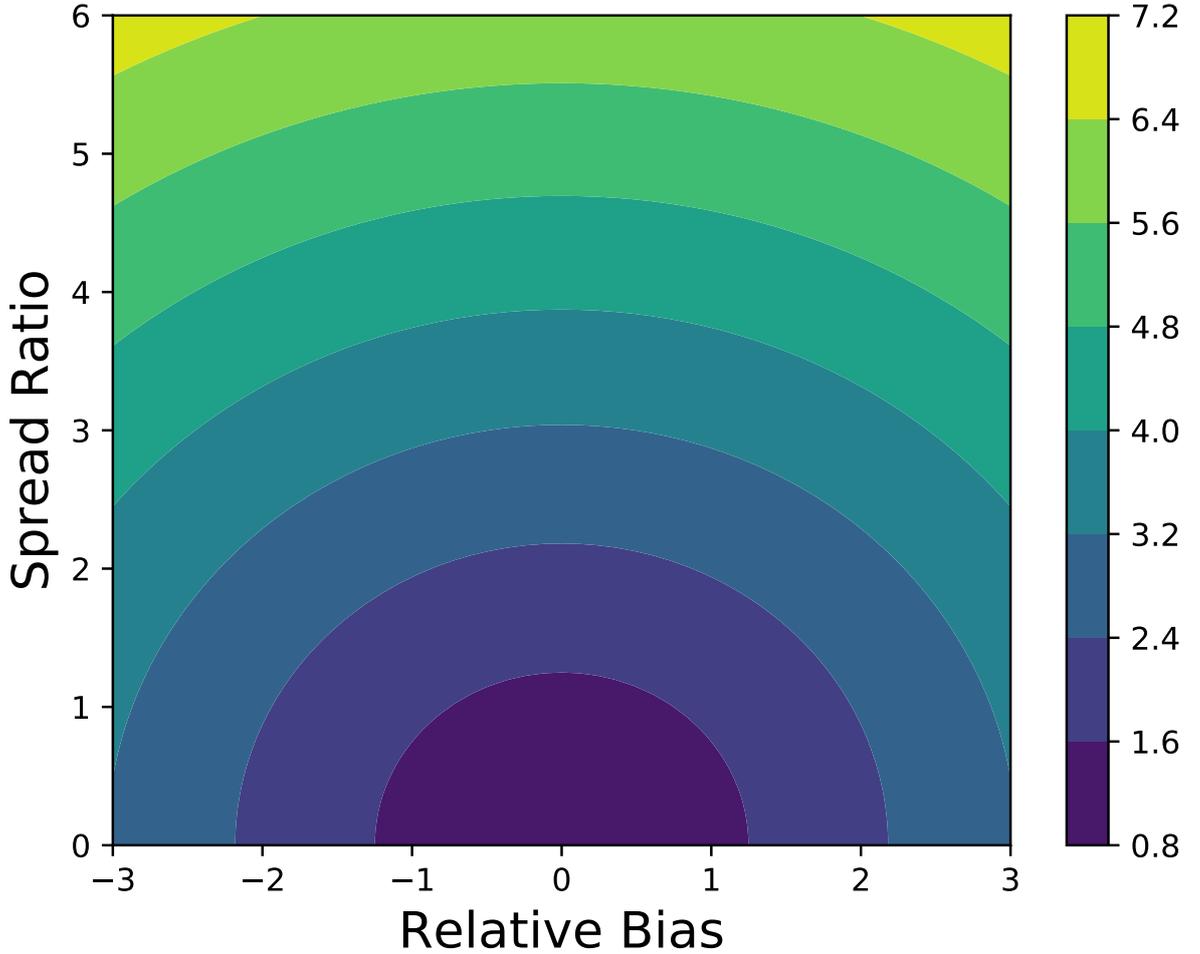$$

see Gneiting & Raftery (2007).

Figure 4.1: $\frac{1}{\sigma_Q} RMSE(P, Q)$ as a function of relative bias $b = \frac{\mu_P - \mu_Q}{\sigma_Q}$ and spread ratio $r = \frac{\sigma_P}{\sigma_Q}$.

The RMSE discussed here should not be confused with the RMSE of the ensemble mean, which is often shortened as 'the RMSE' in the language of operational NWP centres. It is a proxy score for the MSE of the ensemble mean, defined as

$$MSE_{\mathrm{mean}}(P, u) \coloneqq (\mathbb{E}_P[U] - u)^2 = (\mu_P - u)^2. \tag{4.9}$$

Considering the ensemble mean as a deterministic forecast in itself, $MSE_{\mathrm{mean}}(P, u)$ and therefore its associated RMSE can be seen as a deterministic score. Compared to Equation 4.3, $MSE_{\mathrm{mean}}(P, u)$ lacks the contribution from the ensemble variance $\sigma_P^2$. The alternative formulation thus makes the RMSE independent from the standard deviation ('spread') of the ensemble. This enables convenient comparison between the RMSE and the ensemble spread, which is routinely done in NWP operations.

### 4.1.3 Continuous Ranked Probability Score

The CRPS is a widely used metric that evaluates the full ensemble distribution of a continuous scalar variable and penalises unsharp distributions. It is the integral of the squared difference between the CDFs of the forecast and the verification:

$$CRPS(P, u) = \int_{-\infty}^{\infty} (F(x) - H_u(x))^2 \, \mathrm{d}x, \tag{4.10}$$

where $F(x)$ is the CDF of $P$ and $H_u(x)$ is the Heaviside function at the verification value $u$.

An equivalent expression for the CRPS, often known as the 'kernel representation' (Gneiting & Raftery 2007), is available for distributions $P$ whose first moments are finite:

$$CRPS(P, u) = \mathbb{E}_P\left[|U - u|\right] - \frac{1}{2}\mathbb{E}_P\left[|U - U'|\right], \tag{4.11}$$

where $U$ and $U'$ are independent random variables drawn from the distribution $P$ (Gneiting & Raftery 2007). A proof of equivalence is provided in Appendix A.1.

Gneiting & Raftery (2007) noted that the CRPS is a strictly proper score over a very general class of distributions, namely the class of Borel probability measures whose first moments are finite. For the special case of normal distributions $P = \mathcal{N}(\mu_P, \sigma_P^2)$, an explicit formula for $CRPS(P, u)$ is available:

$$CRPS(P, u) = \frac{\sigma_P}{\sqrt{\pi}}\left(-1 + \sqrt{\pi}\frac{u - \mu_P}{\sigma_P}\operatorname{erf}\left(\frac{u - \mu_P}{\sqrt{2}\sigma_P}\right) + \sqrt{2}\exp\left(-\frac{1}{2}\left(\frac{u - \mu_P}{\sigma_P}\right)^2\right)\right), \tag{4.12}$$

where $\varphi(x) := \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)$ is the probability density function (PDF) of a standard normal random variable and $\Phi(x) := \int_{-\infty}^{x} \varphi(x') \, \mathrm{d}x'$ is its CDF. Appendix A.2 provides a derivation of such.

Equation 4.12 can be integrated over a normal $\mathcal{N}(\mu_Q, \sigma_Q^2)$ kernel to yield a

formula for the expected score $CRPS(P, Q)$:

$$CRPS(P, Q) = \int_{-\infty}^{\infty} CRPS(P, u)\, \varphi\left(\frac{u - \mu_Q}{\sigma_Q}\right) \frac{1}{\sigma_Q}\, \mathrm{d}u$$

$$= \int_{-\infty}^{\infty} \sigma_Q \frac{r^2}{\sqrt{\pi}} \left(-1 + \sqrt{\pi} x \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) + \sqrt{2} \exp\left(-\frac{1}{2}x^2\right)\right) \varphi\left(rx + b\right)\, \mathrm{d}x,$$

$$(4.13)$$

where the substitution $x = \frac{u - \mu_P}{\sigma_P}$ is used, and $r$ and $b$ are as in Equations 4.6 and 4.7. Appendix A.3 demonstrates that the integral can be simplified as

$$CRPS(P, Q) = \frac{\sigma_Q}{\sqrt{\pi}} f(b, r) \tag{4.14}$$

where

$$f(b, r) = -r + \sqrt{2(1 + r^2)} \exp\left(-\frac{b^2}{2(1 + r^2)}\right) + \sqrt{\pi}\, b \operatorname{erf}\left(\frac{b}{\sqrt{2(1 + r^2)}}\right). \tag{4.15}$$

Note that provided the verifying distribution $Q$ is fixed, the qualitative properties of $CRPS(P, Q)$ are fully determined by the function $f(b, r)$. This formula for $CRPS(P, Q)$ agrees exactly with the one obtained by Leutbecher & Haiden (2020), who used the kernel representation of the CRPS (Equation 4.11) as the starting point of their derivation.

To verify that the CRPS is indeed a strictly proper score over the class of normal distributions, consider

$$\frac{\partial f}{\partial b} = \sqrt{2(1 + r^2)} \left(-\frac{b}{1 + r^2}\right) \exp\left(-\frac{b^2}{2(1 + r^2)}\right) + \sqrt{\pi} \operatorname{erf}\left(\frac{b}{\sqrt{2(1 + r^2)}}\right)$$

$$+ 2b \exp\left(-\frac{b^2}{2(1 + r^2)}\right) \frac{1}{\sqrt{2(1 + r^2)}} \tag{4.16}$$

$$= \sqrt{\pi} \operatorname{erf}\left(\frac{b}{\sqrt{2(1 + r^2)}}\right)$$

which is zero if and only if $b = 0$. Since

$$\frac{\partial^2 f}{\partial b^2} = \sqrt{\frac{2}{1 + r^2}} \exp\left(-\frac{b^2}{2(1 + r^2)}\right) > 0, \tag{4.17}$$

$f(b, r)$ attains a unique minimum at $b = 0$ for any given $r$. This suggests that on average, and provided that distributions are normal, the best CRPS for any given spread ratio is

attained when the ensemble is unbiased. Now, substituting $b = 0$ into Equation 4.15 and differentiating with respect to $r$, we have

$$\frac{\partial f}{\partial r}(0, r) = -1 + \sqrt{\frac{2}{1 + r^2}}\, r \tag{4.18}$$

which implies $\frac{\partial f}{\partial r}(0, r) = 0$ if and only if $r = 1$ (note that $r$ cannot be negative, since it is a ratio of standard deviations). With

$$\frac{\partial^2 f}{\partial r^2}(0, r) = \sqrt{\frac{2}{1 + r^2}} - \sqrt{2}r^2 \left(1 + r^2\right)^{-\frac{3}{2}} = \sqrt{2}\left(1 + r^2\right)^{-\frac{3}{2}} \tag{4.19}$$

so that $\frac{\partial^2 f}{\partial r^2}(0, 1) = \frac{1}{2} > 0$, we can now conclude that $(b, r) = (0, 1)$ is the unique global minimum of $f(b, r)$. Therefore, for any verifying distribution $Q$ that is normal, $CRPS(P, Q)$ is uniquely minimised when $P = Q$, assuming that $P$ is also normal. This makes the CRPS a strictly proper score over the class of normal distributions, which is graphically confirmed in Figure 4.2.

## 4.2   Derivation of the CRPS-RMSE relationship

So far we have seen the basic mathematical properties of the CRPS and the RMSE. Since the former is proper while the latter is improper, it is generally impossible to draw a one-to-one correspondence between the two. However, if we restrict our attention to reliable predictions ($P = Q$, or in other words $(b, r) = (0, 1)$) of normally distributed variables, then the scores reduce to

$$CRPS(P, P) = \frac{1}{\sqrt{\pi}}\sigma_P \tag{4.20}$$

and

$$RMSE(P, P) = \sqrt{2}\sigma_P, \tag{4.21}$$

so that there exists a bijective relationship between the two:

$$CRPS(P, P) = \frac{1}{\sqrt{2\pi}}RMSE(P, P). \tag{4.22}$$

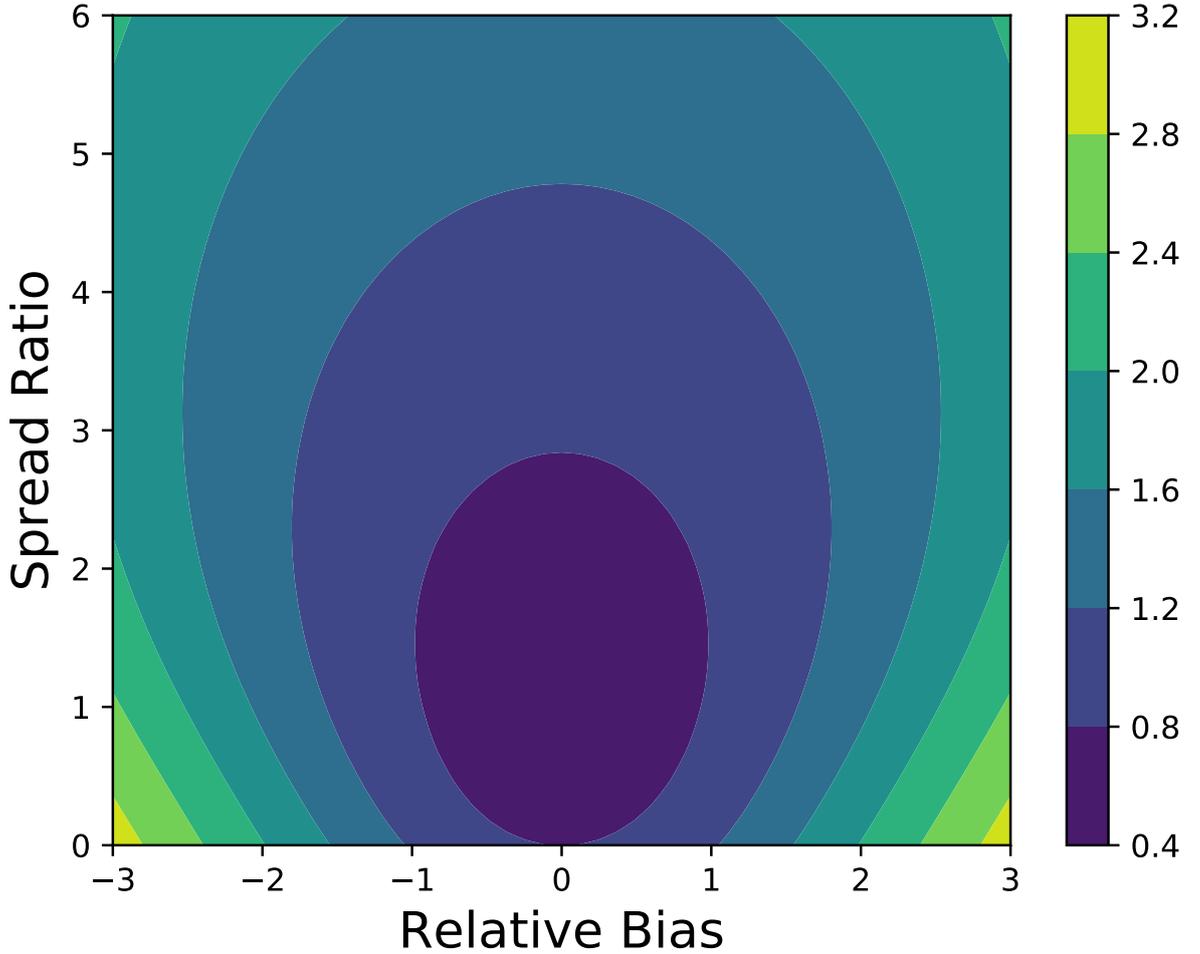What Equation 4.22 suggests is that on average, the CRPS and the RMSE are

Figure 4.2: $\frac{1}{\sigma_Q}CRPS(P,Q)$ as a function of relative bias $b = \frac{\mu_P - \mu_Q}{\sigma_Q}$ and spread ratio $r = \frac{\sigma_P}{\sigma_Q}$.

related through a multiplicative factor of $\frac{1}{\sqrt{2\pi}}$ as far as reliable predictions of normally distributed scalar variables are concerned. The average, as defined in Sub-section 4.1.1, refers to aggregation over a large number of reliable predictions which share the same distribution $P$. The condition — as to what can be included in the average so that Equation 4.22 holds — can be slightly relaxed here: since the scores for individual events are invariant to translations[2], one can achieve the same CRPS-RMSE relationship by aggregating reliable predictions of scalar variables which are normally distributed and share a certain variance $\sigma_P^2$, but which need not share the same mean $\mu_P$. Nevertheless, standard NWP practice is to aggregate these scores across dimensions defined *a priori* such as grid points and forecast start-dates, rather than by variance of the predictive distribution. How can Equation 4.22 be modified to accommodate this?

---

[2]These scores do not depend on the mean $\mu_P$ of the predictive distribution alone but only through its difference with the verification value $u$, see Equations 4.3 and 4.10.

It is important to bear in mind that in the notation $S(P, Q)$ for a given score $S$, there is an implied conditioning on the predictive distribution being $P$, since $S(P, Q)$ is the average of $S(P, u)$ over many $P$-distributed predictions. $Q$ in this notation refers to the distribution of the verification value $u$, but it is also conditional upon the predictive distribution being $P$. To derive a formula for an aggregated score, it is therefore necessary to include information about the heteroscedasticity of $P$ — the relative frequency of occurrence of different predictive distributions $P$. If we assume that $P$ is from a certain family of parametric distributions, then such heteroscedasticity can be interpreted as a (joint) meta-distribution $\Theta$ of the parameter(s) $\theta$ of $P$. Then the aggregated score $S^*$ can be written as

$$S^* = \mathbb{E}_\Theta \left[ S(P, Q) \right] = \int_\Omega S(P, Q) \lambda(\theta) \, d\theta, \tag{4.23}$$

where $\lambda$ is the PDF of the meta-distribution, and $\Omega$ is the parameter space to which $\theta$ belongs.

Generally, $S(P, Q)$ depends on $\theta$ as well as the parameters of $Q$, although the conditionality of $Q$ on $P$ suggests that the parameters of $Q$ also depend on $\theta$. Without further information on how $\theta$ and $Q$ relate, it is impossible to simplify Equation 4.23, let alone derive a relationship between various aggregated scores. However, in the case of reliable predictions, the information is supplied by the equality $P = Q$. We thus have

$$S^* = \mathbb{E}_\Theta \left[ S(P, P) \right] = \int_\Omega S(P, P) \lambda(\theta) \, d\theta. \tag{4.24}$$

Substituting $CRPS$ for $S$ and using Equation 4.20, this gives rise to

$$CRPS^* = \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sqrt{\pi}} \sigma_P \lambda(\mu_P, \sigma_P) \, d\mu_P \, d\sigma_P = \frac{1}{\sqrt{\pi}} \mathbb{E}_\Sigma \left[ \sigma_P \right], \tag{4.25}$$

where it has been assumed that $\mu_P$ and $\sigma_P$ in the meta-distribution are independent, and where $\Sigma$ denotes the marginal distribution of the standard deviation $\sigma_P$. Bearing in mind that the RMSE is a convenient proxy for the score MSE so that the expectation (in $\Theta$) is only taken after Equation 4.21 is squared, we similarly have

$$RMSE^* = \sqrt{\mathbb{E}_\Theta \left[ RMSE(P, P)^2 \right]} = \sqrt{\int_0^\infty \int_{-\infty}^\infty 2\sigma_P^2 \lambda(\mu_P, \sigma_P) \, d\mu_P \, d\sigma_P} = \sqrt{2\mathbb{E}_\Sigma \left[ \sigma_P^2 \right]}$$
$$\tag{4.26}$$

under the same assumption of independence.

Equations 4.25 and 4.26 thus provide expressions for the CRPS and the RMSE aggregated under heteroscedastic conditions, where $P$'s parameters can vary from grid point to grid point, and from one forecast start-date to another. These expressions assume the reliability of the ensemble, the normality of its distribution, and the independence of ensemble mean and standard deviation in the frequency distribution of the parameters. Combining these expressions and denoting

$$h := \frac{\mathrm{Var}_\Sigma\left[\sigma_P\right]}{\left(\mathbb{E}_\Sigma\left[\sigma_P\right]\right)^2} \tag{4.27}$$

for the relative heteroscedasticity, we have

$$\frac{CRPS^*}{RMSE^*} = \frac{1}{\sqrt{2\pi}}\frac{\mathbb{E}_\Sigma\left[\sigma_P\right]}{\sqrt{\mathbb{E}_\Sigma\left[\sigma_P^2\right]}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1+h}}, \tag{4.28}$$

whence the inequality

$$\frac{CRPS^*}{RMSE^*} \leq \frac{1}{\sqrt{2\pi}} \tag{4.29}$$

holds. In the limit where the standard deviation $\sigma_P$ is homoscedastic, i.e. $\mathrm{Var}_\Sigma\left[\sigma_P\right] \to 0$, the inequality becomes an equality, and Equation 4.22 is recovered.

## 4.3 Verifying the relationship on an idealised 2D turbulence model

The CRPS-RMSE relationship for reliable predictions of normally distributed random variables (Equation 4.28) is numerically tested in a perfect-model experiment involving idealised 2D turbulence. The doubly periodic model discussed in Section 3.1 (including the details of the stochastic forcing and the numerical dissipation) is reused, with a resolution of $k_t = 1024$ so that the mesoscale $k^{-\frac{5}{3}}$ spectrum may be reasonably well-resolved (Figure 3.1). Physically, it corresponds to a grid with $2k_t = 2048$ grid points in each direction.

### 4.3.1 Experimental design

A control integration of Equation 3.1 is considered as the truth. When the turbulence is fully developed and reaches a statistically stationary state, $M_0 = 4$ independent perturbed simulations are generated from the control, one of which is taken as the verification[3] and the remaining $M := M_0 - 1 = 3$ as ensemble members. All $M_0$ of them are perturbed in the same way, using the procedure laid out in Sub-section 3.1.2, but with different random seeds. The initial perturbation is introduced at a single small-scale wavenumber $k_p = 512 = \frac{1}{2}k_t$ outside the small-scale forcing and dissipation ranges. The control integration is not used in any way other than to generate the perturbed simulations, which is analogous to weather forecasting as the true atmospheric state is unknown and therefore cannot be used in computing verification metrics. The perturbed simulations are integrated for a fixed time-period of $T = 150$ non-dimensional units, allowing the error to almost fully saturate by the end of it.

The experiment is repeated for $N = 30$ 'start-dates'. This can be thought of as $N_1 = 5$ years, among which the control integrations are fully independent, and $N_2 = \frac{N}{N_1} = 6$ start-dates per year initialised at intervals of $0.1T$.

The choice of a relatively small $M$ and large $N$ is inspired by Leutbecher (2019). This paper suggests that if the CRPS is adjusted using

$$CRPS_\infty^* := \frac{M}{M + 1} CRPS^* \tag{4.30}$$

to remove the effects of the ensemble size being finite, then a reduction in the number of ensemble members used for numerical experimentation returns more robust results than a reduction in the number of start-dates, provided that the constraints in computational cost are similar.

The experimental design guarantees a reliable ensemble, since the verification is statistically indistinguishable from the remaining $M$ simulations generated from the control integration. As such, Equation 4.28 is expected to hold subject to $P$ being a

---

[3]The verification is known as the analysis in NWP terminology.

normal distribution, and $\mu_P$ and $\sigma_P$ being independent.

The scalar variables of interest chosen for this study are the velocity components $u$ and $v$. For each start-date and grid point, the CRPS and the MSE are computed for both velocity components in physical space. The computation for the CRPS is performed using the algorithm set out by Hersbach (2000). These metrics are then aggregated over $\Lambda := G \times S \times D$, where $G$ represents the set of $2048^2$ grid points, $S$ the 30 start-dates and $D$ the two canonical directions ($u$ and $v$), but remain as functions of the forecast lead time. Isotropy of the turbulence enables the scores for $u$ and $v$ to be combined without changing the quality of the results. Although the full dataset is included in the calculation of the scores, the results are not expected to change if we sub-sample in any of the aggregated dimensions, since the turbulence is homogeneous, isotropic and statistically stationary. In real-world weather forecasts, however, the effects of the turbulence's inherent heteroscedasticity in relation to those of inhomogeneity, anisotropy and non-stationarity could be more subtle.

When the metrics are aggregated, the quantity $\mathbb{E}_\Lambda\left[S(P, u_i)\right]$ is computed for each lead time, where $S$ can be $CRPS$ or $MSE$, and where $u_i$ represents a generic velocity component. The law of iterated expectations guarantees

$$\mathbb{E}_\Lambda\left[S(P, u_i)\right] = \mathbb{E}_\Theta\left[\mathbb{E}_P\left[S(P, u_i)\right]\right] = \mathbb{E}_\Theta\left[S(P, P)\right] = S^*, \tag{4.31}$$

the last two equalities of which result from the definition of $S$ and Equation 4.24 respectively. In this way, $CRPS^*$ and $RMSE^*$ (the square-root of $MSE^*$) can be empirically computed, which should satisfy Equation 4.28 subject to the assumptions. To account for the finite ensemble size, the aggregated CRPS is corrected by Equation 4.30 before being compared with the aggregated RMSE.

## 4.3.2 Results

For notational purposes in this sub-section, we denote the start-date by $t_0$, and write $U(t, t_0, \mathbf{x}, \mathbf{e}_1)$ for $u(t, t_0, \mathbf{x})$ and $U(t, t_0, \mathbf{x}, \mathbf{e}_2)$ for $v(t, t_0, \mathbf{x})$. A subscript $f$ attached to $U(t, t_0, \mathbf{x}, \mathbf{e}_i)$, $u(t, t_0, \mathbf{x})$ or $v(t, t_0, \mathbf{x})$ (where $i = 1$ or 2) indicates a forecast, in which case

the variable is understood to be a random variable with distribution $P$. The absence of the subscript indicates the verification, which is also interpreted as a random variable but with distribution $Q$.

Figure 4.3 illustrates the growth of the error energy spectrum. More precisely, it is the spectrum of the ensemble-mean error energy aggregated over all grid points and start-dates, i.e. the spectral decomposition of

$$\mathbb{E}_{G \times S} \left[ \mathbb{E}_P \left[ \frac{1}{2} \left( (u_f(t, t_0, \mathbf{x}) - u(t, t_0, \mathbf{x}))^2 + (v_f(t, t_0, \mathbf{x}) - v(t, t_0, \mathbf{x}))^2 \right) \right] \right]. \tag{4.32}$$

In two spatial dimensions[4] and where the ensemble is reliable ($P = Q$), this is equivalent to the spectral decomposition of

$$\begin{aligned}
&\mathbb{E}_{G \times S \times D} \left[ \mathbb{E}_P \left[ (U_f(t, t_0, \mathbf{x}, \mathbf{e}_i) - U(t, t_0, \mathbf{x}, \mathbf{e}_i))^2 \right] \right] \\
&= \mathbb{E}_\Lambda \left[ MSE(P, U) \right] \\
&= MSE^* \\
&= RMSE^{*2},
\end{aligned} \tag{4.33}$$

where Equations 4.2 and 4.31 have been used in the first two equalities respectively. As such, Figure 4.3 may also be interpreted as the evolution of the power spectrum of $RMSE^*$. Initially, the growth rate is more or less uniform in spatial scale and the error spectrum peaks at the small scale. Later, as the $k^{-\frac{5}{3}}$ range saturates, a synoptic-scale peak emerges in the $k^{-3}$ range, and the growth slows down. These observations are consistent with those reported earlier in Chapter 3.

Like $RMSE^*$, it is possible to spectrally decompose $CRPS^*_\infty$. To compute $CRPS^*_\infty$ for a wavenumber or range of wavenumbers, one simply picks out the associated waves in spectral space, transforms them to physical space, then aggregates the score over $\Lambda$ and applies Equation 4.30. Such $CRPS^*_\infty$ may be compared to $RMSE^*$ for the same wavenumber(s) using Equation 4.28. Here, the verification metrics are decomposed into the planetary scale ($k \in [1, 8]$), synoptic scale ($k \in [9, 64]$), mesoscale ($k \in [65, 512]$)

---

[4]The equivalence between Expressions 4.32 and 4.33 is not extendable to higher spatial dimensions, because it only happens in two dimensions that the factor $\frac{1}{2}$ for the kinetic energy is also the factor used to compute the average over $D$. In higher dimensions, the ensemble-mean error energy can be related to the MSE of velocity components via a constant multiplicative factor.
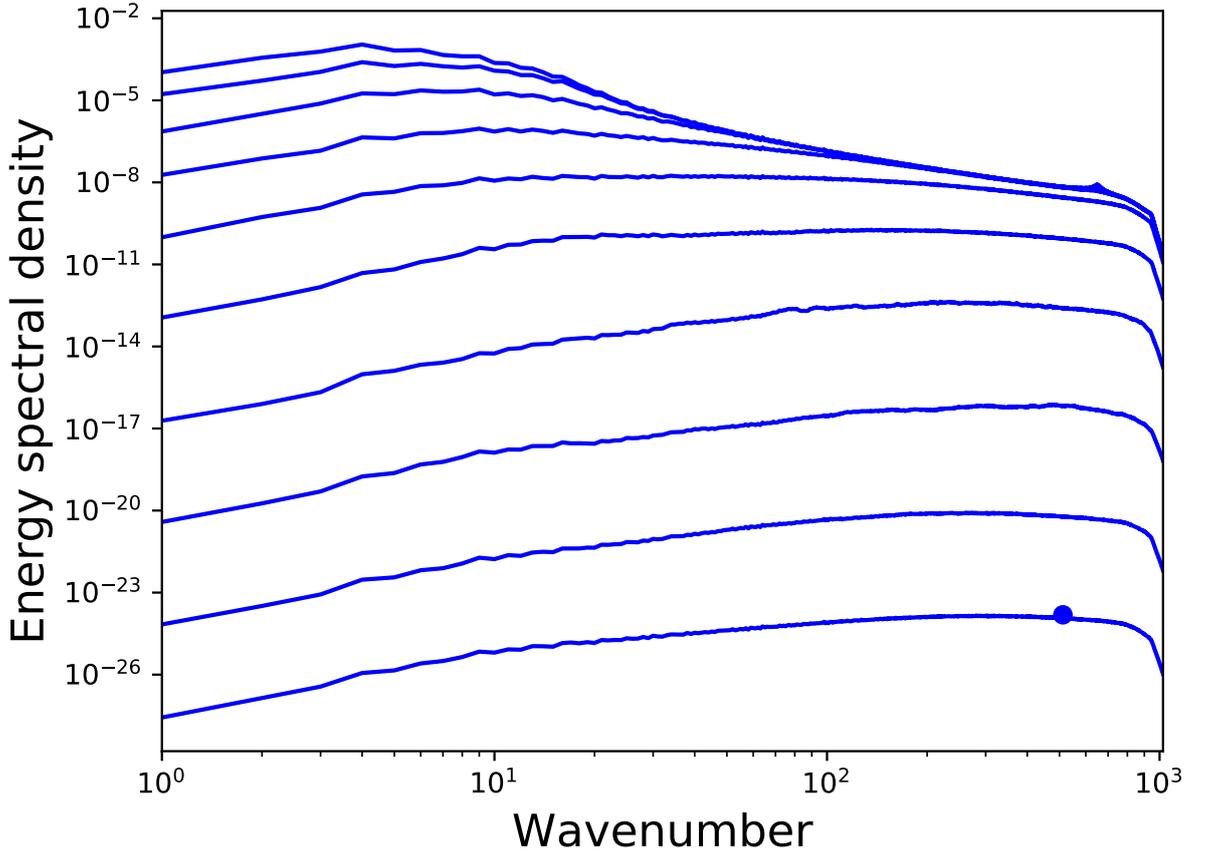
Figure 4.3: Growth of the ensemble-mean error energy spectrum, or equivalently the power spectrum of $RMSE^*$ (blue curves from bottom to top, plotted at intervals of $0.1T$), whose initial condition is indicated by the blue dot.

and sub-mesoscale ($k \in [513, 1024]$). The evolution of these metrics is shown in Figure 4.4. Generally speaking, they grow steadily (in exponential terms) by about ten orders of magnitude throughout the simulations. In Figure 4.4(a), $RMSE^*$ is normalised by $\sqrt{2\pi}$ so that, according to Equation 4.28, the curves for the CRPS and the RMSE would coincide if $P$ were normal, $\sigma_P$ were homoscedastic, and $\mu_P$ and $\sigma_P$ were independent. While the agreement appears to be reasonably good at first glance, the discrepancy between the two curves is non-trivial between $t = 30$ ($0.2T$) and $t = 90$ ($0.6T$) although it remains within an order of magnitude. To enable closer examination of the discrepancy, the ratio of the two curves is plotted (Figure 4.4(b)). Evidently, the discrepancy is stronger at the smaller scales. Beyond $t = 90$ ($0.6T$), as the mesoscale error saturates and the error growth at larger scales slows down, the CRPS and the normalised RMSE agree fairly well, especially in the meso- and sub-mesoscale.

Figure 4.5 shows these curves and their ratio for the full field without decomposition into wavebands. In addition, the evolution of the ratios for the $N = 30$ individual
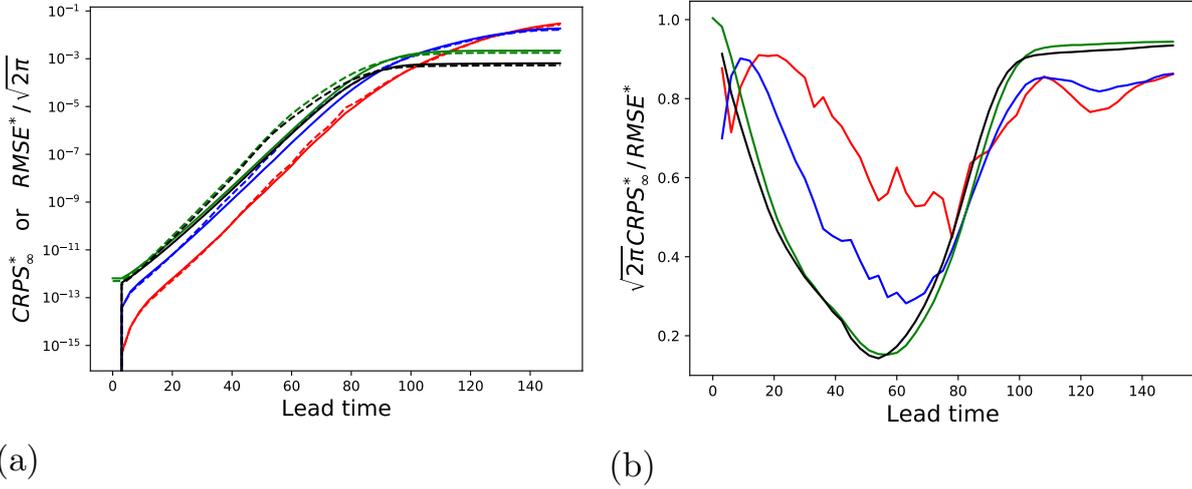
(a)                                            (b)

Figure 4.4: (a) $CRPS^*_\infty(t)$ (solid) and $\frac{1}{\sqrt{2\pi}}RMSE^*(t)$ (dashed) for the planetary (red), synoptic (blue), meso- (green) and sub-mesoscale (black), up to $T = 150$. (b) The ratio $\sqrt{2\pi}\frac{CRPS^*_\infty(t)}{RMSE^*(t)}$ between the solid and dashed curves of (a) for the respective colours.

start-dates, i.e. with $\Lambda = G \times D$ instead of $G \times S \times D$, is shown in the thin red curves. Considerable variation in this ratio across the 30 cases is seen, not only during intermediate lead times when the CRPS-RMSE discrepancy is the largest, but also towards the end of the integration when the errors at all but the few largest scales have saturated. Furthermore, the ratio for the mean (the $\Lambda = G \times S \times D$ case) consistently favours the lower end of the distribution of such ratio for individual start-dates. It is not clear why this is so.



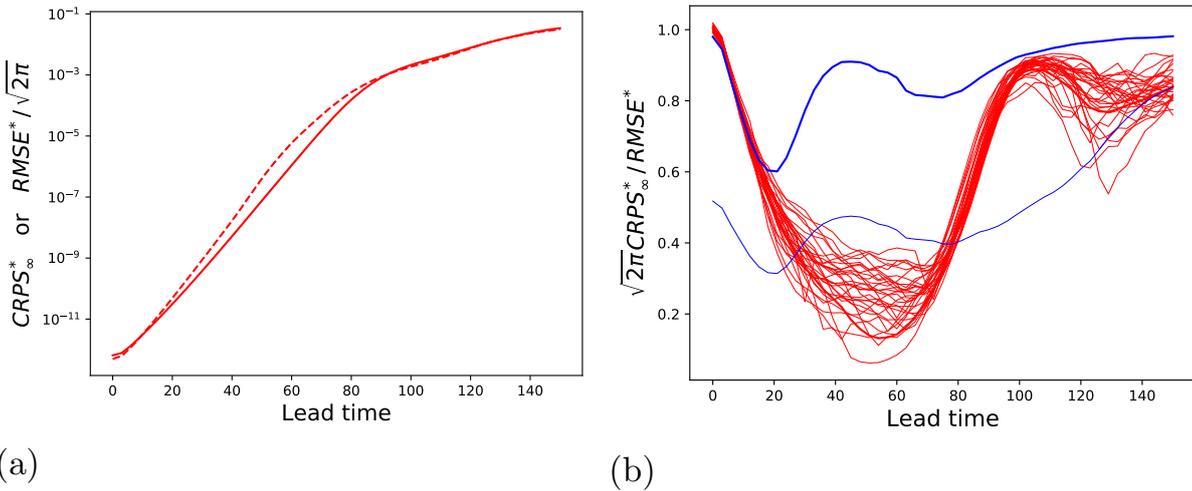(a)                                            (b)

Figure 4.5: As in Figure 4.4, but for the full field without the scale-decomposition (thick red curves). The additional thick blue curve in (b) shows $\frac{1}{\sqrt{1+h}}$ as a function of $t$ for the set of $M_0 = 51$ simulations. The additional thin curves in (b) show $\sqrt{2\pi}\frac{CRPS^*_\infty(t)}{RMSE^*(t)}$ but for $\Lambda = G \times D$ (i.e. for specific start-dates), for the $M = 3$ ensemble (red) and the $M = 50$ ensemble (blue).

According to Equation 4.28, the red curves in Figure 4.5(b) are expected to

coincide with $\frac{\mathbb{E}_\Sigma[\sigma_P]}{\sqrt{\mathbb{E}_\Sigma[\sigma_P^2]}} = \frac{1}{\sqrt{1+h}}$ if the ensemble is normally distributed, and if $\mu_P$ and $\sigma_P$ are independent. Computing this ratio involves evaluating the ensemble's standard deviation $\sigma_P$, but the sample size ($M = 3$ or $M_0 = 4$) is too small to estimate $\sigma_P$ robustly. To mitigate this, a larger set of $M_0 = 51$ simulations is run to estimate the ensemble's heteroscedasticity. This is done for only a single start-date ($N = 1$) due to limited computational resources. As shown in Figure 4.5(b), $\frac{1}{\sqrt{1+h}}$ exhibits two local minima throughout the integration, the more extreme of which corresponds to a relative heteroscedasticity of $h \approx 1.8$.

It is tempting to conclude that heteroscedasticity fully accounts for the discrepancy between $CRPS_\infty^*(t)$ and $\frac{1}{\sqrt{2\pi}}RMSE^*(t)$ up to a lead time of $0.1T$, and that the ensemble is not normally distributed thereafter. But the argument would only stand if the ratio $\sqrt{2\pi}\frac{CRPS_\infty^*(t)}{RMSE^*(t)}$ for $\Lambda = G \times D$ (i.e. for individual start-dates) were to behave similarly between the $M = 3$ and $M = 50$ ensembles, and this is not the case (the thin red curves versus the thin blue curve in Figure 4.5(b)). If the single realisation of $\sqrt{2\pi}\frac{CRPS_\infty^*(t)}{RMSE^*(t)}$ for the larger ensemble is deemed representative of an infinite-size ensemble, then we may conclude that the combined effects (the gap between the thick and thin blue curves) of the ensemble's non-normality and the possible dependence between $\mu_P$ and $\sigma_P$ are felt from the very beginning of the simulation and do not vary a lot throughout, except near the end when they gradually reduce. It seems reasonable to speculate that the $M = 3$ ensemble is too small to accurately capture and manifest such effects, thereby showing an apparent normality at small lead times.

The possible dependence between $\mu_P$ and $\sigma_P$ is investigated by computing their correlation over the large sample of $2048^2$ grid points and the two velocity components. This is done for the set of $M_0 = 51$ simulations, so only one start-date is available. Figure 4.6 shows the Pearson correlation coefficient of $\mu_P$ and $\sigma_P$ as a function of lead time. While there is some indication that correlation is being built up as the simulation progresses, it remains rather small and never exceeds 0.15 in magnitude. Yet, the apparent smallness of the correlation coefficient alone does not provide sufficient indication for conclusive statements about the observed correlation to be made, since no candidate distribution of $\mu_P$ and $\sigma_P$ is available and therefore an appropriate statistical test is lacking. Moreover, even if $\mu_P$ and $\sigma_P$ were found to be uncorrelated, it would not necessarily imply indepen-

dence. Hence major questions remain regarding the validity of the $\mu_P$-$\sigma_P$ independence assumption. The validity of the other assumption, however, can be rigorously checked, as will be discussed in the next sub-section.
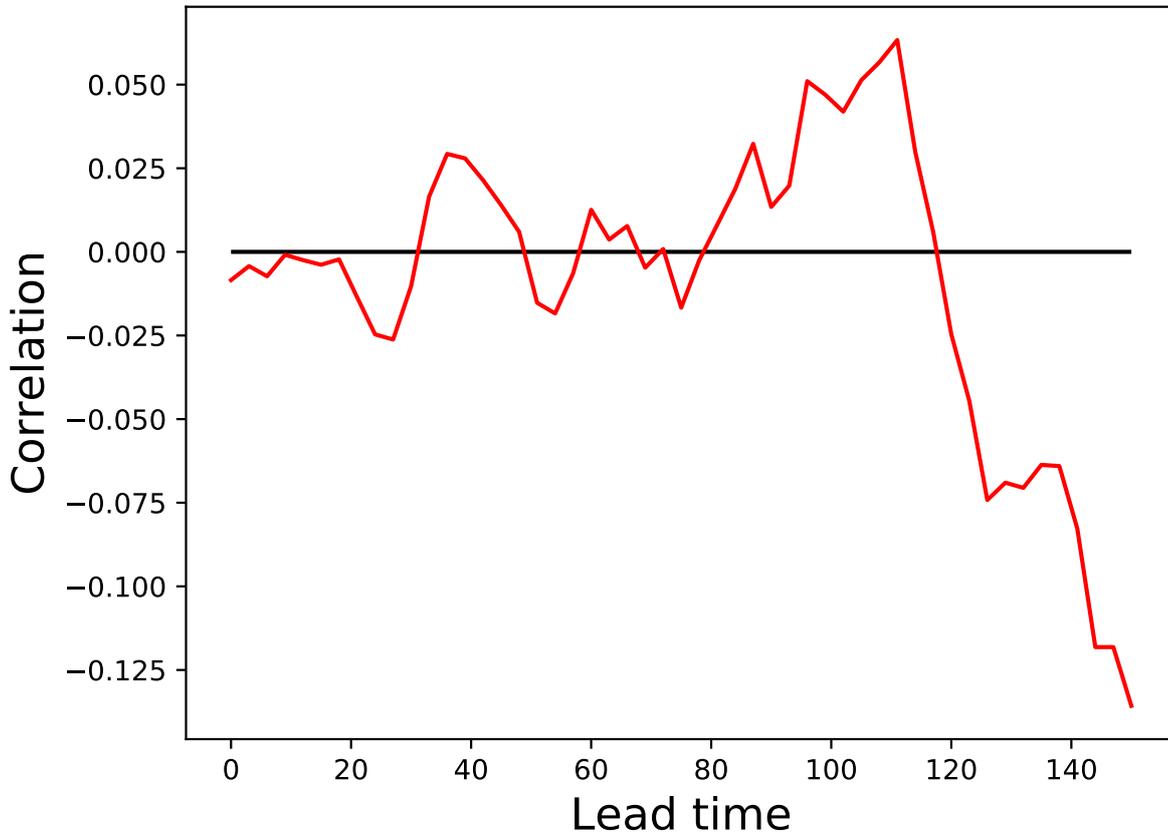


Figure 4.6: The red curve shows the evolution of the Pearson correlation coefficient of $\mu_P$ and $\sigma_P$ over the sample of $2048^2$ grid points and the two velocity components for the set of $M_0 = 51$ simulations. A black horizontal line indicating the level of zero correlation is inserted for reference.

### 4.3.3 Non-normality of the ensemble distribution

In this sub-section we explore whether the larger ($M_0 = 51$) set of simulations is indeed not normally distributed by means of a statistical test. The Jarque-Bera test is chosen for this purpose (Bowman & Shenton 1975, Jarque & Bera 1987). The test statistic

$$ JB = \frac{n}{6} \left( \hat{S}^2 + \frac{1}{4} (\hat{K} - 3)^2 \right) \tag{4.34} $$

is a function of the sample size $n$, the sample skewness $\hat{S}$ and the sample kurtosis $\hat{K}$. Under the null hypothesis of normality, $JB$ is expected to follow a chi-squared distribution with two degrees of freedom ($\chi_2^2$) in the limit of large samples ($n \to \infty$). It is known that at

least $n = O(1000)$ samples are needed for $JB$ to reasonably converge to $\chi^2_2$ (Bowman & Shenton 1975).

Since our sample size $n = M_0 = 51$ is much smaller than the threshold required for the convergence, the distribution of the test statistic under normality is approximated via a Monte-Carlo simulation. $2048^2 \times 51$ standard normal random variables are divided into groups of 51, and $JB$ is evaluated for each group. A histogram showing the distribution of $JB$ across the $2048^2$ groups is shown in Figure 4.7, which indicates that the distribution of the test statistic for small samples is fatter-tailed than $\chi^2_2$.
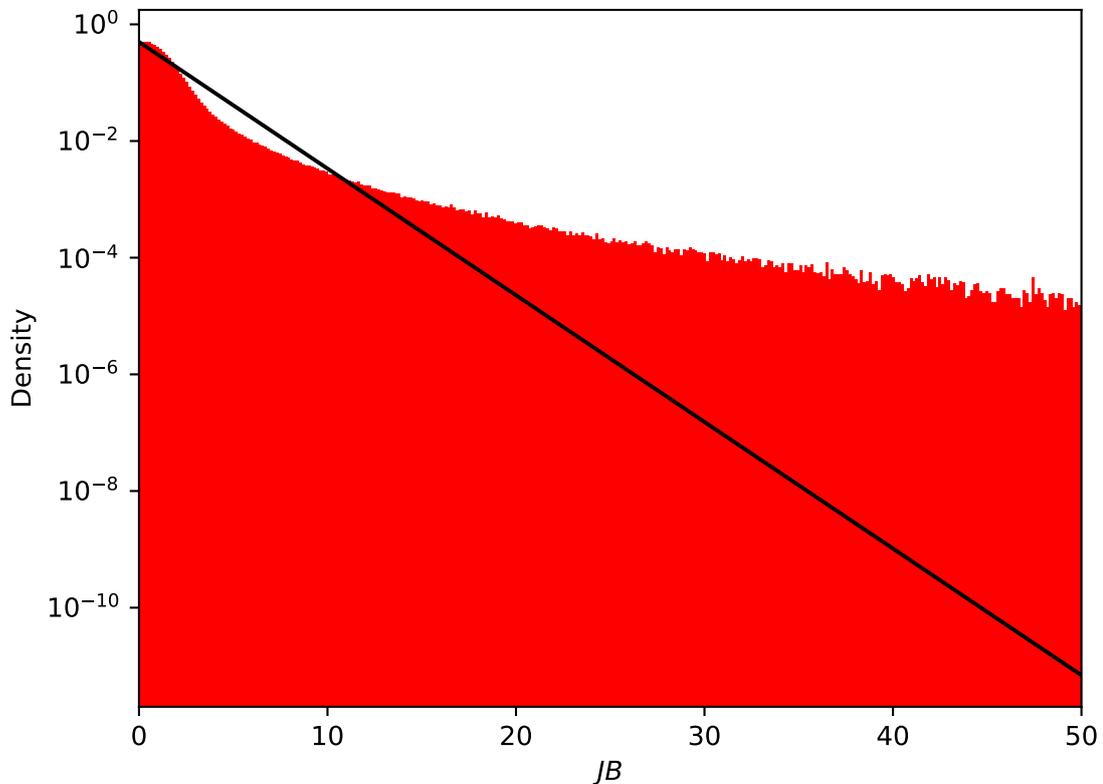


Figure 4.7: A histogram of $JB$ for small ($n = 51$) samples of standard normal random variables obtained via a Monte-Carlo Simulation. The histogram, in red, is normalised to form a probability density. The black line shows the PDF of a $\chi^2_2$ distribution. Note that only samples with $JB \leq 50$ are shown, and the vertical axis is logarithmic.

Having computed the null-hypothesis distribution of $JB$, the test statistic is evaluated at various lead times for the $M_0 = 51$-member ensemble of idealised 2D turbulence. Its distribution over $\Lambda = G \times D$ is compared against the null-hypothesis distribution. The two distributions are consistently different, as Figure 4.8 suggests, and the difference is found to be statistically significant (with a $p$-value indistinguishable from zero at com-

puter precision) after applying the two-sample Kolmogorov-Smirnov test (Wilks 2019). The same conclusion has been reached individually for the sample skewness $\hat{S}$ and the sample excess kurtosis $\hat{K}-3$, whose distributions are shown in Figures 4.9 and 4.10 respectively. Therefore, it is almost certain that the ensemble distribution is non-normal at all lead times, and the difference between $\sqrt{2\pi}\,\frac{CRPS^*_\infty(t)}{RMSE^*(t)}$ and $\frac{1}{\sqrt{1+h}}$ can indeed be attributed to the non-normality of the ensemble distribution, although not necessarily solely (since the possible dependence between $\mu_P$ and $\sigma_P$ may also contribute to the difference). Yet it remains unknown in quantitative terms how non-normality may affect the CRPS-RMSE ratio.



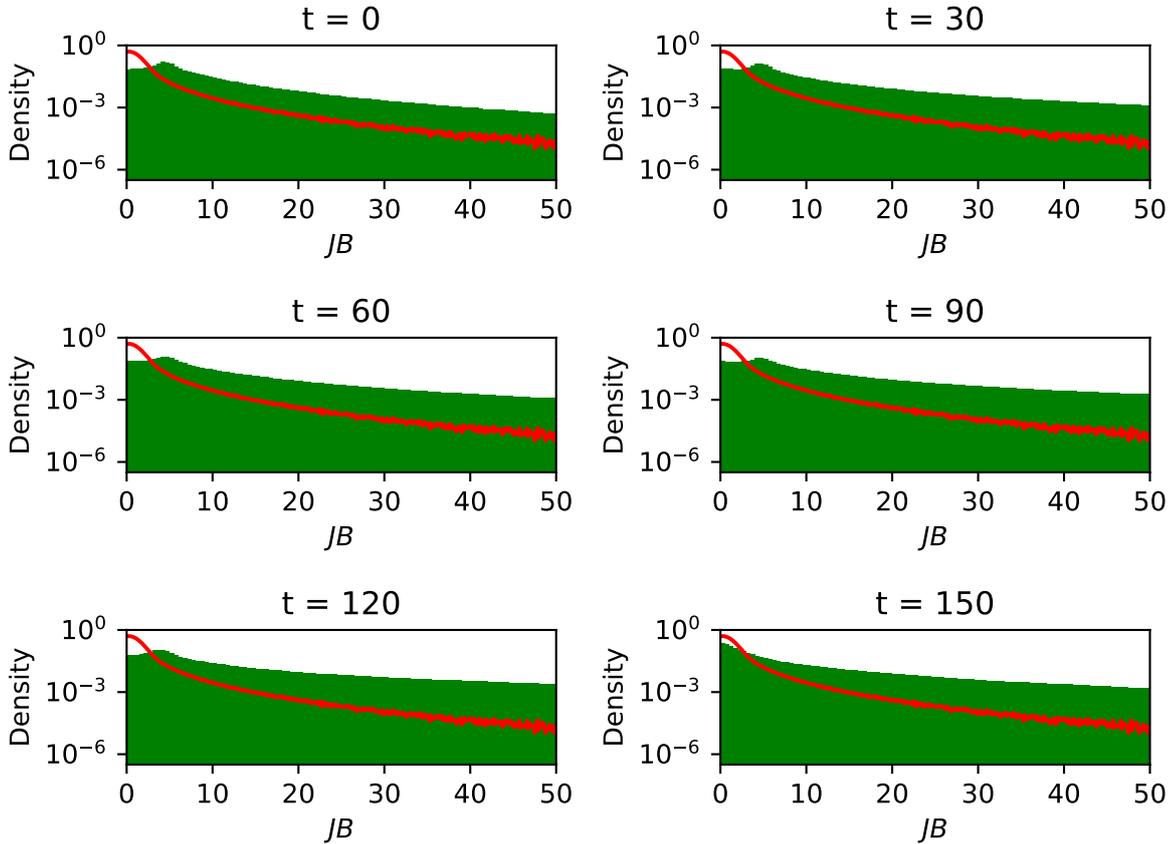Figure 4.8: Histograms of $JB$ taken over $\Lambda = G \times D$ for the $M_0 = 51$-member ensemble simulation of idealised 2D turbulence (the verification member included), for lead times up to $T = 150$ at intervals of $\Delta t = 30 = 0.2T$, in green and in the form of probability densities. The red curves indicate the PDF of $JB$ for $n = 51$ samples under the null hypothesis of normality (as in the red histogram of Figure 4.7).
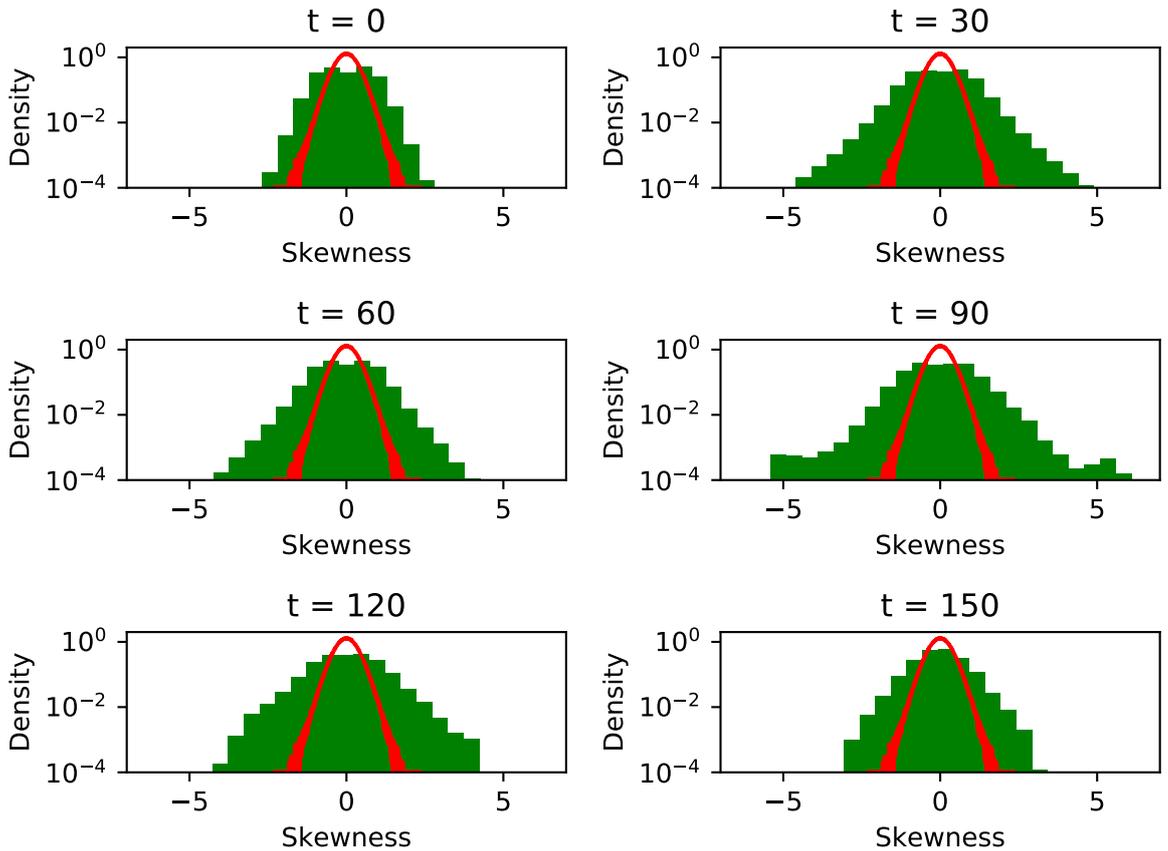
Figure 4.9: As in Figure 4.8, but for the sample skewness $\hat{S}$ instead of $JB$.

## 4.4 Discussion and summary

In this chapter, we have derived a functional relationship between two forecast verification metrics: the CRPS and the RMSE (Sections 4.1 and 4.2). The CRPS is a standard probabilistic score that rewards forecasts that are both sharp and reliable. In fact, there exists a decomposition of the aggregated CRPS into three parts representing respectively the sharpness of the predictive distribution, the reliability of the ensemble, and the verifying distribution's uncertainty which does not depend on the predictive distribution (Hersbach 2000). On the other hand, the RMSE is the sum of the ensemble variance and the squared error of the ensemble mean. In some contexts, only the latter contribution is included in the definition of RMSE, which makes it a deterministic verification metric since the ensemble mean can be interpreted as a deterministic prediction in its own right. The fact that the CRPS and the RMSE can be functionally related provides a common theoretical foundation between deterministic and probabilistic verification. The relationship, which has only been derived for reliable predictions of normally distributed variables, comes in
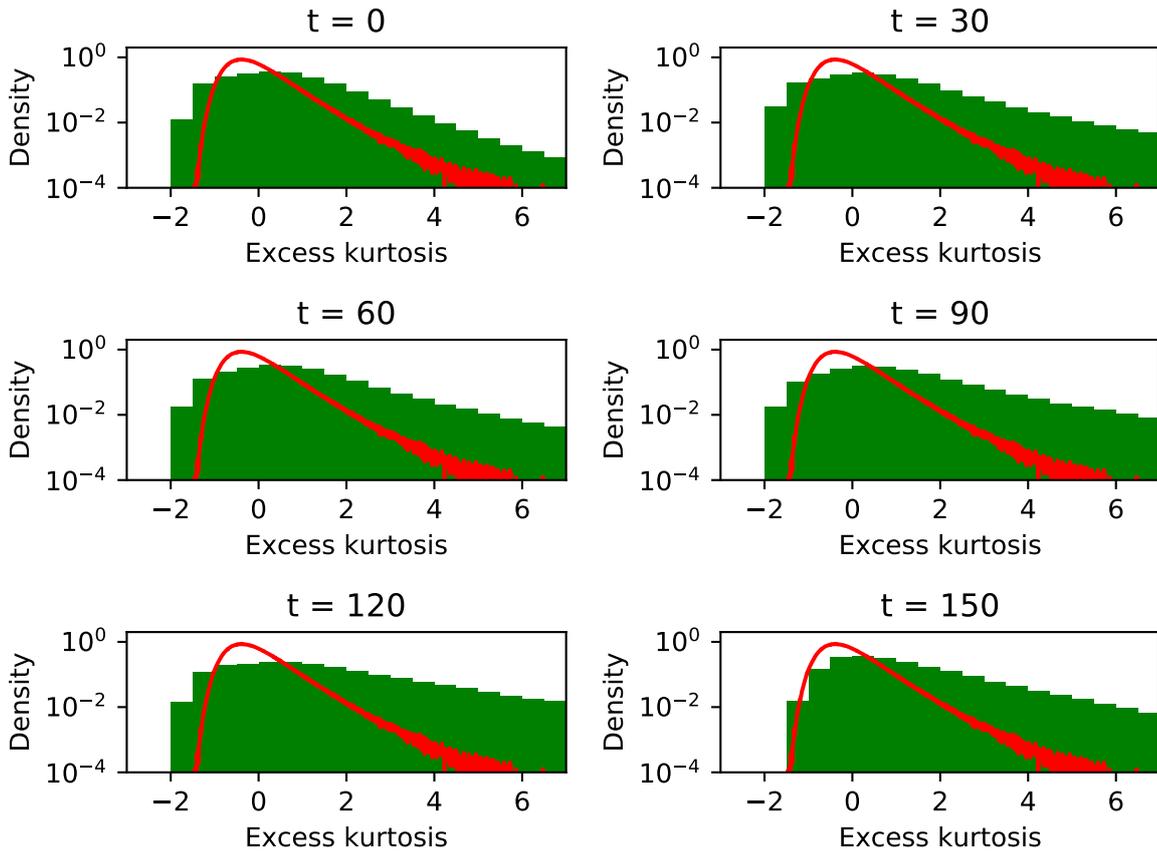
Figure 4.10: As in Figure 4.8, but for the sample excess kurtosis $\hat{K} - 3$ instead of $JB$.

the form

$$CRPS^* = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1+h}} RMSE^*, \tag{4.35}$$

where $h$ is the relative heteroscedasticity of the ensemble's standard deviation $\sigma_P$ as defined in Equation 4.27. The heteroscedasticity refers to the variability of $\sigma_P$ across the dimensions over which the CRPS and the RMSE are aggregated. In addition to reliability and normality, the relationship also requires that the ensemble mean $\mu_P$ be independent of $\sigma_P$ over the aggregated sample.

The relationship has been tested on simulations of idealised 2D turbulence (Section 4.3), in which ensembles are reliable by the experimental design. Heteroscedastic effects are present, yet they have already been minimised by virtue of the turbulence being idealised. Had such heteroscedastic effects been absent, a simpler form of the relationship

$$CRPS^* = \frac{1}{\sqrt{2\pi}} RMSE^* \tag{4.36}$$

would have held, thus making the CRPS a constant multiple of the RMSE. Equation

4.36 turns out to be a reasonably good approximation of the CRPS-RMSE relationship recorded in the numerical simulations. Deviations from this equation due to heteroscedasticity, non-normality and the $\mu_P$-$\sigma_P$ dependence remain within an order of magnitude, yet they are substantially larger than the discrepancies reported by Leutbecher & Haiden (2020). The effects of heteroscedasticity are estimated by running a larger ensemble. The results from the larger ensemble also suggest a possibility that the $\mu_P$-$\sigma_P$ independence assumption may not hold in the numerical simulations, although we are unable to confirm or refute it. The last remaining source of discrepancy, namely the non-normality of the ensemble distribution, has been independently verified by means of a statistical test. This is hardly surprising, since non-normality is a known feature of 2D turbulence (Farge et al. 1999). There is some evidence suggesting that the effects of non-normality could be obscured by a small ensemble size, but further work would be needed to demonstrate this rigorously.

The CRPS-RMSE relationship may be applied on any scalar meteorological variable in the real world, provided that the distribution of the variable is not overly non-normal. Inhomogeneity and anisotropy of the atmospheric flow imply that the results will depend on the domain and direction of aggregation. Using this relationship, the results of Chapter 3 on error growth and predictability properties in the infinite-resolution limit may now be translated into statements about the CRPS. It would be particularly helpful in predicting the evolution of the CRPS for future model resolutions with the aid of Lorenz's 1969 model and this relationship. A realistic estimate of heteroscedasticity in atmospheric flows would further improve such predictions.

# 5 Conclusions and outlook

Predicting the weather accurately is a fundamental desire of human civilisation so that advance preparations can be made to mitigate against losses brought by weather hazards. While there had been such attempts over the course of history using various rudimentary methods, it was not until the mid-20<sup>th</sup> century that weather forecasting was first made possible on an industrial scale and in a systematic way backed by rigorous science. The continued flourishing of knowledge in atmospheric-oceanic dynamics as both a mathematical and physical discipline, coupled with scientific and computational advances in the modelling of the Earth system, has since then contributed to substantial progress in NWP in terms of reducing errors and enhancing predictability. However, the foundational understanding about error growth and predictability as a discipline in its own right is less complete and has been somewhat overlooked in the recent decades compared to the attention it received before then. This thesis serves to fill several of these gaps by discussing the relevant theoretical topics in detail.

First, a renewed understanding of the inherent finite-time barrier of atmospheric predictability is presented (Chapter 2). This barrier makes it impossible to skilfully predict the instantaneous weather more than two or three weeks ahead no matter how small the initial error is. Through a careful investigation of the analytic properties of the 2D Navier-Stokes equations in relation to Lorenz (1969)'s classic result on limited predictability, this study proposes that the theory of the predictability horizon as understood by the meteorological community may only be valid so long as the inertial range of the energy spectrum is concerned. Were it possible to make the initial error so small in scale that it fell beyond the inertial range (i.e. within the viscous range), then further decreasing the scale of the error would make it possible to extend the range of predictability indefinitely,

thereby removing the predictability barrier. Hence, the predictability limit may be seen as a result of the stringent requirements on the initial error magnitude as one goes deeper into the inertial range.

In the second topic, the applicability of the classic results of error growth and predictability to the hybrid $k^{-3}$-$k^{-\frac{5}{3}}$ energy spectrum is discussed. The issue is becoming more relevant as global NWP models begin to resolve the shallower $k^{-\frac{5}{3}}$ range of the hybrid spectrum. Much of the existing theory on error growth and predictability properties rests on an implicit assumption that the basic-state energy spectrum consists of only one inertial range. As demonstrated in Chapter 3, extending these results to the hybrid spectrum is not so straightforward because the dynamics pertaining to the two constituent inertial ranges interact with each other. The $k^{-3}$ range at the synoptic scale is found to largely suppress the theoretically expected fast error growth characteristic of a $k^{-\frac{5}{3}}$ spectrum in the first decade of wavenumbers beyond the spectral break, irrespective of the scale of the initial error. It is only beyond that decade of wavenumbers that the fast mesoscale error growth is expected to become visible. Physically, this corresponds to an effective resolution of approximately 20 kilometres or a grid resolution of a few kilometres, which will probably be reached by global NWP ensembles in the not-too-distant future. With the hybrid spectrum, the predictability horizon on the global scale could be extended by several days to a week when compared with Lorenz (1969)'s original estimates (Table 3.2).

The third topic considers the relationship between forecast verification metrics. In the introductory Chapter 1, it was mentioned that the predictability horizon depends on the choice of metric and the prescribed threshold of skilful prediction. In Lorenz (1969)'s model, the metric in question is the error variance, which is closely related to the RMSE, and the threshold (in his original formulation) is about 0.815 times the variance of the basic-state flow. It is from these assumptions that he reached the famous conclusion of the two-week predictability limit. Over the years, the two-week limit has been found to be generalisable to other popular verification metrics and thresholds, such as the CRPS with the threshold being the expected score of the climatological distribution. The generalisability is not trivial. It suggests that the verification metrics are somehow related. Chapter 4 shows that there is indeed a quantifiable relationship between the RMSE and the CRPS when predictions are reliable and normally distributed. To first order, the two

quantities are related through a multiplicative constant. When they are aggregated over heteroscedastic environments, an extra factor accounting for the flow's heteroscedasticity comes into play, but subject to an extra assumption on the meta-distribution of the predictive distribution's moments. The relationship is tested on simulations of idealised 2D turbulence, and discrepancies are discussed in light of the assumptions behind the derived relationship.

It is important to recognise the limitations of this study while drawing conclusions. The primary limitation throughout the work of this thesis is the use of idealised models. These include Lorenz (1969)'s error growth model, the incompressible 2D Navier-Stokes equations, and the 2D barotropic vorticity model. While these models are closely related to each other, they represent a significant idealisation of Earth's atmosphere whose motion is governed by Equations 1.1 and 1.4–1.7. First and foremost, atmospheric flows are 3D. The approximation of two-dimensionality adopted in these models reflects the quasi-2D nature of large-scale atmospheric motions as a result of Earth's rotation and the atmosphere's stratification. Yet, key to the discussion of predictability is the limiting behaviour of error growth at small scales, where vertical motions become important and the approximation breaks down. Although it is generally inappropriate to make inferences on small-scale flows with 2D models, the issue could be mitigated in the context of error growth and predictability properties by modelling the mesoscale $k^{-\frac{5}{3}}$ spectrum appropriately, since the result by Rotunno & Snyder (2008) suggests that these properties are more a function of the energy spectrum than of the model dynamics *per se*. Indeed, the resemblance of the results of Chapter 3 to the error growth pattern of Judt (2018)'s simulations using a global convection-permitting NWP model illustrates this point. Another effect of assuming flows to be 2D (and incompressible) is the dropping out of thermodynamic effects from the equations. Physical processes such as convection and fronts can no longer be explicitly represented in these models, although they are implicitly and collectively represented in the $k^{-\frac{5}{3}}$ spectrum.

These idealised models also represent a simplification of Earth's geometry and rotational effects. Even if atmospheric flows were horizontal and thus 2D, the two-dimensionality would refer to flows on a 2D manifold embedded in 3D space (the surface of a sphere). Topologically, this is not equivalent to a doubly periodic domain in Carte-

sian geometry which has been assumed throughout these models. It would therefore have been more appropriate to characterise horizontal waves using spherical harmonics, the counterpart of Fourier modes proper to the surface of a sphere. Moreover, these models do not account for the apparent forces associated with Earth's varying rate of rotation across latitudes (the $\beta$-effect). These forces contribute to anisotropy of atmospheric flows, which gives rise to the variation of error growth and predictability properties across latitudinal zones (Judt 2020). The models considered in this thesis are too simple to capture such variation, but rather considered these properties in a bulk sense. Also, they have not accounted for seasonal changes in error growth characteristics that subtly affect the predictability horizon (Selz 2019). These changes could have possibly been represented in the numerical simulations of Chapter 3 by allowing the forcing amplitude $\hat{A}$ to vary, thereby mimicking the shift in location of the $k^{-3}$-$k^{-\frac{5}{3}}$ spectral break from one season to another. Furthermore, inhomogeneity of atmospheric flows arising from Earth's complex orography may not be well-represented in these models. For example, the scale-selective forcing in the 2D barotropic vorticity model does not reflect the continuous spectrum of orographic forcing.

It is also important to bear in mind that the results of Lorenz's model are conditional upon the quasi-normal turbulence closure scheme and a rather primitive mechanism of incorporating the non-linear effects of turbulence. While the introduction of a closure scheme is inevitable, other schemes such as the eddy-damped quasi-normal Markovian closure (Burgess & Shepherd 2013) might yield more realistic results. The non-linear effects were treated in a better way in the more recent works of Durran & Gingrich (2014) and Sun & Zhang (2020), an approach that has been decided against in this study in favour of a more convenient analysis of the mathematical properties of the model. It remains to be seen how the alternative treatment might have impacted the results.

Another limitation of the present study is the assumption of normal distributions in the derivation of the CRPS-RMSE relationship. Some scalar meteorological variables cannot be normally distributed. An example of such is rainfall, since it cannot take negative values. The normality assumption limits the applicability of the results to these variables. Although the relationship for normal distributions may to some extent be followed in those cases, it is not clear how the discrepancy can be quantified in terms of

measures of non-normality such as skewness or excess kurtosis.

In terms of the results, the relationship of the CRPS or the RMSE with other verification metrics has not been discussed in a wider context. It would be useful to systematically find out how the most commonly used verification metrics in NWP operations are related to each other. This could then provide an overview of the strengths and weaknesses of individual scores, and help forecasters and model developers choose the scores best suited to their purposes of identifying specific types of peculiarities.

It is recommended that future theoretical work on weather predictability be expanded along the lines of the limitations listed above. In my view, it would be most interesting above all to derive the counterpart of Lorenz's model for spherical harmonics and observe how the dynamics of error growth could be better described using a more appropriate geometry. On top of this, there could be further improvements by incorporating the effects of planetary rotation and introducing a vertical dimension, and with a better treatment of turbulence closure and non-linearity. Such a more realistic version of Lorenz's model could help characterise the next level of essential aspects of error growth beyond what has already been found in this thesis. While these characteristics are expected to be not very different from those observed in direct aqua-planet simulations, the analysis of qualitative discrepancies between the two could further substantiate the understanding of error growth mechanisms. The sensitivity of the error growth picture to initial error profiles could also be studied in the context of these slightly more complex models.

The effects of model errors is another interesting area to explore. In an earlier work (Leung 2017), initial errors and model errors were shown to interact in such a way that in principle, whichever error at the larger spatial scale would dominate the growth of the overall error. Yet, the effects of small-scale convective parametrisations cannot be ignored, since they can change the shape of the mesoscale spectrum (Wang & Sardeshmukh 2020) and therefore induce qualitative changes to the error growth behaviour, especially within these small scales. Such effects on error growth may best be studied in simplified models, such as those in which convection can be represented by an additional term controlled by a stochastic switch (Würsch & Craig 2014).

Having said all this, the results contained in this thesis suggest that the study of theoretical predictability is an inseparable part of the development of NWP. As is often the case, simple models help provide guidance on future research directions involving state-of-the-art models at a relatively modest cost. In the NWP context, they specifically have the potential to advance scientific understanding of the dynamical mechanisms limiting weather predictability. This thesis is a humble contribution to such, but much work is yet to be done to consolidate this new knowledge. With further work along similar directions, it is hoped that the understanding of predictability would become more complete, and that it would lead to the eventual development of more skilful NWP models for the benefit of mankind.

# Appendix A

# Proofs and derivations of certain formulae related to the Continuous Ranked Probability Score

## A.1 Equivalence of the CRPS and its kernel representation

The equivalence between the original definition of the CRPS (Equation 4.10) and its kernel representation (Equation 4.11) follows from the following two lemmata demonstrated by Baringhaus & Franz (2004).

**Lemma A.1.** *Let $X$ and $Y$ be independent real-valued scalar random variables with finite expectations. Let $\tilde{F}$ be the CDF of $X$ and $\tilde{G}$ be the CDF of $Y$. Then*

$$\mathbb{E}_{X,Y}\left[|X - Y|\right] = \int_{-\infty}^{\infty} \tilde{F}(z)\left(1 - \tilde{G}(z)\right)\,\mathrm{d}z + \int_{-\infty}^{\infty} \tilde{G}(z)\left(1 - \tilde{F}(z)\right)\,\mathrm{d}z. \qquad (A.1)$$

*Proof.* Since $X$ and $Y$ are independent, we may write

$$\begin{aligned}
|X - Y| &= \int_{-\infty}^{\infty} \left(\mathbb{I}(X \leq z < Y) + \mathbb{I}(Y \leq z < X)\right)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \left(\mathbb{I}(X \leq z)\mathbb{I}(Y > z) + \mathbb{I}(Y \leq z)\mathbb{I}(X > z)\right)\,\mathrm{d}z,
\end{aligned} \qquad (A.2)$$

where $\mathbb{I}$ is the indicator function taking the value one if the condition in its argument is satisfied and zero otherwise. Taking the joint expectation of Equation A.2 and applying Fubini's theorem[1] yields

$$
\begin{aligned}
\mathbb{E}_{X,Y}\left[|X-Y|\right] &= \int_{-\infty}^{\infty} \mathbb{E}_{X,Y}\left[\mathbb{I}(X \le z)\mathbb{I}(Y > z) + \mathbb{I}(Y \le z)\mathbb{I}(X > z)\right] \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \left(\mathbb{E}_X\left[\mathbb{I}(X \le z)\right]\mathbb{E}_Y\left[\mathbb{I}(Y > z)\right] + \mathbb{E}_Y\left[\mathbb{I}(Y \le z)\right]\mathbb{E}_X\left[\mathbb{I}(X > z)\right]\right) \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \tilde{F}(z)\left(1 - \tilde{G}(z)\right) \mathrm{d}z + \int_{-\infty}^{\infty} \tilde{G}(z)\left(1 - \tilde{F}(z)\right) \mathrm{d}z
\end{aligned}
$$
(A.3)

as required. ∎

**Lemma A.2.** *Let $X_1$, $X_2$, $Y_1$ and $Y_2$ be independent real-valued scalar random variables with finite expectations. Let $X_1$ and $X_2$ be identically distributed with CDF $\tilde{F}$, and $Y_1$ and $Y_2$ be identically distributed with CDF $\tilde{G}$. Then*

$$
\mathbb{E}_{X_1,Y_1}\left[|X_1 - Y_1|\right] - \frac{1}{2}\mathbb{E}_{X_1,X_2}\left[|X_1 - X_2|\right] - \frac{1}{2}\mathbb{E}_{Y_1,Y_2}\left[|Y_1 - Y_2|\right] = \int_{-\infty}^{\infty} \left(\tilde{F}(z) - \tilde{G}(z)\right)^2 \mathrm{d}z.
$$
(A.4)

*Proof.* Apply Lemma A.1 to the pairs $(X_1, Y_1)$, $(X_1, X_2)$, $(Y_1, Y_2)$ of random variables separately. Elementary calculations give the result. ∎

Equation 4.11 can now be established by applying Lemma A.2 to $X_1 = U$, $X_2 = U'$, and both $Y_1$ and $Y_2$ being random variables with the Dirac delta distribution at $u$. In this case, $\tilde{F} = F$ and $\tilde{G} = H_u$. The third term on the left-hand-side of Equation A.4 drops out since $Y_1 = Y_2$ almost-surely by virtue of them being point masses. Hence we have

$$
\mathbb{E}_P\left[|U - u|\right] - \frac{1}{2}\mathbb{E}_P\left[|U - U'|\right] = \int_{-\infty}^{\infty} (F(x) - H_u(x))^2 \mathrm{d}x,
$$
(A.5)

which shows the equivalence between the CRPS and its kernel representation.

---

[1]Fubini's theorem allows the order of integration to be swapped provided that the absolute value of the integrand integrates to a finite value.

## A.2   CRPS of a single event with a normal predictive distribution

The explicit expression for $CRPS(P, u)$ for normally distributed $P = \mathcal{N}(\mu_P, \sigma_P^2)$ shown in Equation 4.12 can be obtained through repeated integration by parts. Denoting $\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ for the PDF of a standard normal random variable and $\Phi(x) := \int_{-\infty}^{x} \varphi(x') \, dx'$ for its CDF, and writing $u' := \frac{u - \mu_P}{\sigma_P}$, we have

$$
\begin{aligned}
&CRPS(P, u) \\
&= \int_{-\infty}^{u} \left(\Phi\left(\frac{x - \mu_P}{\sigma_P}\right)\right)^2 dx + \int_{u}^{\infty} \left(1 - \Phi\left(\frac{x - \mu_P}{\sigma_P}\right)\right)^2 dx \\
&= \int_{-\infty}^{u'} (\Phi(y))^2 \sigma_P \, dy + \int_{u'}^{\infty} (1 - \Phi(y))^2 \sigma_P \, dy \\
&= \sigma_P \left( u' \left(\Phi(u')\right)^2 - \int_{-\infty}^{u'} 2y\Phi(y)\varphi(y) \, dy - u' \left(1 - \Phi(u')\right)^2 + \int_{u'}^{\infty} 2y \left(1 - \Phi(y)\right)\varphi(y) \, dy \right) \\
&= \sigma_P \left( u' \left(\Phi(u')\right)^2 - u' \left(1 - \Phi(u')\right)^2 + \int_{u'}^{\infty} 2y\varphi(y) \, dy - \int_{-\infty}^{\infty} 2y\Phi(y)\varphi(y) \, dy \right) \\
&= \sigma_P \left( u' \left(2\Phi(u') - 1\right) + \int_{u'}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \, d\left(y^2\right) - \int_{-\infty}^{\infty} \Phi(y) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \, d\left(y^2\right) \right) \\
&= \sigma_P \left( u' \left(2\Phi(u') - 1\right) + 2\varphi(u') + \int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}} \Phi(y) \, d\left(e^{-\frac{1}{2}y^2}\right) \right) \\
&= \sigma_P \left( u' \left(2\Phi(u') - 1\right) + 2\varphi(u') - \int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}y^2} \varphi(y) \, dy \right) \\
&= \sigma_P \left( u' \left(2\Phi(u') - 1\right) + 2\varphi(u') - \int_{-\infty}^{\infty} \frac{1}{\pi} e^{-y^2} \, dy \right) \\
&= \sigma_P \left( \frac{u - \mu_P}{\sigma_P} \left(2\Phi\left(\frac{u - \mu_P}{\sigma_P}\right) - 1\right) + 2\varphi\left(\frac{u - \mu_P}{\sigma_P}\right) - \frac{1}{\sqrt{\pi}} \right) \\
&= \frac{\sigma_P}{\sqrt{\pi}} \left( -1 + \sqrt{\pi}\frac{u - \mu_P}{\sigma_P} \operatorname{erf}\left(\frac{u - \mu_P}{\sqrt{2}\sigma_P}\right) + \sqrt{2}\exp\left(-\frac{1}{2}\left(\frac{u - \mu_P}{\sigma_P}\right)^2\right) \right).
\end{aligned}
$$

$$\text{(A.6)}$$

The last two lines of Equation A.6 are equivalent expressions of $CRPS(P, u)$ by observing an elementary property of the error function $\operatorname{erf}(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-y^2} \, dy$ that

$$
\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) = 2\Phi(z) - 1. \tag{A.7}
$$

## A.3 Expected CRPS for normal predictive and verifying distributions

The integral

$$\int_{-\infty}^{\infty} \sigma_Q \frac{r^2}{\sqrt{\pi}} \left( -1 + \sqrt{\pi} x \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) + \sqrt{2} \exp\left(-\frac{1}{2}x^2\right) \right) \varphi\left(rx + b\right) \, dx \qquad (\text{A.8})$$

in Equation 4.13 can be simplified to provide an analytic expression for $CRPS(P,Q)$, the expected CRPS for normal predictive and verifying distributions. The integral will be decomposed into three contributions according to the terms inside the outermost parentheses of the integrand. These contributions will be evaluated one by one. To begin,

$$\int_{-\infty}^{\infty} \sigma_Q \frac{r^2}{\sqrt{\pi}} \left(-\varphi\left(rx+b\right)\right) \, dx = -\sigma_Q \frac{r}{\sqrt{\pi}} \int_{-\infty}^{\infty} \varphi\left(rx+b\right) \, d(rx+b) = -\sigma_Q \frac{r}{\sqrt{\pi}}. \qquad (\text{A.9})$$

We also have

$$\begin{aligned}
&\int_{-\infty}^{\infty} \sigma_Q \frac{r^2}{\sqrt{\pi}} \sqrt{2} \exp\left(-\frac{1}{2}x^2\right) \varphi\left(rx+b\right) \, dx \\
&= \sigma_Q \frac{r^2}{\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 + r^2 x^2 + 2brx + b^2)\right) \, dx \\
&= \sigma_Q \frac{r^2}{\pi} \exp\left(-\frac{1}{2}\left(b^2 - \frac{b^2 r^2}{1+r^2}\right)\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(1+r^2)\left(x + \frac{br}{1+r^2}\right)^2\right) \, dx \quad (\text{A.10}) \\
&= \sigma_Q \frac{r^2}{\pi} \exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right) \sqrt{\frac{2\pi}{1+r^2}} \\
&= \sigma_Q r^2 \sqrt{\frac{2}{\pi(1+r^2)}} \exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right).
\end{aligned}$$

As for the integral

$$\int_{-\infty}^{\infty} \sigma_Q r^2 x \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \varphi\left(rx+b\right) \, dx, \qquad (\text{A.11})$$

we proceed by first seeking an anti-derivative $A(x)$ of $x\varphi\left(rx+b\right)$, so that Expression A.11 can be written as

$$\sigma_Q r^2 \left(A(\infty)\operatorname{erf}(\infty) - A(-\infty)\operatorname{erf}(-\infty)\right) - \sigma_Q r^2 \int_{-\infty}^{\infty} A(x) \, d\left(\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right). \qquad (\text{A.12})$$

To compute $A$, we have

$$A(x) = \int_{-\infty}^{x} x' \varphi \left(rx' + b\right) \, \mathrm{d}x' = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} x' \exp\left(-\frac{1}{2}(rx'+b)^2\right) \, \mathrm{d}x'$$

$$= \int_{-\infty}^{rx+b} \frac{1}{\sqrt{2\pi}} \frac{y-b}{r} \exp\left(-\frac{1}{2}y^2\right) \frac{1}{r} \, \mathrm{d}y \tag{A.13}$$

$$= -\frac{1}{\sqrt{2\pi}r^2} \exp\left(-\frac{1}{2}(rx+b)^2\right) - \frac{b}{r^2}\Phi(rx+b),$$

so that $A(\infty) = -\frac{b}{r^2}$ and $A(-\infty) = 0$. Substituting these into Expression A.12, the contribution from $\sigma_Q r^2 \left(A(\infty)\operatorname{erf}(\infty) - A(-\infty)\operatorname{erf}(-\infty)\right)$ leaves us with only $-b\sigma_Q$, whereas

$$-\sigma_Q r^2 \int_{-\infty}^{\infty} A(x) \, \mathrm{d}\left(\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$$

$$= \sigma_Q r^2 \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}r^2} \exp\left(-\frac{1}{2}(rx+b)^2\right) + \frac{b}{r^2}\Phi(rx+b)\right) \, \mathrm{d}(2\Phi(x)-1)$$

$$= \frac{2\sigma_Q}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(rx+b)^2\right) \varphi(x) \, \mathrm{d}x + 2b\sigma_Q \int_{-\infty}^{\infty} \Phi(rx+b)\varphi(x) \, \mathrm{d}x \tag{A.14}$$

$$= \frac{\sigma_Q}{\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 + r^2x^2 + 2brx + b^2)\right) \, \mathrm{d}x + 2b\sigma_Q \int_{-\infty}^{\infty} \Phi(rx+b)\varphi(x) \, \mathrm{d}x.$$

In light of Equation A.10, the first term on the right-hand-side of Equation A.14 equals $\sigma_Q \sqrt{\frac{2}{\pi(1+r^2)}} \exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right)$. The following lemma will be useful for computing the second term:

**Lemma A.3.** *For any $r, b \in \mathbb{R}$,*

$$I(r,b) := \int_{-\infty}^{\infty} \operatorname{erf}\left(\frac{rx+b}{\sqrt{2}}\right) \varphi(x) \, \mathrm{d}x = \operatorname{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right). \tag{A.15}$$

*Proof.* Using $\frac{\mathrm{d}}{\mathrm{d}z}\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\exp\left(-z^2\right)$, we have

$$\frac{\partial I}{\partial b} = \int_{-\infty}^{\infty} \frac{2}{\sqrt{\pi}} \exp\left(-\frac{1}{2}(rx+b)^2\right) \frac{1}{\sqrt{2}} \varphi(x) \, \mathrm{d}x$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 + r^2x^2 + 2brx + b^2)\right) \, \mathrm{d}x \tag{A.16}$$

$$= \sqrt{\frac{2}{\pi(1+r^2)}} \exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right)$$

in light of Equation A.10. Since $I(r,0) = 0$ (as the integrand is in that case an odd

function), we can write

$$
\begin{aligned}
I(r, b) &= \int_0^b \frac{\partial I}{\partial b}(r, b')\, \mathrm{d}b' \\
&= \sqrt{\frac{2}{\pi(1+r^2)}} \int_0^b \exp\left(-\frac{1}{2}\frac{b'^2}{1+r^2}\right)\, \mathrm{d}b' \\
&= \sqrt{\frac{2}{\pi(1+r^2)}} \int_0^{\frac{b}{\sqrt{2(1+r^2)}}} \sqrt{2(1+r^2)}e^{-u^2}\, \mathrm{d}u \\
&= \sqrt{\frac{2}{\pi(1+r^2)}} \sqrt{2(1+r^2)}\frac{\sqrt{\pi}}{2}\, \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right) \\
&= \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right)
\end{aligned}
\tag{A.17}
$$

as required. ∎

The second term on the right-hand-side of Equation A.14 is therefore

$$
\begin{aligned}
& 2b\sigma_Q \int_{-\infty}^{\infty} \Phi(rx+b)\varphi(x)\, \mathrm{d}x \\
&= b\sigma_Q \int_{-\infty}^{\infty} \left(1 + \mathrm{erf}\left(\frac{rx+b}{\sqrt{2}}\right)\right)\varphi(x)\, \mathrm{d}x \\
&= b\sigma_Q \left(1 + \int_{-\infty}^{\infty} \mathrm{erf}\left(\frac{rx+b}{\sqrt{2}}\right)\varphi(x)\, \mathrm{d}x\right) = b\sigma_Q \left(1 + \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right)\right).
\end{aligned}
\tag{A.18}
$$

In this way, we can write Expression A.11 as

$$
\begin{aligned}
& \int_{-\infty}^{\infty} \sigma_Q r^2 x\, \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\varphi(rx+b)\, \mathrm{d}x \\
&= -b\sigma_Q + \sigma_Q \sqrt{\frac{2}{\pi(1+r^2)}}\exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right) + b\sigma_Q \left(1 + \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right)\right) \\
&= \sigma_Q \sqrt{\frac{2}{\pi(1+r^2)}}\exp\left(-\frac{1}{2}\frac{b^2}{1+r^2}\right) + b\sigma_Q\, \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right).
\end{aligned}
\tag{A.19}
$$

Substituting this and Equations A.9 and A.10 into Expression A.8 and therefore Equation 4.13, we finally arrive at

$$
CRPS(P, Q) = \frac{\sigma_Q}{\sqrt{\pi}}f(b, r)
\tag{A.20}
$$

where

$$
f(b, r) = -r + \sqrt{2(1+r^2)}\exp\left(-\frac{b^2}{2(1+r^2)}\right) + \sqrt{\pi}\,b\, \mathrm{erf}\left(\frac{b}{\sqrt{2(1+r^2)}}\right).
\tag{A.21}
$$

# Bibliography

Abbe, C. (1901), 'The physical basis of long-range weather forecasts', *Monthly Weather Review* **29**, 551–561.

Baringhaus, L. & Franz, C. (2004), 'On a new multivariate two-sample test', *Journal of Multivariate Analysis* **88**, 190–206.

Bauer, P., Quintino, T., Wedi, N., Bonanni, A., Chrust, M., Deconinck, W., Diamantakis, M., Düben, P., English, S., Flemming, J., Gillies, P., Hadade, I., Hawkes, J., Hawkins, M., Iffrig, O., Kühnlein, C., Lange, M., Lean, P., Marsden, O., Müller, A., Saarinen, S., Sarmany, D., Sleigh, M., Smart, S., Smolarkiewicz, P., Thiemert, D., Tumolo, G., Weihrauch, C. & Zanna, C. (2020), The ECMWF scalability programme: progress and plans, Technical Memorandum 857, European Centre for Medium-Range Weather Forecasts.

Bauer, P., Thorpe, A. & Brunet, G. (2015), 'The quiet revolution of numerical weather prediction', *Nature* **525**, 47–55.

Bierdel, L., Selz, T. & Craig, G. C. (2017), 'Theoretical aspects of upscale error growth through the mesoscales: an analytical model', *Quarterly Journal of the Royal Meteorological Society* **143**, 3048–3059.

Bierdel, L., Selz, T. & Craig, G. C. (2018), 'Theoretical aspects of upscale error growth on the mesoscales: idealized numerical simulations', *Quarterly Journal of the Royal Meteorological Society* **144**, 682–694.

Bistafa, S. R. (2018), 'On the development of the Navier-Stokes equation by Navier', *Revista Brasileira de Ensino de Física* **40**, e2603.

Bjerknes, V. (1904), 'Das problem der wettervorhersage, betrachtet vom standpunkte der mechanik und der physik', *Meteorologische Zeitschrift* **21**, 1–7. Translated and edited by E. Volken and S. Brönnimann — *Meteorologische Zeitschrift* **18** (2009), 663–667.

Boer, G. J. & Shepherd, T. G. (1983), 'Large-scale two-dimensional turbulence in the atmosphere', *Journal of the Atmospheric Sciences* **40**, 164–184.

Boffetta, G. & Musacchio, S. (2001), 'Predictability of the inverse energy cascade in 2D turbulence', *Physics of Fluids* **13**, 1060–1062.

Bowman, J. C. (1996), 'On inertial-range scaling laws', *Journal of Fluid Mechanics* **306**, 167–181.

Bowman, K. O. & Shenton, L. R. (1975), 'Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and $b_2$', *Biometrika* **62**, 243–250.

Buizza, R. & Leutbecher, M. (2015), 'The forecast skill horizon', *Quarterly Journal of the Royal Meteorological Society* **141**, 3366–3382.

Buizza, R., Miller, M. & Palmer, T. N. (1999), 'Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System', *Quarterly Journal of the Royal Meteorological Society* **125**, 2887–2908.

Burgess, B. H. & Shepherd, T. G. (2013), 'Spectral non-locality, absolute equilibria and Kraichnan-Leith-Batchelor phenomenology in two-dimensional turbulent energy cascades', *Journal of Fluid Mechanics* **725**, 332–371.

Charney, J. G. (1947), 'The dynamics of long waves in a baroclinic westerly current', *Journal of Meteorology* **4**, 135–162.

Charney, J. G. (1971), 'Geostrophic turbulence', *Journal of the Atmospheric Sciences* **28**, 1087–1095.

Charney, J. G., Fjørtoft, R. & von Neumann, J. (1950), 'Numerical integration of the barotropic vorticity equation', *Tellus* **2**, 237–254.

Christensen, H. M., Lock, S.-J., Moroz, I. M. & Palmer, T. N. (2017), 'Introducing

independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme', *Quarterly Journal of the Royal Meteorological Society* **143**, 2168–2181.

Dalcher, A. & Kalnay, E. (1987), 'Error growth and predictability in operational ECMWF forecasts', *Tellus* **39A**, 474–491.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK.

Devaney, R. L. (1989), *An Introduction to Chaotic Dynamical Systems*, 2$^{nd}$ edn, Addison-Wesley Publishing Company, Reading, Massachusetts, USA.

Durran, D. R. & Gingrich, M. (2014), 'Atmospheric predictability: why butterflies are not of practical importance', *Journal of the Atmospheric Sciences* **71**, 2476–2488.

Eady, E. T. (1949), 'Long waves and cyclone waves', *Tellus* **1**, 33–52.

English, S., McNally, T., Bormann, N., Salonen, K., Matricardi, M., Horanyi, A., Rennie, M., Janisková, M., Di Michele, S., Geer, A., Di Tomaso, E., Cardinali, C., de Rosnay, P., Muñoz Sabater, J., Bonavita, M., Albergel, C., Engelen, R. & Thépaut, J.-N. (2013), Impact of satellite data, Technical Memorandum 711, European Centre for Medium-Range Weather Forecasts.

Epstein, E. S. (1969), 'Stochastic dynamic prediction', *Tellus* **21**, 739–759.

Euler, L. (1757), 'Principes généraux du mouvement des fluides', *Mémoires de l'académie des sciences de Berlin* **11**, 274–315. Adapted by U. Frisch (2008, arXiv:0802.2383) following an English translation by Thomas E. Burton.

European Centre for Medium-Range Weather Forecasts (2016), 'ECMWF Strategy 2016–2025: the Strength of a Common Goal', available at `https://www.ecmwf.int/sites/default/files/ECMWF_Strategy_2016-2025.pdf` (accessed: 17$^{th}$ July, 2020).

Farge, M., Schneider, K. & Kevlahan, N. (1999), 'Non-Gaussianity and coherent vortex simulation for two-dimensional turbulence using an adaptive orthogonal wavelet basis', *Physics of Fluids* **11**, 2187–2201.

Franklin, W. S. (1898), 'Review of P. Duhem, *Traité Élémentaire de Mécanique Chimique fondée sur la Thermodynamique* (Paris, 1897), Volumes I and II', *Physical Review* **6**, 170–175.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007), 'Probabilistic forecasts, calibration and sharpness', *Journal of the Royal Statistical Society: Series B* **69**, 243–268.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association* **102**, 359–378.

Haiden, T., Janousek, M., Bauer, P., Bidlot, J., Dahoui, M., Ferranti, L., Prates, F., Richardson, D. S. & Vitart, F. (2015), Evaluation of ECMWF forecasts, including 2014–2015 upgrades, Technical Memorandum 765, European Centre for Medium-Range Weather Forecasts.

Haiden, T., Janousek, M., Vitart, F., Ferranti, L. & Prates, F. (2019), Evaluation of ECMWF forecasts, including the 2019 upgrade, Technical Memorandum 853, European Centre for Medium-Range Weather Forecasts.

Held, I. M., Pierrehumbert, R. T., Garner, S. T. & Swanson, K. L. (1995), 'Surface quasi-geostrophic dynamics', *Journal of Fluid Mechanics* **282**, 1–20.

Hersbach, H. (2000), 'Decomposition of the Continuous Ranked Probability Score for ensemble prediction systems', *Weather and Forecasting* **15**, 559–570.

Holm, D. D. (2015), 'Variational principles for stochastic fluid dynamics', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **471**, 20140963.

Horn, L. H. & Bryson, R. A. (1963), 'An analysis of the geostrophic kinetic energy spectrum of large-scale atmospheric turbulence', *Journal of Geophysical Research* **68**, 1059–1064.

Hou, T. Y. & Li, R. (2007), 'Computing nearly singular solutions using pseudo-spectral methods', *Journal of Computational Physics* **226**, 379–397.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M. & Raynaud, L. (2010), Ensemble of data assimilations at ECMWF, Technical Memorandum 636,

European Centre for Medium-Range Weather Forecasts.

Jarque, C. M. & Bera, A. K. (1987), 'A test for normality of observations and regression residuals', *International Statistical Review* **55**, 163–172.

Judt, F. (2018), 'Insights into atmospheric predictability through global convection-permitting model simulations', *Journal of the Atmospheric Sciences* **75**, 1477–1497.

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.

Kalnay, E. (2002), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK.

Kolmogorov, A. (1941), 'The local structure of turbulence in incompressible viscous fluid for very large Reynolds' numbers', *Doklady Akademiia Nauk SSSR* **30**, 301–305.

Kraichnan, R. H. (1967), 'Inertial ranges in two-dimensional turbulence', *The Physics of Fluids* **10**, 1417–1423.

Lea, D. J., Mirouze, I., Martin, M. J., King, R. R., Hines, A., Walters, D. & Thurlow, M. (2015), 'Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere-land-ocean-sea ice model', *Monthly Weather Review* **143**, 4678–4694.

Leith, C. E. (1974), 'Theoretical skill of Monte Carlo forecasts', *Monthly Weather Review* **102**, 409–418.

Leray, J. (1933), 'Etude de diverses équations intégrales non linéaires et de quelques problèmes que pose l'hydrodynamique', *Journal de Mathématiques Pures et Appliquées* **12**, 1–82.

Leray, J. (1934), 'Essai sur le mouvement d'un liquide visqueux emplissant l'espace', *Acta Mathematica* **63**, 193–248.

Leung, T. Y. (2017), Role of Initial and Model Errors in Uncertainty of Weather Forecasts,

MRes thesis, Imperial College London and University of Reading.

Leung, T. Y., Leutbecher, M., Reich, S. & Shepherd, T. G. (2019), 'Atmospheric predictability: revisiting the inherent finite-time barrier', *Journal of the Atmospheric Sciences* **76**, 3883–3892.

Leung, T. Y., Leutbecher, M., Reich, S. & Shepherd, T. G. (2020), 'Impact of the mesoscale range on error growth and the limits to atmospheric predictability', *Journal of the Atmospheric Sciences* **77**, 3769–3779.

Leutbecher, M. (2019), 'Ensemble size: how suboptimal is less than infinity?', *Quarterly Journal of the Royal Meteorological Society* **145 (Suppl. 1)**, 107–128.

Leutbecher, M. & Haiden, T. (2020), Understanding changes of the continuous ranked probability score using a homogeneous Gaussian approximation. *Quarterly Journal of the Royal Meteorological Society*, Early View, doi: 10.1002/qj.3926.

Leutbecher, M. & Lang, S. T. K. (2014), 'On the reliability of ensemble variance in subspaces defined by singular vectors', *Quarterly Journal of the Royal Meteorological Society* **140**, 1453–1466.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H. M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R. M., Goddard, J., Haiden, T., Hogan, R. J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P. K., Subramanian, A., Vitart, F., Wedi, N. & Weisheimer, A. (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.

Lilly, D. K. (1989), 'Two-dimensional turbulence generated by energy sources at two scales', *Journal of the Atmospheric Sciences* **46**, 2026–2030.

Lilly, D. K. (1990), 'Numerical prediction of thunderstorms — has its time come?', *Quarterly Journal of the Royal Meteorological Society* **116**, 779–798.

Lindborg, E. (1999), 'Can the atmospheric kinetic energy spectrum be explained by two-

dimensional turbulence?', *Journal of Fluid Mechanics* **388**, 259–288.

Lorenz, E. N. (1963), 'Deterministic nonperiodic flow', *Journal of the Atmospheric Sciences* **20**, 130–141.

Lorenz, E. N. (1969), 'The predictability of a flow which possesses many scales of motion', *Tellus* **21**, 289–307.

Lynch, P. (2006), *The Emergence of Numerical Weather Prediction: Richardson's Dream*, Cambridge University Press, Cambridge, UK.

Lynch, P. (2008), 'The origins of computer weather prediction and climate modeling', *Journal of Computational Physics* **227**, 3431–3444.

Maltrud, M. E. & Vallis, G. K. (1991), 'Energy spectra and coherent structures in forced two-dimensional and beta-plane turbulence', *Journal of Fluid Mechanics* **228**, 321–342.

Massart, S. (2019), A new hybrid formulation for the background error covariance in the IFS: evaluation, Technical Memorandum 856, European Centre for Medium-Range Weather Forecasts.

Meteorological Office (n.d.), 'D-day: The role of the Met. Office', available at `https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/archive/archive-treasures/d_day_the_role_of_the_met_office.compressed.pdf` (accessed: 19th June, 2020).

Molteni, F., Buizza, R., Palmer, T. N. & Petroliagis, T. (1996), 'The ECMWF Ensemble Prediction System: methodology and validation', *Quarterly Journal of the Royal Meteorological Society* **122**, 73–119.

Nastrom, G. D. & Gage, K. S. (1985), 'A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft', *Journal of the Atmospheric Sciences* **42**, 950–960.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J. & Sandu, I. (2017), 'Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF

ensemble', *Quarterly Journal of the Royal Meteorological Society* **143**, 408–422.

Orszag, S. A. (1970), 'Analytical theories of turbulence', *Journal of Fluid Mechanics* **41**, 363–386.

Owens, R. G. & Hewson, T. D. (2018), *ECMWF Forecast User Guide*, European Centre for Medium-Range Weather Forecasts.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M. & Weisheimer, A. (2009), Stochastic parametrization and model uncertainty, Technical Memorandum 598, European Centre for Medium-Range Weather Forecasts.

Palmer, T. N., Döring, A. & Seregin, G. (2014), 'The real butterfly effect', *Nonlinearity* **27**, R123–R141.

Persson, A. (2005), 'Early operational Numerical Weather Prediction outside the USA: an historical introduction. Part 1: Internationalism and engineering NWP in Sweden, 1952–69', *Meteorological Applications* **12**, 135–159.

Rabier, F. (2005), 'Overview of global data assimilation developments in numerical weather-prediction centres', *Quarterly Journal of the Royal Meteorological Society* **131**, 3215–3233.

Richardson, D. S. (2000), 'Skill and relative economic value of the ECMWF ensemble prediction system', *Quarterly Journal of the Royal Meteorological Society* **126**, 649–667.

Richardson, D. S., Cloke, H. L. & Pappenberger, F. (2020), 'Evaluation of the consistency of ECMWF ensemble forecasts', *Geophysical Research Letters* **46**, e2020GL087934.

Richardson, L. F. (1922), *Weather Prediction by Numerical Process*, Cambridge University Press, Cambridge, UK. Reprinted by Dover Publications (1965) with a new introduction by S. Chapman; reprinted by Cambridge University Press (2006) with a new introduction by Peter Lynch.

Robertson, A. W., Vitart, F. & Camargo, S. J. (2020), 'Subseasonal to seasonal prediction

of weather to climate with application to tropical cyclones', *Journal of Geophysical Research: Atmospheres* **125**, e2018JD029375.

Robinson, J. C. (2001), *Infinite-Dimensional Dynamical Systems: An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*, Cambridge University Press, Cambridge, UK.

Robinson, J. C. (2013), 'Attractors and finite-dimensional behaviour in the 2D Navier-Stokes equations', *ISRN Mathematical Analysis* **2013**, 291823.

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.

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.

Selz, T. (2019), 'Estimating the intrinsic limit of predictability using a stochastic convection scheme', *Journal of the Atmospheric Sciences* **76**, 757–765.

Selz, T. & Craig, G. C. (2015), 'Upscale error growth in a high-resolution simulation of a summertime weather event over Europe', *Monthly Weather Review* **143**, 813–827.

Skamarock, W. C. (2004), 'Evaluating mesoscale NWP models using kinetic energy spectra', *Monthly Weather Review* **132**, 3019–3032.

Smagorinsky, J. (1969), 'Problems and promises of deterministic extended range forecasting', *Bulletin of the American Meteorological Society* **50**, 286–312.

Sun, Y. Q. & Zhang, F. (2016), 'Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect', *Journal of the Atmospheric Sciences* **73**, 1419–1438.

Sun, Y. Q. & Zhang, F. (2020), 'A new theoretical framework for understanding multiscale

atmospheric predictability', *Journal of the Atmospheric Sciences* **77**, 2297–2309.

Sutton, G. (1955), 'Weather forecasting: the future outlook', *Nature* **176**, 993–996.

Thompson, P. D. (1953), 'On the theory of large-scale disturbances in a two-dimensional baroclinic equivalent of the atmosphere', *Quarterly Journal of the Royal Meteorological Society* **79**, 51–69.

Thompson, P. D. (1957), 'Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns', *Tellus* **9**, 275–295.

Toth, Z. & Kalnay, E. (1993), 'Ensemble forecasting at NMC: the generation of perturbations', *Bulletin of the American Meteorological Society* **74**, 2317–2330.

Vallis, G. K. (1985), 'Remarks on the predictability properties of two- and three-dimensional flow', *Quarterly Journal of the Royal Meteorological Society* **111**, 1039–1047.

Vallis, G. K. (2017), *Atmospheric and Oceanic Fluid Dynamics*, 2$^{nd}$ edn, Cambridge University Press, Cambridge, UK.

Vallis, G. K., Shutts, G. J. & Gray, M. E. B. (1997), 'Balanced mesoscale motion and stratified turbulence forced by convection', *Quarterly Journal of the Royal Meteorological Society* **123**, 1621–1652.

van Niekerk, A., Shepherd, T. G., Vosper, S. B. & Webster, S. (2016), 'Sensitivity of resolved and parametrized surface drag to changes in resolution and parametrization', *Quarterly Journal of the Royal Meteorological Society* **142**, 2300–2313.

Vannitsem, S. & Toth, Z. (2002), 'Short-term dynamics of model errors', *Journal of the Atmospheric Sciences* **59**, 2594–2604.

Waite, M. L. & Snyder, C. (2009), 'The mesoscale kinetic energy spectrum of a baroclinic life cycle', *Journal of the Atmospheric Sciences* **66**, 883–901.

Waite, M. L. & Snyder, C. (2013), 'Mesoscale energy spectra of moist baroclinic waves', *Journal of the Atmospheric Sciences* **70**, 1242–1256.

Wang, C. & Fiedler, P. C. (2006), 'ENSO variability and the eastern tropical Pacific: a review', *Progress in Oceanography* **69**, 239–266.

Wang, J.-W. A. & Sardeshmukh, P. D. (2020), 'Challenges in improving the representation of mesoscale kinetic energy in NWP models', paper presented at the 100th Annual Meeting of the American Meteorological Society, available at `https://ams.confex.com/ams/2020Annual/webprogram/Paper367785.html` (accessed: 19th August, 2020).

Wheeler, M. C., Zhu, H., Sobel, A. H., Hudson, D. & Vitart, F. (2017), 'Seamless precipitation prediction skill comparison between two global models', *Quarterly Journal of the Royal Meteorological Society* **143**, 374–383.

Wilks, D. S. (2019), *Statistical Methods in the Atmospheric Sciences*, 4th edn, Elsevier, Amsterdam, Netherlands.

Würsch, M. & Craig, G. C. (2014), 'A simple dynamical model of cumulus convection for data assimilation research', *Meteorologische Zeitschrift* **23**, 483–490.

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: Dynamic Meteorology and Oceanography* **69**, 1287492.

Zhang, C. (2005), 'Madden-Julian Oscillation', *Reviews of Geophysics* **43**, RG2003.

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.